Understanding complexities in Biochemical Pathways related to Human Diseases - A Computational and Mathematical Approach

Thesis Submitted to AcSIR for the Award of the Degree
of
DOCTOR OF PHILOSOPHY
in
Biological Sciences

By

Saikat Chowdhury

Registration Number: **10BB13A26041**

Under the guidance of

Dr. Ram Rup Sarkar
(Research Supervisor)

Chemical Engineering and Process Development Division
CSIR-National Chemical Laboratory, Pune 411008

December, 2018

DECLARATION

I hereby declare that the thesis entitled "***Understanding complexities in Biochemical Pathways related to Human Diseases - A Computational and Mathematical Approach***" submitted for the degree of Doctor of Philosophy in Biological Sciences to the Academy of Scientific & Innovative Research (AcSIR), has been carried out by me at the Chemical Engineering and Process Development Division of CSIR-National Chemical Laboratory, Pune under the supervision of **Dr. Ram Rup Sarkar**. Such material as has been obtained by other sources has been duly acknowledged in the thesis. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other Institution or University.

Saikat Chowdhury.

Saikat Chowdhury

Ph. D candidate

Date: 19/12/2018

Place: CSIR-NCL, Pune

सीएसआयआर–राष्ट्रीय रासायनिक प्रयोगशाला
(वैज्ञानिक तथा औद्योगिक अनुसंधान परिषद)
डॉ. होमी भाभा मार्ग, पुणे – 411 008. भारत
CSIR-NATIONAL CHEMICAL LABORATORY
(Council of Scientific & Industrial Research)
Dr. Homi Bhabha Road, Pune - 411008. India

Certificate

This is to certify that the work incorporated in this Ph.D. thesis entitled *"Understanding complexities in Biochemical Pathways related to Human Diseases - A Computational and Mathematical Approach"* submitted by Mr. *Saikat Chowdhury* to the Academy of Scientific and Innovative Research (AcSIR) in fulfillment of the requirements for the award of the Degree of **Doctor of Philosophy in Biological Sciences**, embodies original research work under my supervision/guidance. I further certify that this work has not been submitted to any other University or Institution in part or full for the award of any degree or diploma. Research material obtained from other sources has been duly acknowledged in the thesis. Any text, illustration, table etc., used in the thesis from other sources, have been duly cited and acknowledged.

Saikat Chowdhury 19/12/2018
(Student)

Mr. Saikat Chowdhury
Senior Research Fellow
Chemical Engineering and Process Development
CSIR-National Chemical Laboratory
Dr. Homi Bhaba Road
Pune - 411008

Ram Rup Sarkar 19/12/2018
(Supervisor)

Dr. Ram Rup Sarkar
Principal Scientist
Chemical Engineering and Process Development
CSIR-National Chemical Laboratory
Dr. Homi Bhaba Road
Pune - 411008

Communication Channels
NCL Level DID : 2590
NCL Board No. : +91-20-25902000
EPABX : +91-20-25893300
: +91-20-25893400
FAX
Director's Office : +91-20-25902601
COA's Office : +91-20-25902660
COS&P's Office : +91-20-25902664
WEBSITE
www.ncl-india.org

I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me.

- Isaac Newton

Acknowledgments

The work presented in this thesis would not have been possible without my close association with many people. I take this opportunity to extend my sincere gratitude and appreciation to all those who made this Ph.D. thesis possible.

First and foremost, I would like to express my sincere gratitude to my Ph.D. supervisor Dr. Ram Rup Sarkar for introducing me to this exciting field of research, and for his dedicated help, advice, inspiration, encouragement and continuous support, throughout the journey of my Ph.D. Under his guidance, I have successfully overcome many difficulties and learned a lot. I appreciate all his contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. The joy and enthusiasm he has for his research were contagious and motivational for me, even during tough times in the Ph.D. pursuit. I hope his constant guidance in future will also help me to drive my motivations in the pursuit of scientific knowledge and discoveries.

I would also like to thank the members of my Doctoral Advisory Committee (DAC), Dr. Sarika Maitra Bhattacharyya, Dr. Anu Raghunathan, and Dr. Leelavati Narlikar. They have given various useful suggestions to improvise the quality of my research works and guided me to fulfill the objectives of my thesis. I would also like to thank Dr. Chetan Gadgil for his critical inputs on my research works, which had further helped me to think about the problems of my research projects from different perspectives.

I would also like to thank our director Prof. Ashwini Kumar Nangia, past director Prof. Sourav Pal, the HODs of CEPD Division Dr. Sunil Joshi, Dr. S. S. Tambe, and Dr. V. V. Ranade for providing me the infrastructure to perform my research works in CSIR-National Chemical Laboratory. I also thank our Finance and Accounts section, Student Academic Office (SAO), CEPD office, DIRC, and other departments of CSIR-NCL for providing me the administrative support and necessary infrastructures.

My sincere thank to DST-INSPIRE, Govt. of India for providing me the JRF and SRF fellowships during the entire period of my Ph.D.

My journey in research started under the guidance of Dr. Somdatta Sinha in CSIR-CCMB as an INSA summer research fellow and later as a Project JRF in her group. I gratefully acknowledge her support and guidance, which she provided to me at the initial stage of my journey in research. I also like to thank the senior lab mates, Dr. Aridaman, Dr. Ashutosh, and Dr. Priya of Dr. Sinha's lab for their support and motivation, which I required most at the beginning of my journey in Ph.D.

I would also like to thank my teachers Prof. Sudip Kundu, Prof. Subhasish Mukherjee, Prof. Sukla Ghosh and other professors of the Biophysics department of the University of Calcutta for motivating me to do Ph.D. after finishing my post-graduation.

I also acknowledge the co-author of my manuscript Dr. Rachana Pradhan for her support to successfully complete the project and writing the manuscript. I am thankful to my all lab mates,

Abhishek, Piyali, Rupa, Sutanu, Noopur, Vidhi, Arpit, and Jitesh for their continuous support and valuable suggestions in my research works. Their presence in the lab always made the workplace more enjoyable, which sometimes helped me to divert my mind from the overwhelming workloads and personal problems. I also acknowledge my delights to work with my junior lab mates Prasun, Sanjana, Rochi, Mudita, and Anil. I am also thankful to the summer students and project trainees, Shomeek, Astha, Sinal, Shubhangi, Siddhi, Shubika for their active participation and helping hands to accelerate my research works.

I would also like to especially thank my lab mates Piyali and Rupa for correcting the proofs of my research articles and the thesis chapters. I also acknowledge their continuous supports and encouragements at various occasions of my PhD.

I am also thankful to my roommates, Souvik, Abhik, Aniruddha, Manoj, Ujjwal, and Alam for helping me on various occasions during my stay in Pune. Our endless debates and discussions on world cinema, sports, and politics are the unforgettable moments of my life. I am also greatly in debt to my senior Swarup da for providing me with accommodation in his flat after I came to Pune.

I remember my university and college friends and seniors, Debasish, Debdatta, Pritam, Malancha, Rajeswari, Jagriti, Sana, Rajat, Arup, Samik, Swapratim da, Anindya da, who were always with me whenever I needed their company, suggestions and help.

Last but not the least, I am greatly thankful to my parents, Mrs. Swapna Chowdhury and Mr. Ramendra Krishna Chowdhury, my uncle Mr. Ajay Choudhury, my aunt Mrs. Sukti Choudhury, my sister Rimi and her husband Arnab, my cousin Debasmita, Barun uncle and Shukla aunty, my childhood friend Bubai, and my other family members for continuously encouraging and spiritually helping me to achieve the desired goals of my life.

Also, I am extremely grateful to all the other people, I came across in NCL who equally helped me during this journey.

Saikat Chowdhury

Table of Contents

List of Figures ..... **i**

List of Tables ..... **iii**

Abstract ..... **iv**

List of Publications ..... **vi**

Chapter 1 ..... **1**

1 GENERAL INTRODUCTION AND SCOPE OF THE THESIS **1**

1.1 BIOCHEMICAL PATHWAYS AND HUMAN DISEASES 1

1.1.1 Biochemical Pathways - a Conceptual Framework to Understand Different Human Diseases 2

1.1.2 Interconnectivities of the Biochemical Networks and Disease Pathogenesis 3

1.1.3 Biochemical Reaction Networks in Cancer Cells 5

1.2 IMPORTANCE OF CELL SIGNALLING NETWORKS IN ONCOGENESIS 6

1.3 TARGETING SIGNALING PATHWAYS IN CANCER STEM CELLS 9

1.4 CURRENT CHALLENGES OF STUDYING SIGNALING PATHWAYS IN TARGET-BASED CANCER THERAPEUTICS 10

1.4.1 Identification of Druggable Proteins in Target-Based Cancer Therapeutics 10

1.4.2 Limitations of the Complete Suppression of Cancer Signaling Network 13

1.4.3 Emergence of Drug Resistant Cancer Cells in Targeted Therapy 14

1.5 SIGNALING PATHWAYS AND ITS APPLICATIONS IN ONCOGENESIS 16

1.6 SCOPE AND SPECIFIC OBJECTIVES OF THE THESIS 21

1.6.1 Outline of the Thesis 23

Chapter 2 ..... **26**

2 MATERIALS AND METHODS **26**

2.1 AVAILABLE RESOURCES OF PATHWAY DATA 26

2.2 DATA COLLATION PROCEDURES OF SIGNALING NETWORKS 27

2.2.1 Reconstructions of the Signalling Networks 29

2.3 CONSTRUCTION OF PATHWAY KNOWLEDGEBASE 29

2.3.1 Database Architecture and Implementation 29

2.4 TOPOLOGICAL ANALYSES OF SIGNALING NETWORKS 39

2.4.1 Network Parameters Calculated From the Adjacency Matrix 39

2.4.2 Network Parameters of Hedgehog and Notch Signaling Pathways 41
2.5 SEMI-DYNAMIC ANALYSES USING BOOLEAN LOGIC 42
2.5.1 General Definitions Used in Logic Theory 45
2.5.2 Logical Steady-state Analyses (LSS) of Hedgehog Pathway 47
2.5.3 Simulation of the Logical Model of Notch Signaling Pathway 49
2.5.4 Methodologies used in the Simulation of Neurogenesis and Gliomagenesis 50
2.5.5 Marker Proteins of aNSC, GSC and GBM Models 52
2.5.6 Determination of the Phenotypes and Cellular States 53
2.5.7 Calculation of Normalized Frequencies of Cellular States, Shannon Entropy and Activity Ratio (AR) Scores 54
2.5.8 Calculation of Phenotype Cost Function 60
2.5.9 Selection of Patient Cohorts and Preparation of RNAseq Sample Data Sets 64
2.5.10 Differential Expression Analyses of mRNA transcripts 65
2.5.11 Logical Simulations using the Transcriptomics Data Generated from TCGA-LGG and TCGA-GBM 65
2.5.12 Methodologies used for Drug Target Screening 68
2.6 SEMI-DYNAMIC ANALYSES USING QUATERNARY LOGIC 73
2.6.1 Truth Table of Quaternary Logic-Based Operation 73
2.6.2 Modifications of the Logical Equations Constructed Using Negation Rule 73
2.6.3 Extended Quaternary States Update Scheme (ExQSUS) 75
2.6.4 Quaternary States Selection Rules 76
2.6.5 Implementation and Execution of Quaternary Logic 77
2.6.6 Simulations of the Signaling Networks using ExQuLogic Method 78
2.6.7 Computation of the Level of Accuracy of Predicting the Activity Profile 78
Chapter 3 ..... **80**
3 DEVELOPMENT OF A NEW PATHWAY DATABASE WITH INTEGRATED COMPUTATIONAL PLATFORM 80
3.1 INTRODUCTION 80
3.2 A BRIEF HISTORY OF THE EXISTING SIGNALING DATABASES 82
3.3 STATE-OF-THE-ARTS 84
3.3.1 Currently Available Active Pathway Databases 85
3.3.2 Comparative Study on Pathway Information 87

3.3.3 Comparative Study Based on Technical Features 92
3.4 RATIONALES FOR DEVELOPING NEW DATABASE 93
3.5 BIOPYDB: A NEW HUMAN CELL SIGNALLING DATABASE 95
3.5.1 Data Storage System 97
3.5.2 Web Interface of BIOPYDB 99
3.5.3 Resource of Biochemical Pathways 99
3.5.4 Resource of Proteins/Genes Involved in Biochemical Pathways 101
3.5.5 Resource of Protein-Protein Interaction Data 102
3.5.6 Resource for Visualizing the Protein-Disease Mapping 103
3.5.7 Computational Platform for Pathway Data Analyses 104
3.6 DISCUSSION 112
3.7 CONCLUSION 114
Chapter 4 ..... **116**
4 UNDERSTANDING OF HEDGEHOG SIGNAL TRANSMISSION DYNAMICS IN CANCER CELLS 116
4.1 INTRODUCTION 116
4.2 STATE-OF-THE-ART 118
4.3 RATIONALE AND OBJECTIVES 119
4.4 HYPOTHESES AND ASSUMPTIONS 123
4.5 RESULTS 124
4.5.1 Reconstruction of Human Hedgehog Signaling Pathway 124
4.5.2 Topological Analyses of the Reconstructed Hedgehog Signaling Network 130
4.5.3 Semi-Dynamic or Logical Analyses of Hedgehog Signaling Pathway 139
4.6 DISCUSSION 159
4.7 CONCLUSION 165
Chapter 5 ..... **166**
5 UNDERSTANDING THE ROLE OF NOTCH SIGNALING PATHWAY IN THE GROWTH OF GLIOBLASTOMA CELLS UNDER STEADY STATE 166
5.1 INTRODUCTION 166
5.2 CURRENT CHALLENGES 167
5.3 HYPOTHESES AND OBJECTIVES 169

5.4 RESULTS....170
5.4.1 Reconstructed Map of Notch Signaling Network....170
5.4.2 Topology of Notch signaling network....174
5.4.3 Logical Analyses of Notch Signaling Network....179
5.4.4 Validations of GBS and NNS Models....180
5.4.5 Simulation of GSI Non-Responsive Scenario....182
5.4.6 Variations in the number of activator and inhibitor nodes in the NNS, GBS, and GSI scenarios....182
5.4.7 Identification of Alternative Drug Targets....186
5.5 DISCUSSION....188
5.6 CONCLUSION....191
***Chapter 6*....192**
6 IMPLICATION OF NOTCH PATHWAY ANALYSES IN THE EVOLUTIONS OF INTRA-TUMOR HETEROGENEITY AND RISK PREDICTION OF GLIOBLASTOMA....192
6.1 INTRODUCTION....192
6.2 CURRENT CHALLENGES AND OBJECTIVES UNDERTAKEN....194
6.3 FORMULATIONS OF THE PROBLEM AND WORK FLOW....195
6.4 RESULTS....196
6.4.1 Mechanistic Models of aNSC, GSC, and GBM Tumor Development Illustrate Cell Proliferation and Differentiation Dynamics....196
6.4.2 Network Motifs of Notch Pathway in the Regulation of Cellular States....202
6.4.3 Increased Activities of JAK2/STAT3, RBPJ, YY1, γ-Secretase Complex and P53 Mutation Promoted Low-Grade GBM to High-Grade....205
6.4.4 Origin of the Evolution of Intra-Tumor Heterogeneity and Distinct Sub-Types of GBM Tumor....207
6.4.5 Understanding the Bias in the Outcomes of Different Cellular States under Different Conditions....210
6.4.6 Applications of Phenotype Predictor Scores to Predict the Appearances of Cellular States......213
6.4.7 Retrospective Studies of the Low and High- Grade GBM Patient's Cohort....217
6.4.8 A Case Study of Screening and Ranking of Potential Drug Targets in High-Grade GBM Cells....223

6.5 DISCUSSION 226
6.6 CONCLUSIONS 232
Chapter 7 ..... **233**
7 IMPLICATIONS OF THE RECONSTRUCTED INTEGRATED CELL SIGNALING NETWORKS TO UNDERSTAND THE GROWTH AND METASTASIS OF GLIOBLASTOMA **233**
7.1 INTRODUCTION 233
7.1.1 Challenges in the Selection of Binary States of the Nodes 236
7.1.2 Methodological Challenges of the Constructions of Logical Equations 238
7.1.3 Alternative Solutions to Overcome the Methodological Challenges of Boolean Logic 239
7.1.4 Proposed Approach 241
7.1.5 Hypothesis and Objectives 246
7.2 RESULTS 247
7.2.1 Case Studies of the Previously Published Logical Models 247
7.2.2 Application of ExQuLogic-based Simulation Techniques on Other Datasets 262
7.2.3 Activity Profiles of the Genes/Proteins in the Integrated Logical Model of Hedgehog, Notch, WNT, and EGFR Pathways 265
7.3 DISCUSSION 277
7.4 CONCLUSIONS 279
Chapter 8 ..... **281**
8 CONCLUDING REMARKS AND FUTURE WORKS **281**
8.1 APPLICATIONS AND FUTURE DIRECTIONS 282
8.1.1 Applications of the Drug-targets in Cancer Therapy 282
8.1.2 Implementation of Predictive Model for Tumor Risk Prediction 282
8.1.3 Applications of ExQuLogic 282
8.1.4 Third-party use of BIOPYDB Database 284
9 APPENDICES **285**
10 REFERENCES **338**

List of Figures

Figure 1: Coupling of different molecular interaction networks in cancer cells. 5
Figure 2: Significance of the understanding of cell signaling pathways in oncogenesis. 8
Figure 3: General experimental strategy to identify drug-targets in cancer cells. 11
Figure 4: Alternate molecular routes mediated by multiple cell signaling pathways in cancer cells. 14
Figure 5: Available databases used for pathway data collation. 26
Figure 6: Flow-chart of pathway nomenclature system used in BIOPYDB database. 31
Figure 7: Flow-chart for pathway image and textual data preparation. 37
Figure 8: Workflow for translating signaling network to logical hyper-graph simulation. 44
Figure 9: Evolution of the existing and active human signaling pathway databases. 83
Figure 10: Signaling pathway databases and the available pathway data types. 86
Figure 11: Comparative statistics of pathway information available in databases. 87
Figure 12: Pathway ontology used by INOH and Pathway Ontology portal. 88
Figure 13: Cross-referencing strategies of pathway data. 90
Figure 14: Schematic diagram of the database architecture used in BIOPYDB. 98
Figure 15: Snapshot of the frontend page of BIOPYDB database. 100
Figure 16: Flow-chart describing different operations executed in BIOPYDB 105
Figure 17: Hedgehog signaling activation mechanisms in OFF and ON states 117
Figure 18: Reaction cascades in Vismodegib sensitive and resistive cancer cells. 120
Figure 19: Reconstructed human cell specific Hedgehog signaling pathway. 125
Figure 20: Comparison of the number of proteins available in other database and our model. 126
Figure 21: Network picture of Hedgehog signaling pathway 131
Figure 22: Topological parameters of Hedgehog signaling network. 132
Figure 23: All-pair directed shortest paths of the Hedgehog signaling proteins. 137
Figure 24: Comparative analyses of Glioblastoma scenarios. 142
Figure 25: Comparative analyses of Colon Cancer scenarios. 147
Figure 26: Comparative analyses of Pancreatic Cancer scenarios. 149
Figure 27: GBM model validation and signaling motif behind anti-SMO drug resistivity. 151
Figure 28: Colon cancer model validation and signaling motif behind anti-SMO drug resistivity. 155
Figure 29: Pancreatic cancer model validation and signaling motif behind anti-SMO drug resistivity. 157
Figure 30: Reconstructed map of the human Notch signaling pathway. 171
Figure 31: Logical states of the nodes observed in differential expression analyses and simulations of NNS and GBS models. 181
Figure 32: Logical states of the nodes observed in experimental analyses and simulations of

NNS, GBS, and GSI models. 183
Figure 33: Comparison between Normal, Glioblastoma, Gamma Secretase inhibition and the two proposed drug target scenarios. 184
Figure 34: Logical states of the nodes observed in experimental analyses and simulations of NNS, GBS, GSI, TS1, and TS2 models. 188
Figure 35: Simulation results of the aNSC, GSC, and GBM models. 197
Figure 36: Activity Ratio (AR) scores of the input molecules. 199
Figure 37: Activity profiles and the temporal dynamics of the marker proteins observed in general GBM model simulation. 200
Figure 38: Active network modules in Notch pathway. 204
Figure 39: Comparative analyses of the normalized frequencies of cellular states observed in general GBM and Grade-IV models. 206
Figure 40: Full State transition graph (STG) of general GBM model. 207
Figure 41: Partial stage transition graph (STG) of the general GBM model. 208
Figure 42: Activity patterns of the pathway molecules observed in all the periodic attractor states corresponding to Grade-IV tumor state in the general GBM model simulation. 209
Figure 43: Violin plots of the total phenotype cost function or total costs calculated for different cellular states. 212
Figure 44: Distributions of the Phenotype Predictor Scores. 214
Figure 45: Comparative statistics of the normalized frequency distributions of different cellular states in master aNSC and general GBM models. 220
Figure 46: Simulation outcomes of drug targets screening analyses 226
Figure 47: Simulation outcomes of Neural and Glioma stem cells development and the emergence of different grades of the tumor cells. 228
Figure 48: The rheostat model proposed for Notch signaling network. 230
Figure 49: Activity profiles of the select proteins of T-cell pathway model. 250
Figure 50: JNK activation profile in the different cellular systems. 252
Figure 51: Simulation output of ERBB family receptor proteins mediated pathway in Glioblastoma cells. 255
Figure 52: Logical steady-states of the proteins in Hedgehog pathway model. 261
Figure 53: Quaternary states of the target genes and temporal dynamics of the corresponding phenotypes of integrated pathway model. 270
Figure 54: Network view of the whole reconstructed signaling network and the identified cell cycle progression and EMT regulatory module. 273
Figure 55: Examples of the signaling motifs identified in the cell cycle progression and EMT regulatory module. 275
Figure 56: Flow-chart of the decision-making steps to determine risk of GBM. 283

List of Tables

Table 1: Statistics of Glioblastoma Patient Cohorts 27

Table 2: Information of the marker proteins mapped with different phenotypes 53

Table 3: The Truth Table of Quaternary Logic 73

Table 4: Binary Logic-Based Truth Table of the Modified NOT Gate Equation 74

Table 5: Truth Table of the Extend ExQuLogic 74

Table 6: BIOPYDB Database Statistics 104

Table 7: Significant Proteins Extracted from Connectivity Analysis 134

Table 8: Comparison of the Percentage of Accuracy between Experimental and Simulation Results. 152

Table 9: Comparative Statistics of the Reconstructed Notch Pathway Data 172

Table 10: Significant Proteins of Notch Signaling Pathway Identified through Topological Analyses 176

Table 11: Mean Values with 95% CI of the Phenotype Predictor Scores of all Cellular States. 216

Table 12: Statistics of the TCGA Glioblastoma Patient Cohorts. 218

Table 13: Differential Expressions Genes in Glioblastoma Patient Cohorts. 219

Table 14: Delay Difference and Significant Correlation Observed Between Grade-IV Trajectory and Pathway Molecules. 224

Table 15: Calculated ranks of the effective drug targets for suppressing Grade-IV tumor cells. 225

Table 16: Percentage of Accuracy Obtained in BL- and ExQuLogic methods 263

Abstract

The present investigations of the thesis have been performed to understand the complex mechanisms of large-scale biochemical pathways, mainly cell signaling networks and its complex interactions with gene regulatory processes in cancer cells using *in-silico* models. The main aim is to identify the crucial nodes or active regulatory reaction motifs in the entire network, which can be targeted to suppress the activities of the signaling networks in the proliferating and metastatic tumor/cancer cells, which could be further considered as important drug-targets in targeted cancer therapeutics. However, developing *in-silico* models of large-scale signaling networks to accurately capture the dynamics of pathway species in a cellular condition is a specifically challenging task. Hence, to overcome such problems, at first the large cell signaling networks have been curated from various sources. A new open source biochemical database BIOPYDB (*http://biopydb.ncl.res.in/biopydb*) has been developed in which the curated pathways have made available for the general research purposes. The database is equipped with the computational tools for performing network, logical, and dynamic analyses of the pathways. Availabilities of these computational tools including various unique features have made this database as a useful resource for pathway data searching, reconstructions, modeling, and simulation tasks within a single platform. Following this, the important oncogenic signaling pathways, such as Hedgehog (HH), Notch, WNT, and EGFR, etc. have been modeled and simulated using semi-dynamic, discrete logical modeling approach in the current investigation. Also, the static graph theoretic analysis has been utilized to understand the topology of the signaling networks in various conditions and subsequently identify the "Hubs" genes/proteins in the entire network.

The broad objectives of the thesis are to understand the molecular mechanisms of the tumorigenic processes in cancer cells, find out the minimal number of potential drug-targets, and most importantly predict the risk of the development of tumorigenic cells from the normal cells or tissues (e.g., brain) using *in-silico* logical models of the signaling network.

In order to fulfill these objectives, at first individual logical models of the oncogenic, developmental cell signaling networks, Hedgehog (HH) and Notch pathways have been developed. The HH model has been able to predict the molecular mechanisms or the active reaction motifs, which comprise of the crosstalk proteins of other signaling pathways, such as

RAS, ULK3, ERK1/2, TWIST in three cancer/tumor cell models *viz.* Glioblastoma, Colon, and Pancreatic cancers. It has been predicted that due to these active reaction motifs, the suppression of Smoothened (SMO) receptor protein, which is known for activating HH pathway, fails to stop the cell cycle progressions of cancer cells in target-based anti-cancer therapy. Furthermore, the topological analyses of the reconstructed HH network has been able to extract the important "Hubs" in the network, from which few select combinations have been proven useful to stop the SMO inhibitor resistant cancer cells. Similar, analyses have been also performed on the reconstructed Notch signaling model to identify the alternative drug target(s), which will be able to suppress the growth of Glioblastoma Grade-IV tumor cells. Two novel combinations of drug targets (a) NICD1 & MAML and (b) NICD1 & HIF1A have been identified, which have shown more powerful impact to shut down the active Notch signaling and its crosstalk pathways in the GBM cells as compare to the known drug-target Gamma-Secretase enzyme of this pathway.

In order to understand Notch pathway in the development of GBM cells starting from the normal adult neural stem cells (aNSCs) or Glioma stem cells (GSCs), different logical models have been developed for each cell types (aNSCs, GSCs, Glioblastoma). The general Glioblastoma (GBM) model has been able to nicely simulate the developmental dynamics of the development of different tumor grades or sub-types (i.e. low and high grade) as well as the molecular differences between the tumor sub-clones. It has been observed that Notch signaling can play central role to tune the dynamics of aNSCs and GSCs, and thus develop the normal differentiated cells (neurons, astrocytes) as well as different sub-types and distinct molecular sub-clones of GBM cells. This analysis has been specifically helpful to explore the causes behind the drug resistivity, intra-tumor heterogeneity, and tumor sub-types of GBM cells.

Lastly, an integrated signaling pathway model combining major oncogenic pathways, such as HH, Notch, WNT, EGFR, PI3K/AKT, and etc. has been developed to study the gene expression patterns and identify novel drug-targets in the GBM cells. In this work, newly developed extended quaternary logic (ExQuLogic)-based modeling technique has been used to simulate and identify the signaling pathway activities in GBM cells. The combinations of the proteins from different signaling pathways *viz.* SMO, GLI1, GLI3, NICD1, HEY1, NRARP, and SNAI1 have been proposed as the minimal number of drug targets for the treatment of advanced metastatic GBM cells.

List of Publications

Research Articles

Related to thesis:

1. **S. Chowdhury**, R. N. Pradhan, R. R. Sarkar (2013) Structural and logical analysis of a comprehensive hedgehog signaling pathway to identify alternative drug targets for glioma, colon and pancreatic cancer. **PLoS ONE,** 8: e69132. doi:10.1371/journal.pone.0069132

2. **S. Chowdhury**, R. R. Sarkar (2013) Drug targets and biomarker identification from computational study of human notch signaling pathway. **Clinical and Experimental Pharmacology,** 3: 2161-1459.1000137.

3. **S. Chowdhury**, R. R. Sarkar (2015) Comparison of human cell signaling pathway databases - evolution, drawbacks and challenges, **DATABASE The Journal of Biological Databases and Curation (Oxford University Press)**, 2015: bau126, DOI:10.1093/database/bau126.

4. **S. Chowdhury**, N. Sinha, P. Ganguli, R. Bhowmick, V. Singh, S. Nandi, R. R. Sarkar (2018) BIOPYDB: A dynamic human cell specific biochemical pathway database with advanced computational analyses platform**. Journal of Integrative Bioinformatics**, DOI: (10.1515/jib-2017-0072)

5. **S. Chowdhury**, R. R. Sarkar (2018) Exploring the mechanisms of Notch pathway during the Emergence of Intra-tumor Heterogeneity in Glioblastoma development. *Manuscript under Review.*

6. **S. Chowdhury**, R. R. Sarkar (2018) A novel quaternary logic-based approach to analyze the gene expression patterns in signaling pathways. *Manuscript under review.*

Other publications:

7. P. Ganguli, **S. Chowdhury**, R. Bhowmick, R. R. Sarkar (2015) Temporal protein expression pattern in intracellular signaling cascade during t cell activation: a computational study. **Journal of Bioscience**, 40(4), 769 - 789 (2015), DOI: 10.1007/s12038-015-9561-1

8. P. Ganguli, **S. Chowdhury**, S. Chowdhury, R. R. Sarkar (2015) Identification of Th1/Th2 regulatory switch to promote healing response during Leishmaniasis: a computational approach, **EURASIP Journal on Bioinformatics and Systems Biology**, 2015:13, DOI: 10.1186/s13637-015-0032-7.

9. N. Sinha, **S. Chowdhury**, R. R. Sarkar (2016) Structural insight of NICD-MAML interactions: virtual screening, docking and molecular dynamics study for identification of potential inhibitor, **Letters in Drug Design & Discovery**, 13(4), 301 - 313

10. N. Sinha, **S. Chowdhury,** R. R. Sarkar (2017) Deciphering Structural Stability and Binding Mechanisms of Potential Antagonists with Smoothened Protein, **Journal of Biomolecular Structure and Dynamics** (*in press*), DOI:10.1080/07391102.2017.1372310.

Patents

1. R.R. Sarkar, **S. Chowdhury** (2013). In silico method to identify the important biomarkers and combinatorial oncoproteins in target based cancer therapy. **US20160154927A1** filed on 21st July 2013. Patent Pending.

2. R. R. Sarkar, **S. Chowdhury** (2014). Identification of minimal combination of oncoproteins in Notch pathway to suppress human glioblastoma. **US20160125127A1** filed on October 29, 2014. Patent Pending.

3. R. R. Sarkar, P. Ganguli, **S. Chowdhury** (2015) Identification of minimal combinations of molecules that act as probable immunostimulators against Leishmaniasis. **WO2017081703A1** filed on December 11, 2015. Patent Pending.

Chapter 1

1 GENERAL INTRODUCTION AND SCOPE OF THE THESIS

1.1 BIOCHEMICAL PATHWAYS AND HUMAN DISEASES

Biochemical pathway is an umbrella term commonly used for describing all sorts of chemical reaction cascades, which occur in living cells and tissues [1-5]. Often the chemical reaction cascades of the biochemical pathways, which are relayed from the reactants to the products catalyzed by different enzymes and multiple regulators, are referred as *"biochemical reaction network"* or simply *"biochemical network"* [6,7]. In terms of chemical and functional properties, the biochemical pathways or networks can be divided into three major categories *viz.* signalling pathway, gene regulatory network, and metabolic pathway [5,8]. Signaling pathways or signal transduction networks are mainly defined as the transductions of the covalent modifications (phosphorylations, dephosphorylations, ubiquitination) and conformational changes of protein molecules upon external and internal stimuli (ligands, hormones, changes in protein concentration, ionic flux, etc.) in the cells [9,10]. Gene regulatory networks (GRN) are commonly defined as the intricate networks of gene regulatory elements (GRE), such as promoters, transcription factors, co-activators, co-repressors, rRNA, miRNA, etc., which play important role in the transcriptions and translations of the transcripts, post-translational modifications and sub-cellular localization of the synthesized proteins, etc. [11-13]. On the other hand, the metabolic pathways are generally defined as the series of chemical reactions in which small molecules or metabolites act as the substrates of several enzymes and undergo complex chemical modifications processes to optimize the requirements of mass and energy of biological cells/tissues [14].

Current applications of the analyses of different biochemical pathways involve identification of the important genes, enzymes, metabolites, and other cellular components in biomass production [15,16], cell divisions [17], cell migrations [18], etc.; revealing the missing components in chemical reaction cascades by genome comparisons and pathway alignments [19,20]; bridging the gaps between genome - proteome - phenome [17]; exploring the alternate and optimal routes for the synthesis of various cellular components [21]; discovering important component(s) in the pathway for novel drug target identifications [22]. The most important applications of biochemical pathway analyses has been observed in the area of the pathogenesis of severe diseases, such as cancer, immunological disorder, neurodegenerative disease, diabetes, cardiovascular disease, etc. [23-28].

1.1.1 Biochemical Pathways - a Conceptual Framework to Understand Different Human Diseases

The interdependencies between various key components of biological systems - cells, genes, proteins, mRNA, microRNA, metabolites, inorganic molecules, free radicals, etc. can be brought together into a single conceptual framework of the biochemical pathway or network. Hence, to decipher various complex biological phenomena, such as embryogenesis, organogenesis, and most importantly the aetiology of various diseases, understanding of such intricate process has been a long-standing goal of the researchers [29-33]. Besides, biochemical pathways are the cardinal in the field of life science research to understand how a living cell or tissue will react and develop in different environmental (e.g., radiation, stimulus, heat, temperature, etc.) and internal fluctuations (e.g., mutations, copy number changes, hyper- or hypo-methylation, chromosomal aberrations, etc.). Due to its high importance in cellular and organ development, study of different biochemical pathways or networks has also become indispensable to infer the causalities of genetic, epigenetic, and environmental changes with the malfunctions of various cells/tissues and the pathogenesis of various human diseases [34-38]. However, the

understanding of the relationships between genetic variations, biochemical pathways, and human diseases had begun after the British physician Sir Archibald Garrod in 1908 hypothesized that the inactivity or lack of specific enzymes in certain biochemical pathways are the main cause behind the human disease alkaptonuria [39]. After this study, several experiments focused on the deregulations of various metabolic pathways related to amino acid metabolism [40], urea cycle [41], glucose metabolism [42], and others have been discovered as the potential regulators behind various severe human diseases including cancers. Successively, the whole genome sequence of human and other organisms as well as high-throughput omics-based experiments have provided us a plethora of information about the genetic, molecular, cellular makeup of normal human cells/tissues and helped us to understand how certain deregulations in such profiles (i.e., transcription, methylation, structural changes) can trigger anomalies in the biochemical pathways and cause different diseases [43-46].

1.1.2 Interconnectivities of the Biochemical Networks and Disease Pathogenesis

However, it should be noted that even though the biochemical pathways are categorised into three separate groups, the interconnectivities and interdependencies of these pathways are very much essential in cellular programming and disease pathogenesis [47-49]. Most importantly, the pathways do not work alone and the components of each pathway may remain in multiple layers and have different time scales but they always work in concert to maintain optimal activities of the cell or tissue [50-52]. Often, during the pathogenesis of several human diseases, the regular patterns of connectivity observed in the molecular reaction network get altered and thus cause the irregular activities of the biochemical pathways. For example, hyperglycemia-induced glucose metabolism pathway including lipid perioxidation process and glutamate toxicity have been found to induce the increased activities of the apoptotic and VEGFR signalling pathways in retinal capillary endothelial cells of the patients' suffering in diabetic retinopathy (DR) [53]. Enrichment of HER2 (ERBB2)

receptor protein in the luminal and myoepithelial cell membrane over-activates the ERBB signaling pathway, which eventually disrupts the regular gene transcription networks and triggers the over-expressions of various oncogenes and simultaneously the down-regulation of several tumour suppressor genes in breast cancer cells [54]. On the other hand, the mutations in the genes (*IDH1* and *IDH2*) encoding the enzyme isocitrate dehydrogenase have been found to be associated with the synthesis of 2-hydroxyglutarate - a potential oncometabolite, which causes deregulations in the cellular metabolisms of glioblastoma brain tumour cells [55]. Mutations of *IDH* genes are also linked with hyper-methylation in Glioma cells, which eventually perturbs the normal gene transcription networks in the pathogenic glioma tumour cells [55]. Also, in several inflammatory/autoimmune diseases, it has been found that the stimulation of molecular reaction cascades via different serine/threonine protein kinase can over-activate the AMP-activated protein kinase (AMPK), which in turn can regulate its downstream metabolic pathways to trigger anti-inflammatory responses. Due to the presence of such strong coupling activity of AMPK signalling pathway and its downstream metabolic pathways, stimulation of AMPK has been exploited in the treatment of various chronic inflammatory diseases [56]. Apart from these, there are other numerous human diseases, in which the confluences of different types of biochemical pathways have been observed as the important governing factor [57,58]. Also, it has been observed that in various types of cancer cells the coupling of the molecular reaction networks is so strong that even the perturbations of the major cancer causing genes, proteins, or oncometabolites cannot stop the deregulations of the biochemical network [59-61]. Most often, the cancer cells can open or close several molecular pathways to minimize the selection pressure induced by the external inhibitors/drug molecules, and thus escape from the applied cancer treatment strategies [61].

Discrete studies, performed in the past, have proven that during carcinogenesis different molecular reaction cascades can interact with each other and trigger

abnormal behaviours (e.g., increased proliferations, nutrient uptakes, metastasis, and drug resistivity, etc.) of the cancer cells [62-64]. The genetic, epigenetic, and environmental changes induced in the cancer cells during carcinogenesis are intercepted by either signalling and/or gene regulatory and/or metabolic pathways [65]. The abnormalities impose in any of the network are then transferred to the other networks, and thus the entire cellular system becomes corrupted [61,66,67]. An elaborate description of the entire process is depicted in the following schematic diagram (**Figure 1**).

1.1.3 Biochemical Reaction Networks in Cancer Cells

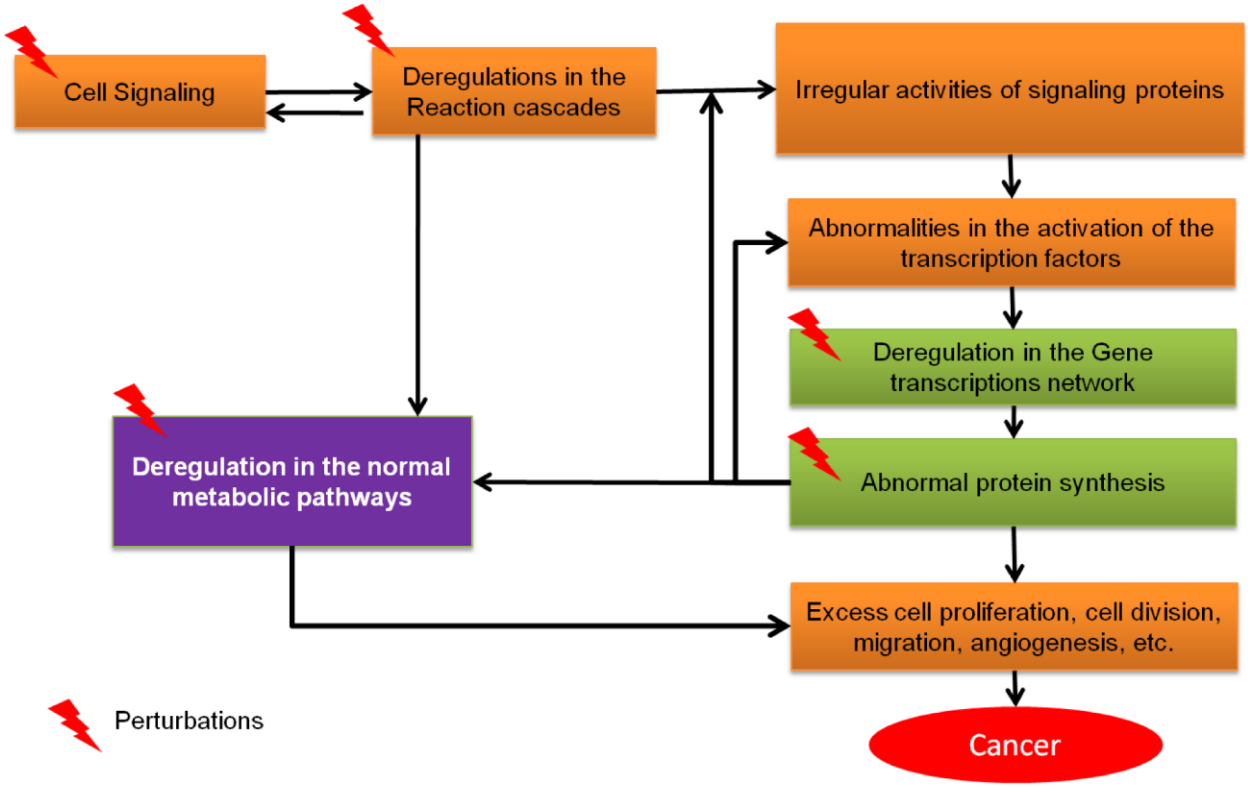

Figure 1: Coupling of different molecular interaction networks in cancer cells.
Here, the perturbations impose in the depicted molecular networks (i.e., signaling, gene transcription, and metabolic) can be due to the abnormal changes in genetic (mutations, copy number gain or deletion), epigenetic (DNA methylation, acetylation), and environmental factors (hypoxia, radiation, carcinogens).

In this schematic diagram (**Figure 1**), it is depicted that how different molecular reaction networks are connected with each other in a concerted manner within a cell, and how the external or internal perturbations caused due to genetic, epigenetic, and

environmental changes on any of the molecular reaction network can trigger the outcomes of cancer/tumour cells. It can be seen that the anomalies created in either of the cell signalling or gene transcription networks can affect both of its normal activities, which may eventually turn out to be the causes of excess cell proliferations and divisions, migration, angiogenesis, and so on. Also, the abnormalities incurred in the cells during carcinogenesis can also affect the normal metabolic pathways, which in turn can also trigger the rapid cell proliferations and other hallmarks of cancer cells. Most often, the oncogenic deregulations are imposed in the metabolic pathways due to the mutations, over-activations, or lack of activities of the enzymes responsible for metabolising the carcinogenic components consumed by the cells and tissues [68,69]. Sometimes, such deregulations in the metabolic pathways can initiate the synthesis of new metabolites, which can be carcinogenic in nature [55]. However, in both the cases, the main driving factor is the deregulations of the metabolic enzymes, which is mainly caused due to the induction of the external and internal perturbations on either of signalling or gene regulatory networks [70].

1.2 IMPORTANCE OF CELL SIGNALLING NETWORKS IN ONCOGENESIS

The abnormal activations of various cell signalling and its associated gene regulatory networks have been proven as the principal regulators of the predisposition of different human cancers [71-73]. However, despite encountering complexities at various levels, researchers have been able to discover unknown regulatory mechanisms of cell signalling and gene regulations system in cancer cells [74]. Characterizations of the subset of intracellular cell signaling networks, which are differentially activated in a specific type of cancer cells, have been proven helpful for acquiring the decipherable understandings of the mechanism of cancer cell development and the existing intra-tumor heterogeneity, analyzing the effects of tumor microenvironment on vascularization or angiogenesis, discovering the molecular mechanisms of drug resistivity and so on [75-78]. It has been observed that different subsets of the whole cell signal transduction networks are involved in the

outcomes of specific phenotypes [79]. Also, earlier studies have classified the cell signaling pathways in various categories, such as developmental, cell proliferations, cell migration, apoptosis, immunological, and so on [8]. Later studies have proven that deregulations in any of these pathways in the normal (matured or stem) cells can cause serious abnormalities in the regular cell division, apoptosis, migration processes, which in turn transform the cells towards oncogenesis [80,81]. For example, the gain-of-function mutation in Smoothened receptor protein of the developmental Hedgehog signaling (HH) pathway can over-activate its downstream effectors proteins (GLI1, GLI2, and GLI3), which are mainly the transcription factors of several cell cycle progression genes, such as *CYCLIN-D*, *CYCLIN-A*, etc. and cancer causing genes, such as *MYC*, *FOXM1*, etc. [82]. Over activations of HH signaling proteins and its downstream target oncogenes have been found in many cancers and tumors cells, such as glioblastoma, medulloblastoma, colon cancer, pancreatic cancer, basal cell carcinoma, etc. [82,83]. On the other hand, the cancer suppressor proteins or genes, such as P53, PTEN have been found in inactive or down-regulated state during oncogenesis [84]. Inactivation or the loss-of-function mutation of these genes most often deregulate the inhibitory effects from the activators of the developmental, cell proliferation pathways, and thus create positive effects in the rapidly proliferating cancer cells [85]. Simultaneously, there are various signaling pathways in the cancer cells, which can over-activate the expressions of various anti-apoptotic genes/proteins, such as BCL2, IAP, etc. and turn-off the normal cell death or apoptotic pathways [86,87]. Also, the signaling pathway, such as WNT pathway has been found to over-express the epithelial-to-mesenchymal transition gene *SNAI1*, which has been observed as the main regulator of cell migration and metastasis [88]. Other signaling pathways, stimulated by growth factor hormones, such as estrogen growth factor (EGF), insulin growth factor (IGF), etc. have been found to be associated with the activation of abnormal mitogenic or cell division process [89,90]. Signaling pathways also have the abilities to modify the epigenetic regulations in the cancer cells. For example, the TGFβ/SMAD signaling pathway has been found to

activate the Histone Deacetylases (HDAC) protein in Glioblastoma tumor cells and associated with the poor survival rate [91]. Epigenetic cascades initiated by the activation of HDAC proteins have profound effect in the regulations of the transcription of the genes related to other cellular activities, such as cell division, proliferations, metastasis, drug resistivity, etc. [92,93].

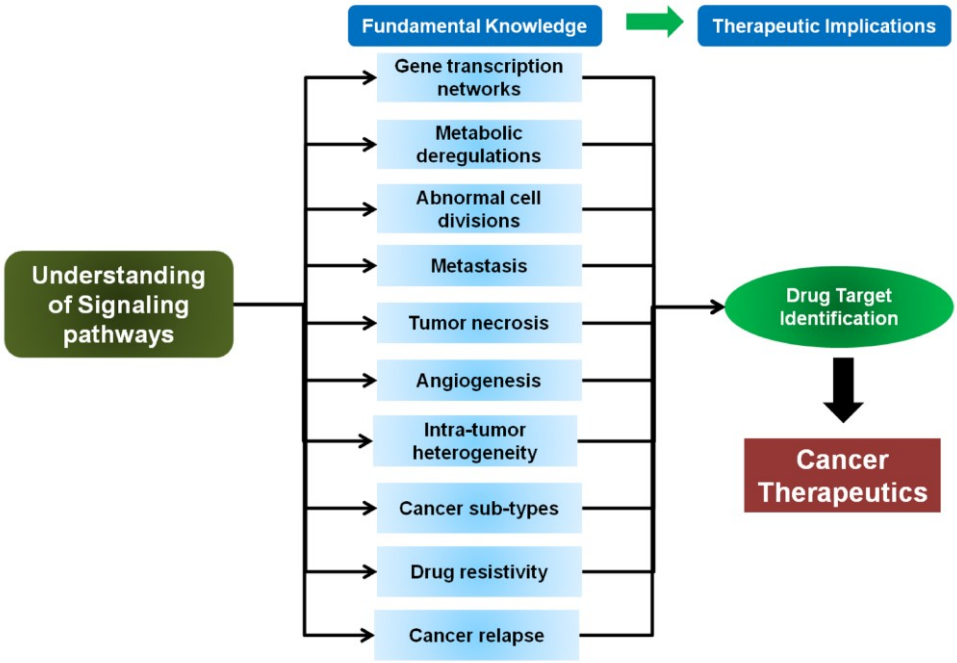

Figure 2: Significance of the understanding of cell signaling pathways in oncogenesis.

There are numerous other examples available in the literature, in which activation of signaling pathways have been observed to be associated with the abnormal gene regulations and faulty metabolisms in cancer cells [94-98]. Due to the higher importance in oncogenesis, targeting different cell signaling pathways has opened up various new possibilities in target-based cancer therapy (**Figure 2**) [99,100]. Different small molecule inhibitors or drugs, such as Vismodegib (inhibits Hedgehog signaling pathway protein Smoothened), Geftinib and erlotinib (EGFR pathway inhibitors), Gleevec or imatinib (BCR/ABL mediated pathway inhibitors), Bevacizumab (VEGF pathway inhibitors), Infliximab (TNF pathway inhibitors) etc. have been synthesized in the laboratories to target the growth of various cancer cells *in-vivo* [101]. Hence, the understandings of various cell signalling pathways are utmost important to not

only gain insights of the general mechanisms of oncogenesis, but also to discover novel drug-targets for the purpose of cancer therapeutics.

1.3 TARGETING SIGNALING PATHWAYS IN CANCER STEM CELLS

The rapid proliferation of cancer stem cells, which are most often difficult to eradicate by surgery, chemotherapy, and radiations therapies, can be neutralized by targeting different active cell signaling pathways [102,103]. It has been observed that due to the presence of various active developmental cell signaling cascades, such as Hedgehog, Notch, and WNT pathways, the cancer stem cell's (CSC's) can still maintain its self-renewal, tumorigenic progress, drug resistivity, etc. during and after the cancer therapies [104,105]. As a result, even after performing the successful surgery and chemotherapy of the tumor cells, there is always a higher probability of the recurrence of cancer cells from these drug resistant cancer stem cells in the future [106]. Hence, to deal with it, different small molecule inhibitors are synthesized, which can specifically target the developmental cell signaling pathways in cancer stem cells and block the active cell signaling cascades [107,108]. For example, the small molecule inhibitor PRI-724 has been tested as an effective modulator of WNT signaling pathway in colorectal cancer stem cells and it is currently in the clinical trials [109]. This inhibitor has been found to inhibit the CREB-binding protein, which is one of the important co-activator of the transcription factor β-catenin protein [109]. WNT signaling pathway, which activates the β-catenin protein as the transcription factor of various stem cell marker genes, thus get down-regulated in the rapidly proliferating cancer stem cells after the transcription activity of β-catenin is blocked by the suppression of its important co-activator protein CREB [110]. Similarly, the inhibitions of other developmental cell signaling pathways, such as Hedgehog, Notch, JAK/STAT, etc. have been also proven effective to suppress the uncontrolled proliferation of other cancer and tumor stem cells [99,111,112]. Targeting cancer stem cells to suppress the relapse of cancer cells can be more effective and precise as the signaling proteins responsible in these pathways have been observed to be explicitly

expressed in the cancer stem cells as compared to its surrounding differentiated normal cells [113,114]. The expression patterns of the receptor proteins of these signaling pathways are also different than its neighboring cells, which can be also seen as beneficial for small molecule inhibitors to target specifically the cancer stem cells [115]. Hence, the receptor proteins specifically expressed in cancer stem cells can act as the potential target biomarkers in the target-based cancer therapeutics [116].

1.4 CURRENT CHALLENGES OF STUDYING SIGNALING PATHWAYS IN TARGET-BASED CANCER THERAPEUTICS

Despite the success of suppressing cell signalling pathways in different types of cancer (matured and stem) cells, target-based cancer therapeutics still face difficulties at various levels [117]. It is still not clear to us that how the cell signaling pathways regulate the switching mechanism of a normal cell to cancerous states. Sometimes, the comprehensive information about the signaling molecules or the physicochemical nature of the molecular interactions is not available to understand the root causes behind the oncogenesis processes. Furthermore, complex pattern of the interconnectivities and interdependencies exist in signaling pathways have made the entire system more opaque to decipher. As a consequence, the researchers have been encountering serious problems to identify the main regulatory molecules as probable drug-targets in the future [117,118]. The following are the descriptions of few major challenges, which should be taken into the consideration in target-based cancer therapeutics.

1.4.1 Identification of Druggable Proteins in Target-Based Cancer Therapeutics

One of the most feasible targets in cancer cell signaling network is the enzymes or receptor proteins mainly responsible in oncogenesis [119,120]. In order to be considered as a potential drug-target for cancer therapeutics, the protein molecule must be druggable. There are various properties, a druggable protein must have. For example, the protein should be easily targetable by external drug molecule and biologically feasible to target by external drug molecules. It has been observed that

most of the receptor proteins and cytoplasmic enzymes (mostly kinase proteins) are biologically feasible to be considered as potential anti-cancer drug-targets. Also, the druggable proteins must have structural folds, which can favor the interaction with small molecule inhibitors. Besides, the target protein must possess a suitable binding pocket, which can be easily accessible to the external inhibitor molecules [121,122]. Last but not the least, the target protein should be highly important for the cancer cell signaling network, and targeting of which should be highly lethal to the cancer cells, but less toxic to the surrounding supporting cells and tissues [123,124]. The proteins, which have all these properties, are considered as the most ideal candidate for anti-cancer, target-based treatment strategy. However, the identification of such druggable proteins from the dense molecular signaling networks of cancer cells is sometimes the most difficult job to accomplish [121-124].

Although, the advancement of high throughput omics-based experiments, protein assays, microarray data etc. have made it possible to identify the active genes/proteins and the corresponding signaling pathways in specific cancer cells (**Figure 3**), but the identification of suitable drug-target(s) from the set of identified differentially active or inactive genes/proteins, which are involved in multiple signaling pathways and cellular activities, have been still considered as an unresolved problem [123,125].

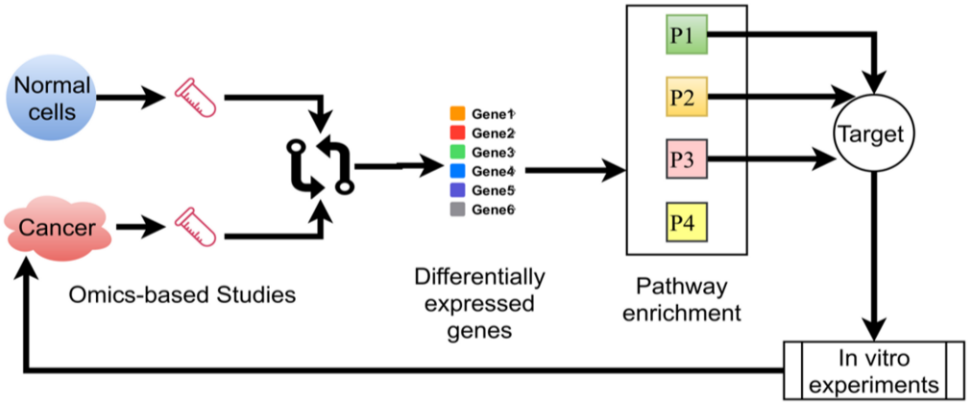

Figure 3: General experimental strategy to identify drug-targets in cancer cells.

The complex regulatory networks of gene expression, interconnections with

metabolic reaction networks, several feedback loops, etc. have made the entire signaling network very hard to decipher. The vast amounts of data originating at each level of the networks have also created several layers of complexities to interpret the working mechanisms of cell signaling network and identify the druggable drug-targets [126,127].

There are various examples, in which popular omics-based experimental strategies have been used to identify the novel drug targets in cancer cell signaling networks. The genomics study (using microarray and next-generation sequencing-by-synthesis techniques) performed on 24 different types of human pancreatic cancer cell lines by *Kinzler et al.* (2008) have shown that there are on an average 63 genetic mutations (mostly point mutations) associated with the pancreatic cancer cells and around 12 different cell signaling and regulatory networks are deregulated due to these genetic alterations [128]. From this study, the authors have been able to extract few important genes (e.g., *TP53, SMAD4, KRAS, APC, CREBBP*, etc.), which are significantly deregulated in the signaling pathways (e.g., KRAS signaling, TGF-Beta signaling, apoptotic pathway, WNT pathways, etc.) and could be used as future drug targets.

In another experiment, performed by the same research group using the same integrated genomics approaches on 22 human Glioblastoma multiforme (GBM) tumor samples, it has been observed that multiple candidate cancer genes (i.e., *TP53, MDM2, RB1, CDK4, CDKN2A*, etc.) from multiple signaling pathways, such as TP53 pathway, RB1 pathway, PI3K/PTEN pathway have shown altered expression levels in the GBM tumor cells [129]. In the successive experiments, several other research groups have also performed similar integrated genomics analyses on other types of cancer cells, such as colon, breast, basal cell cancers and identified the significantly expressed genes/proteins including the active signaling pathways in the cancer cells as potential targets [130-132]. However, these integrated genomics based approaches performed throughout the last few decades on different cancer or tumor cells have

proven that the progression and spread of cancer cells is not due to a mutation of a single gene/protein, rather it is the abnormalities caused by large number of factors (genes, proteins, miRNA, etc.) and pathways [131]. Out of the large number of identified active genes/proteins, the omics-based studies are not capable to precisely pin-point the proteins or pathways targeting of which would have been most destructive and lethal for the cancer cells. That is why the identification and ranking of most suitable druggable proteins from the differential gene expression studies have been still considered as a major challenging problem in the area of target-based cancer therapeutics.

1.4.2 Limitations of the Complete Suppression of Cancer Signaling Network

The active and dense signaling networks, which are often entangled with gene regulatory and metabolic networks in the cancer cells, are highly difficult to suppress by applying stress and external drug molecule [133]. There are multiple alternative routes, feedback mechanisms, alternate splicing variants of the signaling proteins, etc. exist in the signaling networks, which confer the entire system much more resilient against external and internal perturbations. For example, EGFR receptor mediated signaling pathway has been found active in many cancer cell types (e.g., glioblastoma, lung cancer) and it can trigger the transcriptions of various oncogenes. Being present in the cell membrane, the receptor protein EGFR have been considered as suitable candidate for many anti-cancer therapies. Drug molecules, such as Geftinib, Erlotinib have been synthesized to inhibit this protein in the lung cancer cells. Although, these drug molecules have shown promising results in the early stage of cancer therapy, in the later stages of therapy it has been found that the cancer cells may show resistivity against these drugs [134]. After a close inspection by dissecting the root molecular mechanisms in cancer cells, it has been revealed that the downstream kinase proteins and the main target transcription factor of EGFR pathway is not only activated by the EGFR receptor mediated pathway, but also by other receptor mediated signaling pathways, such as FGFR, FZD, C-KIT, PTPRF,

ITGB, etc. in the cancer cells (**Figure 4**) [135]. The alternate routes leading to the activation of the cell signaling pathways in cancer cells always have been a major problem in the complete suppression of oncogenic signaling networks in cancer cells.

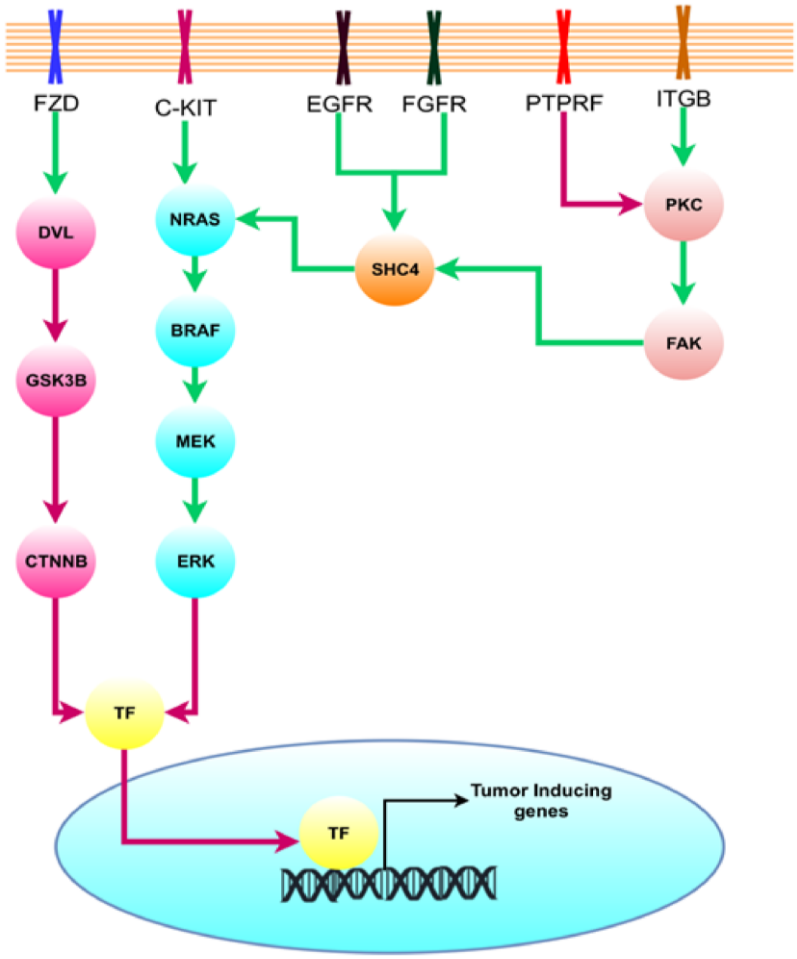

Figure 4: Alternate molecular routes mediated by multiple cell signaling pathways in cancer cells.
This schematic diagram clearly depicts the alternative routes through which the main target transcription factor (TF) of EGFR pathway can be activated, even in the absence of upstream signal from the EGFR receptor. Hence, if a cancer cell is targeted in the EGFR receptor, then the alternative pathways, such as WNT/FZD or C-KIT mediated signaling pathways can compensate the absence of the activation signal outgoing from the EGFR receptor. This alternate pathway activation theory also holds true while targeting other receptor proteins in the cancer cells.

1.4.3 Emergence of Drug Resistant Cancer Cells in Targeted Therapy

Apart from the identification of suitable druggable proteins and the remaining challenges of the complete suppressions of cancer signaling networks, emergence or relapse of drug-resistant cancer cells and cancer stem cells is also a major challenge in

the field of target-based anti-cancer therapies. There are various experimental evidences available in the literature in which the emergence of resistant cancer cells against popular anti-cancer drugs has been reported. For example, the small molecule inhibitor vismodegib, a known FDA approved drug for the anti-cancer therapy of basal cell carcinoma, had shown promising outcomes in targeting the cancer stem cells in which activation of Hedgehog (HH) signaling pathway has been reported [136]. The receptor protein Smoothened (SMO), the activator of HH pathway is the main target protein of the drug vismodegib. However, despite showing its promising results in the suppression of HH pathway in cancer cells, later studies have proven that a population of vismodegib-resistant cancer cells may relapse in the cancer patients, who have undergone vismodegib therapy [137]. It has been also observed that the receptor protein SMO can also have mutations either within its drug binding site or distantly apart from the active site, and thus the SMO variants becomes resistant against the drug vismodegib [137]. The mutations in the signaling proteins thus make the signaling pathways resistant against small molecule inhibitors and thus create more difficult challenges to the targeted therapy. Furthermore, there are various others factors in cancer cells, which have been shown to create serious challenges in front of the success of target-based anti-cancer therapies. Existence of the intra-tumor heterogeneity in the malignant tumor is one such problem.

It has been observed that the same type of tumor or cancer cells may have different molecular signatures and can have differential effects against target-based inhibition therapy [138]. A previous study performed on glioblastoma tumor cells using single-cell and bulk RNA-Seq analyses have revealed that GBM tumor cells consist of multiple heterogeneous tumor sub-clones, which are genetically different from each other. The RNA-Seq read counts mapped on these tumor sub-clones have revealed that different sub-set of tumor sub-clones can express different sub-sets of receptor proteins, such as EGFR, PDGFRA, FGFR1, ERBB2, ERBB3, KIT, FZD3,

NOTCH1, etc. [139]. This result also proves that the signaling pathways active in these genetically different sub-clones of GBM tumor cells have different active cell signaling pathways, and thus the transcriptional and proteosomal activities of these cells may also differ from each others. Hence, it can be also concluded that the target-based anti-GBM tumor therapy which has been developed only to target EGFR receptor protein may not be effective to eradicate completely all the tumor cells from the GBM tumor ecosystem. Thus, the tumor sub-clones expressing other developmental signaling pathways mediated by NOTCH, FZD3 receptor proteins can be remained unharmed even after performing anti-EGFR target-based therapy.

Moreover, the existence of surrounding normal cells and tissues in the tumor microenvironment can also pose challenges to the target-based cancer therapy [140]. However, as compare to the conventional anti-cancer therapies (chemo and radio), target-based anti-cancer therapy has less side effects and higher precision to kill the cancer cells. Although, the existence of similar types of normal cells with almost similar genetic signatures in tumor microenvironment can still create difficulties in target-based therapeutics. For example, the molecular and genetic characteristics of adult neural stem cells (aNSCs) and Glioma stem cells, both of which are present in sub-ventricular zone of human brain, have been found almost similar to each other [141,142]. Hence, finding the dissimilarities in almost similar gene/protein expression patterns in cancer cells should be the main priority to achieve high success rate in target-based anti-cancer therapies.

1.5 SIGNALING PATHWAYS AND ITS APPLICATIONS IN ONCOGENESIS

Before the completion of human genome project, the mutant genes/proteins which were discovered as responsible factors for driving the cancer cells had been strongly advocated as the potential drug-targets for the treatment of various types of cancers [143]. Literally speaking, it was a "gold rush" for the researchers to identify new genes or proteins, which could be readily targeted for suppressing the cancer

cell proliferation, metastasis, angiogenesis, and so on [143]. During this time period, researchers had used mostly the "reductionist approach" to understand the nature of oncogenesis and the involvement of various individual mutant genes/proteins [144]. Eventually, many research groups had started reporting the importance of several mutated genes/proteins behind the root causes of different types of cancers [145-147]. Simultaneously, individual cell signaling pathways and gene regulatory networks had been also identified and implicated in the cancer research as the conceptual framework to draw the connections between various types of genetic, epigenetic, environmental anomalies with the abnormal activities observed in the biochemical reaction networks of cancer cells [148,149].

After the successful completion of human genome project in the last decade of twentieth century, the researchers had started to understand the complex landscape of genes, proteins, miRNA, and other cellular components in the pathogenesis of various types of cancer cells [150]. The conceptual framework of signaling pathways, which were used only for visual inspection of comparing the flow of signaling cascades in the normal and cancer cells, had gained more importance in the study of holistic understanding of the cellular mechanisms. The advancements of high-throughput sequencing techniques, array-based and sequence-based omics studies, and others experimental techniques had started flooding large volume of omics data and soon different areas of life science research, especially cancer biology entered into the "fourth paradigm" of science, which is mainly driven and dominated by the large biochemical datasets [151]. With the accessibility of freely available large datasets, such as NCBI, EBI, and DDBJ, the researchers from other discipline of science had started to explore the complex pattern in the biochemical data more explicitly. Instead of using the data to develop empirical and hypothesis driven models of different biological phenomena, researchers had started to explore and use these data to understand the basic and natural laws of the biological world [152].

The main idea behind searching the new laws in the biological world by using the

big biochemical data was to gain the holistic view of various biological systems, such as cells, tissues, organs, and in fact the whole organism [153]. In order to analyze such big data, researchers have started implementing various tools and techniques of mathematics, physics, chemistry, and computer sciences to develop the *"theoretical/mathematical"* or *"computational"* or *"in-silico"* models of the biological systems. As a result various interesting properties of different biological systems have been identified by analyzing and simulating the *in-silico* models. For example, the seminal paper published by *Barabasi et al.* (2000) had shown that the organization of the metabolic networks is not only conserved in different organisms, but also follows similar scale-free topology, which is the main reason behind the robust and error-free properties of the biochemical networks [6]. In the later studies, performed on protein-protein interaction, gene co-expression, and signaling networks of cancer cells, have also proven that scale-free property may be an important factor of the increased robustness and fault-tolerance capacity of cancer cells against any external perturbations [22,154,155]. For example, the topological analyses of the protein-protein interaction network of Glioblastoma tumor cells have revealed that the constructed network follows scale-free property. After dividing the whole network into different sub-networks or modules, and enriching them with different biological functions, it has been revealed that two novel protein molecules, CSK21 and PP1A play significant roles to connect the two important sub-networks, which comprise of the proteins CDC2-PTEN-TOP2A-CAV1-P53 and CDC2-CAV1-RB-P53-PTEN, respectively [156]. The proteins involved in these two sub-networks are important for cell cycle regulations, and from the topological study, it has been identified that CSK21 and PP1A have significant role in the development of glioblastoma. The Real-time quantitative reverse transcription-PCR and immunohistochemical staining of these two proteins in glioblastoma tumor samples have revealed that these two proteins are significantly over-expressed in the GBM tumor cells, which in turn makes these two proteins as the suitable candidates for anti-GBM therapy.

Several other *in-silico* experiments of different types of molecular networks have also shown promising outputs in the identification of primary regulatory factors behind the pathogenesis of various types of tumors [157-160]. In the later studies, it has been also identified that other topological parameters, such as degree, clustering coefficient, Betweenness centrality, Closeness centrality, Eigen-vector centrality, modularity, etc. computed from the signaling, protein-protein interactions, gene co-expression, methylation co-occurrence networks, etc. have also higher importance in the discovery of crucial nodes or "*Hubs*" (genes, proteins, miRNA) in the networks responsible for cancer pathogenesis [161-163]. The success of implementing the network or topological analyses in the study of cancer and other disease pathogenesis and drug-target identifications studies have opened up a new direction to study the systems (i.e., cells, tissue, organs) as a whole rather than understanding it part by parts via reductionist approach. The concept of "*systems biology*" approach has been constantly gaining its momentum and has been applied to decipher various complex phenomena of human diseases including cancers.

It has been strongly argued by various scholars that system modeling by *in-silico* mathematical models is utmost necessary for understanding the complex phenomena of cancer pathology, identifying novel drug-targets, classifying the patients into different categories or sub-classes, or developing new strategies of personalized precision medicines in cancer therapeutics, etc. [164,165]. Several computational modeling techniques, which have been used to model and simulate the cell signaling and different other pathways in cancer cells so far, can be broadly categorized into three classes *viz.* (i) structure or topology based static model, (ii) discrete and continuous dynamic model, and (iii) rule based hybrid discrete and continuous dynamic models [166-169]. However, in each of the mathematical technique, there exist certain pros and cons [170]. Discussion of each of the modeling techniques is beyond the scope of this chapter, but in the subsequent chapters, the appropriate implications of these modeling techniques in cell signaling networks have been

documented. Various review articles have been written on these techniques, which can be useful to gain significant understandings of the *in-silico* modeling techniques, used in theoretical and computational cancer biology studies [171,172].

In short, it can be stated that the understanding of the complex phenomena of cancer pathogenesis, which in some cases have been proven difficult to analyze using the conventional experimental techniques, the *in-silico* computational models have been able to decipher those complex phenomena more conveniently, efficiently and explicitly [166]. Extraction of functionally important modules, identification of novel drug targets, discovery of alternate signal activation paths, etc. from the large-scale, dense cell signaling networks in cancer cells have been made possible by using topological or graph theoretical analyses [173,174]. The steady-state behaviors of several genes, proteins, and pathway species in cancer signaling networks have been analyzed and computed using semi-dynamic, discrete Boolean or logical modeling approaches [175]. On the other hand, to observe the transient and steady-state quantitative changes of the expressions or concentrations of pathway species in cancer cell signaling networks have been computed by implementing *in-silico* dynamic models of signaling and gene regulatory networks [172]. However, implementation of dynamic continuous model of signaling pathways has been found difficult to model large-scale signaling networks, especially in which the enzyme kinetic equations, rate parameter values are mostly unknown [176]. In such cases, application of semi-dynamic logical modeling technique has been proven useful as it does not depend on extensive details and information of the parameter values of the signaling network [177]. There are other mathematical techniques, such as flux-balance analyses (FBA), partial differential equations (PDE), agent-based modeling (ABM), etc. have been also implied in the understanding of several signaling, gene regulatory, and metabolic networks in cancer cells [176].

1.6 SCOPE AND SPECIFIC OBJECTIVES OF THE THESIS

Understanding of the role of various developmental and receptor tyrosine kinase (RTK) mediated cell signaling pathways with their associated gene regulatory mechanisms in cancer cells, especially the Glioblastoma tumor cells is the main focus of the present investigation. As mentioned in the **Sections 1.1.1** and **1.2**, different biochemical pathways including cell signaling networks are the important components to construct the conceptual framework for understanding the differences of the biochemical events in normal and cancer cells. It is also mentioned in these sections that the unresolved questions of cancer pathogenesis, such as the reasons behind the irregular cell divisions, progression towards metastasis, evolution of the intra-tumor heterogeneity, and drug resistivity and relapse of the cancer cells can be answered by deciphering the complex nature of signaling cascades and measuring its influences in the gene regulatory mechanisms in cancer cells. Therefore, reconstructions and analyses of the important signaling networks, such as Hedgehog, Notch, WNT, EGFR, and other related pathways followed by finding the differences of the chemical cascades between normal astrocytes cells (healthy control) and gliomagenic tumor cells bring the valuable information to decode the complex etiologies of gliomagenesis. The inhibitors of Hedgehog, Notch, and other pathways have been used to suppress the growth of Glioblastoma (GBM) tumor cells, but most often the tumor cells have been able to find out alternatives routes to eliminate the drug induced perturbations and evolve as drug resistant [178,179]. Hence, the important question which is asked in the current investigation is that how the signaling pathways in cancer cells adapt the stress induced by the external target-based drug molecules during the targeted therapy.

Besides, the GBM tumor cells depict higher level of inter-tumor and intra-tumor heterogeneities. Depending on the rate of proliferations and histopathological characteristics, the GBM tumor cells are classified into four major grades, *viz.* Grade-I, Grade-II, Grade-III, and Grade-IV. The Grade-I, II are considered as low-grade

GBM (LGG), and the Grade-III, IV are considered as high-grade primary GBM (HGG) [180]. In the population, such types of inter-tumor heterogeneities or different grades of GBM tumors are commonly observed. The groups of patients suffering in LGGs have shown high median survival year and better prognosis after therapy, whereas HGGs are more difficult to treat and have very less median survival year (less than 1.5 years) as well as depict poor rate of prognosis [181]. There are evidences in which it has been shown that the LGG tumor cells can sometimes progress and transform into HGG and are called as secondary GBM. Discrete experimental studies have been able to show the differences in the molecular characteristics between the GBM sub-types, but how such differences are appeared in the tumor sub-types, which are originating from the common Glioma stem cells (GSCs) in the brain tissue are hitherto unknown [182]. Even the tumor cells of HGG are also not same and possess higher level of molecular heterogeneities [183]. In the **Section 1.4.3**, it is mentioned that the existence of such intra-tumor heterogeneity of GBM cells poses serious challenges to the target-based anti-GBM therapy [183]. Indeed, if the molecular expression patterns and the corresponding signaling pathways significantly differ within the same sub-type of GBM cells, then finding the right molecular target or druggable protein to kill all the tumor cells becomes a challenging issue. It has been also mentioned in this section that the emergence of drug-resistant tumor cells after performing target-based anti-GBM tumor therapy is also occurred due to the presence of the appearances of intra-tumor heterogeneity [184]. Hence, in order to understand the involvements of various large-scale, non-linear cell signaling pathways and its crosstalk with various gene regulatory networks behind the emergence of inter-tumor and intra-tumor heterogeneities of GBM cells, visual inspections and comparisons of the reaction cascades in normal and tumor cells are not sufficient enough.

In **Section 1.5** it has been discussed that how the implementation of computational modeling of signaling pathways is beneficial to decode the patterns in

the signaling pathways in various biological systems including the cancer cells. In this present investigation, identification of the molecular reactions motifs in GBM tumor cell signaling networks, responsible for the emergence of inter- and intra-tumor heterogeneities have been targeted to accomplish. Apart from this, using *in-silico* mathematical models and simulations, the activities of various cell signaling pathways are modeled in GBM cells.

The main objective of this work is to use the knowledge of signal transduction mechanisms in the development of mathematical model and try to identify the alternative druggable proteins for suppressing the active oncogenic cell signaling pathways in GBM tumor cells. Besides, one of the major objectives of the thesis is to develop the computational frameworks, new algorithms, and simultaneously provide the processed datasets for the reconstructions, modeling, and simulations of signaling pathways, which can be further used for rigorous hypothesis testing and designing new experimental strategies to deal with various types of human diseases, such as cancer.

1.6.1 Outline of the Thesis

In **Chapter 2**, the materials and methods used in the present investigations for the curation of cell signaling pathways, modeling, and simulations of GBM and other cancer cells are elaborately discussed. In the **Sections 1.1.2** and **1.1.3**, it is mentioned that the signaling pathways and its associated gene regulatory machineries inside the cells (both normal and cancer) are strongly connected with each other. Hence, to consider the strong cross connections and interconnectivities of the cell signaling networks in the development of useful mathematical models of GBM cells, at first the comprehensive knowledge of the signaling pathways including all its all crosstalk reactions and the information about its target genes are required. This chapter has elaborately discussed about the procedures and resources required for pathway data curation and modeling. For simulation of such large-scale reconstructed signaling networks in tumor cells, the theoretical understandings and the formulae required

for computing the topological parameters of the networks are also provided here. To develop the mathematical models and simulations of the large-scale signaling networks, all the present investigations in the thesis have used graph theoretical and logical modeling approaches.

In **Chapter 3**, a comprehensive review of the existing challenges of pathway data curation from the current cell signaling databases are discussed [8]. The limitations of the computations of pathway related data in these databases are also discussed in this chapter [8]. Also, the comprehensive descriptions of our newly launched database-cum-pathway data analyses platform BIOPYDB, which has been specifically designed for resolving the current limitations of the existing databases, are also provided in this chapter. The database can be accessed by following the link: http://biopydb.ncl.res.in/biopydb

In **Chapter 4**, the descriptions of the observations obtained from the mathematical model developed for simulating the activities of Hedgehog signaling and its cross-talks with other cell signaling and gene regulatory mechanisms in GBM, Colon, and Pancreatic cancer cells are provided [82].

In **Chapter 5**, the simulation outcomes of active Notch signaling and its cross-talks with other cell signaling and gene regulatory mechanisms in GBM cells are described [185].

In **Chapter 6**, the descriptions and the outcomes of the mathematical models developed for the analyses and simulation of inter-tumor and intra-tumor heterogeneities of GBM tumor cells are discussed. In this chapter, it is shown that how different reaction modules in the reconstructed Notch signaling and its cross-talk networks play important roles in the development of both adult neural stem cells (aNSCs) and Glioblastoma tumorigenic stem cells (GSCs) in the sub-ventricular zone of human (SVZ) brain. The developmental dynamics of GSCs from the origin of aNSCs as well as the development of matured differentiated tumor cells or

astrocytomas from GSCs in human brain is simulated and further studied for the downstream analyses. A phenotype predictor score is proposed in this work, which is able to successfully compute the frequency distributions of the differentiated, proliferating, apoptotic, and quiescence cells, etc. in the normal and tumorigenic stem cells niche. In further downstream studies, it is shown that this new phenotype predictor scoring technique is particularly helpful to predict the future risk of the development of different grades of Glioblastoma (GBM) tumor in human brain. In this work, it is also shown that how different types of heterogeneous molecular clones of the same grade of tumor originate in the GBM tumor ecosystem. Also, a novel technique using Fast Fourier Transformation (FFT) analyses is described in this work, which is specifically helpful to pin-point the most crucial nodes (or drug-targets) in the dynamic or active signaling network in cancer cell.

In **Chapter 7**, the descriptions of a newly developed extended quaternary logic (ExQuLogic)-based simulation technique to model the cell signaling pathways in various types of cellular systems are provided. Also, the ExQuLogic-based simulation technique is applied on the developed integrated pathway model of Hedgehog, Notch, EGFR, WNT, PI3K-AKT, JAK/STAT, HIF1A cell signaling networks and subsequently the model outputs are studied for the analysis of the growth of GBM cells and identification of novel drug targets.

In **Chapter 8**, the conclusions and future directions of the thesis are provided.

Chapter 2

2 MATERIALS AND METHODS

2.1 AVAILABLE RESOURCES OF PATHWAY DATA

Different types of databases or resources *viz.* "Pathway databases," "Protein-Protein Interactions (PPI) databases," "Omics databases," "Disease databases," and most importantly the "Literature resources" were used for pathway reconstructions, modeling and simulation purposes (**Figure 5**). Around 24 different cell signaling databases were identified, which are currently at active states and distribute the pathway related data (e.g., pathway image, molecular interactions, molecule names, sequence, etc.) freely to the users (**Appendix Table 1**). The protein-protein interactions (PPI) databases, such as MINT, STRING, APID, PIPS, DIP, and BIOGRID were also found to be the useful resources for the collation of pathway data (**Appendix Table 1)**[1].

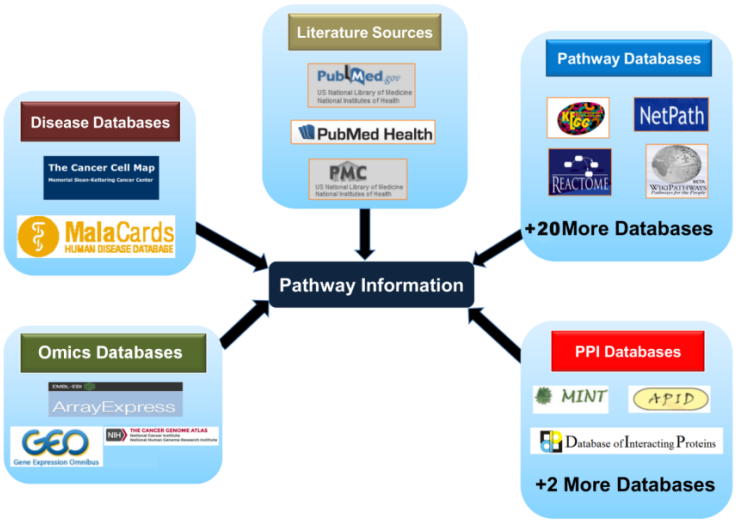

Figure 5: Available databases used for pathway data collation.

[1] The materials of this chapter have been taken verbatim from our previously published articles (a) Chowdhury et al., *PLoS ONE*, 2013 (b) Chowdhury and Sarkar, *Clin Exp Pharmacol*., 2013 (c) Chowdhury et al., *J Integr Bioinform*, 2018.

"The Cancer Genome Atlas (TCGA)" data portal was also used to study the next-generation sequencing-based transcriptomics (mRNA-Seq) data of Glioblastoma (GBM) patients. TCGA has studied a large group of GBM patients, which is divided into two cohorts, *viz.* Low-grade GBM (LGG) and High-grade GBM (HGG). The following are the statistics of the two patient cohorts (**Table 1**).

The raw counts of mRNA sequencing data are freely available through Genomics Data Commons (GDC) data portal (https://portal.gdc.cancer.gov/). There were three types of tumor cells considered in the mRNA sequencing analyses of the TCGA-LGG and TCGA-GBM patient cohorts, *viz.* (i) Primary tumor, (ii) Recurrent tumor, and (iii) Solid normal tumor cells.

Table 1: Statistics of Glioblastoma Patient Cohorts

Cohorts	Gender				mRNA-Seq Data			
	Male	Female	Unknown	Total	PTM	RTM	SNT	Total
Low Grade GBM	285	230	1	516	513	16	0	529
High Grade GBM	366	230	121	617	156	13	5	174

PTM = Primary tumor; RTM = Recurrent tumor; SNT = Solid normal tumor

A comprehensive review of the available signaling pathway databases, which was used for pathway data collation, is provided in *Chapter 3* (see **Section 3.3**).

2.2 DATA COLLATION PROCEDURES OF SIGNALING NETWORKS

The intra-cellular, biochemical interactions data available in Cell signaling and PPI databases (**Appendix Table 1**) were used for the reconstructions of signal transductions pathways. At first, the comprehensive maps of the signalling pathways were constructed by using the available pathway (or network) data provided in the cell signaling databases. All the 24 cell signaling databases were searched to compile the information (i.e., molecular species, chemical reactions, physical interactions, translocations, etc.) of a selected pathway (e.g. Hedgehog) and a master data table

was created. Followed this, the redundant molecular species and biochemical interactions from the non-redundant datasets were separated out. The redundant molecular interactions data collated from most of the databases were considered as the core reaction module of the pathway. After that, the non-redundant molecular interactions data was integrated into the core reaction module and thus a coherent network of the pathway of interest was reconstructed.

The pathway reactions which had at least one literature evidences were considered during pathway reconstructions. To expedite this process, PPI databases were also used to cross-verify the pathway species and reactions included in the reconstructed pathway diagrams. Thus, following these filtering processes, it was made sure that all the molecular species and interactions included in the reconstructed pathway are correct and do not have any ambiguity.

Furthermore, the molecular species and the interactions were annotated on the basis of their sub-cellular locations in the cells. In case of multiple locations observed for a single molecular species, the molecule in both the locations was indexed using the prefixes (e.g.," NUC_" for nuclear, "CYT_" for cytoplasmic locations, and so on) to distinguish it in the reconstructed pathway. At this step, all possible homologs of the protein molecules in human and their activities in the pathway were also searched. Thus, all the molecular species in the pathway were indexed according to their observed sub-cellular locations. After that, the connections (i.e. chemical reactions, physical interactions, translocations, etc.) between the molecular species within and outside of the sub-cellular locations were constructed. Pathway constructions by literature search had helped to find out the missing links within the reaction network, which were not addressed by the cell signaling pathway databases. The pathway reconstruction procedure was not restricted only in the core reaction module, but also spread to connect various cross-talks and feedback reactions involved with the core signaling and gene regulatory networks of the pathway. The transcribed genes and their encoded proteins of the reconstructed signaling

pathways were connected with various cellular phenotypes, diseases, and other signaling networks.

2.2.1 Reconstructions of the Signalling Networks

The reconstructions of the signaling networks, such as Hedgehog, Notch, WNT pathways including their crosstalk pathways and gene regulatory networks, mentioned in *Chapters 4-7* were performed by collating data from different signal transduction and PPI databases (**Appendix Table 1**). The pathway data were extracted from more than 500 research articles. *CellDesigner Ver. 4.2.* was used for pathway design and annotations of the molecular species of the signaling pathways [186]. The pathway species were also annotated according to their sub-cellular locations of the cell using standard notations available in CellDesigner. The annotation procedures mentioned in the previous section (**Section 2.2**) was followed during pathway reconstructions.

2.3 CONSTRUCTION OF PATHWAY KNOWLEDGEBASE

The open source database "BIOPYDB (**BIO**chemical **P**athwa**Y** **D**ata**B**ase)" was prepared to store and distribute the reconstructed pathways for the general research purposes. Following are the descriptions of the database architecture, which was used for constructing the database.

2.3.1 Database Architecture and Implementation

2.3.1.1 Pathway Nomenclature and Ontology

It was found that a consensus naming system with well defined vocabulary and a hierarchical tree of pathway based ontology will be of great importance for indexing, searching, analyzing the pathways systematically in the database. It was observed that the ontology tree defined by BioPortal for annotating the rat, mouse and human genes with different pathway terms/names would be most appropriate in this case [187].

Hence, the similar pathway ontology tree of BioPortal was used to index the signaling and diseases pathways into the database. New syntaxes and vocabularies were introduced for this purpose. Three different functional (Developmental, Immunological and Cell proliferation pathways) and one biological context dependent (Disease Pathways) nodes from BioPortal ontology tree were placed at the top of the hierarchy of the newly developed pathway ontology tree. This newly developed ontology tree further served the purpose of indexing the reconstructed signaling pathways in the database. These topmost level nodes were further linked with the child nodes. For example, under the node of Immune Signalling pathways, the cytokine signalling pathways constitute a separate sub-category (child node), which was further classified into distinct Interleukin families like IL-2 family, IL-10 family, etc. Each of these interleukin families were further assigned with different interleukin signalling cascades, which were having similar functionalities. More number of different categories in the subsequent updates of this database will be also included. To assign the names of the new pathways and make them distinct from each other within a same family and sub-family, the pathways were named according to either ligand (e.g. IL2, IL4) or the main receptor protein (e.g. TLR, T-cell receptor or TCR) molecule, which triggers that particular signalling cascade. The following vocabularies (a, b, c) were used for the nomenclature of the signalling pathways involved in various cellular functions (a, b) and disease pathogenesis (c):

a) <u>X</u> Ligand(s) Stimulated Signaling Pathway(s)

b) <u>Y</u> Receptor(s) Mediated Signaling Pathway(s)

c) Deregulated Signaling Pathways in <u>Z</u>

Where, X, Y, and Z represent the name of the Ligand or the family of the Ligands, Single or a family of Receptors molecule, and name of the Disease respectively.

If there were multiple ligands/receptors associated with the signalling pathway, then the logical operator 'AND' was used in the vocabularies ('a' and 'b') to separately mention the names of each of the ligands/receptor molecules. Introduction of such

detail classification and nomenclature system is useful for searching/browsing a specific class/family of pathway(s) in the database. It is also helpful to understand and compare the different pathways activated/stimulated under different extra-cellular stimuli. The information about the biological context specific pathways (e.g. the disease pathways) was also indexed here by using a controlled vocabulary (c), which could be further useful to differentiate the deregulated or malfunctioned pathways from its normal counterparts and their involvements into a particular disease. A flow-chart describing the entire process of implementing and processing the pathway nomenclature system in the BIOPYDB database is depicted in **Figure 6**.

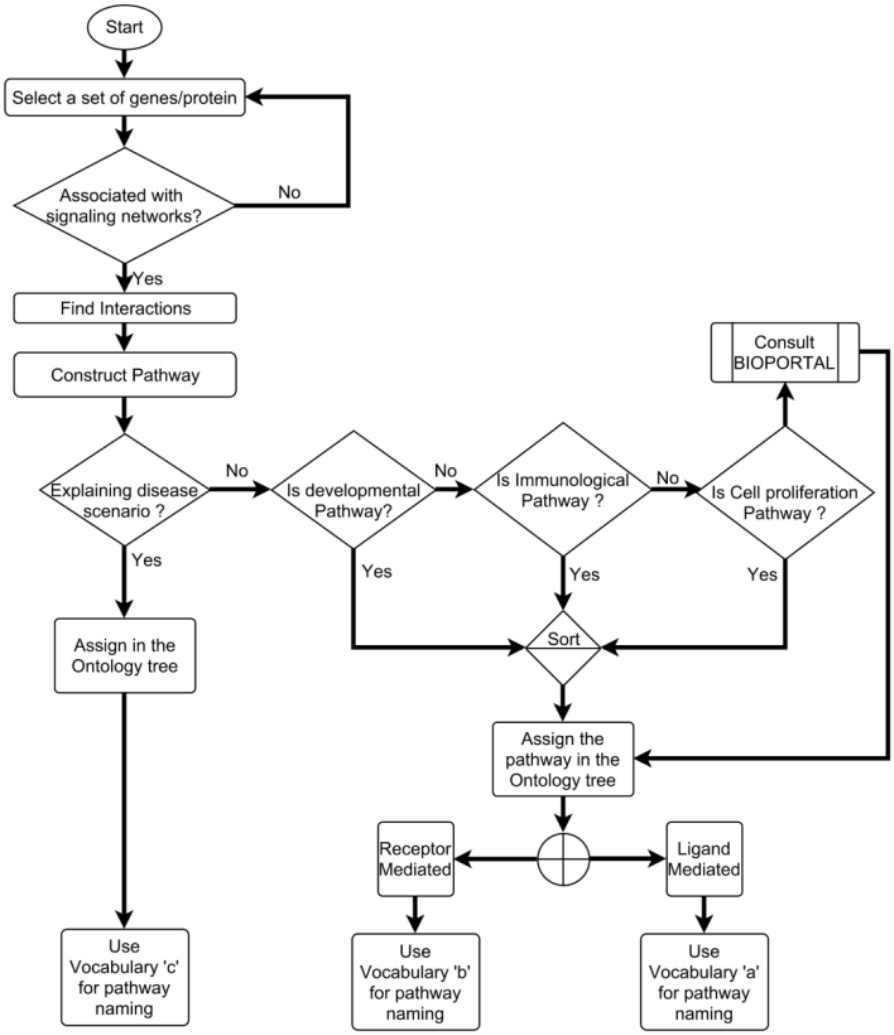

Figure 6: Flow-chart of pathway nomenclature system used in BIOPYDB database.
This flow chart was used for naming the pathways according to the vocabulary we introduced in our database. We also consulted BioPortal pathway ontology tree for making the pathway nomenclature system concordant with the international standard.

2.3.1.2 Assignment of Unique Pathway Identification Number

A standard protocol for assigning pathway identification number (PID) for each of the reconstructed pathways was also implemented in the database. Following are the descriptions of the entire protocols.

1. At first, the newly reconstructed pathway, which was decided to be included in the database, was allocated in the appropriate position in the BIOPYDB ontology tree. The ontology constituted the pathways in four main classes: *viz.* "Developmental", "Immunological", "Cell proliferation", and "Disease pathways" (see flow-chart of **Figure 6)**.

2. Followed by the selection of the classification of the new pathway, a unique number (or database ID) was assigned against the pathway ontology. For example, the ontologies, *viz.* "Developmental Pathway", "Immunological pathways", "Cell proliferation pathways", and "Disease pathways", were assigned by the IDs BIOPYDB1000, BIOPYDB2000, BIOPYDB3000, and BIOPYDB4000, respectively.

3. After that the following strategies to assign a new PID of our pathway of interest in the ontology tree was considered. An example is given bellow:

The Cytokine signalling, T-cell signal transduction, TLR mediated pathways are all come under Immunological pathways (ID: BIOPYDB2000)". Hence, these pathways were assigned as the "child" pathways under the PID of BIOPYDB2000. All the corresponding pathways associated under this PID were given new PIDs as BIOPYDB2001, BIOPYDB2002, and BIOPYDB2003, respectively. Moreover, there are many pathways, which come under the category of Cytokine signalling pathway (PID: BIOPYDB2001). Hence, the pathways under this category, such as IL1, IL2, Interferon pathways etc. were given new PIDs as BIOPYDB2001.1, BIOPYDB2001.2, and BIOPYDB2001.3, etc.

Furthermore, the signalling pathways activated by the IL2 family molecules were classified into the family of "IL2 Family Ligand Stimulated Signalling Pathways (ID:

BIOPYDB2001.2)". This family of pathway corresponds to the signaling pathways governed by the ligands IL2 and IL4. Hence, these child pathways under "IL2 Family Ligand Stimulated Signalling Pathways" were named as "IL2 Ligand Stimulated Signalling Pathways (ID: BIOPYDB2001.2.1)" and "IL4 Ligand Stimulated Signalling Pathways (ID: BIOPYDB2001.2.2)".

2.3.1.3 Processing of Pathway Molecule Table

Followed by the nomenclature and ontological assignment, the molecular species (i.e., proteins, genes, metabolites, RNA etc.) involved in the reconstructed pathways were categorized according to their molecular and chemical properties (e.g., protein, carbohydrate, gene, ion, mRNA/rRNA, etc.). Then, the molecule was indexed by its corresponding BIOPYDB Pathway ID in the database. An abbreviated name, which was used in the published literatures or UNIPROT database, was assigned to each of the protein molecule.

Furthermore, to show the molecular complexes formed by proteins and other molecules in the pathway, the following vocabulary was used.

*Let's consider the following reactions (i.e., **E1, E2,** and **E3**) from a signalling event in which the complexes are formed as products.*

$$E1: A + B = A : B \mid A := Ligand; B := Receptor; E1 \in Ligand\ vs\ Receptor\ Interaction$$

$$E2 : A : B + C = A : B : C \mid A : B = Complex; C := Molecule; E2 \in Hetero\,Complex\,Formation$$

$$E3 : M + M = M : M \mid M := Molecule; E3 \in Dimerization\,Pr\,ocess$$

Different types of complexes formed in the signalling cascades by various chemical reactions were indexed in the molecule list by using the vocabularies A:B; A:B:C, and M:M etc.

2.3.1.4 Hyper-Linking and Annotation of Pathway Molecules

Hyper-linking and annotation of each protein molecule of all the pathways in a

database is the most time consuming process, as both the processes require more manual interventions and efforts to execute. However, in BIOPYDB, this entire time-consuming process was minimized by utilizing an intelligent pathway update engine, which can take minimal pathway information as inputs and simultaneously build the pathway database automatically. A custom script was written for developing the pathway update engine, which runs at the backend of the database and automatically processes the hyperlinks of each of the molecular species with other popular databases (**Appendix Table 1**).

Moreover, each protein molecule of a pathway was annotated with the interactive access of PPI databases *viz.* PIPs [188] and STRING [189] to get the information of other protein molecules interacting with that particular protein of interest. Furthermore, each molecular species of the reconstructed pathways were automatically hyperlinked with the popular database GeneCards [190], which is known for providing the gene/protein-disease association data. Apart from these, the customized script running in the automated update engine can also annotate the protein molecules with the database TiGER [191] from which the tissue specific expression of the proteins/genes are freely available. The entire custom script for the automated update processes was written in PHP, Perl, and Python languages.

2.3.1.5 Processing of Molecular Interactions Table

The molecular interactions or reactions present in each pathway of BIOPYDB database were manually curated from the experimental data available in literatures and were successively stored in a master table called "Interactions" table. All the interactions of this master table were linked with the corresponding pathway ID assigned for each pathway in the database ontology tree. The corresponding literature reference of each interaction was hyper-linked with the associated Pubmed ID. The molecular and chemical properties (e.g., phosphorylation, ubiquitination, dimerization, etc.) of different types of biochemical reactions were tabulated with

each of the interaction in the 'Interactions' table. The sub-cellular location (i.e., extracellular space, membrane, cytoplasm, etc.) in which the molecular interactions were found evident was also tabulated in this table. To maintain the direction of the reaction cascades, each of the reactions was considered as the directed edges connecting a source molecule (i.e., reactant) with another target molecule (i.e., product). The molecular interactions of the reconstructed pathways were automatically hyper-linked with iRefIndex database [192], which provides a standardized indexing of the non-redundant molecular interactions found in the popular protein-protein and molecular interactions databases, such as BioGRID [193], IntAct [194], HPRD [195] etc. Mapping the interactions with the iRefIndex database is very much useful for the users to verify the interactions with other popular PPI databases.

2.3.1.6 Processing of Disease, Reference and Mathematical Model Tables

The protein molecules of the reconstructed signaling pathways were mapped with various diseases and thus a protein-disease network was reconstructed. The protein-disease mapping data was mainly curated from the external database - MalaCards: Human Disease Database [196]. The disease data was fetched from this database for each protein included in a particular BIOPYDB pathway ID. It should be noted that the protein-disease data provided in BIOPYDB is solely owned by the MalaCards database and BIOPYDB is providing only an advanced, automated interface to fetch and process the data in a more user-friendly process.

BIOPYDB also maintains a dedicated database of all the relevant references or published literature related to each pathway in the "Reference" table. The reference table was generated automatically by fetching the citation information from Pubmed database. The users can get this entire "Reference" list by accessing the pathway-browsing interface of BIOPYDB. Thus, the literature mining for getting the information of a particular pathway will be much easier to the experimental

biologists. Furthermore, to obtain the information of previously published mathematical models of different biochemical pathways, BIOPYDB had also made a database of the literature references related to the mathematical models (e.g., Graph theory, Boolean, Ordinary or Partial Differential equations, etc.) and tabulated in the "Model" table. The Pubmed id or DOI was also provided for each published mathematical model.

2.3.1.7 Processing of Pathway Image and Textual Data

BIOPYDB distributes the pathway images and textual data to the users in various file formats. Users can download the data without any restrictions and charges. However, for commercial uses, the users are required to take appropriate license from us. The entire processes of pathway data processing steps followed in the database are described in **Figure 7** for better clarification. The entire process can be broadly categorized into two categories: (i) "Pathway Image Processing" and (ii) "Textual Data Processing". Both were automatically processed from the BIOPYDB main data sources (i.e., Molecules list, Interactions list, etc.). The pathway images were processed in JPEG, SVG, and PNG formats. Two types of pathway images were produced *viz*. *Structured or Hierarchical* and *Unstructured or Network*. The Structured pathway image of a pathway was produced by allocating the pathway species in a hierarchical fashion, in which the ligands and receptor molecules (present in the extracellular and membrane regions) of a pathway were placed at the left (in left to right hierarchy) or top (in top to bottom hierarchy) in the pathway image. The pathway components found in the cytoplasm and in nucleus were placed accordingly in the subsequent hierarchical levels (i.e., Extracellular region, Membrane, Cytoplasm, Nucleus) respectively. Another hierarchical location 'Output' was introduced in the pathway image to signify the target genes/proteins produced at the end of the signaling cascades. All the pathway molecules belong to the hierarchical location 'Output' were placed at the rightmost or at the bottom of the *Structured or Hierarchical* image. Though this region does not indicate any physical

sub-cellular location, but to distinguish the pathway molecules from the target proteins, inclusion of such location in the pathway image will be very much useful for better simplification.

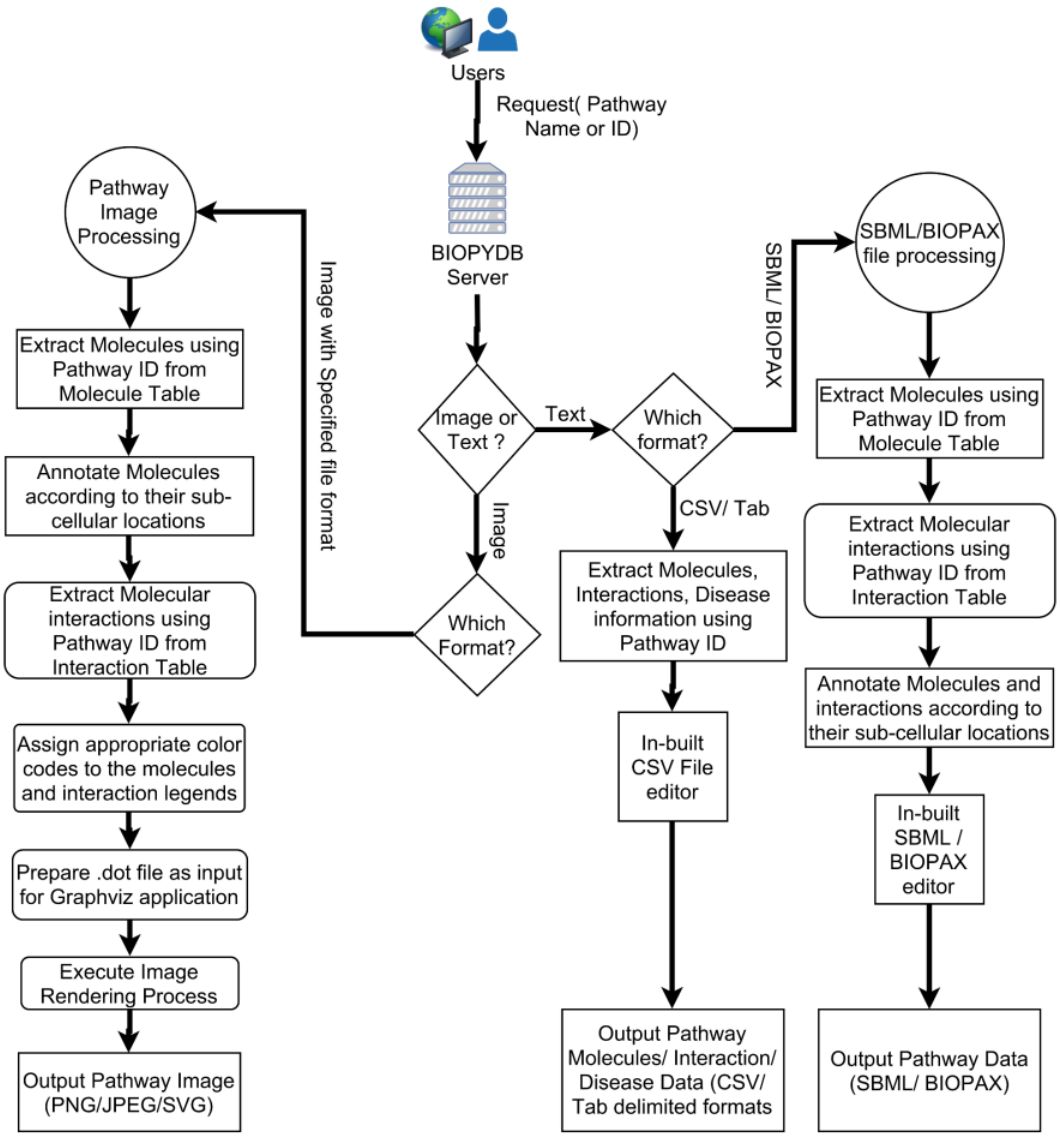

Figure 7: Flow-chart for pathway image and textual data preparation.
The pathway image and textual data provided to the users are processed in real-time upon user's request. Here, the idea is to reduce the effort of database developers to continuously update the pathway images and textual information after every update in the pathway database. Using this process, the users always get the updated pathway data and the developers do not need to manually compile and keep the pathway images and textual data in the database archive for storage and distribution.

The sub-cellular locations, required to maintain this hierarchy, were fetched from the "Molecules" table generated at the beginning of the pathway curation. On the

other hand, the connections/edges between two pathway molecules were fetched from the "Interactions" list. Different color codes and shapes were used to classify different types of molecular interactions enlisted in this table. Thus, an *in-silico* structured pathway image, resembling the internal cellular environment was reproduced automatically in the database interface. The unstructured pathway image (i.e., network) was also generated to analyze the topological properties of the biochemical signaling pathways. Such type of network is viewed as the *directed graph* in which the direction of the interaction between the two species of a pathway (i.e., Source and Target) was generated from the information provided in the "Interactions" table. Both of these pathway images (i.e., hierarchical and network) were generated automatically by using the popular graph drawing software: Graphviz [197] and the in-house script written in PHP.

On the other hand, the textual data (i.e., molecule names, interactions, disease names, etc.) of biochemical pathways were generated in BIOPYDB through an automated fashion by fetching the raw data from the database tables' *viz.* "Molecules", "Interactions", and "Diseases". The textual data for each pathway were processed in simple flat files i.e., either in TXT or CSV formats or in XML based SBML (Systems Biology Mark-up Language) and BIOPAX (Biological Pathway Exchange) formats [198,199]. SBML and BIOPAX files were computer readable, vocabulary based, widely used standard pathway sharing file formats, which were useful for pathway visualization and performing mathematical modeling. The dynamic engine running at the backend of BIOPYDB web services automatically processes these verities types of textual data. To generate runtime BIOPAX format of the biochemical pathways, we used open source, third-party tool Sig2BioPAX for converting the biochemical reactions into BioPAX level 3, OWL format [200]. Using such automated process of generating pathway images and textual data, the data in BIOPYDB can be readily updated and thus does not depend on any manual intervention by the database administrators and curators.

2.4 TOPOLOGICAL ANALYSES OF SIGNALING NETWORKS

The topological properties of the signaling networks were analyzed by considering the pathway as a directed graph or network. The pathway species (i.e. molecules) were considered as "nodes" and the interactions/reactions between a pair of molecules as "arcs". Following this assumption, the adjacency matrix (***A***) of the entire signaling network was constructed, whose elements (A_{ij}) were computed based on the following criteria ***(Eq. 1).***

$$A_{ij} = \begin{cases} 1, iff\ the\ node\ 'i'\ interacts\ with\ the\ node\ 'j' \\ 0, otherwise \end{cases}$$

(Eq. 1)

2.4.1 Network Parameters Calculated From the Adjacency Matrix

In-Degree (K_{in}): It refers the total number of nodes that are directly acting on a particular node in the network [82,201].

Out-Degree (K_{out}): The total number of interactions (activations or inhibitions) that are acting by a particular node on the other nodes in the network [82,202].

Degree (K_i): It refers the total number of in-degree and out-degree of a particular node [82,202]. The total degree (Ki) of a node ***"i"*** is calculated as shown in ***(Eq. 2)***.

$$K_i = K_{in} + K_{out}$$

(Eq. 2)

Eigenvector centrality: It refers that a node in a network will be more central if it is connected to many central nodes in the network [82,203]. According to Newman [203], the centrality x_i of a node ***"i"*** is directly proportional to the cumulative sum of the centralities of its neighbors: x_j.

$$x_i \propto \sum_{j=1}^{n} A_{ij} x_j \qquad \textit{(Eq. 3)}$$

Or,

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

(Eq. 4)

Where, $\frac{1}{\lambda}$ is the proportionality constant.

Now, if we consider ***(Eq. 4)*** as a vector equation, then we can write that,

$$\lambda x = A x$$

(*Eq. 5*)

Where, $x = (x_1, x_2, x_3, \ldots\ldots, x_n)$ is the Eigenvector of the adjacency matrix ***A*** having highest positive Eigen value λ.

Betweenness centrality: It is the ratio of the number of shortest paths that pass through the node to the total number of shortest paths of all the nodes to all the other nodes. It signifies that how a node is important in the shortest paths of all the other nodes of the network [82,204]. The following formula was used for calculating the Betweenness centrality (B_x) of a node (***x***) in the network ***(Eq. 6)***.

$$B_x = \frac{\sum_{x \neq y \neq z} \sigma_{yz}(x)}{\sigma_{yz}}$$

(***Eq. 6***)

Here, $\sigma_{yz}(x)$ represents the number of shortest paths passing through the node **(*"x"*)** from all other pairs of nodes *"**y**"* and *"**z**"*. Whereas, σ_{yz} represents the total number of shortest paths between the nodes *"**y**"* and *"**z**"*.

Closeness centrality: It is the inverse of the total number of shortest paths of a node to all the other nodes in the network. Therefore higher closeness centrality of a node implies the lower shortest paths to the all other nodes in the network [82,162]. The following formula was used for calculating the Betweenness centrality (C_x) of a node (***x***) in the network ***(Eq. 7)***.

$$C_x = \frac{1}{\sum_y d(y,x)}$$

(Eq. 7)

Shortest Path (L_{ij}): It refers the minimum number of intermediate links or connections that have to traverse from one node *"**i**"* to another node *"**j**"*. *Dijkstra's* shortest path algorithm was used to compute all possible shortest paths between every pair of nodes in the network [82,205].

2.4.2 Network Parameters of Hedgehog and Notch Signaling Pathways

The reconstructed pathways of Hedgehog and Notch signaling networks (discussed in *Chapters 4* and *5*) were converted into the adjacency matrices, which were then converted into *".net"* format (default input file format of the network manipulation and visualization software "Pajek" [206]). The constructed networks of Hedgehog or Notch pathways were directed graphs, which were analyzed using the software Pajek and Gephi 0.8.1 [207]. The graph theoretical parameters, such as Degree (IN-, OUT-, TOTAL-DEGREE), Eigenvector centrality were computed in Gephi. The network figures of Hedgehog and Notch signaling networks shown in *Chapter 4* and *Chapter 5* were also drawn in this software. The *.net* files of the two networks were also used in "igraph" (a software package of "R") to calculate "All pairs shortest paths" [208]. Pajek was also used to calculate the *Betweenness* and *Closeness centrality* scores. The Matlab (R2012b, The MathWorks) function *"SPY"* was used to visualize "all pairs shortest paths" matrix.

The topological analyses of the integrated signaling network discussed in *Chapter 7* were performed in the Cytoscape (version 3.6.1) software environment [209]. The clusters in the network were identified by EAGLE algorithm implemented Clusterviz app available in Cytoscape [210]. The cluster, which contains the nodes related to cell cycle progression and EMT was considered as "cell cycle progression and EMT" module. The Cytoscape app MCDS was used to extract the strongly connected

components (SCC) in the whole and the select modules of the network [211]. The identification of the alternate routes (or directed shortest paths) from the input nodes leading towards cell cycle progression and EMT were performed by using PESCA app in Cytoscape [212].

2.5 SEMI-DYNAMIC ANALYSES USING BOOLEAN LOGIC

Semi-dynamic analyses of signaling pathways were performed using logical modeling approach. After preparing the knowledgebase of each of the chemical reaction of the signaling network, the pair-wise binary molecular reactions were translated into logical hyper-graph. All the logical hyper-graphs of a reconstructed signaling network were comprised of universal *"AND"*, *"OR"*, and *"NOT"* gates. The "AND" gate was used for representing the interdependencies of two or more than upstream molecules (i.e. reactants) on the downstream molecules (i.e., products) in the signaling network. On the other hand, the "OR" gate was used if the upstream molecules or reactants of a signaling cascade were independent with each other and produce the downstream product. The enzymes or repressor molecules responsible for slowing down or inhibiting the molecular reactions in the network were represented by "NOT" gate. The implementations of universal logic gates to construct the logical hyper-graph from the binary reactions list were straight-forward as it only required the biological interpretations of the reaction events and did not depend on the enzyme kinetics, rate equations, reaction parameters, and initial concentration of the reactants. The following criteria were applied while translating the chemical reactions of the signaling network into logical equations.

$$Rule1: A(t+1) = B(t), \text{ if } B \equiv Reactant \ \& \ A \equiv Product$$
$$Rule\ 2: A(t+1) = B(t)\ AND\ C(t), \text{ if } B \ \& \ C \text{ both simultaneously require for the production of } A$$
$$Rule\ 3: A(t+1) = B(t)\ OR\ C(t), \text{ if } B \text{ or } C, \text{ any of the molecule can produce } A$$
$$Rule\ 4: A(t+1) = NOT\ D(t), \text{ if } D \text{ is the inhibitor of } A$$

(Eq. 8)

Using the logical rules defined in ***(Eq. 8),*** the chemical reactions were translated

into logical hyper-graphs. For example, in signaling network, if a single reactant molecule ***B*** was found responsible for the production of the product ***A***, then ***Rule 1*** of ***(Eq. 8)*** was applied for the synthesis of logical hyper-graph**.** Here, the ***Rule 1***, measures the logical activity of the product molecule ***A*** at time ***(t+1)*** depending on the logical state of the reactant molecule ***B*** at time ***"t"***. On the other hand, if the reactants ***B*** and ***C*** both were observed to be simultaneously required for the production of ***A***, then the logical rule: ***Rule 2*** (using universal logic gate ***"AND"***) was applied for the synthesis of logical hyper-graph of the corresponding reaction. Similarly, ***Rule 3*** was used for the chemical reactions in which the reactants (e.g., ***B*** and ***C***) were independent on each other and can produce the product ***A***. The ***Rule 4*** was used in the scenario, where a molecule (i.e., ***D***) was experimentally observed as an inhibitor of a product ***A*** in a chemical reaction.

Example 1:

In order to understand the logical model construction and simulation, an example using a toy reaction model is provided in ***Figure 8****. In this figure, a toy signaling network is considered, in which the nodes* ***A, B, C, D, E, F,*** *and* ***G*** *are connected by chemical reactions. Here, the molecules* ***A*** *and* ***B*** *are reacting with each other and forming another molecule* ***C****. After that* ***C*** *participates with the downstream reactants* ***D*** *and* ***F****, which further produces the molecules* ***E*** *and* ***G****. Also, there is a positive feedback loop from G to D and negative feedback loop from* ***E*** *to* ***F*** *in the reaction cascade. The entire reaction cascade now can be translated into a digital circuit diagram using universal* ***AND****,* ***OR*** *and* ***NOT*** *gates. After that the corresponding logical equations of the circuit are formulated.*

The truth table generated from the logical equations corresponding to the signaling network is shown in ***Figure 8****. To generate the truth table, here it was considered that the initial state of the node "A" as True (or 1). The tuple of the binary states (0 or 1) of all the nodes at initial time point (i.e., at "0th time-point") is called as the initial state (S_0= [1, 0, 0, 0, 0, 0, 0]) of the system. Theoretically, there are total 2^7 numbers of distinct initial states*

possible in the toy network shown in ***Figure 8****.*

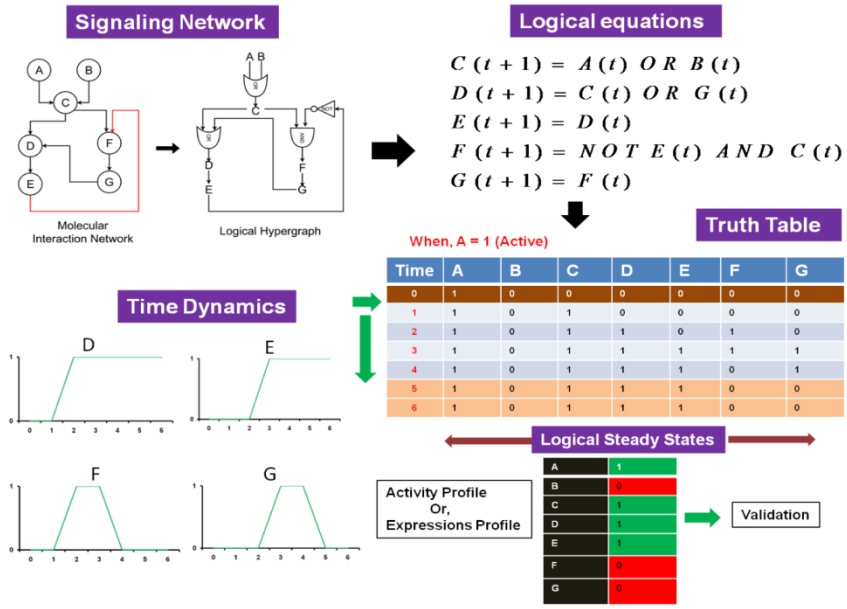

Figure 8: Workflow for translating signaling network to logical hyper-graph simulation. Here, a toy model of a molecular interaction network is translated into logical hyper-graph model, which is then translated into corresponding logical equations. The logical state of the node is either 0 (OFF) or 1 (ON). The truth table of the network is generated by considering A = 1 (ON or Active) state, and the logical states of other nodes are computed based on the considered logical equations. The computation is performed until the steady state is observed (here at time = 5) for all the molecules. The temporal dynamics of the molecules can be also analyzed as well as the logical expressions of all the molecules at the corresponding state scenario. Both temporal and steady state data can be validated either time course or static gene/protein expression profile.

However, to reduce the calculations in this toy example model, only one initial state is considered here, in which the node "A" is at active (or 1) state and the rest of the other nodes are at inactive (or 0) states. If the model assumes the initial state (S_0), then as per the asynchronous logical model update rule, the states of all the nodes in the toy network will be updated to the next time steps subsequently and reach the steady-state at Time = 5. At this time point (Time = 5), transitions all the nodes are stopped and in the next time step (Time = 6), the same states are repeated (S_5 = [1, 0, 1, 1, 1, 0, 0]). Therefore, the time step of five (Time = 5) can be considered as steady-state time point and the state S_5 as the steady-state. The state transitions path from S_0 --> S_1 --> S_2 --> S_3 --> S_4 --> S_5 is considered as state transition graph (STG). The full STG of the system can be constructed if all possible initial states (2^7) of the system are considered here. The steady-state(s) (S_5) of STG are called the "attractor" and the

other transient states (S_0, S_1, S_2, S_3, and S_4) are called as the "basin-of-attraction" [213]. There are two types of attractors found in STG, viz. "Fixed-point or stable focus" and "Cyclic or periodic attractors". In cyclic attractor, all the attractor states are repeated in definite time point with period of length greater than one.

*The temporal dynamics of all the proteins/nodes of the network can be also extracted from the transition states and can be plotted as time dynamics graph (**Figure 8**). The steady-states of all the proteins at S_5 (**Figure 8**) can be also extracted and validated with the experimental, high-throughput omics data (e.g., microarray, RNA-Seq, Mass-Spec, etc.). However, before validation with the experimental omics datasets, the expression values of the genes/proteins are required to be binarized (0 or 1). There are various statistical methods (e.g., t-statistics based differential expressions, k-means, BASC A, BASC B, Coarse-scale qualitative modeling approaches, etc.) available in the literature, which can be used for binarizing static and time-course microarrays [214-217].*

2.5.1 General Definitions Used in Logic Theory

Dependency matrix: This matrix defines the influence of a node *'i'* to another node *'j'* in the network. The matrix is also termed as influence matrix of the Boolean network, in which the influences (activation or inhibition effects) of any arbitrary node to all other nodes in the network are represented. The rows and columns of Dependency matrix (***D***) are the nodes of the network and the element of the matrix (D_{ij}) is calculated by using following criteria ***(Eq. 9***).

$$D_{ij} = \begin{cases} 0 & iff\ the\ node\ 'i'\ has\ no\ effect\ on\ node\ 'j' \\ 1 & else if\ 'i'\ directly\ activates\ 'j' \\ 2 & else if\ 'i'\ directly\ inhibits\ 'j' \\ 3 & else if\ 'i'\ indirectly\ activates\ 'j' \\ 4 & else if\ 'i'\ indrectly\ inhibits\ 'j' \\ 5 & else\ 'i'\ acts\ as\ both\ activator\ and\ inhibitor\ of\ 'j' \end{cases}$$

(*Eq. 9*)

The dependency matrix can be further used for calculating the total number of

upstream activators and inhibitors of an arbitrary node in the network after the system reaches at steady state level. The matrix can be also realized for determining the total number of downstream activated and inhibited molecules of an arbitrary node in the network after the system reaches at steady state level. The following definitions were used for determining these parameters from dependency matrix (***D***).

Upstream activator species of a node (X): The upstream activator species of any arbitrary node is defined as the total number of direct and indirect upstream activators of that node ***(Eq. 10)***. The upstream activator of node ***'i'*** is calculated from the dependency matrix (***D***) defined in **(*Eq. 9*)**.

$$Upstream\ activator\ species\ of\ a\ node\ (X) = \sum_{j=1}^{N=Total\ Nodes} [(D_{jX} \equiv 1) + (D_{jX} \equiv 3)] \qquad \textbf{(Eq. 10)}$$

Downstream species activated by a node (X): The total numbers downstream pathway species activated by any arbitrary node are calculated as the sum of total number species or molecules directly and indirectly activated in the downstream of that node. Following equation is used for calculating this parameter in a Boolean network ***(Eq. 11).***

$$Downstream\ species\ activated\ by\ a\ node\ (X) = \sum_{j=1}^{N=Total\ Nodes} [(D_{Xj} \equiv 1) + (D_{Xj} \equiv 3)]$$

(Eq. 11)

Upstream inhibitor species of a node (X): The upstream inhibitor species of any arbitrary node is defined as the total number of direct and indirect upstream inhibitors of that node. Following equation is used for calculating this parameter in Boolean network ***(Eq. 12)***.

$$Upstream\ inhibitor\ species\ of\ a\ node\ (X) = \sum_{j=1}^{N=Total\ Nodes} [(D_{jX} \equiv 2) + (D_{jX} \equiv 4)]$$

(Eq. 12)

Downstream species activated by a node (X): The total numbers downstream pathway species inhibited by any arbitrary node are calculated as the sum of total

number species or molecules directly and indirectly inhibited in the downstream of that node. Following equation is used for calculating this parameter in a Boolean network ***(Eq. 13)***.

$$Downstream\ species\ inhibited\ by\ a\ node\ (X) = \sum_{j=1}^{N=Total\ Nodes} [(D_{Xj} \equiv 2) + (D_{Xj} \equiv 4)]$$

(Eq. 13)

2.5.2 Logical Steady-state Analyses (LSS) of Hedgehog Pathway

In the logical formulations of the reconstructed Hedgehog pathway (*Chapter 4*), all the pathway species or proteins (Ligands, Receptors, kinase or Transcription factor) and cellular responses (Cell Proliferation, Cell cycle progression, WNT Pathway) were considered as "*nodes*". Their logical states can be either be "0" (OFF) or "1" (ON), which means depending on the cellular function and/or location, the proteins may be active (ON) or inactive (OFF), respectively. The entire simulation of Boolean models was performed in CellNetAnalyzer [126] and the following steps were followed during the logical simulation.

i) Selection of input and output proteins. In order to construct the logical model for Hedgehog signaling network, at first the input and output nodes were at first considered. Mainly the proteins, which did not have upstream connection in the reconstructed pathway map, were considered as input proteins. Similarly, the proteins, which were the downstream effectors of input proteins, were considered as output proteins in the logical model. Though there were few exceptions in the model considered while choosing the initial states of the proteins for three cancer scenarios (Glioblastoma, Colon and Pancreatic cancers) with respect to the normal pathway scenario. The three ligands of HH pathway SHH, IHH, and DHH were considered as time invariant (i.e., fixed logical state) and their logical states were kept at constant active state as "1" or "ON" during the simulation of cancer scenarios. But in normal scenario, their expression or logical state were considered as time variant, which were dependent on the logical states of the upstream

activators (i.e., BMP_RUNX3, CDO, BOC, etc.). The logical states of the input proteins were considered from various literature sources (**Appendix Table 6**), EBI-ArrayExpress Atlas, and also from various signaling and cancer databases (**Appendix Table 1**).

ii) Construction of Boolean or logical equations. The Boolean equations of all the nodes of Hedgehog signaling network were formulated after consulting the biological evidences of each reactions from various literature sources (**Appendix Table 5**). There are total 96 Boolean equations and 63 nodes including cellular responses presented in the model. The detail information about model formulation procedure is provided in **Section 2.5**.

iii) Simulation. The protocols for model simulations were described in ***Example 1*** of **Section 2.5** [126,218]. In order to create cancer the specific model of Hedgehog pathway , the logical states (‘‘0’’ as ‘‘OFF’’ or ‘‘1’’ as ‘‘ON’’) of the select proteins, which were found to be associated with cancer pathogenesis, were kept time-invariant during the simulations. ‘Gain-of-function’ (‘‘ON’’ or ‘‘1’’) states of the oncogenic proteins like RAS, ERK12, TWIST etc., and the ‘‘loss of functions’’ (‘‘OFF’’ or ‘‘0’’) of few tumor suppressor proteins, like GAS1, SUFU, NUMB, SNO were considered while creating different cancer scenarios in the logical simulation (**Appendix Table 6**). Also, in order to simulate the scenarios in temporal space, the time scales were accounted into the Boolean equations written in the master model (**Appendix Table 5)**. It is known that in a signaling network a few reactions are specifically dependent and can be only executed after the executions of other reactions. Such scenarios were considered in the model by assigning ‘‘Time Scale’’ of every equation. In the developed model of HH pathway, the logical equations representing the productions of GLI1, PTCH1 and HHIP proteins from the transcription factors and their translocations to the specific sub-cellular locations were assigned in "time scale 2".

The developed logical model was also simulated using both the synchronous and asynchronous updates schemes available in CellNetAnalyzer. However, it should be mentioned that after attaining the logical steady states, it was observed that the simulation results of both the update schemes were showing same dynamics. Hence, in the current investigation of the simulation of Hedgehog pathway, the logical steady states obtained in synchronous update scheme were only shown and used for further downstream analyses.

2.5.3 Simulation of the Logical Model of Notch Signaling Pathway

The master model of the logical equations of Notch pathway (*Chapter 5*) was developed by following the similar protocols used in Hedgehog pathway modeling (**Appendix Table 7)**. The constructed master model was simulated in CellNetAnalyzer to simulate the normal activities of Notch pathway. The initial states of the input proteins used for the simulation is provided in **Appendix Table 8** [126]. In normal Notch pathway scenario (NNS), only the reactions of the core Notch pathway was simulated, in which the Notch pathway ligands were considered as active (i.e., ON/TRUE). The non-canonical and cross-talk reactions in this scenario were not activated in the simulation of NNS. On the other hand, the expressions of the mRNAs of the proteins found to be differentially expressed (up or down regulated) in the transcripts of Grade-IV tumor cells derived from GBM patients were considered to create the Glioblastoma Scenario (GBS) [219].

In order to create the GAMMA-SECRETASE inhibition scenario (GSI), at first one of the important component proteins (PRESENILIN1) of this enzyme complex in the NNS scenario was suppressed (or OFF) by making its logical state as time-invariant. This scenario was created to observe the effect of GSI in the canonical Notch pathway. It was observed that the inhibition is effective to suppress the target genes/proteins of Notch pathway. This scenario was denoted as the GSI sensitive cellular model. However, the main objective of this current investigation was to

create the GSI resistant scenario, in which the suppression of GAMMA-SECRETASE protein was found ineffective to inhibit Notch target genes of the GBM cells in the previous reports. Hence, to create that scenario, all the cross-talks and non-canonical Notch pathway were made constitutively active in the GBS scenario and then the GAMMA-SECRETASE enzyme was suppressed by targeting one of its component proteins PRESENILIN1. This method was successful to create the GSI resistant scenario for GBM cells as the inhibition of GAMMA-SECRETASE enzyme was compensated due to the over-activations of cross-talks and non-canonical pathway.

In order to create the drug treatment scenarios (TS), random attacks (single and double knock-outs) on the signaling proteins in the GSI resistant scenario were performed. The two target proteins (TS1 and TS2) with maximum suppression effects on Notch target genes were identified. The total number of upstream activators and downstream activated proteins were also got reduced by these potential targets in the GSI resistant GBM scenario. The detail information of the calculation of total number of upstream activators and the downstream activated nodes of each protein in the Notch pathway under different simulation scenarios are provided in **Section 2.5.1**.

2.5.4 Methodologies used in the Simulation of Neurogenesis and Gliomagenesis

The entire methodological procedures used in *Chapter 6* to simulate the normal neurogenesis and tumorigenic development starting from adult neural stem cells (aNSCs) are described in this section. To understand the neurogenesis of aNSCs in sub-ventricular zone (SVZ) of human brain, Notch signaling pathway and its cross-talk reactions with other pathway molecules (i.e., JAK2/STAT3, HIF1A, P53, RAS, PI3K/AKT, WDR12, and JIP1) were considered to construct a comprehensive logical dynamic model (**Appendix Table 9**). The preliminary data for constructing the core pathway model was taken from the logical dynamic model of Notch pathway discussed in *Chapter 5* (**Appendix Table 7**) [185]. Based on this pathway model, the

entire logical model was then modified and restructured to simulate the developmental dynamics of adult neural stem cells (aNSCs) and the mutated Glioblastoma stem cells (GSCs). The logical rules defining the dependencies of the pathway molecules were derived from related literatures and the chemical regulations of the constructed signaling network.

The dynamics of the pathway molecules were updated synchronously and the following rules ***(Eq. 14)*** & ***(Eq. 15)*** were considered while constructing the dynamic Boolean model of Notch and its cross talks reactions were considered in this work. The initial states (1/ON or 0/OFF) of the input nodes were chosen randomly by generating random number for each input node from uniform distribution $(Unif(0,1))$ in the range of 0 to 1 and setting the cutoff at 0.5 ***(Eq. 15)***.

$$X_i(t+1) = X_i(0) \quad iff \ \forall X_i \in Input\ Nodes$$
$$= f_i(X_j^{1}(t), X_j^{2}(t), X_j^{3}(t), \ldots, X_j^{k_i}(t)) \quad iff \ \forall X^{k_i}{}_j \in Intermediate\ and\ Output\ Nodes$$

(Eq. 14)

$$With, initial\ conditions: X_i(0) = 1 \quad iff\ Unif(0,1) \geq 0.5$$
$$= 0 \quad otherwise$$
$$Where, k_i = Total\ number\ of\ Input\ Nodes\ of\ node\ X_i$$

(Eq. 15)

The pathway consists of 117 molecules out of which 53 were the input molecules (**Appendix Table 10**). Here, input molecules refer the molecules that do not possess any predecessor molecules (i.e., no upstream regulators) and during the signaling event they can be at either up-regulated (ON) or down-regulated (OFF) state ***(Eq. 15)***. Hence, the possible highest numbers of expression patterns of these input molecules were 2^{53} (~9.0071993 X 10^{15}). In presence of such enormous number of possibilities, mainly occurred due to the variations in the expressions of several extrinsic and intrinsic molecules, it is indeed obvious that adult and inactive NSCs (aNSCs) in the neurogenic niche of SVZ have to make right decision to opt any of the cellular states/phenotypes either by proliferation and differentiation during the developmental process. The aNSCs can also choose to stop its cell division process

(i.e., undergo cell cycle arrest or quiescent state) or undergo natural cell death process (i.e., apoptosis).

Hence, to capture such huge possibilities at steady-state, a robust simulation technique to stimulate the developmental dynamics of aNSCs with all input states (~ 9.0071993 X 10^{15}) was required. However, in practical scenario, logical dynamic simulations by considering all the input conditions and finding all possible attractors at equilibrium state were not feasible in terms of required computational cost. Hence, to reduce the computational cost, a fraction (total 10^6) random, non-redundant initial input sequences were generated by using uniform random number distribution and subsequently assigned binary states ON (1) or OFF (0) to each input node of the reconstructed model ***(Eq. 14)***. The entire random input sequences (10^6) were then further divided into 100 separate simulation batches (10,000 random sequences) and simulations are performed for all the 100 batches. Boolean functions shown in ***(Eq. 14)*** for each node/molecule are provided in **Appendix Table 9**. All the nodes considered in the logical model were further classified into four sub-classes: ***(i) Inputs, (ii) Intermediates, (iii) Targets, and (iv) Phenotypes***. A detailed description of all the nodes included in the model is given in **Appendix Table 10**.

2.5.5 Marker Proteins of aNSC, GSC and GBM Models

The marker proteins whose expressions were analyzed to observe the dynamics of different cellular states or phenotypes in the aNSC, GSC and GBM models were enlisted, with appropriate references, in **Table 2**. The expressions of these marker proteins are denoted in the model simulation by discrete binary states **0** or **1**. The occurrences of a particular phenotypic/cellular state in the model are mapped with a Boolean function (f) of the corresponding marker proteins (M) of that phenotype (P) and therefore its expression values (or activity patterns) will be varied in the discrete binary domain of $\{0, 1\}$.

$$f : M \rightarrow P$$

$$f := Boolean\ Function$$

$$P := Phenotypes \in \{p_1, p_2, p_3, \ldots\} \in \{0,1\}$$

$$M := Marker\ Proteins \in \{m_1, m_2, m_3 \ldots\} \in \{0,1\}$$

2.5.6 Determination of the Phenotypes and Cellular States

There were total five phenotypic functions considered in the developed dynamic model, which were dependent on the expression of specific marker proteins of normal and tumorigenic brain cells. These five phenotypes were i) Apoptosis, ii) NSC Renewal, iii) NPC Differentiation, iv) ASPC Differentiation, and v) GBM Development. Depending on the distribution of input signals, different marker proteins were expressed with different distribution, which in turn would regulate the expressions of different phenotype in the binary domain $\{1, 0\}$. Hence, theoretically, it can be considered that in total 2^5 or 32 single and combinations of phenotypes were possible to be occurred with equal probability in the attractor distribution space.

Table 2: Information of the marker proteins mapped with different phenotypes

Cellular Phenotypes	Marker Proteins	References
Apoptosis	Pro-apoptotic: PUMA, NOX, BAD, BAX	[220,221]
	Anti-apoptotic: FLIP, IAP, BCL2	
Neural Stem Cells Renewal	CYCLIN-D3, CYCLIN-D1, CDK, HES1, HES5	[222-224]
Neural Progenitor Cells	NESTIN, NEUROD, β-TUBULIN [225]-III	[226]
Astrocyte Progenitor Cells	GFAP	[227]
Glioblastoma Stem Cells	CYCLIN-D3, CYCLIN-D1, CDK, HES1, HES5, FLIP, IAP, BCL2	[228]
Glioblastoma tumor Cells	CYCLIN-D3, CYCLIN-D1, CDK, C-MYC, TENASCIN-C, GFAP	[229,230]

The probability to reach any of the phenotype would be random and unbiased if the logical model was purely random. Also, in case of a random logical network model, it can also be assumed that the flow of signal transduction within the signaling network would be purely unbiased and can lead the reaction cascades towards the development of any phenotype under a set of specific intra and extra-

cellular stimuli. However, in reality, the topology of the constructed logical model was not completely random and also unbiased in nature as the logical rules defined for the Notch signaling network was mainly taken from experimental evidences. The normal functioning of Notch signaling network was specifically oriented towards the neural stem cell renewal and its differentiation into neural progenitor cells [231]. Similar to the experimental findings, it was also observed that at the time of Notch pathway simulation, only a few phenotypic states occurred in the attractor space and those states were denoted as "Cellular states". A cellular state can be an individual phenotype (such as Apoptosis) or a combination of multiple phenotypes (e.g., NSC Renewal/NPC Differentiation). The following attributes were also considered during the model simulation to define two typical cellular states: Quiescent cells and GSC Renewal.

$$Cellular\ States := Quiescent(t) = 1 \ \ iff \ \forall \ Phenotypes : P_i(t) \in \{0\}; where, i = 1, 2, \ldots, 5$$

$$GSC\ Renewal(t) = NSC\ Renewal(t)\ and\ not\ Apoptosis(t)$$

Quiescent state was a distinct cellular state, which was found at ON state only when all the phenotypes were at OFF state. In this state all other cellular activities, such as proliferation, differentiation, and apoptosis were found at dormant stage or at lower rate [232]. Similarly, GSC renewal will be at ON state if the phenotype Apoptosis was at OFF state and the "NSC Renewal" state was at ON state.

It can be also possible that the attractor states obtained in the simulation studies at LSS can be either "Fixed-point" or "Periodic". In the fixed-point attractor state, the binary expressions of the cellular state and the pathway molecules will show homeostatic behavior, whereas in the periodic state it will be oscillatory in nature.

2.5.7 Calculation of Normalized Frequencies of Cellular States, Shannon Entropy and Activity Ratio (AR) Scores

In the earlier example (***Example 1*** of **Section 2.5**), there are total 7 (seven) nodes in the toy Boolean network (**Figure 8**), and out of which 2 (two) nodes (***A*** and ***B***) have

no upstream connector nodes. Hence, these two nodes are considered as inputs in the network, which can have total $2^2 = 4$ initial conditions (A=0, B=0; A=0, B=1; A=1, B=0; and A=1, B=1). Therefore, starting from all of these initial conditions, it was easier to observe the steady-state expressions of the other nodes in the system and generate the state transition graph (STG) of the system.

However, the entire calculation becomes complicated for a system with large number of input nodes. For example, if a Boolean network consists of total 50 input nodes, then the entire input space will be $2^{50} = \sim 1.1258999 \text{ X } 10^{15}$), which is practically impossible to simulate with the limited computational resource. Hence, to reduce the computational cost and time, a heuristic search algorithm (simulation based) is used here to find the maximum number of attractors of this logical model. The basic criterion of this algorithm is to run the simulation by taking random samples of non-redundant, finite set of initial states (say $N = 10^6$) from the overall population (~ $1.1258999 \text{ X } 10^{15}$) of input states.

Similar strategy was also used in this study. The random initial states were further divided into 100 separate batches (i.e. each batch will contain 10,000 random input sequences) and the frequencies of all the attractor states (or cellular states) were calculated for each batch of simulation. The frequencies of each cellular state were further normalized by dividing its total frequency observed in a simulation batch with respect to the total number of random initial conditions (i.e., 10,000) used for simulation. It should be noted that degeneracy of the observed attractor states with respect to a particular cellular state is possible, which means multiple sets of similar attractor states can be mapped with a particular cellular state depending on the expression state(s) of the marker protein(s) in those attractor distribution $\Pi(A)$.

Let us consider that such attractor distributions were calculated from the initial input states distribution $\Pi(I)$ in any of the simulation batch. Hence, it can be said that the distribution of the cellular states $\Pi(C)$ is also dependent on the distribution

of the attractor states $\Pi(A)$ and can be mapped with Boolean function $f : \Pi(A) \rightarrow \Pi(C)$.

Let us also consider that in an arbitrary batch of simulation (B_i), there exists a set of cellular states $C_i = \{C_i^1, C_i^2, C_i^3, ..., C_i^m\}$. Here, '$i$' is the simulation batch number and 'm' is the total number of observed cellular states or the size of maximum information content possible to be observed in the attractor distribution $\Pi(A)$ generated in the simulation draw B_i.

If the probability mass functions (i.e., information) of all the elements of the sets of cellular states (C_i) drawn from the simulation batches (B = 100) were taken into the consideration, then according to the "principle of maximum entropy" the probability mass function with highest information entropy will be chosen as the "proper one" for further data analyses [233]. The information entropy of a given batch of simulation can be calculated by calculating the Shannon entropy score of the observed cellular states.

Definition 1: The Shannon entropy H(C_i) of the i^{th} instance of a simulation containing the distribution of cellular states $C_i = \{C_i^1, C_i^2, C_i^3, ..., C_i^m\}$ ***is the negative logarithm of the probability mass function of the observed cellular states in that instance of simulation (Eq. 16).***

$$Hence,\ Shannon\ Entropy = H(C_i) = E[-\ln(\mathrm{P}(C_i))]$$
$$= -\sum_{m=1}^{m} \mathrm{P}(C_i^m).\log_2 \mathrm{P}(C_i^m)$$

(Eq. 16)

Theorem 1: If $\Pi(A_1), \Pi(A_2), \Pi(A_3), ..., \Pi(A_n)$ ***are the attractor distributions generated from initial input states*** $\Pi(I_1), \Pi(I_2), \Pi(I_3), ..., \Pi(I_n)$ ***and their corresponding normalized frequency distributions of cellular states are*** $\Pi(C_1), \Pi(C_2), \Pi(C_3), ..., \Pi(C_n)$***, then the simulation instance containing highest Shannon entropy score H(C_i) will posses maximum numbers of different types of***

cellular states $C_i = \{C_i^1, C_i^2, C_i^3, ..., C_i^m\}$.

Proof: Let us assume that *i=1, 2, 3,...,n* are the different instances of Boolean simulations performed under a set of equal number of initial expressions vectors (say *N*), in which the expressions of the elements (i.e. input molecules) are randomly distributed. Let us consider the distribution of the random initial expressions sequences is $\Pi(I_i)$. Let us also consider that each of the simulation instances is producing a distribution of attractor states $\Pi(A_i)$, which can be mapped with a distribution of cellular states $f : \Pi(A_i) \rightarrow \Pi(C_i)$. Here, $C_i = \{C_i^1, C_i^2, C_i^3, ..., C_i^m\}$ is the normalized frequency of the individual cellular states (total numbers *m*) observed in the attractor distribution $\Pi(A_i)$ of i^{th} simulation instance. It could be possible that the total number of individual cellular states *(m)* observed in a particular simulation instance may not be equal for all the simulation instances.

Let us assume that the probabilities of each of the cellular states in a given simulation instances are $P(C_i^1), P(C_i^2), P(C_i^3), ..., P(C_i^m)$.

If there are $p^1, p^2, p^3, ..., p^z \in \{0, 1\}$ numbers of fixed phenotypes considered in the model, which are mapped with the cellular states $C_i = \{C_i^1, C_i^2, C_i^3, ..., C_i^m\}$, then theoretically there exist maximum $M \ (= 2^z)$ numbers of cellular states and the maximum Shannon entropy score by ***definition 1*** will be $Max\left(H(C)\right) = -\sum_{m=1}^{M} P(C^m).\log_2 P(C^m)$.

The maximum Shannon entropy of any arbitrary simulation instance *'i'* is,

$$\text{Max}\,(H(C_i)) = -\sum_{m=1}^{M} P(C_i^m).\log_2 P(C_i^m).$$

$$Here, -P(C_i^m).\log_2 P(C_i^m) \geq 0$$

$$Since, \quad \log_2 P(C_i^m) \le 0 \quad \forall \quad P(C_i^m) > 0$$

$$\therefore For\ any\ arbitrary\ number\ of\ cellular\ states\ 'm',$$

$$\therefore \quad -\sum_{m=1}^{m} P(C_i^m).\log_2 P(C_i^m) \ge 0$$

$$Therefore, when\ m \to \max(m),$$

$$-\sum_{m=1}^{\max(m)} P(C_i^m).\log_2\{P(C_i^m)\} \le Max\,(H(C_i)).$$

Hence, it is proven that at the level of maximum Shannon entropy score, there exists maximum number of distinct cellular states (i.e., *max(m)*).

Activity Ratio Score: This novel scoring technique is a metric to determine the contribution of an arbitrary pathway molecule (X_i) in the cell-signaling network to drive the cellular dynamics towards a particular cellular state. This scoring technique is specifically useful for extracting the important proteins from the set of input proteins, which are helping the signaling cascade to reach the specific phenotype. In the pathway simulation, given a set of total random input sequences (N), if S_k is the total number of random input sequences which direct the model simulation towards a particular cellular state (C_i^m), then the activity ratio of any arbitrary input molecule (X_i) will be calculated by using the following equations ***(Eq. 17)*** & ***(Eq. 18)***.

$$Activity\ Ratio = (AR)_{X_i \in X | C^j \in C} = Log_2\left(e^{\gamma_{X_i}}\right)_{X_i \in X | C^j \in C}$$

(Eq. 17)

$$(\gamma)_{X_i \in X | C_j \in C} = \frac{1}{|S_k|}\left\lceil \sum_{l=1}^{|S_k|}(Y_l : X_i = 1) - \sum_{l=1}^{|S_k|}(Y_l : X_i = 0)\right\rceil_{X_i \in X | C_j \in C}$$

(Eq. 18)

$$Let\ us\ assume, \quad C := Set\ of\ Cellular\ States = \{C^1, C^2, C^3, \ldots, C^m\}$$

$$m = Total\ Number\ Phenotypes\ observed$$

$$X_i \in X := Set\ of\ Input\ Molecules = \{X_1, X_2, X_3, \ldots, X_{N1}\}$$

$$N1 = Total\ Input\ proteins$$

$$Y := Set\ of\ Input\ sequences = \{Y_1, Y_2, Y_3, \ldots, Y_{|S_k|}\}\ reaching\ to\ cellular\ state\ C^j$$

$$Where, |S_k| = Total\ number\ of\ input\ sequences\ reaching\ to\ a\ particular\ celllar\ states\ out\ of\ N\ random\ sequences$$

$$i.e.\ |S_k| \subseteq N$$

Here, the proportion $(\gamma)_{X_i \in X | C_j \in C}$ of an input protein (X_i) in a given cellular state (C_i) is defined as the ratio of the difference of the number of times the input protein is up-regulated and down-regulated to the total number random input sequences directed towards a specific cellular state in the simulation.

Lemma 1: In an arbitrary simulation instance, started from a finite set of random initial conditions, the value of AR score of any input node is bounded in the range of -1.44 to +1.44.

Proof: Let us assume that in an arbitrary simulation instance there are $|S_k|$ numbers of input sequences $(Y = \{Y_1, Y_2, Y_3, \ldots, Y_{|S_k|}\})$ out of N random sequences generating the cellular state C_i.

Hence, it can be written as,

$$\sum_{l=1}^{|S_k|} (Y_l : X_i = 1) + \sum_{l=1}^{|S_k|} (Y_l : X_i = 0) = |S_k|$$

$$Now,$$

$$\text{if}, \sum_{l=1}^{|S_k|} (Y_l : X_i = 0) \to 0, \sum_{l=1}^{|S_k|} (Y_l : X_i = 1) \leq |S_k|$$

$$\Rightarrow \frac{1}{|S_k|} \left(\sum_{l=1}^{|S_k|} (Y_l : X_i = 1) \right) \leq 1$$

$$else, \sum_{l=1}^{|S_k|} (Y_l : X_i = 1) \to 0, \sum_{l=1}^{|S_k|} (Y_l : X_i = 0) \leq |S_k|$$

$$\Rightarrow \frac{1}{|S_k|} \left(\sum_{l=1}^{|S_k|} (Y_l : X_i = 0) \right) \leq 1$$

From ***(Eq. 17)*** it can be written that,

$$\therefore (\gamma)_{X_i \in X | C_j \in C} = \frac{1}{|S_k|} \left\lceil \sum_{l=1}^{|S_k|} (Y_l : X_i = 1) - 0 \right\rceil \leq 1$$

Or else, we can write that,

$$(\gamma)_{X_i \in X | C_j \in C} = \frac{1}{|S_k|} \left\lceil 0 - \sum_{l=1}^{|S_k|} (Y_l : X_i = 0) \right\rceil$$

$$\Rightarrow -(\gamma)_{X_i \in X | C_j \in C} = \frac{1}{|S_k|}\left\lceil \sum_{l=1}^{|S_k|} (Y_l : X_i = 0) \right\rceil \leq 1$$

$$\Rightarrow -1 \leq (\gamma)_{X_i \in X | C_j \in C}$$

$$\therefore -1 \leq (\gamma)_{X_i \in X | C_j \in C} \leq +1$$

$$and \quad -1.44 \leq (AR)_{X_i \in X | C^j \in C} = Log_2\left(e^{\gamma_{X_i}}\right)_{X_i \in X | C^j \in C} \leq +1.44$$

In the input list, if a particular protein is essential (activator) for a particular cellular state, then its expression would be constitutively 1 in the set of all random sequences S_K. In that case, the numerator of the proportion $(\gamma)_{X_i \in X | C_j \in C}$ will be equal to $|S_k|$ and maximum and thus the value of the proportion will be equal to +1. Similarly, for a protein, which is inhibitor (or negative inducer) of a particular cellular state, it can be shown that the proportion will be equal to -1. The Log2 of the exponential of this proportion value is simply taken as scaling factor to stretch the distributions of the scores in the region of -1.44 to +1.44. This scaling is specifically useful for gaining high resolutions among the activity patterns of a large set of input proteins in the development of a particular cellular state. If AR score of a protein is either +1.44 or -1.44 in a given cellular state, then it can be said that the specific protein is highly essential and positive or negative inducer for that particular cellular state.

2.5.8 Calculation of Phenotype Cost Function

The emergence route of intra-tumor heterogeneity and the sub-clones of GBM tumor cells from a common origin of mutated cells can be best analyzed with the help of tumor cell evolution phylogeny [234]. If the overall expressions of all the pathway molecules at t^{th} time are considered as a 'state' of the cell $Z_t = \{M_t^1, M_t^2, M_t^3, ...M_t^m : m = Total\ molecules\}$, then the entire developmental routes through which different states $\{Z_t, Z_{t+1}, ...Z_{t+T} : T = Total\ simulation\ time\}$ reach at the attractor states (singleton or periodic) are called as 'State Transition Graph' or STG [235]. The nodes of a STG are the states and edges are directed which defines the transition of one state to another state at every Boolean update. Hence, State Transition graph is able to depict the expression dynamics of all the pathway

components and the transformation of the cells towards different cell types, starting from a common origin of tumor initiating cells (states). It is observed that for reaching at a particular attractor state, STG has to pass through few transition/intermediate states or nodes $\{S_t, S_{t+1}, S_{t+2}, ...\}$ and then land into either fixed-point node $\{S_{t+s}\}$ or cyclic attractor nodes $\{L_{t+1}, L_{t+2}, L_{t+3}, ...L_{t+l}\}$. It is considered that the transition from one state to another state of a cell is also associated with a signaling cost function (i.e. for chemical reactions/any physical processes) and it affects the overall state transition rate. Hence, it can be also assumed that the probability of a particular attractor/cellular state (C_i) in equilibrium state starting from initial states depends on the overall distance (i.e. transition steps) and the total number of molecular transitions to reach at that particular state. Hence, to quantify the average signaling cost functions (Φ_{C_i}) in each simulation batch for each cellular state is considered to be the functions of total transitions steps and periodicity (in case of cyclic attractor), and the total number of molecular alterations performed starting from initial state to the attractor state. This function is defined in ***(Eq. 19)***.

$$\Phi_{C_i} = \frac{\phi(C_i)}{P_{C_i}}$$

(Eq. 19)

$$Where,\ \phi(C_i) = \frac{\sum_{t=0}^{s} d_{S_t \to S_{t+1}}}{s} + \frac{\sum_{t=0}^{l} d_{L_t \to L_{t+1}}}{l}$$

$Where, s = Total\ number\ of\ transition\ states\ require\ to\ reach\ at\ the\ steady\ state;$

$l = 1 \quad for\ Singleton\ attractor\ \&\ \sum_{t=0}^{l} d_{L_t \to L_{t+1}} = 0$

$> 1\ for\ Periodic\ attractor\ (=\ Periodicities\ of\ the\ cyclic\ attractor)$

$\mu = Total\ mutational\ costs\ of\ the\ entire\ model$

$d_{S_t \to S_{t+1}}\ \&\ d_{L_t \to L_{t+1}} = Hamming\ distance\ between\ two\ states\ in\ the\ transient\ and\ periodic\ paths$

$P_{C_i} = Total\ number\ of\ occurrences\ of\ C_i\ in\ a\ simulation\ batch$

Calculation of "Hamming distance" between two successive nodes (or cellular states) in STG is particularly interesting to know the temporal changes or evolution

of the expression dynamics of the pathway molecules in the successive time points [236]. Let us assume that $S_t = \{X^1_t, X^2_t, X^3_t \ldots\ldots, X^N_t\}$ & $S_{t+1} = \{X^1_{t+1}, X^2_{t+1}, X^3_{t+1} \ldots\ldots, X^N_{t+1}\}$ are the expression vectors referring to the any arbitrary successive states in STG. The expression of the elements (X^k_j) lies in binary domain $\{0, 1\}$. The Hamming distance $(\left|d_{S_j \to S_{j+1}}\right|)$ can be calculated by using the following ***(Eq. 20)***.

$$\left|d_{S_j \to S_{j+1}}\right| = \sum_{k=1}^{N} (X^k_{j+1} - X^k_j)$$

(Eq. 20)

Moreover it is also assumed that, apart from the signaling cost a mutated model in which few molecules are constitutively activated or suppressed will also induce mutational cost (Ω) in the cells at the time of its transitions. The mutational cost function for a specific model is dependent on the total number of induced mutations (μ) and the change in the total number of added and omitted cellular states in the mutated model with respect to the non-mutated model $(\mu = 0)$.

Let us consider, in the non-mutated model there are total maximum X number of Cellular states $(C^n : C^n_1, C^n_2, C^n_3, \ldots\ldots C^n_X)$ observed and in the mutated model there are total maximum Y number of Cellular states $(C^m : C^m_1, C^m_2, C^m_3, \ldots\ldots C^m_X)$ observed. Suppose in the mutated model there are total μ mutations induced. Hence, the mutational cost (Ω) is defined as follows ***(Eq. 21)***.

$$\Omega = \mu \times \frac{|C^n - (C^n \cap C^m)| + |C^m - (C^n \cap C^m)|}{\left(\left|C^n \cup C^m\right|\right)}$$

(Eq. 21)

$\mu = Total\ induced\ mutations\ in\ the\ mutated\ model$

$\left|C^n \cup C^m\right| = Total\ number\ of\ cellular\ states$

$|C^n - (C^n \cap C^m)| = Total\ number\ of\ cellular\ states\ which\ are\ not\ appeared\ in\ the\ mutated\ model$

$|C^m - (C^n \cap C^m)| = Total\ number\ of\ cellular\ states\ which\ are\ newly\ appeared\ in\ the\ mutated\ model$

Hence, the phenotype cost function (Ψ_{C_i}) is defined as the sum of total signal cost and the mutational cost (if any) for a particular state (C_i) in the STG of a Boolean model. It is defined as follows ***(Eq. 22)***.

$$\Psi_{C_i} = \Phi_{C_i} + \Omega$$

(Eq. 22)

Phenotype Predictor Score: The predictor score $[\Theta(C_i)]$ of a particular cellular state observed in the simulation defines as the ratio of the total number of observed occurrences (P_{C_i}) to the total cost (Ψ_{C_i}) requires for reaching at ***ith*** cellular state (C_i) in the attractor space. Hence, this parameter can be defined as follows ***(Eq. 23)***.

$$\Theta(C_i) = \frac{E(P_{C_i})}{E(\Psi_{C_i})}$$

(Eq. 23)

$$Where,\ E(P_{C_i})\ \&\ E(\Psi_{C_i}) = Expected\ values\ of\ P_{C_i}\ and\ \Psi_{C_i}\ out\ of\ N\ number\ of\ simulation\ batches$$

The probability density of the phenotype predictor score $(\Theta = \frac{P}{\Psi})$ is previously discussed in various literatures [237,238].

Here, $P = N(\mu_P, \sigma_P^2), \Psi = N(\mu_\Psi, \sigma_\Psi^2), \rho = Corr(P, \Psi) \neq \pm 1$, and the probability density function $f(\Theta)$ is given as follows ***(Eq. 24)***.

$$f(\Theta) = \frac{\sigma_P \sigma_\Psi \sqrt{1-\rho^2}}{\pi\left(\sigma_\Psi^2 \Theta^2 - 2\rho\sigma_P\sigma_\Psi\Theta\right)}\left[\exp\left(-\frac{1}{2}SupR^2\right) + \sqrt{2\pi}.R.\Phi(R).\exp\left(-\frac{1}{2}[\sup R^2 - R^2]\right)\right]$$

(Eq. 24)

$$Where, R = R(\Theta) = \frac{\left(\frac{\mu_P}{\sigma_P} - \rho\frac{\mu_\Psi}{\sigma_\Psi}\right)\Theta - \left(\rho\frac{\mu_P}{\sigma_P} - \frac{\mu_\Psi}{\sigma_\Psi}\right)\frac{\sigma_P}{\sigma_\Psi}}{\sqrt{1-\rho^2}\sqrt{\Theta^2 - 2\rho\frac{\sigma_P}{\sigma_\Psi}\Theta + \left(\frac{\sigma_P}{\sigma_\Psi}\right)^2}}$$

$$SupR^2 = \frac{\left(\frac{\mu_P}{\sigma_P}\right)^2 - 2\rho\frac{\mu_P}{\sigma_P}\frac{\mu_\Psi}{\sigma_\Psi} + \left(\frac{\mu_\Psi}{\sigma_\Psi}\right)^2}{1-\rho^2}$$

$$SupR^2 - R^2 = \frac{\left(\frac{\mu_P}{\sigma_P}\frac{\sigma_P}{\sigma_\Psi} - \frac{\mu_\Psi}{\sigma_\Psi}\Theta\right)^2}{\left(\Theta^2 - 2\rho\frac{\sigma_P}{\sigma_\Psi}\Theta + \left(\frac{\sigma_P}{\sigma_\Psi}\right)^2\right)}$$

$$\Phi(R) = \frac{1}{2}erf\left(\frac{R}{\sqrt{2}}\right)$$

The function $f(\Theta)$ follows the general properties of the Cauchy-like distribution [237].

2.5.9 Selection of Patient Cohorts and Preparation of RNAseq Sample Data Sets

Two patient cohorts consist of low-grade (TCGA-LGG) and high-grade Glioblastoma (TCGA-GBM) from "The Cancer Genome Atlas (TCGA)" research networks [239,240] were chosen for the case studies in this present work. HTSeq raw counts files (RNASeq experiment) of the Glioblastoma patients with TP53 mutation were selected to determine the mRNA expression profiles of these two patient cohorts (data downloaded on 27th April, 2017).

The TCGA-Low Grade Glioblastoma (TCGA-LGG) cohort contains tumor samples from Grade-II and Grade-III glioblastoma patients, whereas the TCGA-GBM cohort contains samples from Grade-IV tumor patients. The RNASeq raw counts data (available as TXT files) for each patient is available in the open source GDC Data portal of National Cancer Institute (https://portal.gdc.cancer.gov/). Further information about the workflows related to sample collection; mRNA sequencing and data processing, read alignments, and mRNA quantification etc. are available at *https://docs.gdc.cancer.gov/Data/PDF/ Data_UG.pdf*.

2.5.10 Differential Expression Analyses of mRNA transcripts

The RNASeq raw counts data files corresponding to the patient cohort with TP53 mutation were extracted from the TCGA-LGG (General) and TCGA-GBM (General) patient cohorts. The differential expression analyses were performed by forming the contrasts between primary (TP53 mutated) (i) LGG (total samples = 240) and (ii) GBM (total samples = 57) tumor samples versus solid normal tumor (total samples = 5) samples using "edgeR" statistical package [241]. The differentially expressed (P-value <= 0.001) transcripts of the input proteins of aNSC, GSC, and general GBM model were extracted. The logical states of the extracted input proteins were considered as "1" or "True" in the logical simulation of the case studies, if the transcript was up-regulated in contrast to normal tumor samples, else "0" or "False". The mRNA expression patterns observed in this analysis were considered as the transcriptomics profiles of the two different patient cohorts (primary LGG and GBM).

2.5.11 Logical Simulations using the Transcriptomics Data Generated from TCGA-LGG and TCGA-GBM

The transcriptomics profiles of the input protein molecules observed in both the TCGA-LGG and TCGA-GBM patient cohorts were taken as inputs for the further simulations of aNSC and general GBM model simulations.

2.5.11.1 Methodologies used in the Case-Study of TCGA-LGG Patient Cohort

In the TCGA-LGG patient cohort there were total 5 and 3 protein molecules found to be over and under expressed, respectively. Mutation or down regulation of P53 protein was also considered in the logical simulation of the case-study of TCGA-LGG cohort. Hence, out of total 53 input molecules of the master aNSC model, there were total 9 protein molecules kept frozen (i.e., constitutively expressed at ON or OFF state) and eliminated from the input list for further randomization in the new simulation. Hence, the total number of mutations μ introduced in the master aNSC

model was 9 and the rest 44 input proteins were randomized 10000 times in 10 separate batches. The mean normalized frequency values of each cellular state were calculated from these 10 independent simulation batches, which will be further used for checking the goodness-of-fit with the normalized frequency values observed for the cellular states in the previously performed, master aNSC model. The objective of this study was to assess the effects of these 9 differentially expressed transcripts of the input proteins in the development of adult NSCs within the neurogenic niche of human brain and how much the normalized frequency distributions of different cellular states will vary with respect to the distributions of the cellular states observed in the master aNSC model. The Chi-square goodness of fit test was used to compare the normalized frequency distributions of this new model (i.e., observed normalized frequencies) with the distributions (i.e., expected normalized frequencies) observed in the master aNSC model. The statistics was performed under the following null (H0) and alternate (H1) hypotheses:

H_0 = The normalized frequency values observed in the two model simulations (i.e. expected and observed) are consistent with each other.

*H_1 = The normalized frequency values observed in the two model simulations (i.e. expected and observed) are **not** consistent with each other.*

Following this simulation, the TCGA-LGG transcriptomics data was considered as inputs in the previously developed, master model of general GBM developmental model. In the master GBM model, there were already 4 mutations added (including P53) and 8 mutations were further added as per the TCGA-LGG transcriptomics expression profile (see **Table 13** of *Chapter 6*). Hence, altogether there were total 12 mutations (μ) introduced in the master general GBM model and a new derivative model was developed by keeping these 12 input proteins at constitutively up or down regulated states. These 12 input proteins were kept aside for further randomization and the rest 41 input proteins were randomized 10000 times in 10

separate batches. The mean normalized frequency values of each cellular state (observed normalized frequencies) were calculated from these 10 simulation batches and the Chi-square goodness-of-fit test considering the null (H_0) and alternate (H_1) hypotheses was performed again to assess the effects of the newly introduced mutations on the normalized frequency values of each cellular state (i.e., expected normalized frequencies) found in the master, general GBM model simulation.

2.5.11.2 Methodologies used in the Case-Study of TCGA-GBM Patient Cohort

Similar to the analyses of TCGA-LGG tumor samples cohort, the transcriptomics profile (see **Table 13** of *Chapter 6*) observed for the TCGA-GBM sample cohort was also considered at first as inputs in the master aNSC and general GBM models. There were total 12 and 6 proteins found to be up and down-regulated, respectively, in the differential expression analyses of TP53 mutated primary GBM versus normal solid tumor tissues. Therefore, in total 19 mutations ($\mu = 12 + 6 + 1$) of the input proteins were considered in the input in the master aNSC model and a new model using the TCGA-GBM transcriptomics data on master aNSC model was developed. Here, the extra 1 mutation was added for the P53 mutation. Similar to the previous analyses, these 19 input proteins were kept constitutively at ON or OFF state as per the differentiation expression results (see **Table 13** of *Chapter 6*) and the rest 34 input proteins out of 53 were further randomized 10000 times in 10 separate simulation batches. Similar Chi-square goodness-of-fit test was performed to assess the similarities of the normalized frequency values of each cellular state observed in master aNSC and new model simulations.

On the other hand, another set of simulation was performed to examine the effects of the differentially expressed transcripts of the TCGA-GBM tumor samples cohort (see **Table 13** of *Chapter 6*) on the master general GBM model. Here, the master general GBM model had 4 mutations (TP53, JAK2, STAT3, and RBPJ) and in the TCGA-GBM cohort another 18 proteins were found to be differentially expressed.

Therefore, in total there were 22 mutations ($\mu = 18 + 4$) added in the new simulation of general GBM model for the case-study. These 22 proteins were kept constitutively expressed (either up or down) based on the transcriptomics profile generated from differential expression analyses (see **Table 13** of *Chapter 6***)** and the rest 31 out of 53 input proteins were randomized 10000 times in 10 separate batches. Similar Chi-square goodness-of-fit test was also performed to study the similarities of the normalized frequency values of each cellular state observed in master general GBM model and new model simulations.

2.5.12 Methodologies used for Drug Target Screening

The simulation outcomes of "Grade-IV GBM model" simulation showed significant increase in the number of Grade-IV tumor cells (i.e., the cellular state: "GSC Renewal/ASPC Differentiation/GBM Development") in the attractor space (**Figure 39C**). The periodic state transition dynamic in the steady state level observed in this cellular state was mainly occurred due to the presence of cyclic expression patterns of its corresponding marker proteins (**Table 2**). The expressions of these marker proteins were dependent on the intermediate molecules through which the flow of signaling cascades were transduced from the input to the target molecules during tumorigenesis. Hence the molecules, which showed higher correlations in their expression pattern with the marker proteins, were the key regulators of the flow of tumorigenic signal inside the developing Grade-IV tumor cells. Therefore, it can be hypothesized that a subset of intracellular intermediate molecules, strongly tuned and correlated with the activity pattern of Grade-IV cellular state via the marker proteins are the most significant molecules for the sustainment of this cellular state. Indeed, it is expected that perturbing the expressions (i.e., logical states) of such molecules (individually or in combination) will alter the activity pattern of the Grade-IV cellular state and those proteins will be considered as potential drug targets. However, identification such small subset of molecules out of the large set of modeled pathway molecules is a challenging task, which could be only possible to

resolve through an efficient computational algorithm capable of analyzing multiple time series data simultaneously.

The intermediate molecules which were mutually interconnected and highly correlated with the temporal activity pattern observed for Grade-IV cellular state in the "Grade-IV GBM" model simulation study were required to be extracted [242]. The correlation and delay between a pair of time course data was calculated by using Fast Fourier Transform (FFT) analysis. The delay and pair-wise correlation between the activity patterns of Grade-IV tumorigenic cellular state (i.e. Target signal) and the time-course logical expressions data of all the intermediated molecules (i.e. Query signal) were measured by the following method.

Let us consider that $C^{Grade-IV} = \{c_1, c_2, c_3, ..., c_T\}$ is the time series activity (ON or 1 and OFF or 0) profile of the Grade-IV cellular state observed in the STG of "Grade-IV GBM model" simulation at the discrete time points $t = 1, 2, 3, ..., T$. This time-course data of Grade-IV cellular state was considered as the "Target" signal. Similarly, let us consider that the time-course logical expression (ON/1, OFF/0) profile of any arbitrary molecule $X_i = \{x_1, x_2, x_3, ..., x_T\}$, which was considered as "Query" signal. Both the temporal signals (S) were decomposed into cyclic patterns (i.e., frequency domain) with each frequency $n = 1, 2, 3, ..., T-1$ by following FFT analyses as shown in ***(Eq. 25)*** and ***(Eq. 26)*** [242].

$$C_n = \frac{1}{T}\sum_{n=1}^{T-1} c_t e^{-i2\pi n\frac{t}{T}}$$

(Eq. 25)

$$X_n = \frac{1}{T}\sum_{n=1}^{T-1} x_t e^{-i2\pi n\frac{t}{T}}$$

(Eq. 26)

Amplitude of the cycle with frequency n = 0 was neglected. The amplitudes and phase angles of the cycles with higher frequencies (n > 0) were calculated for both the

signals. The frequency of the cycle (n) for which the amplitude is found at maximum magnitude was at first identified. After that phase angles of the cycles from both the target and query signals (ϕ_C^n, ϕ_X^n) were calculated at that frequency (n) and the difference or delay $\Delta^n{}_{CX} = \phi_C^n - \phi_X^n$ between the two signals was measured. The delay between two signals δ_{CX} was further calculated in the range of $0 \; to \; \frac{T}{n}$ by using the following ***(Eq. 27)*** [242].

$$\delta_{CX} = \frac{\Delta^n{}_{CX}}{\left(\frac{360n}{T}\right)}$$

(Eq. 27)

Lagged Pearson correlation coefficient was also calculated for measuring the strength and association between the two signals or trajectories [242]. In this work the delay and correlation between Grade-IV cellular state and for each pathway molecules were measured pair-wise. There were total six probable outcomes, which were found while comparing all such pairs of trajectories (i.e., Target vs. Query) using this approach. These entire mathematical calculations were done in "DynOmics" package developed in the statistical package "R" [242].

i) The signals were positively correlated (correlation > 0) with no delay (Delay = 0). In this case both target and query signals were superimposed on each other and the phase (or direction) of the signals was same (i.e. positively correlated). Hence, it was considered that the molecular expression pattern (i.e. query signal) was positively influencing the activity pattern or dynamics of Grade-IV tumor state and there was no delay between the time-course profiles. Therefore, the molecule was a strong positive inducer or activator of Grade-IV tumor cells, which means that when the molecule was at ON (i.e., up -regulated/ active) state, the activity of Grade-IV cellular state was also High (or ON).

ii) The signals were positively correlated (correlation > 0) with negative delay (Delay < 0). In this case the phase (or direction) of both the signals was same (i.e., positively correlated), but the initiation of the target signal at initial time point was delayed with respect to the query signal. Hence, it was considered that the molecular expression pattern (i.e. query signal) was positively influencing the activity pattern or dynamics of Grade-IV tumor state, but there was a lag of the Grade-IV time-course activity profile with the respect to that positively influencing molecule. Therefore, the molecule was a positive inducer or activator of Grade-IV tumor cells, which means activation of these molecules, will lead to the higher expression of Grade-IV tumor state.

The signals were negatively correlated (correlation < 0) with positive delay (Delay > 0). In this case the phases of the signals were opposite (i.e. negatively correlated) and the initiation of the target signal was ahead of the query signal at initial time point. Hence, it was considered that the molecular expression pattern (i.e., query signal) was negatively influencing the activity pattern or dynamics of Grade-IV tumor state, but the Grade-IV time-course activity profile was running ahead with the respect to that negatively influencing molecule. Therefore, the molecule was a negative inducer or activator of Grade-IV tumor cells, which means which means when the molecule was at ON (i.e., up-regulated/ active) state, the activity of Grade-IV cellular state was Low (or OFF).

iii) The signals were negatively correlated (correlation < 0) with no delay (Delay = 0). In this case the phases of the signals were opposite (i.e. negatively correlated), but there was no delay at the initiation of the signals at initial time point. Hence, it was considered that the molecular expression pattern (i.e., query signal) was negatively influencing the activity pattern or dynamics of Grade-IV tumor state, but there was no lag between the trajectories of these two signals. Therefore, the molecule was a strong negative inducer or inhibitor of Grade-IV tumor cells which means activations

of these molecules, will inhibit the expression or activation of Grade-IV tumorigenic state.

iv) The signals were negatively correlated (correlation < 0) with negatively delay (Delay < 0). In this case the phases of the signals were opposite (i.e., negatively correlated), but the initiation of the target signal at initial time point was delayed with respect to the query signal. Hence, it was considered that the molecular expression pattern (i.e.. query signal) was negatively influencing the activity pattern or dynamics of Grade-IV tumor state, but there was a lag of the Grade-IV time-course activity profile with the respect to that negatively influencing molecule. Therefore, the molecule was a negative inducer or inhibitor of Grade-IV tumor cells which means activations of these molecules, will inhibit the expression or activation of Grade-IV tumorigenic state.

v) The signals were positively correlated (correlation > 0) with positive delay (Delay > 0). In this case the phase (or direction) of both the signals is same (i.e., positively correlated), and the initiation of the target signal was ahead of the query signal at initial time point. Hence, it was considered that the molecular expression pattern (i.e., query signal) was positively influencing the activity pattern or dynamics of Grade-IV tumor state, but the Grade-IV time-course activity profile was running ahead with respect to that negatively influencing molecule. Therefore, the molecule was a positive inducer or activator of Grade-IV tumor cells which means activations of these molecules, will lead to the higher expression of Grade-IV tumor state. The pathway molecules, which showed significant positive correlation (correlation value ≥ 0.6) and Delay ≤ 3 with the activity trajectory of Grade-IV cellular state were chosen as drug targets for the subsequent perturbation analyses.

Perturbation study: The molecules enlisted in **Table 14** of *Chapter 6* showed positive correlation with the activity profile of Grade-IV cellular state were targeted by freezing the expression states as down-regulated (or OFF) state in the "Grade-IV

GBM" model. On the other hand, the molecules, which were negatively correlated, were kept up-regulated (i.e., ON) in the perturbation study. In **Table 15** of *Chapter 6*, the molecules which had absolute correlation value ≥ 0.75 with P-value < 0.05, Delay ≤ 3 and showed profound effect of reducing the activity of LGG-I, LGG-II, and Grade-IV cellular states while perturbing their expressions in the "Grade-IV GBM" model were extracted rank-wise.

2.6 SEMI-DYNAMIC ANALYSES USING QUATERNARY LOGIC

2.6.1 Truth Table of Quaternary Logic-Based Operation

The truth table of the quaternary logic (**Table 3**) was taken from the truth table used in previous work proposed for constructing quaternary logic gate circuits of universal NOT, OR, and AND gates using Spatial Wavefunction-Switched Field-Effect transistors (SWSFET) [243].

Table 3: The Truth Table of Quaternary Logic

Serial Number	I_1	I_2	$O_1 = I_1$ AND I_2	$O_2 = I_1$ OR I_2	O_3 = NOT O_2
1	0	0	0	0	3
2	0	1	0	1	2
3	0	2	0	2	1
4	0	3	0	3	0
5	1	0	0	1	2
6	1	1	1	1	2
7	1	2	0	3	0
8	1	3	1	3	0
9	2	0	0	2	1
10	2	1	0	3	0
11	2	2	2	2	1
12	2	3	2	3	0
13	3	0	0	3	0
14	3	1	1	3	0
15	3	2	2	3	0
16	3	3	3	3	0

2.6.2 Modifications of the Logical Equations Constructed Using Negation Rule

The implementation of NOT gate in the logical equation which consists of only a single or multiple upstream inhibitors without any activator(s) was considered in the newly developed quaternary logic-equation in a slightly different way than the conventional approach. For example, if a node ***A*** is only inhibited by an upstream inhibitor ***B,*** then the logical equation of the corresponding event will be denoted by

the following logic-equation ***(Eq. 28)***.

$$A(t+1) = A(t) \wedge !B(t)$$

(Eq. 28)

The following truth table of the equation can be obtained from ***(Eq. 28)*** by using binary logic, which is also comparable with the real biochemical event (**Table 4**).

Table 4: Binary Logic-Based Truth Table of the Modified NOT Gate Equation

A(t)	*B(t)*	*A(t+1)=A(t) AND NOT B(t)*
0	**0**	**0**
0	**1**	**0**
1	**0**	**1**
1	**1**	**1**

In the quaternary logic-based operations, the resultant truth table after executing ***(Eq. 28)*** is provided below (**Table 5**).

Table 5: Truth Table of the Extend ExQuLogic

Serial Number	*A(t)*	*B(t)*	*NOT B(t)*	*A(t+1)=A(t) AND NOT B(t)*
1	0	0	3	0
2	0	1	2	0
3	0	2	1	0
4	0	3	0	0
5	1	0	3	1
6	1	1	2	0
7	1	2	1	1
8	1	3	0	1
9	2	0	3	2
10	2	1	2	2
11	2	2	1	0
12	2	3	0	0
13	3	0	3	3
14	3	1	2	2
15	3	2	1	1
16	3	3	0	0

It can be seen from both the truth tables that when the node ***A*** is absent or not expressed in the previous time-step, then the logical state of the node ***A*** at the next time-step will be always at absent or not expressed state, irrespective of the state of ***B*** at the previous time-step. Hence, it can be said that memory of the node ***A*** at t^{th} time-step plays important role in its own update scheme at $(t+1)^{th}$ time-step. The memory function of the inhibited downstream node can be also used in case multiple

inhibitors present in the upstream with no activator node(s). In such instance, the equation will be written as follows ***(Eq. 29)***.

$$A(t+1) = A(t) \wedge !B_1(t) \wedge !B_2(t) \wedge !B_3(t) \wedge .....!B_n(t)$$
$$Where, n = Total\ number\ of\ upstream\ inhibitors$$

(Eq. 29)

2.6.3 Extended Quaternary States Update Scheme (ExQSUS)

The quaternary state update schemes of the nodes used in the conjunction rule (AND operator) of logical equations were modified based on the assumptions of "maximal or dominant effect" of the upstream effectors nodes on the downstream effected node. The algorithm for the extended quaternary states update scheme (ExQSUS) used in this work is elaborately discussed here.

Let us consider that there are ***m*** number of upstream activators $\{A_1, A_2, A_3, A_4, ...A_m\}$ and ***n*** numbers of inhibitors $\{I_1, I_2, I_3, I_4...., I_n\}$ nodes regulating a downstream node $\{D\}$ in the network. Let us also consider that all the activators and inhibitors work in conjunction to activate or inhibit the expression of the downstream node $\{D\}$. Hence, the following logical equation using the universal "AND" and "NOT" operators can be written to describe the update rule of the node $\{D\}$ at ***(t+1)th*** time-point ***(Eq. 30)***.

$$D(t+1) = A_1(t) \wedge A_2(t) \wedge A_3(t)..... \wedge A_m(t) \wedge !I_1(t) \wedge !I_2(t) \wedge !I_3(t)..... \wedge !I_n(t)$$

(Eq. 30)

Let us compute the total number upstream nodes at ***tth*** time-point having quaternary states $X = \{0,1,2,3\}$ as θ_X by using the following equation ***(Eq. 31)***.

Where, $\theta_X = \sum_{i=1}^{m} (A_i \equiv X) + \sum_{j=1}^{n} (I'_j \equiv X),\ X = \{0,1,2,3\}\ and\ I'_n = !I_n(t)$

(Eq. 31)

The stepwise description of the ExQSUS algorithm implemented in the extended

quaternary logic (ExQuLogic)-based operations is provided in **Appendix Text 1**.

The disjunctive (OR operator) rule implemented in ExQuLogic operation were simply performed by using the truth table of quaternary number based algebra (**Table 3**). The logic equation ***(Eq. 32)*** consists of conjunctive, disjunctive, and negation rules were solved by first computing the negation rule (or NOT operator), after that conjunctive rule (or AND operator) using ExQSUS algorithm followed by disjunctive rule (or OR operator).

$$D(t+1) = (A^1_1(t) ..... \wedge A^1_m(t) \wedge !I^1_1(t) ..... \wedge !I^1_n(t)) \vee (A^2_1(t), ... \wedge A^2_r(t) \wedge !I^2_1(t) ..... \wedge !I^2_s(t) ..... \vee (A^w_1(t), ... \wedge A^w_p(t) \wedge !I^w_1(t) ..... \wedge !I^w_q(t))$$

$$D(t+1) = (AND(A^1_1(t) ..... A^1_m(t), NOT(I^1_1(t), ..., I^1_n(t)) \vee (AND(A^2_1(t) ..... A^2_r(t), NOT(I^2_1(t), ..., I^2_s(t)) \vee (AND(A^w_1(t) ..... A^w_p(t), NOT(I^w_1(t), ..., I^w_q(t))$$

$$D(t+1) = OR(AND^1(m,n), AND^2(r,s), ..., AND^w(p,q))$$

$$D(t+1) = OR(Q^1, Q^2, ..., Q^w)$$

(Eq. 32)

2.6.4 Quaternary States Selection Rules

The following quaternary state selection rules are proposed here to select the logical states of the genes/proteins after performing differential expression analyses of transcripts/proteins.

$$\begin{aligned}
&= 3, \ iff\ (P\text{-}value <= \theta_1\ \&\ log\ [FC] \geq \theta_2) \\
&= 2, \ iff\ (P\text{-}value <= \theta_1\ \&\ 0 < log\ [FC] < \theta_2) \\
&\qquad or, \\
&\qquad (P\text{-}value > \theta_1\ \&\ log\ [FC] > \theta_2) \\
&\qquad or, \\
&\qquad (P\text{-}value > \theta_1\ \&\ 0 < log\ [FC] < \theta_2) \\
Quaternary\ state = 1, \quad &iff\ (P\text{-}value <= \theta_1\ \&\ -\theta_2 < log\ [FC] < 0) \\
&\qquad or, \\
&\qquad (P\text{-}value > \theta_1\ \&\ -\theta_2 < log\ [FC] < 0) \\
&\qquad or, \\
&\qquad (P\text{-}value > \theta_1\ \&\ log\ [FC] < -\theta_2) \\
&= 0, \ iff\ (P\text{-}value <= \theta_1\ \&\ log\ [FC] \leq -\theta_2)
\end{aligned}$$

Here, θ_1 and θ_2 are the user defined non-zero, positive values to setup the thresholds for *P*-value (or FDR) and the log of fold change ratio, respectively. Also, it should be noted that the quaternary state "0" is used here for representing the under expressed or down-regulated or knock-out or absent transcripts/genes/proteins of the logical model. On the other hand, the quaternary state "3" represents over-expressions or up-regulation or hyper-activation of the pathway species. The other

two states are used for representing the intermediate states between up (hyper-active) or down (inactive) regulations. The quaternary state "2" can be denoted as modestly or regularly active nodes, whereas the quaternary state "1" can be used for signifying the hypoactive or lesser active nodes in the network. In a differential omics-based expression study, if a transcript/gene/protein is quantized as "1" as per our quaternary state selection rule, then it can be signified that the biomolecule is almost equally expressed in both the groups (e.g., tumor versus normal) of the study and has little contribution to make significant changes in the regulation of its downstream nodes in the network. On the other hand, the quaternary state "0" refers the complete absent (knocked-out) or a very little expression of the nodes in one group of cells as compare to other group.

2.6.5 Implementation and Execution of Quaternary Logic

The execution of quaternary logic-based operation in the conventional computing environment is not straightforward as the modern computing systems are hitherto operated in the regime of binary numbers (0 or 1) system-based logical operations proposed in the truth table of Boolean algebra. None of the existing model or software developed for the simulation of multi-valued logical analyses of biochemical pathways uses the actual truth table of quaternary logic (**Table 3**). In this work, this problem was solved and the executions of quaternary logic-based operations were executed in the conventional computing environment. The truth table for performing quaternary algebra was also implemented by using a special GCC library developed for this purpose. This new library can be compiled with standard C programming environment and the simulations of the logical equations or steady-state analyses of quaternary logic-based operations can be performed without changing the core structure and conceptual framework of the conventional logical equations. That means, the rules applied for implementing the conjunction (AND gate), disjunction (OR gate), and negation (NOT gate) in the conventional logic-based model can be still used in the new computational framework while

performing quaternary logic-based operations.

2.6.6 Simulations of the Signaling Networks using ExQuLogic Method

The logical model of the T-cell receptor signaling pathways was taken from the model published by *Saez-Rodriguez et al.* [244]. The logical equations of the T-cell receptor mediated signaling pathway were simulated in three different time-steps (t = 0, 1, 2) based on ExQuLogic-based logical update rules. The logical model of the ERBB receptor mediated signaling pathway was taken from the model published by *Sahin et al.* [245]. The logical model of the pathway is provided in **Appendix Table 11**. The logical models of Hedgehog and Notch signaling pathways were taken from [82] and [245]. By freezing the logical update rule of the node with respect to time, the perturbation analyses of the logical nodes were performed.

For example, while perturbing the logical state of any arbitrary node ***D*** in the logical model, the logical equation was changed to freeze the update of the node $D(t+1) = f(A_1, A_2, ..., A_n, I_1, I_2, ..., I_n)$ to $D(t+1) = D(t), where\ D(0) = q \in \{0,1,2,3\}$. Here, the quaternary state ***q*** is the altered or perturbed state of the node ***D***.

The logical model of the integrated signaling pathways of Hedgehog, Notch, WNT, ERBB pathways shown in *Chapter 7* were developed by integrating the previous models of the Hedgehog [82], Notch [245], and ERBB pathways [245] in a common logical model followed by its integration with the newly developed WNT pathway model. The WNT pathway contains the comprehensive information of the all the nodes (proteins, complexes, genes) and their reactions, which were manually collected from various resources, such as PPI databases, Cell signaling databases, and most importantly from various literature sources [8]. The logical equations of the integrated pathway model are provided in **Appendix Table 12.**

2.6.7 Computation of the Level of Accuracy of Predicting the Activity Profile

The activity profiles of all or a few select nodes in the logical model refer the

active or inactive states of the nodes (genes/proteins) in an experimental datasets (obtained from differential expression studies) or at the logical steady state (obtained from model simulations). Here, the "active" state of a node in a logical network refers quaternary states "2" and "3" or the binary state "True (1)", whereas "inactive" state signifies either the quaternary states "1" and "0" or the binary state "False (0)". The differential expression of the target nodes (genes/proteins) considered in the activity profile was at first taken from the differential expression studies of the target model scenario (e.g., Glioblastoma) followed by the application of quaternary state selection rule (see **Section 2.6.4**) to convert and discretize the expression values in quaternary states. After discretization, the quaternary states were hot encoded as "Active ($\alpha_{expected}$)" if the states are either "2" or "3" and "Inactive ($\beta_{expected}$)" if the states are either "0" or "1". Similarly, the steady-state expression values of the quaternary states obtained after ExQuLogic-based simulation are also hot encoded as "Active ($\alpha_{predicted}$)" and "Inactive ($\alpha_{predicted}$)" states. The following equation ***(Eq. 33)*** was then used to compute the percentage of accuracy of the predicted activity profile obtained by the ExQuLogic-based operations with respected to the descritized and hot encoded experimental expression values of the transcripts obtained after differential expression studies.

$$Percentage\,of\,accuray = \frac{\sum_{i=1}^{N} Z(i)}{Total\,Number\,of\,encoded\,transcripts/genes/proteins(N)}$$

$$Z(i) = \begin{cases} 1, & iff\ V^{i}_{expected} = V^{i}_{predicted} \\ 0, & otherwise \end{cases}$$

Where, $V^{i}_{expected} \in \{\alpha_{expected}, \beta_{expected}\}, V^{i}_{predicted} \in \{\alpha_{predicted}, \beta_{predicted}\}$

(Eq. 33)

Chapter 3

3 DEVELOPMENT OF A NEW PATHWAY DATABASE WITH INTEGRATED COMPUTATIONAL PLATFORM

3.1 INTRODUCTION[2]

Systems level understandings of biochemical pathways to analyze their roles in governing multiple biological systems is expected to flourish at various directions in the near future using state-of-the-art molecular biology experiments [153,246-248]. A huge amount of unstructured data related to biochemical pathways is now available in literatures as well as in various other resources as raw or partially processed formats [248-252]. However, it is one of the major challenges to the researchers for acquiring and collating such highly dispersed data into a common hub of biological knowledgebase for further analyses [251]. Such platform will help to share the annotations of newly identified biomolecular species (e.g., genes, proteins, RNA, miRNA, etc.) in a biochemical pathway within the academic communities for future experimentations and references [253].

However, initially the construction and analysis of the large and complex reaction network of biochemical pathways was not so easy due to the unavailability of advanced tools and techniques for the functional annotations of various unknown or newly identified genes and proteins in the biochemical pathways. The recent developments of multidisciplinary, integrative approaches using bioinformatics tools to analyze such big data have made these entire processes easier. The advancements in cellular and molecular biology experiments, high-throughput genomics or

[2] The materials of this chapter has been taken verbatim from our previously published articles (a) Chowdhury and Sarkar, *Database (Oxford).*, 2015 (b) Chowdhury et al., *J Integr Bioinform*, 2018

proteomics studies and the successful completion of Human genome project have flourished this field of study by generating plethora of genomics and proteomics data, and introducing new computational algorithms [254-256]. Moreover, the sub-cellular localization data of the cellular components from immunohistochemistry, FISH, live cells microscopy, etc. experiments have made it possible to annotate the pathway molecules according to their cellular locations in the constructed pathway diagram [257]. Eventually many scientific communities, research groups, database developers have got involved in pathway reconstruction by collating experimental observations from published literatures [257,258]. Various databases (e.g., KEGG, REACTOME, PANTHER, NETPATH, NCI-PID, SIGNALINK) with interactive user friendly interface have been developed to facilitate the operations for pathway data retrieval, sharing and storing in proper formats more easily via internet (see **Appendix Table 1**) [259].

These databases are majorly used for pathway enrichment, gene ontology (GO) analyses, which are the downstream analyses of the high-throughput omics-based experiments [97,260-262]. The data available in these databases are also proven useful for the identification of novel drug or drug targets of various diseases. Spatial annotations of the bimolecular entities provided in these databases are also very much useful for the molecular and cell biology experiments to track the expression dynamics of various marker proteins in in-vitro assays. Apart from these studies, such databases have also become valuable resources for *in-silico* experiments to interpret the emerging properties of the cellular networks upon its exposure in various external cues under different pathological conditions [251]. The biochemical reaction networks provided by these databases are routinely used by the computational biologists to generate and subsequently prove various experimental hypotheses through *in-silico* models and simulations [263]. Indeed, the successful simulation of such *in-silico* models depends on the accuracy of the reconstructed pathway data and respective parameters provided by the databases [263,264].

3.2 A BRIEF HISTORY OF THE EXISTING SIGNALING DATABASES

The history of biochemical pathway databases starts from the construction and storage of metabolic pathway maps by various research groups. Initially these metabolic pathway maps were used to refer as "Biochemical Pathways Wall Chart", originally developed by Dr. Gerhard Michal [265]. In the last two decades, the advance of computer science, internet, web browsers and data sharing policy have made it possible to host and share the rendered pathway images via internet. In 1993-94, EcoCyc, a family of Cyc family database launched the first formal representation of metabolic pathways of *E.coli* [266,267]. Subsequently, in the year of 1995-96, the database *"Kyoto Encyclopedia of Genes and Genomes (KEGG)"* launched by *Profs. Minoru Kanehisa, Kyoto University, Japan,* initiated the web hosting of manually curated pathway diagrams of metabolic, genetic and signal transductions networks of different organisms. It was the first initiative, where the binary gene interaction data from genome projects was used to map and group with functional dependencies and subsequently was presented by so called higher level information schema or pathway diagrams [258].

Inspired by this approach, several other academic and commercial groups put their efforts to construct various biochemical pathways from the experimental data [259]. As time proceeds, a group of databases restricted themselves to only curate either metabolic (e.g. BioCyc, MetaCyc) or gene regulatory (e.g. TRED) or cell signaling networks (SPAD, NETPATH) [268-272]. We observed that throughout the last two decades, the progress of these databases was clearly divided into two branches, mainly lead by either Commercial or Academic groups (**Figure 9**).

It is worthy to mention that before 1998-99 (after the publication of KEGG), there were no as such any divisions or progress seen to develop human cell signaling databases for commercial purposes by any commercial groups or companies. Hence, KEGG, SPAD, and STKE/Science Signaling Database (which were made for academic

purposes), are placed in the middle of the database evolution tree in **Figure 9.**

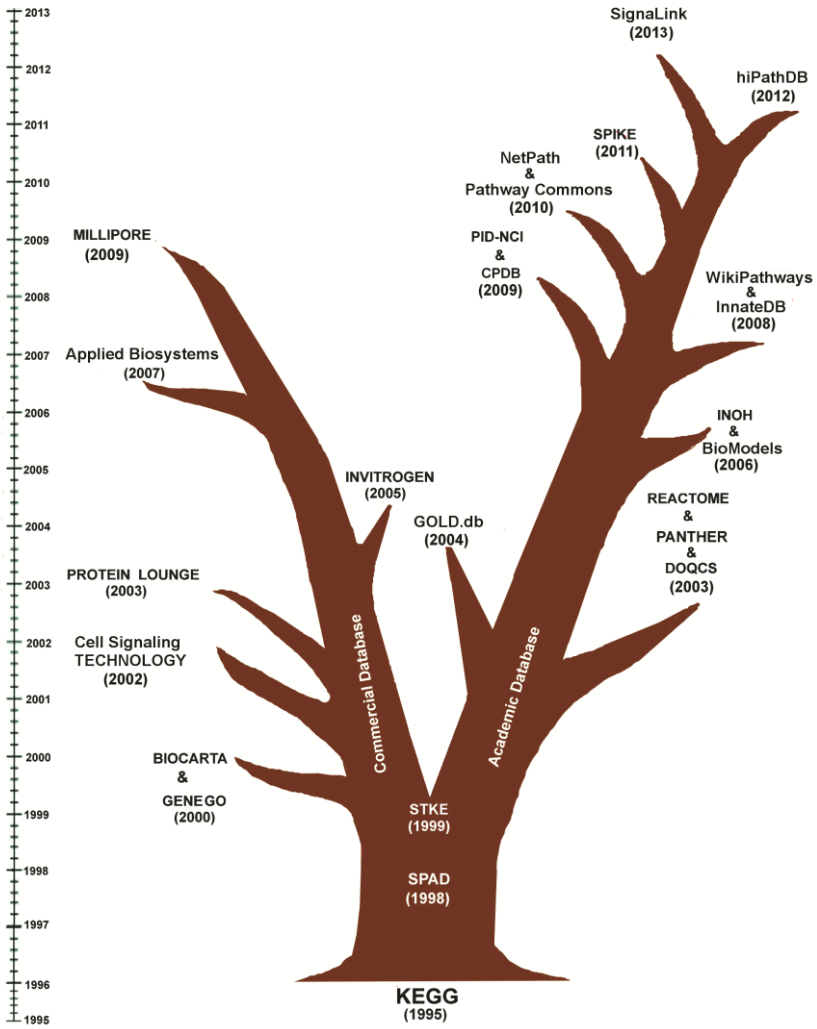

Figure 9: Evolution of the existing and active human signaling pathway databases.
This figure clearly depicts that KEGG, SPAD, and STKE can be thought as the pioneers in the field of the development of human cell signaling databases. In the subsequent years, the evolution of human cell signaling databases is mainly led by the "Commercial Groups" and several "Academic Groups". The first commercial database BIOCARTA and GENEGO were launched in 2000. In the subsequent years several databases, like CELL SIGNALING TECHNOLOGY (2002), PROTEIN LOUNGE (2003), INVITROGEN (2005), APPLIED BIOSYSTEM (2007), MILIPORE (2009) were also started to provide human cell signaling data freely to the users. Integration of signaling pathway components with their corresponding antibodies, drugs, inhibitor molecules in their products website, helps the bench biologists to order and purchase those products readily. On the other hand, the evolution of the other branch of the human cell signaling database is lead by several academic groups across the world. Since 1995, almost each year on an average one or two such database has been launched. The objectives of these databases are wider than the commercial databases and are not only restricted to the pathway data annotation and presentation but also to the analysis of cross talks of multiple pathways, drug target identification, *in-silico* simulation, computer readable pathway data sharing, and pathway analyses and so on.

However, after the year of the launch of STKE in 1999, the history of human cell signaling databases is bifurcated into two branches: Academic and Commercial database. Following 1999, the commercial databases GENEGO/METACORE and BIOCARTA (2000) started to provide signal transduction data to the common users. Successively, the academic databases, such as KEGG, PANTHER, REACTOME, PATHWAY COMMONS, [273] etc. also started to provide similar types of tools to the pathway curators and database developers, and in this context various API services as well as the computer readable pathway data sharing files, like *SBML, BIOPAX, KGML* etc. are distributed for academic purposes to the non-commercial users and software developers at free of cost [274-277]. A detail description of the history and evolutions of these databases is available in our previously review article on human signaling pathway databases [8].

It is also worth to mention that most of the pathway databases have evolved in such a way that it can provide maximum information and facilities to the general users by smoothly sharing the data contents across a large number of users. Moreover, there are many other features, which these databases have included to facilitate the users' experience to interact with the database. However, there are few constrains, which remain unsolved and still demand the attentions of the database developers and as well as the end-users of these databases.

3.3 STATE-OF-THE-ARTS

We observed that all the human cell signaling databases developed till date can be classified according to their mode of collating pathway data and can be grouped into any of the three categories *viz.* "Primary", "Secondary", and "Hybrid" database. Major academic databases KEGG, SPAD, DOQCS, NETPATH, REACTOME, SIGNALINK, SPIKE, BIOMODELS, INOH, and PANTHER and almost all the commercial databases BIOCARTA, GENE GO/METACORE, CELL SIGNALING TECHNOLOGY, PROTEIN LOUNGE, MILLIPORE, APPLIED BIOSYSTEM, and INVITROGEN,

which possess self-curated pathway data fall in the first category. The secondary databases are PATHWAY COMMONS, GOLD.DB, and HIPATHDB, which mainly aggregates the pathway data from the primary databases for various other purposes. On the other hand, the "Hybrid" database, such as WIKI PATHWAYS, NCI-PID, CPDB, and INNATEDB are dependent on self-curated as well as aggregated data from the primary databases. To understand the state-of-the-arts of these databases, we have performed a comparative study on these databases to evaluate the pathway data quality and the available technical features showcase by these databases. In the following sections, we have pointed-out the major outcomes of our review.

3.3.1 Currently Available Active Pathway Databases

We identified around 24 different cell signaling databases, which are currently at active states and distribute the pathway related data (e.g., pathway image, molecular interactions, molecule names, sequence, etc.) freely to the users (**Appendix Table 1**). We observed that various types of cell signaling pathways, such as developmental (e.g., Hedgehog, Notch, etc.), disease-specific (cancer, cardiovascular, etc.), immunological (T cell, B cell signaling, etc.), apoptotic, etc., were freely available in these databases. In total 19 such broad classifications of pathway data types were found in these databases. Hence, to summarize our findings, we constructed a matrix to show the databases and the different types of pathway data available in the respective databases (**Figure 10**).

Furthermore, it was observed that there were only a few databases which store the pathway data compatible with the in-silico model development. BioModels is such kind of database which serves as the repository of the in-silico computational models of the biochemical pathways collated from the published literature. Using a software module Path2Model [278], it automatically generates the mathematical models in the computer-readable formats. It is considered as a useful resource of biochemical pathway models and relevant data, such as kinetic rate parameters,

concentrations of biomolecules, dynamic rate equations, etc. Another database, DOQCS also distributes the in-silico quantitative models for neuron and other signaling pathways. The models provided in this database are collated from literature and are freely available to the users. The kinetic rate parameters are obtained from various experimental resources, such as enzyme assays, binding experiments, the time course of reactions, etc.

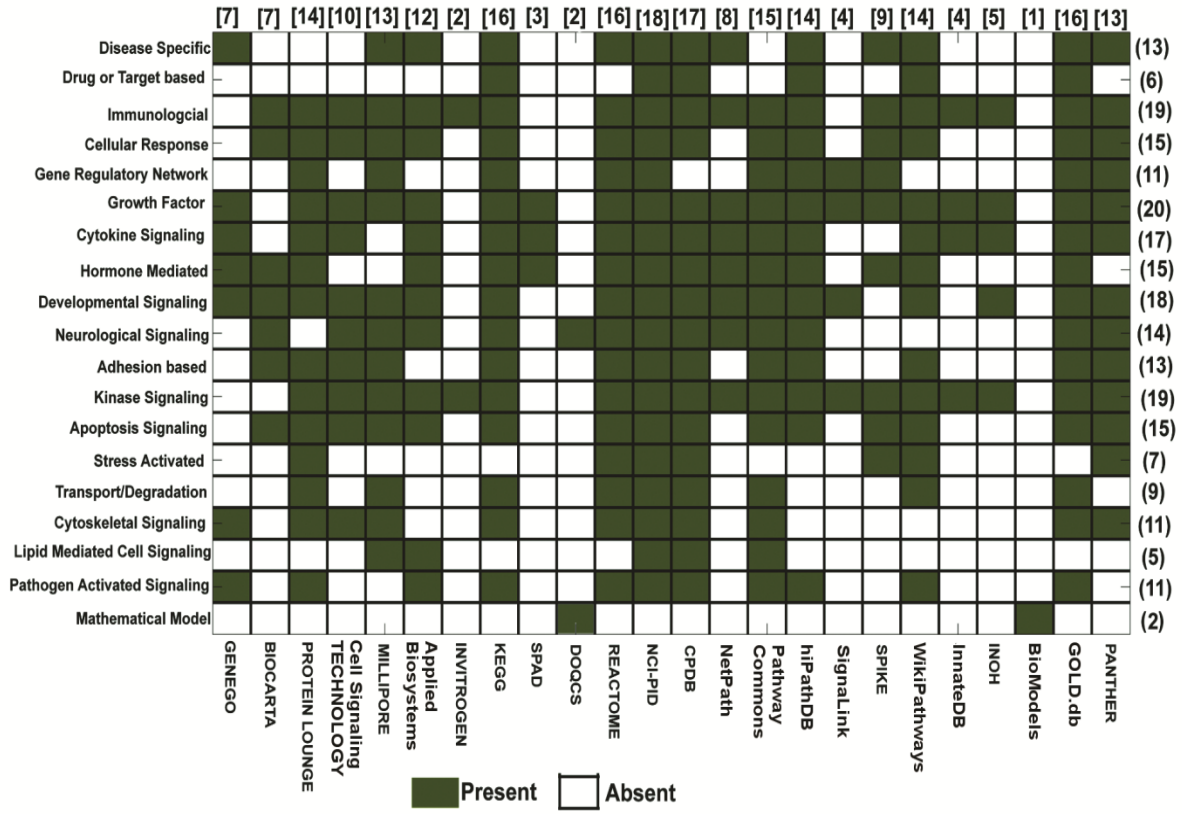

Figure 10: Signaling pathway databases and the available pathway data types.
This figure illustrates a matrix, whose "Rows" and "Columns" represent the number of different types of signaling pathways (Y-axis) available in different databases (X-axis) respectively. Color legends are used to represent the "presence" or "absence" of different types of signaling pathway data (total 19) available in the selected 24 databases. The numbers in the first bracket indicate the number of the databases which contain a particular type of pathway data (represented in row-wise), whereas the numbers in square bracket indicate the total number of different types of pathway data present in a particular database (represented in column wise).

In the following sections, a brief comparative review based on the pathway information and the technical features available in these databases is provided. The elaborate discussion about this topic is also available in our review article [8].

3.3.2 Comparative Study on Pathway Information

3.3.2.1 Pathway Information is not Homogeneous

A case study on Hedgehog and Notch pathway was performed by manually counting the number of molecular species and interactions given in different database (**Figure 11A** and **Figure 11B**).

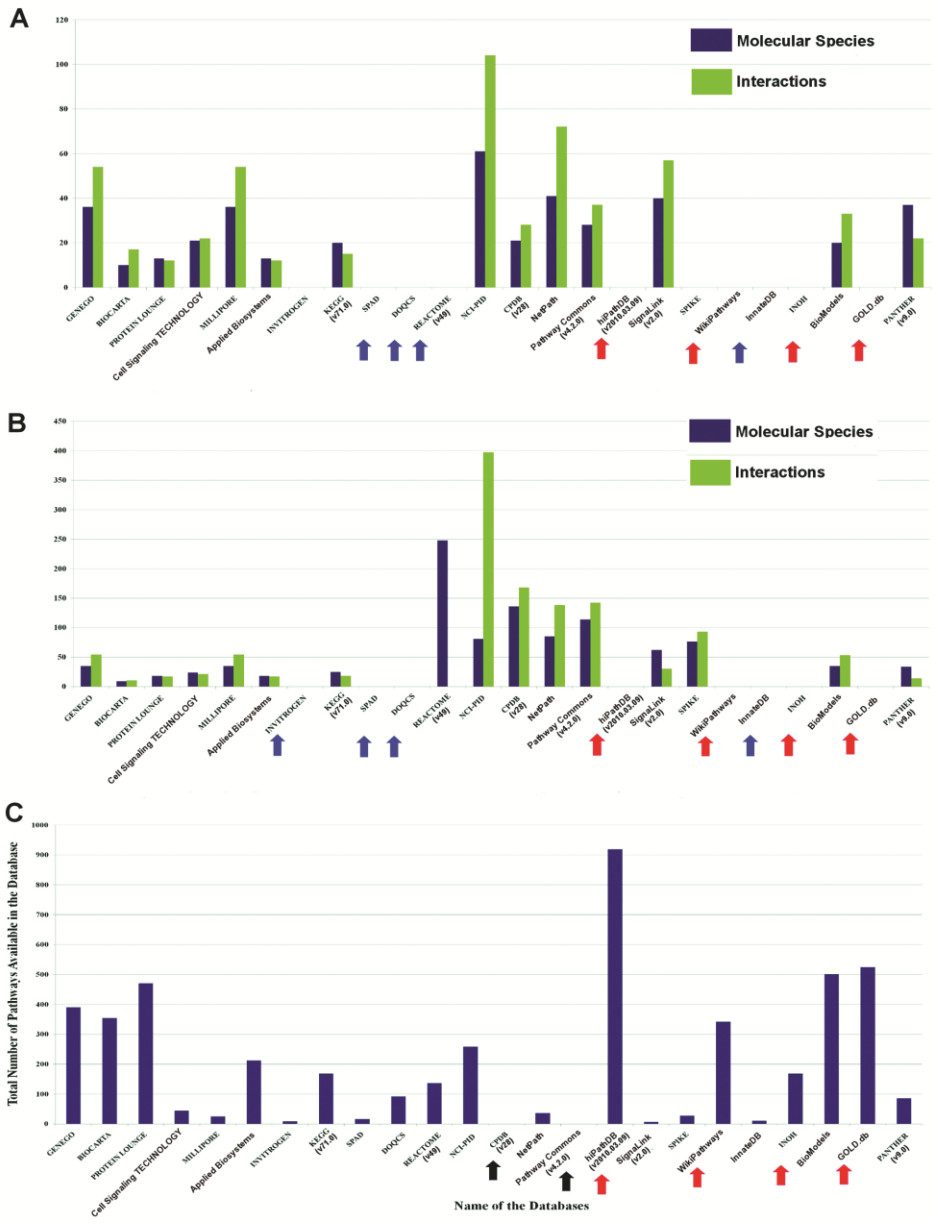

Figure 11: Comparative statistics of pathway information available in databases.
Total number of molecular species and interactions available in **(A)** Hedgehog & **(B)** Notch pathways in different databases are shown here. The blue arrows indicate the databases, which do not possess Hedgehog or Notch pathway data and the red arrows indicate the databases, which do not possess their own curated Hedgehog or Notch pathway data. **(C)** The total numbers of signaling pathways available in different databases are shown here.

We also observed that the total number of signaling pathways (human) in different databases is also seemed to be widely varied in different databases (**Figure 11C**). The metabolic or the biosynthesis pathways were not included for this comparative study.

3.3.2.2 Pathway Annotation and Nomenclature is not Standardized

We observed that although the signaling pathway databases assign specific entry number for each pathway, there is no such specific nomenclature exists for assigning the name of the signaling pathways in the databases.

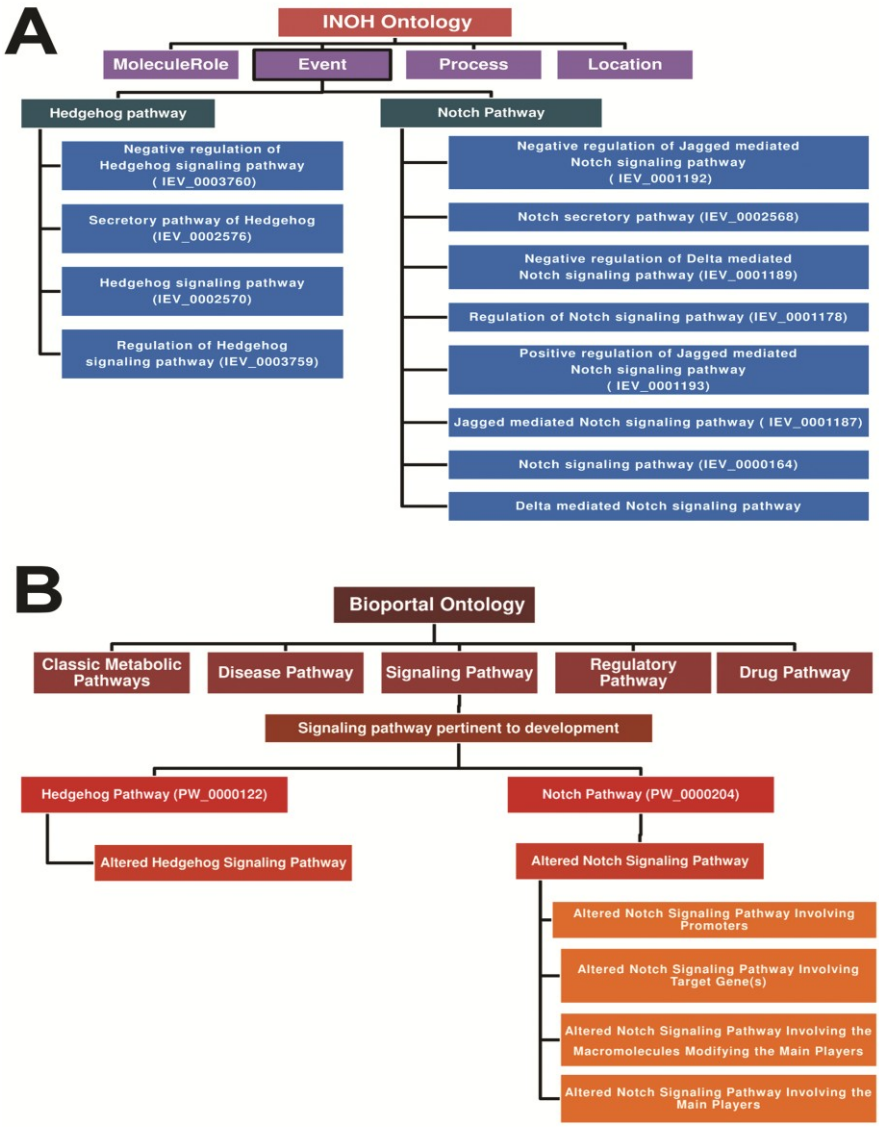

Figure 12: Pathway ontology used by INOH and Pathway Ontology portal.
(A) INOH based ontology and **(B)** Bioportal: Pathway Ontology.

For example, in case of naming Hedgehog pathway, KEGG, PANTHER, WIKI PATHWAYS etc. are using common names, such as "Hedgehog signaling pathway", but the other databases are using different nomenclatures (e.g., BIOCARTA assigns the Hedgehog pathway as “Sonic Hedgehog (Shh) Pathway”; NCI-PID as “Signaling events mediated by the Hedgehog family”, etc.). We noticed that “Ontology (i.e. structured vocabulary or the terms for conceptualization)” based pathway annotation and nomenclature can be a reasonable solution for this problem. By forming the ontology based pathway annotation tree, one can easily assign a specific name as well as the functional annotations for a particular pathway in the database. There are two such types of ontology database exist: INOH [279] and BIOPORTAL: PATHWAY ONTOLOGY [280], which can be used for pathway nomenclature (**Figure 12**).

3.3.2.3 Cross-rreferences of the Signaling Reactions are Mostly Unavailable

We noticed that the data heterogeneity sometimes creates the confusion to choose the appropriate and comprehensive data from a database. Hence, it is therefore the developer's responsibility to minimize all the ambiguities and create the database with high level of authenticity and reliability. We observed that the reference(s) for each included interactions of a particular pathway are most often not available in other popular databases except BIOMODELS, NETAPTH, and REACTOME. As a result, in most cases, the appropriate annotations of the pathway reactions, such as corresponding literature references, its experimental details, reaction stoichiometry, rate parameters, etc., are not easily available. We observed that a large amount of molecular reactions data is still dispersed in many signaling pathway databases without proper annotation.

In this context, iRefIndex, a protein-protein interaction indexing database is useful as it assigns each reaction IDs for all the redundant and non-redundant reactions present in the popular protein-protein interactions databases [192]. Using

the reaction IDs, the pathway reactions or interactions in the pathway databases can be mapped successfully without any redundancy (**Figure 13A**).

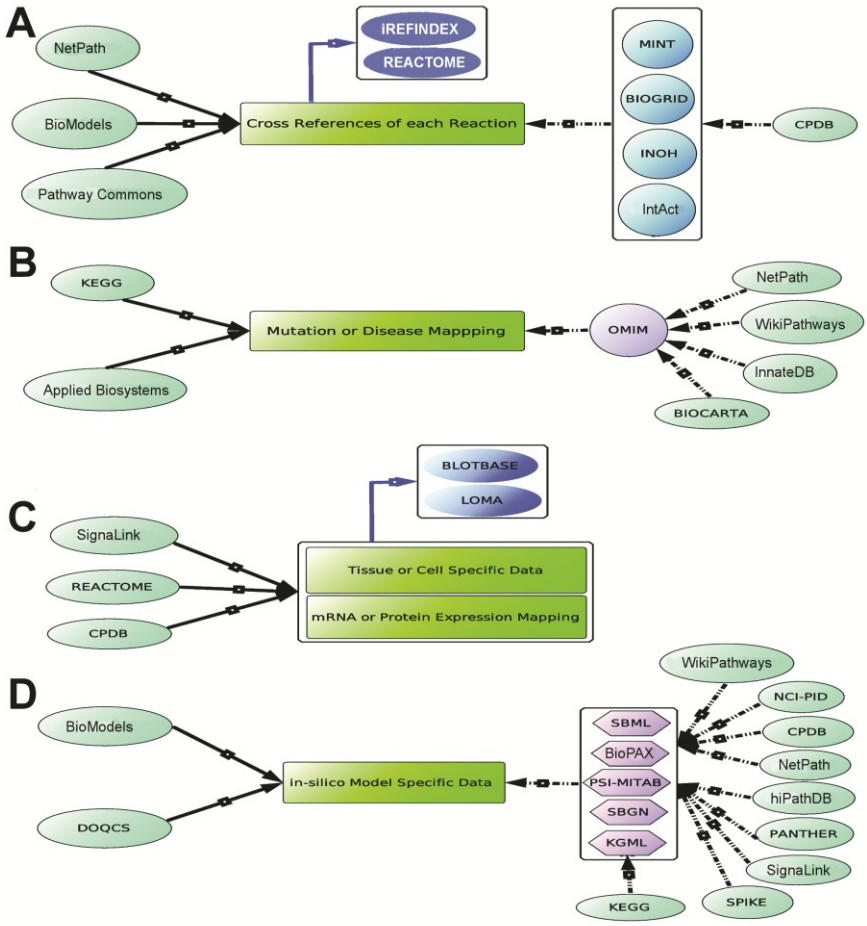

Figure 13: Cross-referencing strategies of pathway data.
The Black, Dotted Black and Blue arrows signify the manual curation, hyper-links and proposed cross-referencing (hyper-linking) strategies respectively. **(A)** Shows the cross-referencing system implemented in different databases. Here, we propose to use iRefIndex, a non-redundant protein interaction database for annotating the interaction. **(B)** Shows the database names which have implemented protein-disease mapping in their databases. **(C)** Shows the databases which map the manually collated tissue specific m-RNA or protein expression data along with the pathway information. Here, we propose to map such data from the third-party databases, like BLOTBASE and LOMA. **(D)** Shows the databases which provide *in-silico* mathematical models of signaling pathways in various formats.

3.3.2.4 Mutation or Disease Related Information are Sparse

We found that there are only two databases, KEGG and APPLIED BIOSYSTEM, which manually annotate the disease specific information with each protein of the pathways. The databases NETPATH, WIKI PATHWAYS, INNATEDB, BIOCARTA hyper-link the pathways with disease specific database OMIM [281]. On the other

hand, REACTOME has nicely categorized the different diseases with various pathways, which are found to be involved in the progression of those diseases in the experimental results. However, it is observed that except few databases, the annotation of protein-disease information is still not much popular and not uniformly adopted in all the signaling pathway databases (**Figure 13B**).

3.3.2.5 Cell or Tissue Specific Pathway Information is most often Found Missing

We observed that SIGNALINK provides the characterized tissue and cancer specific protein expressions data with the pathway information. Another popular database, REACTOME also provides the expressions of the proteins across different tissues linked in a signaling pathway. However, it should be noted that the cell or tissue specific expression provided by REACTOME is provided through another popular microarray expression database: EBI web resource.

Also, the gene enrichment analysis present in CPDB can also be used to analyze the genome-wide gene expression or proteome-wide protein abundance analysis between two different phenotypes using Wilcoxon signed-rank test [282]. Two other databases (not signaling pathway database), *viz.* BLOTBASE [283] and Library of Medical Genomics (LoMA) [284], which provide the northern blot results of different tissues, can also be used to map the proteins/genes within a particular signaling pathway (**Figure 13C**).

3.3.2.6 Unavailability of In-Silico Pathway Models

We observed that there are very few databases, which store the pathway data relevant for the in-silico model development. BIOMODELS serves as the repository of the in-silico computational models of the signaling pathways or the biological processes mined from the published literatures. Using Path2Model [278], it automatically generates the mathematical models and enrich it with the cross references. It is a reliable resource for the model developers as it manually curates the model related data such as kinetic rate parameters, protein concentration,

dynamic rate equations etc. from the published literatures. There is another useful database, DOQCS, which serves the *in-silico* quantitative models for neuron and other signaling pathways. The models provided in this database are collated from peer-reviewed journals and are freely available to the users. The kinetic rate parameters are obtained from various experimental resources, such as enzyme assays, binding experiments, time course of reactions, etc. Both BIOMODELS and DOQCS have authors' assessment system and the users can comment, contact or run the model simulation to test its accuracy in the respective databases (**Figure 13D**).

3.3.3 Comparative Study Based on Technical Features

3.3.3.1 Unavailability of Integrated Computational Tools for in-silico Pathway Analyses

In order to compare different types of online analysis tools available in major databases, this comparison has broadly divided the online tools according their functional modes, which are "Pathway Uploading tools", "Pathway drawing tools" and "Pathway analysis tools". Different databases have found to possess different types tools in their web interface as shown in **Appendix Table 2**. In this table, it is shown that BIOCARTA, WIKI PATHWAYS, REACTOME are only the database, which provide tools to the users or pathway curators to curate and upload the pathway map or pathway related information in their databases. On other hand, to facilitate the pathway drawing KEGG, CPDB, PROTEIN LOUNGE, PATHWAY COMMONS, SPIKE provide various type of pathway drawing tools (desktop or online version) in their databases. On the other hand, the open access pathway drawing tools provided by CPDB, PATHWAY COMMONS etc. are also very useful in this regard. Open source, desktop application, Chisio BioPAX Editor (ChiBE), provided by PATHWAYCOMMONS database is also very useful to edit, visualize and modify the pathway models using BIOPAX format.

3.3.3.2 Multiple File Formats are used for Data Sharing

The types of data present in signaling pathway databases can be broadly

categorized in two parts: Pathway Image and Pathway Description. The pathway image is provided in various file formats, such as PDF, PNG, TIFF, JPEG, SVG, etc. We observed that out of all these formats, SVG has better advantage compared to the other formats as it is xml and vector based image file and therefore much easier to zoom in higher magnification level without losing the image quality. On the other hand the pathway descriptions (molecules, reactions, cellular locations, etc.) are mostly provided in SBML, BIOPAX, PSI-MI, SBGN [285,286] file formats.

3.4 RATIONALES FOR DEVELOPING NEW DATABASE

Followed by the comparisons of pathway data and technical features available in different databases, we observed that different types of signaling pathways are present in different databases and all such data are still heterogeneously distributed across multiple databases (**Figure 10**) [8]. However, it is important to note that all of these databases have evolved in different time points with new ideas of sharing different types of pathway related data, new file formats for data sharing purpose, new applications to display the pathway data (**Figure 9**). Hence, expecting homogenous pathway information with a standard file format from these databases is limited in this case. The systematic reviews on the active cell signalling databases have revealed few limitations of these databases and simultaneously discussed the probable solutions, which could help to overcome the drawbacks [251]. For example, we noticed that the absence of appropriate ontology and standard pathway nomenclature system; heterogeneity in the information of same pathway data across multiple databases; absence of cross references of molecular interactions (i.e. Reaction ID, PPI ID, Literature references etc.); inability to define a standard boundary of pathway reconstruction; unavailability of the information of protein complex formation data (i.e. dimerization, trimerization, dissociation etc.); sub-cellular locations of the molecules and their translocation to various organelles within the cell during signalling events; unavailability of biological context specific (disease, tissue specific, mutation etc.) pathway information; and absence of

advanced computational tools for performing various pathway analyses studies, etc. are the major drawbacks, which are required to be improvised or modified in the current databases [8].

Moreover, hypothesis generation, followed by mathematical formulations of biochemical reactions and its *in-silico* simulations using various mathematical approaches (e.g., topological, logical, and dynamic, etc.) to assess the roles of a biochemical pathway in governing different biological scenarios are also one of the major research interest to the computational biologists [251]. These types of mathematical simulations heavily depend on the data provided by the pathway databases and hence, the availability of computer readable pathway information (i.e., file systems, syntax, schema, reaction parameters, etc.) should also be considered carefully by the data curators [8]. In this scenario, the importance of integration of various computational tools altogether with the database interface could be one of the major up-gradation of these databases from a simple pathway data sharing portal to a pathway data analyses platform [8]. Considering the current scenarios of multi-disciplinary research works in the fields of molecular and computational biology, such type of up-gradation would be always beneficial for a wide spectrum of database users. On the other hand, from the user's perspective, it is worth to mention that the database interface should be more users friendly and interactive for manual as well as automated computer-guided operations [8].

In order to make it possible, the use of advanced and useful database query language (e.g., SQL), appropriate file format (e.g., SBML), and API based web services would be much more effective and useful [251]. However, updating the database through proper annotations of the pathway molecules and reactions with the appropriate sub-cellular locations, and cross referencing with external database sources and literatures are the major problems, which are currently faced by the developers of such databases [8]. As mentioned previously, the deluge of biochemical pathway specific data in various public domains of scientific literatures makes it

almost difficult to the data curators to continuously update the data in their databases regularly. As a result, the pathway reconstructions and simultaneously illustrating the pathway images with the newer information during every update in the database are one of the major challenges faced by the data curators [8]. In summary, the requirements of the advanced platform for pathway data sharing process and the computational analyses tools with more user friendly features for performing various computational tasks are the major demands, which could be included or modified in the existing or in a new pathway database.

3.5 BIOPYDB: A NEW HUMAN CELL SIGNALLING DATABASE

We have taken an initiative of developing a human cell-specific **BIO**chemical **P**athwa**Y** **D**ata**B**ase, "**BIOPYDB**" with the aim of overcoming the current challenges discussed in the previous sections. This database is designed in such a way that after inserting minimal amount of curated raw data into the database, a significant portion of the post-processing tasks is performed automatically and the processed data are stored into the internal database. The raw data of this database are mainly extracted from the manual curation of published experimental articles and from the other popular databases (**Appendix Table 1)**. The post-processing operations (e.g., reconstruction and automated productions of the pathway images, annotations and hyper-linking of pathway components and reactions with other databases, post-processing of computer readable files for pathway data sharing, mapping of protein molecules with different diseases, etc.) are performed automatically with the help of BIOPYDB's in-build dynamic computational algorithms without any manual interventions. Data entry operations such as insertion, deletion, or modification of the pathway data are performed via SQL and the database is based on Relational Database Management System (RDBMS), which allow the entire operations in a more dynamic fashion.

Presently, it is providing the information of 46 different human cells specific,

intra-cellular cell signalling pathways that are involved in various developmental events of cells and tissues, such as cellular growth, tumorigenesis, and immune cell activation, etc. A new pathway ontology and standard nomenclature system of cell signalling pathways are introduced in this database to allocate as well as index the curated pathways according to their biological functions and relevance. Each molecule and reaction of the pathways is automatically hyper-linked with various other resources for further references of the database users. Additionally, disease pathway (currently only Glioma specific pathway is available) is included in this database as a repertoire of biological context based (e.g. disease specific) human cell signalling pathway database. Also, the relationships between different proteins with various diseases (specific to human) are mapped and shown as a network of proteins and diseases. Furthermore, the architecture of the backend of this database is designed dynamically, which could be easily updated and modified after performing insertion, modification or deletions. Biochemical pathway related information such as images, networks, molecules and interactions list, protein-disease mapping, etc. shown in the database webpage are instantly generated from the data stored in the backend of this database and do not require the continuous modification as well as the manual changes in its frontends after any updates in the database. Such dynamic and automated process is specifically helpful to the pathway curators to populate the pathway information without giving any effort for further post-processing operations (see **Section 2.3**).

On the other hand, to develop this database as a platform for performing *in-silico* pathway analyses, useful mathematical tools, such as Network or Topological analyses of pathway networks (using Graph theoretic analysis), Logical analyses (using discrete time, semi-dynamic Boolean equations) and Dynamic analyses (using Ordinary Differential equations) are made available through user friendly interface. The pathway data sharing and analyses platforms are brought together in this database into a single computational framework, which makes it easier for the user

to perform multiple analyses on pathway data with less effort. To access the database from the external computational platform, RESTFul API service is made available for the advanced users and software developers. Using this service, it will be possible to obtain the various types of pathway related data (e.g. list of all pathways, pathway description, pathway image, species and reactions etc.) stored in this database.

In order to welcome the involvement of more number of data curators to collate pathway specific data for this database and to keep the database up to date with current experimental findings, BIOPYDB is also providing the facilities to upload new pathway data and simultaneously analyse it within a single platform. Based on the user's request and approval, the newly uploaded data will be verified by the curators of BIOPYDB and will be stored in the main database as a freely available data to the general users. Such crowd sourcing facility to populate the database will be very much useful to expand and update the database continuously in future. The database is now available in the public domain and common users can also suggest any changes in any of the existing pathway by providing their feedbacks/comments.

In the following sections, we have given the detailed description about the database storage system and architecture, statistics and properties of the curated pathways; implementations and operations of various pathway analyses tools; procedures of using pathway data upload system from the user end; and a brief discussion about the BIOPYDB web interface and its backend (i.e. database schema and objects) structure.

3.5.1 Data Storage System

We used "Relational Database Management System (RDBMS)" in BIOPYDB for managing the entire database architecture. Here, we used MySQL database server to access the stored data using Structured Query Language (SQL). Different types of pathway related data (i.e., molecules, interactions, diseases, etc.) were stored in MySQL as different 'Tables'. These tables are the logical objects of the relational

database within which the relations were established through unique pathway ids or primary key [287]. We stored the pathway specific data (within tables) in such a way, so that after inserting or updating the pathway data, the successive outputs of pathway images, annotation and hyper-linking of pathway molecules with other databases, mapping the molecular interactions with iRefIndex database, or generating the textual data for pathway data sharing etc. will be automatically mapped by the corresponding unique pathway IDs and successively stored in the respective tables of the database. This automated process and the structure-function relationship of BIOPYDB are presented more precisely in **Figure 14**. In this figure, the entire data structure and its relationship with the available features are represented in a simple relational schematic diagram.

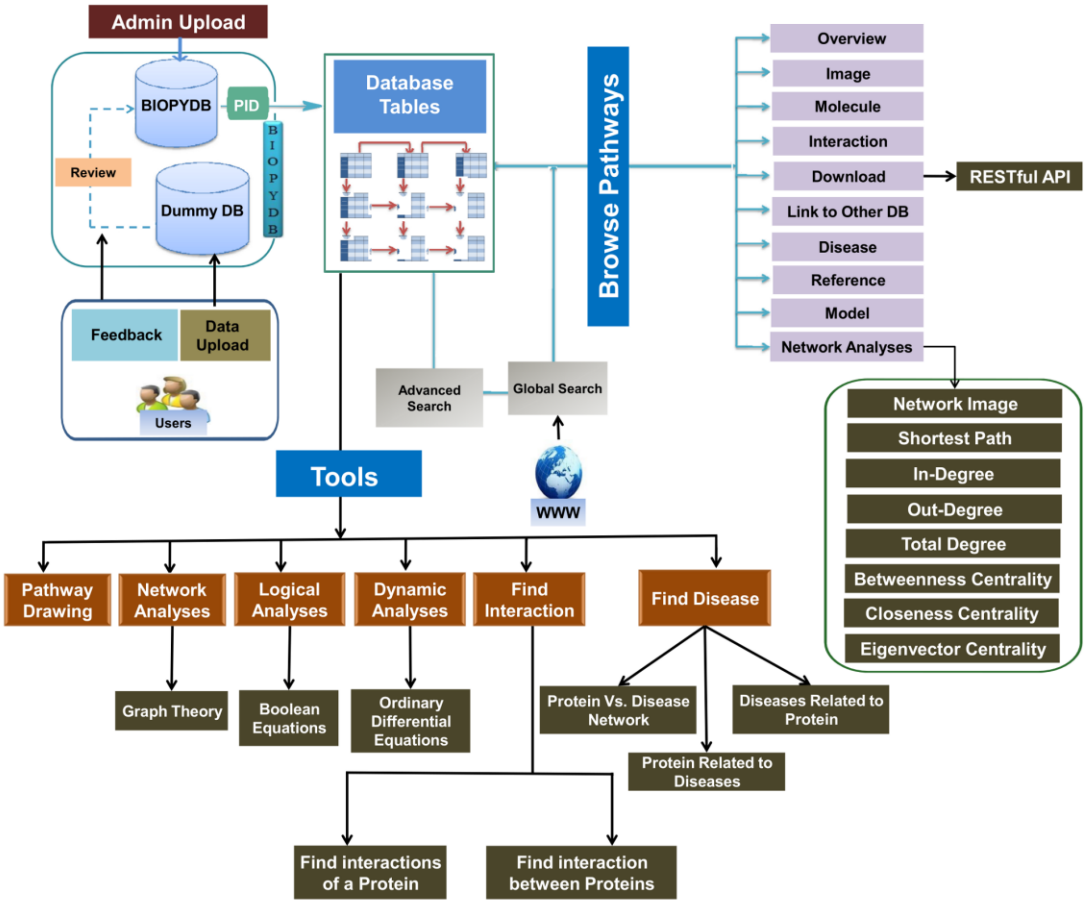

Figure 14: Schematic diagram of the database architecture used in BIOPYDB.

In the backend of this database, it contains the logical objects or 'Tables', which individually store the pathway related data, such as Pathway Name, Pathway Info,

Overview of each of the pathway, the cross-talks of each pathway, Molecules, Interactions, Diseases, information about the developed mathematical models, References, etc. All the tables contain a primary key (Pathway ID), which was used to connect each of the tables while performing the data searching and entry operations in the database. Hence, each and every data in the database, stored in the database tables (pathway molecules, interactions, references etc.), are associated with such unique pathway id. This helps to perform the entire operations (e.g., insertion, deletion, modification etc.) in the database more accurately using structured query language (SQL) with the help of any dynamic programming language, such as PHP. This schema provides this database more fluidity to perform post-processing tasks in an automated fashion by consuming less time and manual interventions.

3.5.2 Web Interface of BIOPYDB

The web interface of BIOPYDB is designed in such a way, so that users can easily interacts and navigate through the database web pages. The "Home" page of the database contains all the necessary tabs like "About BIOPYDB", "Browse pathways", "Downloads", "Tools", "Upload Pathway", "FAQ" and "Contact Us". Home page also provides text "Search" and "Advanced search" options. "Search" option has “auto-suggestion” option enabled, which facilitates more convenient searching of proteins, pathways and diseases. Left panel of Home page contains "News & Updates" to highlight the subsequent updates and the right panel contains "Current Database Statistics" which is automatically updated if there is any changes/insertion or deletion in the database. In **Figure 15**, we have provided the picture of BIOPYDB home page to understand the navigation of different web pages from the frontend.

3.5.3 Resource of Biochemical Pathways

BIOPYDB is currently providing the biochemical, intra-cellular signalling pathway data under four different categories viz. i) Developmental Pathways, ii) Immunological Pathways, iii) Cell Proliferation Pathways, and, iv) Disease Pathway.

In the Developmental Pathways category, two important developmental pathways *viz.* Hedgehog and Notch pathways are included.

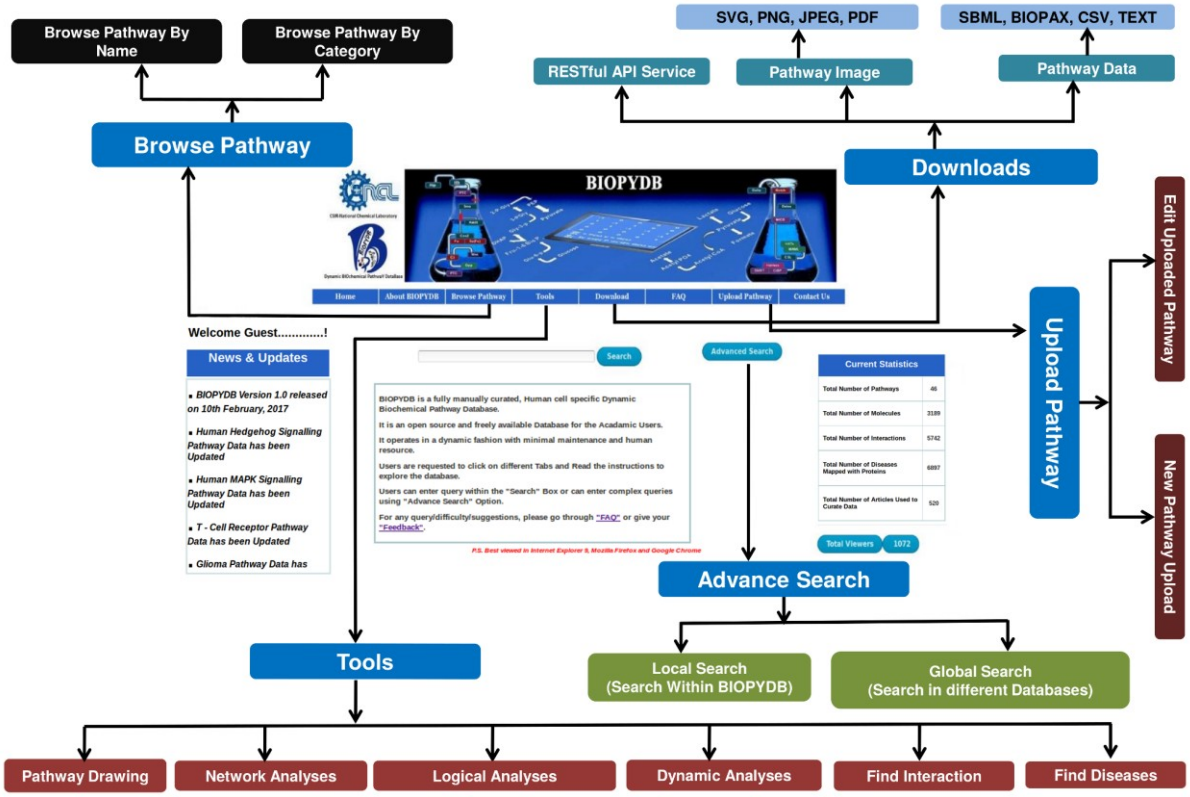

Figure 15: Snapshot of the frontend page of BIOPYDB database.
The frontend is designed in such a way that the users can navigate to various features available in our database. Pathway browse option is available under "Browse Pathway" tab. A global search field along with "Advanced Search" is also available to execute user-specific simple and complex queries, respectively. Using "Download" tab user can also download the pathway images and textual data of the curated pathways available in the database. The information requires for accessing our developed "RESTful API" service is also available in this section. All the tools developed for pathway image drawing, analyses, and searching interaction and disease related information is available in "Tools" section.

Also, BIOPYDB has a rich collection of cell growth regulator pathways eminently represented by the Epidermal Growth Factor Receptor (EGFR/ErbB) mediated Mitogen Activated Protein Kinase (MAPK) pathways under the “Cell Proliferation Pathways” category. The EGFR/ErbB mediated signalling system forms a family of pathways, which is classified here on the basis of ligand-receptor interaction. Apart from the known Growth factors, like the Epidermal Growth Factor (EGF), heparin-binding EGF (HB-EGF) and Neuregulins (NRG1, NRG2, NRG3 and NRG4), a few of

the lesser known ligands of the family like Amphiregulin, Betacellulin, Epiregulin are also included. A unique ligand of neuregulin subfamily, Neu Differentitaion Factor (NDF-β), is also included in this category.

BIOPYDB also contains an exhaustive list of "Immunological Pathways" that includes pathways related to both innate (e.g., TLR pathways) and adaptive immunity (e.g., T cell signalling pathways). It contains a repertoire of cytokine signalling pathways as well, that includes different Interleukins, Interferons, Tumor Necrosis factors, Tumor Growth Factors and Colony Stimulating factors, all of which plays a pivotal role in regulating the immune-regulatory network. To provide the context specific pathways, BIOPYDB also provides the signalling pathways, which are found to be deregulated in a particular disease scenario under "Disease Pathways" category (see **Section 2.3.1.1**). The malfunctioned signalling pathways, which are only stimulated by PDGF, IGF1 and EGF ligands in Glioma, are currently provided in this section. Later, it also intends to provide all other possible pathways, which are responsible for the development of Glioma in human. All these categories of pathways are easily accessible from the "Browse Pathways" tab provided in the home page of BIOPYDB (**Figure 15)**.

To access the information of the pathway of interest, the users are required to click on the corresponding link of the pathways in the pathway browse section. A small description of each pathway, including the pathway statistics (number of molecule, interactions, articles, diseases etc.), will be available in the "Overview" section of the pathway data. Besides, the phenotypic expressions (e.g. cell division, cell growth, apoptosis etc.) and cross talks with other pathways with proper literature references will be also available in this section for each pathway.

3.5.4 Resource of Proteins/Genes Involved in Biochemical Pathways

The users can also use BIOPYDB as a resource of signalling protein(s)/gene(s) involved in the intracellular signal transduction cascades with appropriate literature

evidences. Currently, it contains a total number of 2748 such proteins/genes, which can be easily accessible through the general 'Search' as well as "Advanced Search" option provided in "Home" page (**Figure 15)**. The query result gives a short description of the protein, as well as the link to access its genetic and amino acid sequence data from NCBI-GENE and UNIPROT, respectively [288,289]. Users can also get the links of the proteins from other pathway resources like KEGG, WIKIPATHWAYS, and HUMANCYC databases [268,290,291]. Besides, it provides the hyper-links of the Protein-Protein Interaction databases from the database PIP; Disease related data from GENECARD database, and tissue specific expression pattern data from TiGER database [188,191,292].

3.5.5 Resource of Protein-Protein Interaction Data

We have also made available of the Protein-Protein interaction data with network representation and proper references through the "Find Interaction" application present in "Tools" section (**Figure 15)**. The type of interactions (i.e., chemical nature) of each interaction is also available with each of the interaction. Currently BIOPYDB has included 25 different types of reactions/ interactions/ connections in the signalling networks. The types of interactions are Cholesterol Modification, Physical interaction, Inhibition, Phosphorylation, Activation, Nuclear Translocation, Transcriptional Co-repression, Protein Production, Transcriptional Co-activation, Protein Recruitment, Auto-phosphorylation, Ubiquitination, Transcriptional activation, Complex formation, Stimulation, Dissociation, Dephosphorylation, Homodimerization, Heterodimerization, Deubiquitylation, Calcium exchange, Acetylation, Phospholypase reaction, Proteolytic cleavage, and Enzymatic Reaction [293-300]. We think that the information related to the type of chemical reactions or interactions occurring between two pathway molecules present in a pathway will help the users of BIOPYDB to understand the biochemical processes regulating the signalling cascades inside the cell. Moreover, to further cross verify the interaction data, we have also mapped each interaction of the database with the iRefIndex

database [192], which consists non-redundant molecular interaction data collated from different protein-protein interaction databases (see **Section 2.3.1.5**).

3.5.6 Resource for Visualizing the Protein-Disease Mapping

It is well known that certain malfunctions caused by the mutation of proteins in biochemical pathways could be responsible for various types of diseases in human body [301-303]. Therefore, it is worth to mention that protein-disease mapping of the protein molecules in the signalling pathways is also important to understand the importance of that pathway in disease pathology. To include this feature into the database, we have provided an interactive interface in BIOPYDB to dynamically fetch the information of various diseases from protein-disease mapping database: MalaCards [196] and made possible to display the results in more user friendly way. However, it should be noted that all the information provided in this section is solely owned by MalaCards database and we do not hold any claim related to the protein-disease mapping data. Here, we have simply provided a technical resource for fetching and visualizing protein-disease network as image and tabular formats within the BIOPYDB interface. The diseases associated with each protein of the pathway of interest are automatically hyper-linked with the MalaCards database ID. Currently, BIOPYDB has dynamically mapped around 6897 diseases with the proteins present in the pathways in BIOPYDB database. The disease related information is available through the "Disease" tab under the "Browse Pathway" option (**Figure 15)**. It is also available through the "Find Disease" application available in the "Tools" Section (**Figure 15)**.

Presently, we have included a total number of 3187 molecules (Proteins, Protein complexes, Inorganic molecules, Mutated proteins, Secondary messengers, Phospholipids and Lipid molecules) and 5740 molecular interactions or connections of 46 manually curated pathways in our database (**Table 6**). For accuracy and authenticity of the pathway data, we have also provided the literature reference

(with Pubmed ID) for each and every interaction included in the database. Users can cross verify the pathway data immediately by accessing the hyper-links provided in the "Interactions" tab under the "Browse Pathway" option. We have also made it possible to cross check the pathway data in similar 17 different databases by automatically searching the pathway links (see **Section 2.3.1.6**).

Table 6: BIOPYDB Database Statistics

Data types	Statistics
Total Number of Pathways	**46**
Total Number of Molecules	**3187**
Total Number of Interactions	**5740**
Total Number of Diseases Mapped with Proteins	**6897**
Total Number of Articles Used to Curate Data	**520**

3.5.7 Computational Platform for Pathway Data Analyses

One of the unique features of BIOPYDB is its in-built pathway data analyses platform for performing Network (using Graph theoretic analysis), Logical (using discrete time, semi-dynamic Boolean equations) and Dynamic (using Ordinary Differential Equations or ODEs) analyses of biochemical pathways [82,304,305]. No other similar resources/databases provide this wide ranges of computational tools altogether in a single platform to the common users. Moreover, to make this database more interactive and user friendly for the large community of experimental and theoretical biologists, we have included various technical features and applications to search, retrieve, annotate or analyse the data stored in the database. The entire resource is easily made executable and do not require any separate hardware or software installation in the user's local machine. To view the pathway image and network, users are required to use a SVG compatible web browser, which is now available in any modern web browsers, like Internet Explorer (version > 9), Mozilla,

and Google Chrome etc. A schematic diagram showcasing all the available tools and the stepwise guidelines of their operational protocol are presented in **Figure 16**. A brief description of all the technical features available in BIOPYDB is also provided in **Appendix Table 3**. Following is the brief descriptions of the technical or computational features available in BIOPYDB.

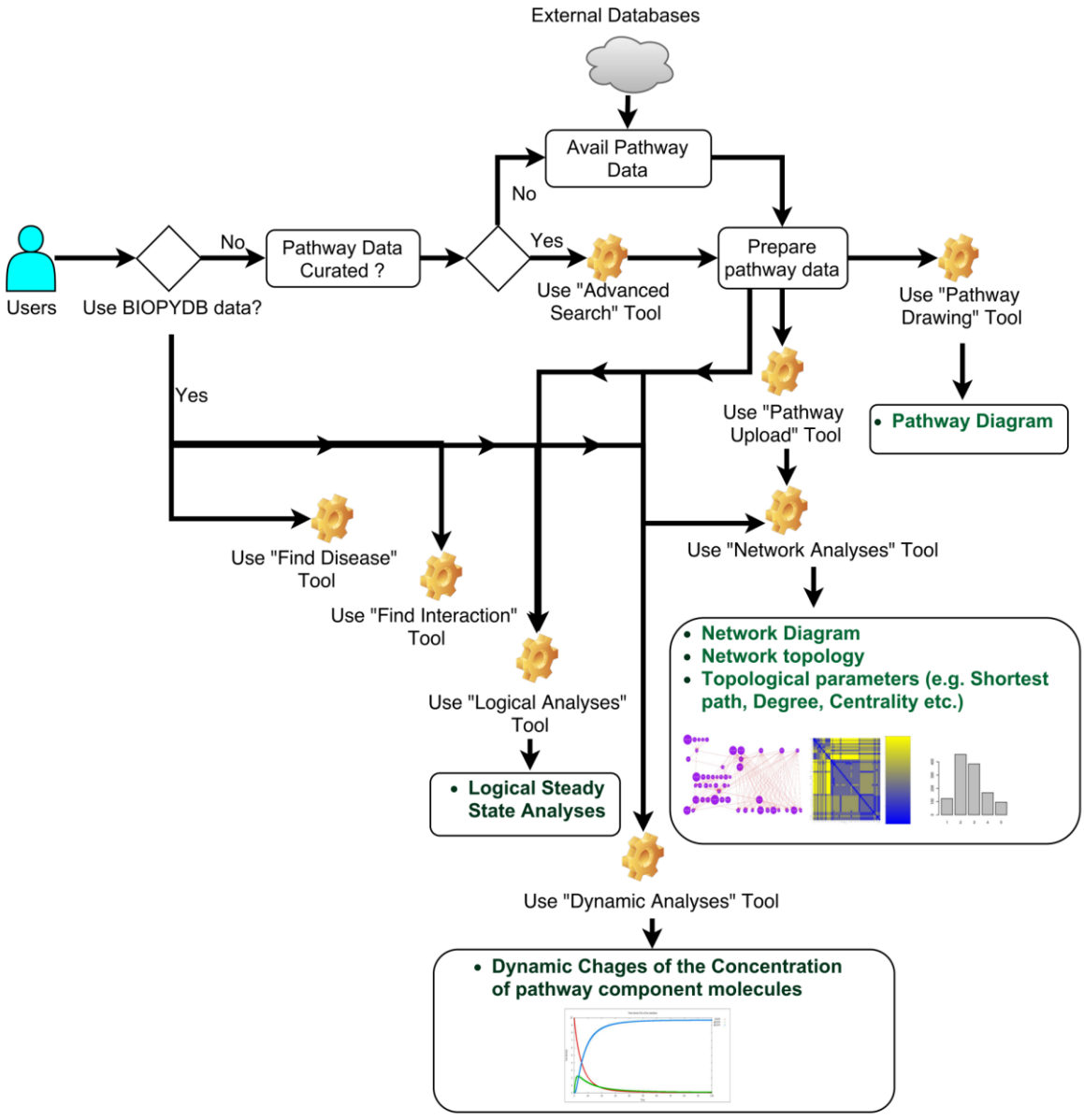

Figure 16: Flow-chart describing different operations executed in BIOPYDB.

3.5.7.1 Searching/Browsing Pathway Data

We have provided the global text search option in the home page of BIOPYDB (**Figure 15)**. The search box includes "auto-suggestion", which suggests the pathway

names, proteins or diseases from the database while typing into the search box and thus expedites and eases the text searching procedures. Users can also browse and select a pathway of interest by using "Browse Pathway" option. The "Browse Pathway" option has two different tabs for browsing the pathways "By Name" or "By Category". After selecting a pathway from any of these options, users will get the pathway details which are provided in 10 different tabs: Overview, Image, Molecule, Interaction, Download, Links to Other DB, Disease, References, Model and finally Network Analysis. Users can click on any tab and get the relevant information about the pathway. We have also included the advanced search option to execute complex and multiple queries through this database (**Figure 15)**. Search queries, such as, pathway name; drugs or inhibitors of a pathway; disease caused by the mutation of proteins; articles searching operations using Pubmed IDs, E.C. number, etc. can be executed in this section. Using this unique application of BIOPYDB, users not only can search in BIOPYDB database, but also can perform search queries to the other 17 similar signalling pathway resources, as well as in PUBMED, ExPASy, BRENDA, METACYC, and GENECARDS databases. We believe that this unique advanced search query options will make this database more attractive to the users for collating diverse sets of biological data from a common web application platform.

3.5.7.2 Pathway Image Drawing

We have introduced this application to provide an automated tool for pathway drawing and annotation. Using this tool the users do not need to manually draw the structured pathway diagram; instead it can automatically annotate and draw the molecular connections between the pathway components based on the information provided by the users. Simultaneously, it will render the pathway image in SVG format and subsequently host it into the web browser. To use this application, users are required to simply insert the molecular entities and their binary interaction or connections in the specified fields mentioned in the web interface of this tool. Users can also specify the different types of inserted molecular entities (e.g., Protein,

Protein complexes, Secondary messenger, Inorganic molecule, Mutated protein and Phospolipid), its sub cellular locations in the cell (e.g., Extracellular region, membrane, cytoplasm, nucleus) or can allocate the molecules in "output" section by using the simple drop down list available in the pathway data upload form. Similarly in the interaction section, we have enlisted a total number of 25 different types of molecular reactions, which can be used by using the drop down menu. For better visualization and image rendering, users are required to provide at least four molecules from four different sub-cellular locations and their interactions in the specified fields. Users can add or delete any molecules or the reaction in this dynamic web page form. After filling up all the information, user has to click on the "Draw" button, which will instantly generate a pathway image in SVG format and the pathway image will appear in the web browser. Here, the different types of molecule and interactions will appear in different colours according to their molecular properties and thus it helps to differentiate the reaction cascades more appropriately in the newly generated image.

3.5.7.3 Network Analyses

We have introduced this application to perform topological analyses of the biochemical pathways using graph theoretical method [82,208,304]. Using this tool, user can calculate various network parameters, such as *Connectivity parameters* (all pairs shortest paths, in-degree, out-degree, total degree and their corresponding distributions), and *Centrality parameters* (Eigenvector, Betweenness and Closeness centrality of a pathway network). All pairs shortest paths (of each molecules of the signalling pathway) are presented by a matrix heatmap, and the other parameters values are shown by bar diagrams. To execute this application, at first the users are required to select the pathway to view the list of the binary interactions of the pathway molecules on which the topological analyses will be performed. After that, the user can directly run the graph theoretical simulation on the provided binary interactions or can add or delete any pathway species and its corresponding

interactions in the provided list and then can run the simulation. The inclusion of this feature into the user interface makes it a useful platform to perform the *in-silico* protein knock-in/out experiments in a signalling network, where one can simultaneously observe the variations of the different network parameters. Each simulation will provide the network picture, average values of the network parameters, and the bar plots of each parameter values for each pathway molecule of the network. Users can also download the network picture and bar plots in SVG, PNG, PDF or JPG formats.

3.5.7.4 Logical Analyses

In order to perform semi-dynamic logical (synchronous, discrete time updates) analyses of a signalling cascade, we have incorporated the Logical analyses tool in BIOPYDB. This application is based on the in-build "Boolean Analysis" simulation engine used for analyzing the activity (i.e., up-regulation and down-regulation) of pathway molecules in the biochemical and gene regulatory networks [82,244,306-308]. This application has three sub-sections, mainly divided for three different purposes. The first section allows the users to perform Boolean analysis of the pathways, which are present in BIOPYDB database. The second sub-sections allows the users to perform Boolean analysis of a new (user defined) pathway and the third sub-section is used for revisiting the previous Boolean analyses data by using the Job ID, which was provided to the user at the first time of performing the simulation.

In order to use this application, at first the users are required to select the appropriate sub-section. If the user selects the first sub-section, a pathway list available in the BIOPYDB database will be appeared in the webpage. After selecting the name of the pathway of interest, the user will be redirected to the next page in which the users are required to provide initial values (binary values i.e. either 1 or 0) of the molecules (or nodes). If the initial values of the nodes are not provided in this section, then it will assign the random binary inputs (either 0 or 1) to the

uninitialized nodes as defaults. The simulation engine running in the backend of BIOPYDB will instantly provide the Boolean equations of the intermediate molecules (i.e. the nodes/molecules which has upstream regulator in the pathway network). It should be noted that the Boolean equations generated in this section for the pathway species may not reflect the exact biological scenario as it is computationally generated from the 'Interaction' table by using the information provided in the binary interactions of the pathway of interest.

User can also easily add new nodes and equations or modify the Boolean equation of the existing node in this interactive web interface. After confirming the initial values and the logical or Boolean equations, users are required to provide the total time steps (not more than 500) up to which the Boolean simulation will be performed. If the user does not provide this field then a default value (time steps 30) will be assigned to the submitted simulation. The user can also select the nodes of interest for which the time variation would be analyzed. After completing these steps, user can now submit the job in BIOPYDB web server. The logical simulation is performed with the help of popular python package BooleanNet [309]. A unique job ID will be provided to the user after submission of the simulation and the outputs will be provided in the next page. The nature of the stability (stable or cyclic attractor) of the simulation outcomes will be provided including the other parameters such as total transition states, number of input nodes, cycle length (if any), attractor length etc. The state transition data can also be downloaded as well as viewed in the web page. Temporal dynamics of the nodes of interest selected in the previous section will also be shown in image and can be downloaded.

Similar analyses can also be performed on a new (or user defined) pathway using the second sub-section of this application. Unique job ID will also be provided in this section, which can be viewed later by providing the job ID in the third sub-section of this application. The results of a performed simulation will be stored in the database for the next six months and until then it can be viewed by this application interface.

3.5.7.5 Dynamic Analyses

We have also added the dynamic analysis tool to simulate the continuous temporal variations of the pathway species under a specific biological condition. This application is based on the numerical analyses of the coupled Ordinary Differential Equations (ODEs) based on the reaction/enzyme kinetics of the biochemical reaction cascades [305,310]. This application can be performed on the existing pathway of BIOPYDB or on a new pathway model provided by the user. In order to perform the dynamic analyses of the existing pathway, users are required to select the pathway name from the drop down list. After that, users will be asked to select the variables (molecular species) and its corresponding initial concentration in the html format used in the application interface. Simultaneously, users are required to provide the reaction kinetic equation (functional forms) of each selected variables/species of the pathway, which should be based on its functional relationships or reaction orders with the other interacting species.

Users can also delete or insert any new molecule in the variable list and can execute the dynamic simulation of any number of variables. The users are also required to provide the numerical values of the rate parameters of the kinetic equations used for modelling the pathway reactions. Moreover, the users will be asked to provide the initial and final time (in seconds) up to which the simulation will be run and the time step at which the data for the time series will be saved. After submitting all these information, the user can submit the dynamic model into the BIOPYDB simulation engine by using the "Submit" button. The simulation outputs will show the time series plot of all the pathway components/variables within the time span provided by the user. On the other hand, to execute the similar dynamic analysis on a completely new pathway model, user can select the "Ordinary Differential Equation of a new pathway" tab and then follow the same procedures described above. Depending on the size of the model (i.e. the number of independent variables and the complexities of the equations), the total time require for executing

the entire simulation may vary. The fourth order Runge-Kutta numerical integration algorithm (implemented through C/C++ programming language) is used here for quick simulation of the dynamic models [310].

3.5.7.6 Find Interaction and Disease

Using this user-friendly application, user can search either all the interacting proteins of a particular protein of interest or the interaction between two specific proteins from BIOPYDB interactions data sets. In each case, it will give the interactions of the protein(s) as a directed network including a table with the appropriate Pubmed and iRefIndex links of each queried interaction. Users can then directly check the literature reference of that queried interaction.

This application is useful to find the diseases associated with a particular protein in the signalling network. This tool is divided into three parts, one for searching mutated proteins associated with a disease, second for searching diseases which are associated with a particular protein of interest and another is for searching the connection between a particular protein and disease from the protein-disease database: Malacards. BIOPYDB provides a beneficial interface for accessing, retrieving and viewing the protein-disease mapping data from this application. The Malacards database accession IDs are provided for each query performed through this tool.

3.5.7.7 Data Download Application

In order to facilitate the data downloading process more efficiently from the database, we have developed a useful application written in PHP for quick data download. It is unique because each time, upon user's request, this application initially generates the downloadable data (pathway image in SVG, PNG, and JPEG format or interaction data in Tab delimited, SBML or BIOPAX formats) from the raw data stored in the database and then serves for further downloading process. It allows the users to obtain an up to date data from the database and thus the

developers of the database do not have to bother about the post-processing and archiving the pathway data every time for data download purpose. This automated process also makes the database maintenance much easier and less time consuming.

Also, to make the database more accessible through various platforms (i.e., browsers, command line or any other third-party software), we have also included the RESTful API service. Under the "Download" section, we have provided the elaborate descriptions of the protocols, which are required to use our API service. Users can access and obtain various pathway related data through browser or command line programming using our API. It can provide the list of pathways with the BIOPYDB accession ID numbers; pathway images in SVG, PNG and JPEG encoded formats; pathway description and statistics encoded in XML format; and information of pathway species, reactions, and disease names encoded in JSON format. A list of URLs is available for obtaining such information through API in the database webpage under "Download" section (**Figure 15**).

3.6 DISCUSSION

It is worth mentioning that in BIOPYDB we have developed a unique repertoire of cell growth regulation, developmental and immune Signalling pathways. This database provides a comprehensive understanding of the signalling pathways and their important roles in regulating the human cells/tissues. These pathways are very much useful to understand the de-regulations involved in diseases related to abnormal cell proliferation, immunity, birth defects and various others pathologic conditions. The exhaustive list of Interleukin pathways helps the users of BIOPYDB to gain a detailed in-sight into the signalling mechanisms of the immune-regulatory network that plays a pivotal role in understanding both infectious diseases and cancer.

In this database all the pathways are manually curated and possess up to date data synced with the current experimental findings. Cross validation of the pathway

data is also available through the same platform with the other similar databases. We have used a new ontology tree, based on the pathway functions, and successively categorized the pathways accordingly. We have introduced a unique pathway nomenclature system this database to eliminate the ambiguities found in the existing databases [8]. In terms of the quality of data, the developers of BIOPYDB have rigorously cross-checked the pathway data (properly supported with existing and up-to-date literature) before including it in the main database. We have collated more number of targets or output proteins/molecules in each of the pathway by deep literature mining, so that a comprehensive picture of the target or output products of biochemical reaction cascades could be shown and linked with various phenotypic outcomes of a cell or tissue. We have also included the intermediate molecules in more numbers so that the cross talks with other pathway molecules could be easily and accurately depicted in the pathway diagrams. For example, to show the Notch signalling events, its cross talks with other important pathway proteins, such as JAK/STAT, HIF1, and P53 are also included. Inclusion of this information is required as these cross talk molecules play major roles in regulating a signalling pathway and hence it would be incomprehensive, if these proteins are not included in the Notch signalling pathway data. The definition and the boundary of the reconstructed pathway diagram is not restricted to the core molecular entity of the newly reconstructed pathway, rather it is expanded as much as possible by including more number of the cross talks through deep literature mining to provide a comprehensive knowledge of the pathway reactions. Due to this reason, the number of molecules of a particular pathway shown in BIOPYDB is higher than any of the similar pathway databases.

Another strong merit of BIOPYDB is that it not only provides the biochemical pathway data, but also provides a common platform to draw new pathway image as well as analyse Topological properties (such as Shortest Path between two proteins, Distribution of shortest paths, In-Degree, Out-degree, Total Degree and their

distributions, Closeness, Betweenness and Eigenvector Centrality), Logical analysis (using discrete, semi-dynamic Boolean equations) and Dynamic analysis (using Ordinary Differential Equations) of the existing as well as new, user defined pathways [82,185,263,304,305]. Moreover, the availability of Protein-Protein and Protein-Disease data (which are displayed as network view) provides an added benefit to this platform, which is not available in any other databases.

Several novel features, such as, automatic annotation or mapping of pathway molecules to other databases, network representation of protein-disease network, advanced data searching across 17 similar pathway and drug databases, user friendly pathway upload and update system for users, instant view of uploaded data by the users, dynamic pathway image drawing, runtime generation of pathway data for download, pathway data analyses platform to perform structural and dynamic analyses, sharing of pathway data in computer readable file formats (SBML and BIOPAX), and its potential usefulness in academic research, experimentalists, computational and theoretical biologists, etc. separates out BIOPYDB from other existing databases. The database is also made accessible through RESTful API service, which also helps the other third party software and database developers to use and analyze the BIOPYDB data.

3.7 CONCLUSION

We have developed BIOPYDB as a useful resource of biological data of human biochemical pathways; annotated proteins/genes that are involved in the pathways; protein-protein interactions, and protein-disease mapping. All the available data is manually curated from several literature resources and is available free of cost through web browser and RESTful API services. BIOPYDB is a resource or repertoire of biochemical pathway data and also provides a method or strategy to make similar resources or a multi-functional platform, which would be easy to update and maintain, user friendly, dynamic. It provides a single computational platform that

would be useful for multiple tasks from data searching to data analyses. In short, BIOPYDB is not only a database, rather it is an information system, integrated with its own manually curated biochemical pathway database, a computational platform for accessing pathway data from the external sources, and providing web-services for *in-silico* pathway data modelling and simulation analyses. The data curation process of BIOPYDB is steadily increasing and in near future it also promises to add more pathway information by including other types of pathways (such as metabolic, gene regulations etc.). The long-term goal of this database is to develop a common pathway data searching and computational analyses platform for performing biochemical pathway based modelling and simulations.

Chapter 4

4 UNDERSTANDING OF HEDGEHOG SIGNAL TRANSMISSION DYNAMICS IN CANCER CELLS

4.1 INTRODUCTION[3]

Hedgehog is an evolutionarily conserved developmental pathway, widely implicated in controlling various cellular responses including various cancers (e.g., basal cell carcinoma, medulloblastoma, glioblastoma, colon, and pancreatic cancers, etc.) [30,301,311]. There are three ligands of this pathway viz. Sonic Hedgehog (SHH), Desert Hedgehog (DHH) and Indian Hedgehog (IHH), which play cardinal role of initiating the Hedgehog signaling cascades in different cells. We have also observed that SMO (Smoothened) receptor protein is the most preferred drug target in Hedgehog pathway to treat various types of cancers [312]. It has been observed that SMO is a membrane-bound G-protein coupled receptor protein (GPCR), which is required to translocate at the membrane of primary cilia for phosphorylating STK36 protein in the downstream region (**Figure 17)**. In the absence of Hedgehog ligands (i.e., SHH, IHH, DHH) this process is inhibited by another membrane bound proteins PTCH1/2 and therefore SMO cannot translocate towards the primary cilia after its synthesis. In this situation, cytoplasmic protein SUFU makes complex products with GLI1/2/3 and triggers proteosomal degradation of GLI transcription factor proteins. Transcription repressor protein GLI3R is produced at this stage, which further translocates to the nucleus and inhibits the transcription of Hedgehog target genes responsible for cell renewal, cell proliferation, migration, etc. Thus, PTCH1/2 acts as repressor of Hedgehog pathway activation as well as tumor

[3] The materials of this chapter has been taken verbatim from our previously published article (a) Chowdhury et al., *PLoS ONE*, 2013

suppressor protein [82,313-315].

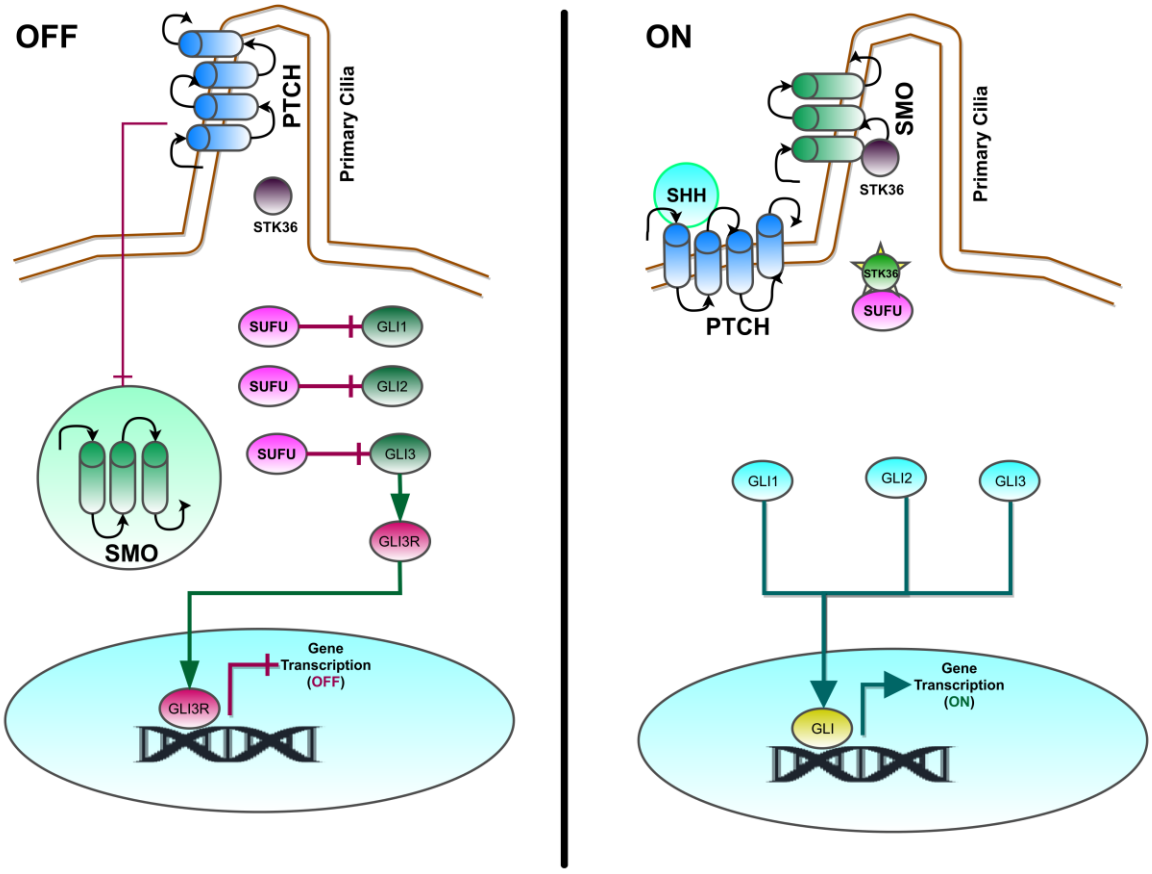

Figure 17: Hedgehog signaling activation mechanisms in OFF and ON states.
This schematic diagram represents the mechanisms describing how SHH or other Hedgehog ligands interact with PTCH and activates the translocation of SMO towards primary cilia. In OFF state, GLI3R (transcription repressor) is produced from GLI3 in cytoplasm, which translocates in nucleus and suppresses the Hedgehog target gene transcription process. In the ON state, GLI1/2/3 proteins are released from SUFU and starts transcription of various genes, such as CYCLIN D2, FOXM1, SFRP, etc.

On the other hand, in presence of Hedgehog ligands, PTCH1/2 proteins make complex products in the cell membrane. In this complex form, PTCH1/2 cannot laterally diffuse towards the primary cilia and thus SMO experiences no inhibition of its translocation towards primary cilia. Subsequently, SMO phsophorylates its downstream kinase protein STK36 in the primary cilia and the phosphorylated STK36 further phsophorylates SUFU at the GLI protein binding site [82,313-315]. As a result, the phosphorylated SUFU cannot further bind with GLI1/2/3 proteins. Thus, the GLI transcription factors become free and further translocate into the nucleus, where these protein starts transcriptions of various target genes, such as CYCLIN D2,

FOXM1, SFRP, JAG2 etc. (**Figure 17)** [82,316]. Controlled regulation of this pathway activates these target genes at certain level and thereby maintains the proper development of cell or tissue [317]. But deregulation of this mechanisms can cause uncontrolled expressions of these target genes and may cause severe outcomes in tissue or organ development [318]. Since, this pathway is also strongly implicated in self-renewal of adult stem cells, therefore the system-component malfunctioning of this pathway can lead to cancer in various cells/tissues of human [83,319].

Moreover, the role of few important proteins has been identified in this pathway, such as PTCH1, SMO, and GLI etc., which are mainly responsible for the malfunctioning of this pathway in various types of cancers [320,321]. We have observed that the follow-up studies by several research groups have developed various therapeutic strategies to inhibit the activation of this pathway by inhibiting these proteins in various cancers [322,323]. However, none of them have achieved complete success to cure a particular cancer caused by the abnormal activation of the Hedgehog pathway. Therefore, it is indeed a growing concern to the cancer research communities to find out alternate/combination of drug targets in Hedgehog pathway to treat various oncogenic anomalies developed in various tissues of human body.

4.2 STATE-OF-THE-ART

A recent review by Li et al. have underscored the importance of combinatorial drug targets to inhibit Hedgehog signaling network in cancer treatment [324]. It is known that activation of cytoplasmic GLI (zinc finger transcription factor) protein, which is the main target transcription factor of this pathway, is initiated by two different mechanisms, *viz.* (i) the ligand dependent mechanism in which the extracellular response i.e., hedgehog ligands directly interact with receptor proteins PTCH1/PTCH2 and activates G-coupled protein SMO (i.e., canonical process) (**Figure 17**), and (ii) the malfunction of the other proteins that are present in the cytoplasm which inhibit or activate GLI protein activity in the absence of hedgehog ligands (i.e.,

non-canonical process). Moreover, certain polymorphic variations in SMO can also constitutively activate its activity in the cell, which in turn over-activate Hedgehog pathway even in the absence of any Hedgehog ligands [312]. We have observed that most of the studies have mainly focused on the development of small molecule inhibitor/drug which is useful to impede the canonical activation mechanism. In order to do that most of these studies are only directed towards the identification of the drug molecules that could suppress either PTCH1 or SMO in the cell membrane [16–19]. We have found that these types of drugs, such as Cyclopamine, Vismodegib, etc. are effective when SMO receptor protein is over-expressed in the primary cilia of the cancer cells [325].

A Phase-I clinical trial performed in the patients with solid tumors (Basal Cell Carcinoma, Pancreatic Cancer, and Medulloblastoma) refractive of current chemotherapeutic drugs had shown promising results while treated with anti-SMO small molecule inhibitor Vismodegib (GDC-0499) [326]. After its successful clinical trials, FDA had approved Vismodegib (GDC-0499) and Sonidegib (anti-SMO molecules) for the treatment of Basal Cell Carcinoma (BCC) patients. In spite of its successful implications in target-based cancer therapy, the follow-up studies have found that a relatively higher number of patients undergoing this target-based therapy would eventually develop resistivity against these anti-SMO molecules [327]. It is observed that genetic mutations in SMO preserve the activity of Hedgehog pathway and thus confer the drug-resistivity of the cancer cells [328]. Apart from targeting SMO receptor in cancer cells, researchers have also tried to inhibit the downstream effector proteins (e.g., GLI) in the Hedgehog pathway to treat cancers. However, such research works are still in the nascent stage and none of the targets or inhibitor molecules are approved for therapeutic usage [329].

4.3 RATIONALE AND OBJECTIVES

Overcoming the drug-resistivity while targeting the aberrantly activated

Hedgehog pathway in the early and advanced stages of cancer cells is one of the major challenges in target-based cancer therapeutics (**Figure 18**) [330]. There is an unmet need to find out alternative drug targets for inhibiting Hedgehog pathway activity in the cancer cells and for which we require to implement the advanced approaches in target identification process.

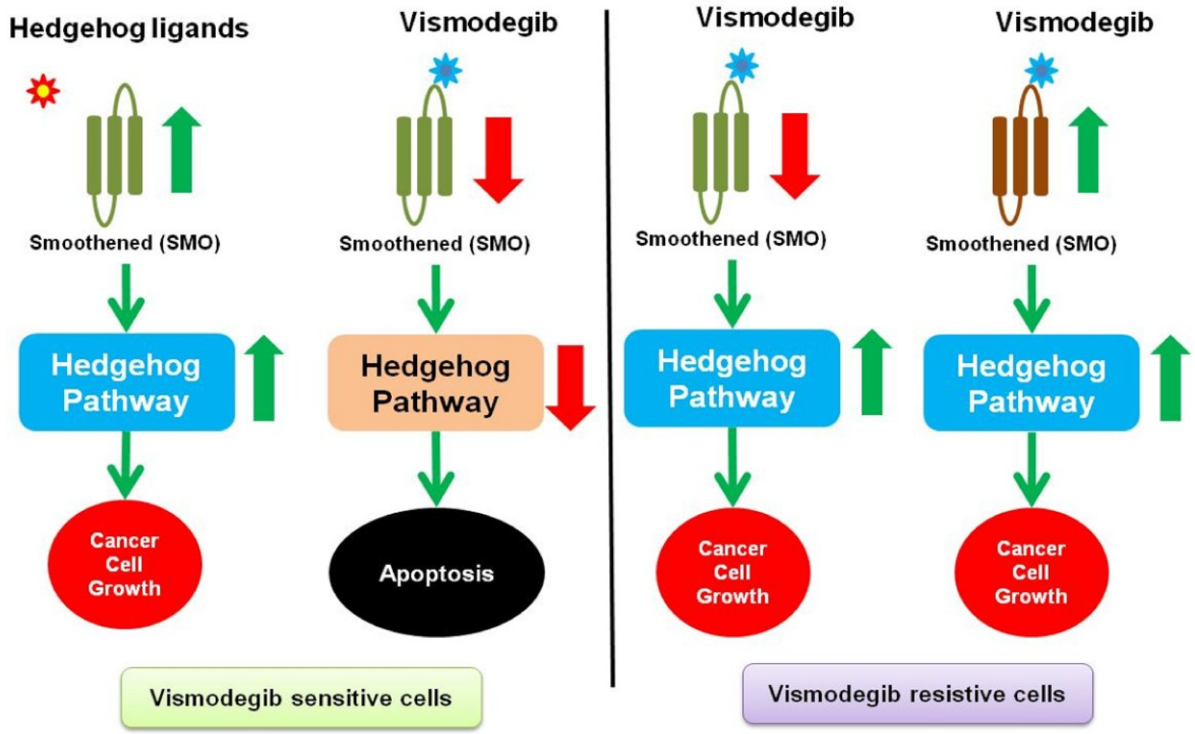

Figure 18: Reaction cascades in Vismodegib sensitive and resistive cancer cells.
In the Vismodegib sensitive cells (with wild-type SMO), cancer cell growth is activated in the presence of Hedgehog ligands, whereas in the presence of anti-SMO molecule Vismodegib, the cancer cells undergo apoptosis. On the other hand, in Vismodegib resistive cells with mutated SMO, the cancer cells do not show any apoptotic effect.

Therefore, the main challenge is to understand the underlying mechanisms in the downstream processes of Hedgehog pathway and subsequently find out suitable target to inhibit a significant portion of the reactions cascades [331]. In this work, we sought for the investigations of the reaction mechanisms involved in Hedgehog signaling network in normal and cancer scenarios, and thereby identification of the suitable drug target or combinations of drug targets as the alternatives of anti-SMO based cancer therapeutics. In our opinion, the identification of alternative drug targets by the conventional experimental approaches sometimes is challenging due

to some limitations in the experimental set-ups. Moreover, the complex gene regulatory networks and most importantly the molecular cross-talks of different pathways, make the target identification processes more challenging for the experimental biologists. The advancements of high-throughput omics-based studies rigorously performed in cancer cells under different inhibition scenarios have generated a plethora of datasets to understand the involvement of several genes, proteins, miRNA, etc. in different cancer pathology. Given the vast amounts of such datasets generated in each level, it has become more challenging to interpret all the information to decode and subsequently use it for identifying new drug targets.

Recent developments in integrative approaches, bioinformatics tools, mathematical and computational methods have become indispensable in understanding and analyzing such high-throughput datasets. Diverse approaches using qualitative and quantitative methods as well as successful implementation of mathematical modeling of signaling pathways can be used to decipher the complexities in signaling cascades [332]. Knowledge-based approaches by developing mathematical models of signaling cascades are proven useful for understanding the reaction mechanisms and identifications of alternative drug targets [317,333]. After a thorough literature review, we have observed that the development of mathematical models of Hedgehog signaling pathway to decipher its roles in cancer pathology is still in its infancy. One of the major challenges which we have observed is the lack of our understanding about the flow of reaction cascades in intra-cellular region. This limitation causes a serious challenge to construct the model for understanding the reaction mechanisms in normal and cancer cells, and successively find out most suitable drug targets from the set of proteins involve in the pathway.

We understand that the up to date information or a comprehensive reaction map of Hedgehog pathway would help us to analyze the entire signaling cascades more accurately. However, we have observed that there is no comprehensive pathway

map available in literature or cell signaling databases for studying the Hedgehog pathway more elaborately. After performing a review of the existing databases, we have noticed that a significant variation exist while comparing the number of molecules and interactions of this pathway reported in the existing databases (**Figure 11**). Also, after assessing the overall scenarios, we have noticed that a suitable combination of alternate drug-targets in target-based cancer therapeutics to inhibit the Hedgehog pathway activity is unknown hitherto. A large-scale analysis of Hedgehog signaling network considering its entire pathway species, cross-talks reactions from other pathway molecules, its associated gene-regulatory network, and its association with different phenotypic outputs is also undone. We have also observed that the state-of-the-arts techniques developed in computational systems biology could be highly useful to simulate the activity of this pathway in normal and cancer cells *in-silico*. Simulation of this pathway could be effective to understand the flow of reactions cascade under disease pathology, which would successively helpful to perform various *in-silico*, automated perturbations (i.e., drug treated) scenarios of multiple proteins.

In this work, the following objectives are undertaken.

- Reconstruction of a comprehensive map of Hedgehog pathway by database searching and literature mining.

- Understanding of the activities or expression patterns of the intracellular components in normal and cancer cells using in-silico simulations. Here, we have considered three different types of cancers or tumors, *viz.* Glioblastoma multiforme, Colon, and Pancreatic cancers.

- Implementation of the state-of-the-art graph or network theory and logical modeling approaches for *in-silico* simulations of the activity of Hedgehog pathway in normal and cancer cells.

- Validation of the expression patterns observed in *in-silico* analyses with the published experimental datasets (e.g., microarray, SDS-PAGE) of cancer cells.

- Identification of the molecular reaction motifs (or modules) in Hedgehog pathway responsible for helping out the cancer cells to escape anti-SMO treatment. Here, comparison of the reaction motifs exist in different cancer cells will be performed.

- Discovery of alternative drug targets in combination with the known drug target protein SMO to treat Glioblastoma, Colon, and Pancreatic cancers.

4.4 HYPOTHESES AND ASSUMPTIONS

We have hypothesized the existence of active signaling reaction modules in Hedgehog pathway, which are responsible for creating drug-resistivity during anti-SMO treatment. Also, our conjecture is that the combinations of potential drug targets can be more effective than targeting the single receptor protein SMO in target-based anti-cancer therapeutics.

There are few assumptions, which we have also undertaken to verify these hypotheses using theoretical modeling and *in-silico* simulation of Hedgehog signal transduction activities in normal and cancer cells. The assumptions are as follows.

Assumption 4.4.1. The signal transduction cascades flowing in the network via different pathway species are considered as the connection between two species. The connection is unidirectional and the directionally is also conserved irrespective of the cellular conditions.

Assumption 4.4.2. The activities of all the pathway species are assumed as discrete and binary in nature, i.e. either ON or OFF. The "ON" state represents the "Active or Over-expressed or Up-regulated" situations of the pathway molecules. On the other hand, the "OFF" state represents as the "Inactive or Under-expressed or Down-

regulated" situations.

Assumption 4.4.3. The logical steady-states of the pathway molecules at the steady-state (i.e., equilibrium) are directly related to the activity or expression of the molecules found in experimental data (e.g., microarray, RNA-Seq, Mass spec, etc.).

Assumption 4.4.4. The gene expression and the synthesis of protein molecules from the transcripts are directly proportional to the activation of the transcription factors and other transcription co-activators of the gene in the signaling network. The events, such as transcriptional regulations of the transcribed mRNA by miRNA, and the post-translational modifications of the newly synthesized proteins are not considered in our simulation.

4.5 RESULTS

4.5.1 Reconstruction of Human Hedgehog Signaling Pathway

In order to fulfill the first objective of this work, at first we have reconstructed the comprehensive and up to date map of the Hedgehog signaling network (**Figure 19**). To the best of our knowledge, the newly reconstructed map is the largest map of Hedgehog pathway till this date.

There are total 57 proteins (52 core Hedgehog pathway proteins and 5 cross-talk proteins from other pathways) and 96 hyper-edges included in the newly reconstructed pathway diagram. A comparative statistics of the total number of pathway species included in our reconstructed map versus the number of species presented in major biochemical signaling databases (e.g., EGG, BIOCARTA, GENE GO, NETPATH and PATHWAY CENTRAL) is depicted in a Venn diagram (**Figure 20**). The pathway information, such as molecule names, interactions, etc., require for reconstructing the map is mainly curated from different databases and literature sources. For detail descriptions of pathway reconstruction procedure, please see **Section 2.2** of **Materials and Methods**.

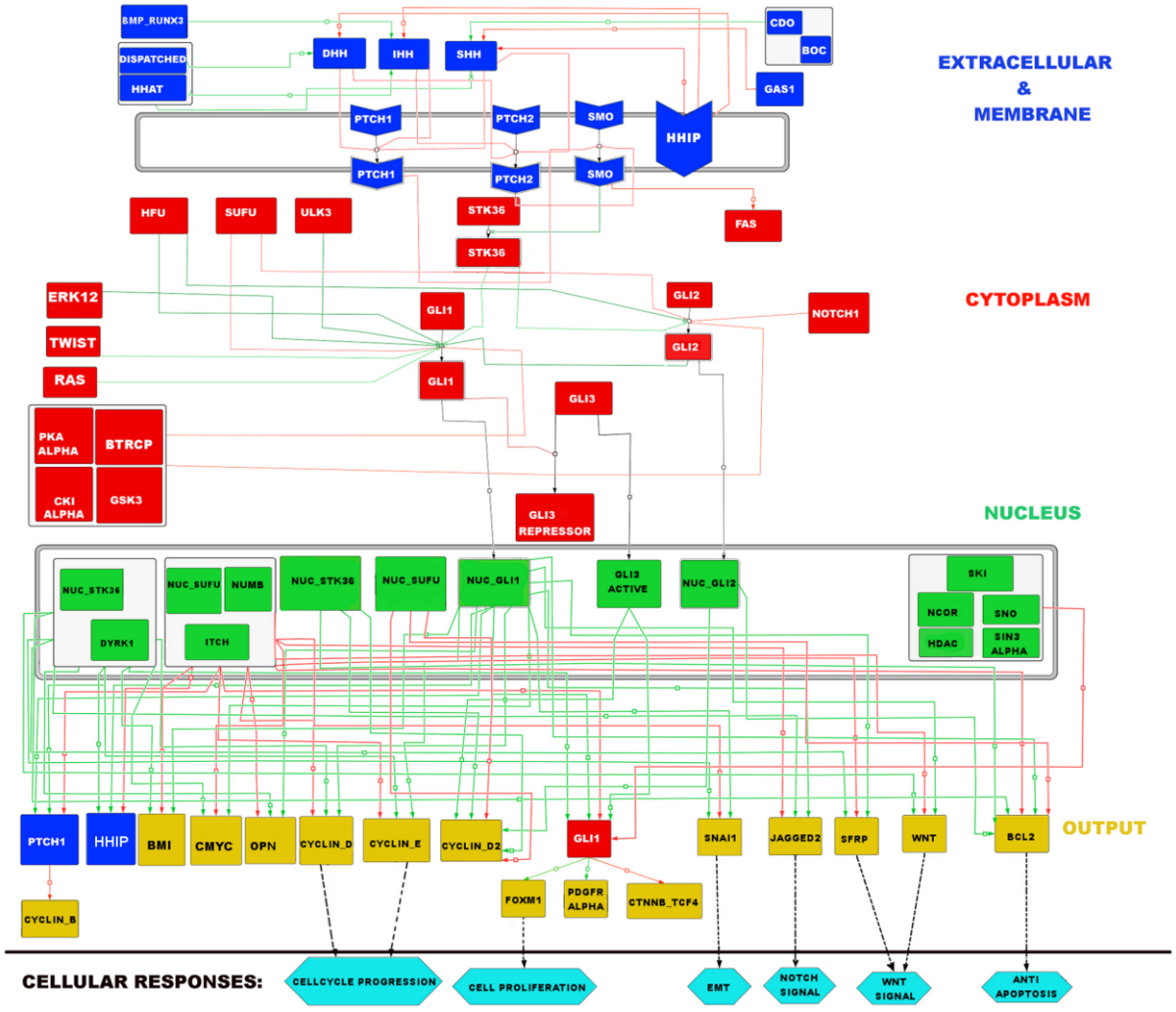

Figure 19: Reconstructed human cell specific Hedgehog signaling pathway.
There are total 57 proteins included in the reconstructed pathway. The green and red arrows indicate Activation/Production and Inhibition process, respectively. The black arrows indicate the nuclear translocation process. All the proteins of this network are allocated into four main regions with different color codes: Extracellular and Membrane (Blue); Cytoplasm (Red); Nucleus (Green); and Output (Yellow). The output proteins are linked with various cellular responses (cross-talks with other pathways or phenotypic expressions) with black dotted arrow.

In order to distinguish the hedgehog signaling proteins according to their cellular locations, we have allocated all the proteins in the reconstructed pathway diagram in four main groups: Extracellular & Membrane, Cytoplasmic, Nuclear and Output/Readouts with four different colors: Blue, Red, Green and Yellow, respectively. The group Output/Readout represents the target genes/proteins of Hedgehog pathway and does not signify any sub-cellular location. The phenotypic outputs or the cross-talks with other pathways which are initiated as a consequence

of Hedgehog pathway are termed as ''Cellular Responses'', which are mainly mapped with the output/readout of the pathway. A detail description of each of the groups and the associated pathway species is provided in the following sections.

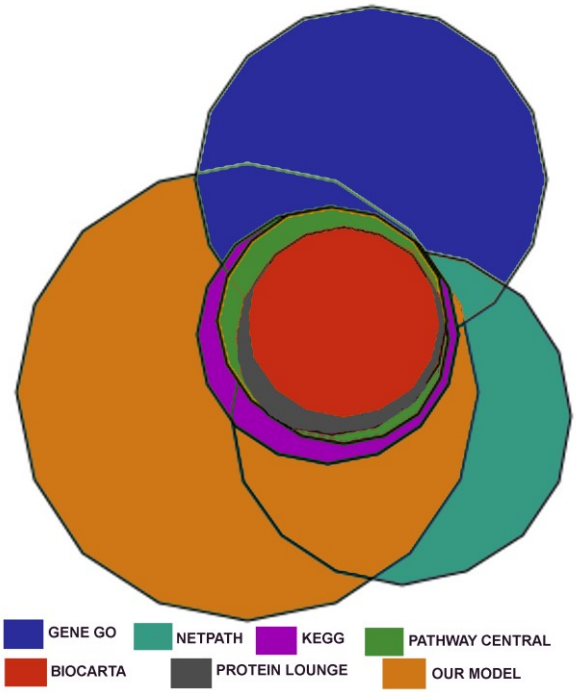

Figure 20: Comparison of the number of proteins available in other database and our model.
This Venn diagram represents a comparative view of number of proteins in our model with existing major databases, KEGG, BIOCARTA, GENE GO, NETPATH and PATHWAY CENTRAL, considered to reconstruct the Hedgehog pathway diagram. The overlapping regions between two circles (i.e., two databases or anyone of the database and our model) are representing the same proteins which have mentioned in both the databases. The large non-overlapping area shown by OUR MODEL signifies the information of the large number of proteins which are not found in any databases and are mostly collated from literature sources.

4.5.1.1 Extra-Cellular and Membrane Region

In this region, we have included three hedgehog ligands: Sonic Hedgehog (SHH), Indian Hedgehog (IHH) and Desert Hedgehog (DHH). These are the ligands that bind to the receptor proteins Patched1 (PTCH1) and Patched2 (PTCH2) of a hedgehog target or responsive cell [334,335]. The inhibitory effect of PTCH1/PTCH2 receptor proteins on the trans-membrane G-coupled protein ''Smoothened (SMO)'' is also depicted in the diagram [336]. In the downstream process the phosphorylation

of Serine/Threonine kinase 36 (STK36) by the activated SMO is included [337]. The consecutive activation of Glioma associated proteins (GLI) by STK36 kinase in the cytoplasm is also shown [337]. This entire process of GLI protein activation in the presence of Hedgehog ligands is known as canonical Hedgehog pathway activation process and the stimulating signal for this process is primarily come from the extra-cellular region, which is finally transduced to the intra-cellular region via receptor protein SMO [338].

There are other molecules (e.g., CDO, BOC, DISPATCHED, HHAT, and BMP_RUNX3 protein complexes) present mainly in the extra-cellular or in the membrane of other cells secreting the Hedgehog ligands [167,339]. These proteins are mainly responsible for the post-translational modifications and dispatching of the Hedgehog ligands in the extra-cellular regions. We have found the literature evidences that in the absence of these proteins in the Hedgehog ligands secreting cells, the activation of Hedgehog signaling is highly affected in the effector cells [337]. There is another membrane bound protein known as Hedgehog interacting protein (HHIP) present in the cell membrane, which is responsible for attenuating the Hedgehog ligands medicated pathway activation and triggering the homeostasis in signal transductions [340]. In this region, we have included total 3 ligands, 6 extracellular proteins and 4 membrane proteins included in Extracellular and Membrane region.

4.5.1.2 Cytoplasmic Proteins

In this region, we have included total 16 protein molecules in the reconstructed pathway diagram (**Figure 19**). We have included all the three isoforms of GLI transcription factors, *viz.* GLI1, GLI2, and GLI3. The inactive form of GLI proteins bound with SUFU and other inhibitors (e.g., PKA ALPHA, BTRCP, GSK3B, and CKIA) are shown in the Cytoplasm. In the inactive state of the pathway, the repressor form of GLI3 protein, known as GLI3R is produced which further

translocates into the nucleus and prohibits gene transcription process [341]. However, by the canonical pathway activation process, the phosphorylated STK36 phosphorylates either SUFU or the bound GLI proteins in the GLI-SUFU complex [337,342]. This phosphorylation event detaches Suppressor of Fused (SUFU) protein from the complex, which eventually gets degraded by proteosomal activities [342]. The phosphorylated GLI proteins in the cytoplasm are transcription active and are able to translocate into the nucleus [30]. The translocation of activated GLI proteins into the nucleus is also shown in the diagram (**Figure 19**). Also, apart from the SHH/IHH/DHH --> SMO --> STK36 --> GLI mediated canonical activation pathway, we have also found other cytoplasmic proteins, which can also phosphorylate the GLI-SUFU complex directly or indirectly. These proteins are Human Fused (HFU), Unc-51-like kinase 3(ULK3), ERK1/2, RAS and TWIST, which are mainly activated by other signal transduction pathways in the cells [343-346]. Such activation process of Hedgehog pathway via cross-talks from other pathway molecules is known as non-canonical Hedgehog pathway activation process, which have also found responsible for abnormal growth and tumorigenesis in different tissues and organs [343-346]. Such ligand-independent hedgehog pathway activation process in Glioma, Colon and Pancreatic cancer scenarios is also seen in the previous research works [345]. On the other hand, we have found that other cytoplasmic proteins *viz.* Protein Kinase-A (PKA), Beta-transducin repeat-containing protein (BTRCP), Casein kinase isoform alpha (CKIA), Glycogen synthase kinase-3 (GSK3) can repress GLI proteins [347-349].

4.5.1.3 Nuclear Proteins

In the nuclear region of the Hedgehog pathway map, we have included 13 molecules which are mainly transcription factors, co-activators or co-repressors. The activated transcription factors GLI1, GLI2 and GLI3 translocate into the nucleus as Nuclear GLI1 (NUC_GLI1), Nuclear GLI2 (NUC_GLI2) and GLI3 active (GLI3_A), respectively [350]. The transcription factors initiate the transcription of various hedgehog target genes with the help of transcription co-activators Nuclear STK36

(NUC_STK36) and Dual specificity tyrosine-phosphorylation-regulated kinase 1 (DYRK1) proteins [351]. Also, there are few transcription co-repressors in the nucleus which act as the inhibitors of the GLI transcription factors. These proteins are Nuclear SUFU (NUC_SUFU), NUMB, ITCH, SKI, Nuclear Receptor Co-repressor (NCOR), SNO, HDAC and SIN3A [352-355]. In nucleus NUC_GLI1 transcription factor transcribes the genes ptch1, hip1, gli1 along with several other responsive genes of this pathway. In order to reduce the complexity in the pathway figure, we did not include any gene or m-RNA in this nuclear region.

4.5.1.4 Outputs/Readouts

This region does not specify any cellular location in the pathway diagram (**Figure 19**). We have included this section separately to identify the target genes or the translated proteins produced at the end of Hedgehog pathway. There were total 15 proteins including GLI1, PTCH1 and HHIP included in this section. The total numbers of proteins shown in this region are highest as compared to any other published human specific Hedgehog pathway map to the best of our knowledge. Besides, all the proteins in this region are colored as yellow, except PTCH1, HHIP and GLI1 to show their feedback mechanism and translocations into the respective cellular locations. Production of PTCH1 and HHIP proteins in this pathway switch ''ON'' a ''negative feedback'' mechanism and thus they prohibit further hedgehog pathway activation by inhibiting SMO or interacting with Hedgehog ligands, respectively [356-358]. On the other hand, the transcription of *GLI1* gene and the synthesis of GLI1 protein as target output of Hedgehog pathway augments the pathway activity and creates a ''Positive feedback'' loop in this network [359].

4.5.1.5 Phenotypic Outputs or Cellular Responses

In order to show the cross connections of the output proteins of Hedgehog signaling with the other pathways or cellular functions, we have kept this section at the end of our pathway figure (**Figure 19**). There are 6 cellular responses included

which are Cell Proliferation, Cell cycle progression, Anti-Apoptosis, Epithelial–Mesenchymal Transition (EMT), Wnt signal, and Notch signal. We have shown the connections of produced proteins with these cellular responses by black dotted arrow in the pathway figure (**Figure 19**).

4.5.2 Topological Analyses of the Reconstructed Hedgehog Signaling Network

The reconstructed diagram of Hedgehog signaling network have helped us to understand the Hedgehog pathways in more details. The pathway is now more informative than ever before - with highest number of pathway species, interactions, feedback loops, cross-talks, and cellular responses, etc. However, with such higher amount of information, the pathway has become more difficult to decipher in various disease scenarios. Hence, to understand the structural integrity or the topological properties of the reconstructed network, we have used state-of-the-arts techniques of graph or network theory [360,361].

Here, we have considered the whole signalling pathway as a network where the reactions (i.e., signal) from the hedgehog ligands traverses from extracellular region to the nucleus of a target cell via various cytoplasmic intermediate proteins (**Figure 21)**. Our hedgehog signaling network has a 'Bow-Tie' like topology and consists of 57 nodes or proteins (52 core and 5 non-core proteins of hedgehog pathway) and 140 directed edges (interactions, regulations or the direction of flow of signal). The network is considered as a directed graph to maintain the directionality of the flow of reactions from the extra-cellular region to intra-cellular region.

We have assigned the size of the nodes in this network according to the total number of connections or degree values. GLI1 in cytoplasm has highest number of total degree in the network; therefore the size of this node in the network is largest among all the other nodes. It is also clear from this figure that the hedgehog signals from the inputs (extracellular and membrane proteins) converged to the particular proteins (GLI1 and GLI2) in cytoplasm to activate it and after its activation these

proteins send the signals (actually translocate into the nucleus) to activate the production of the various target genes/proteins (like OPN, BCL2, GLI1, HHIP etc.) at the downstream of hedgehog pathway.

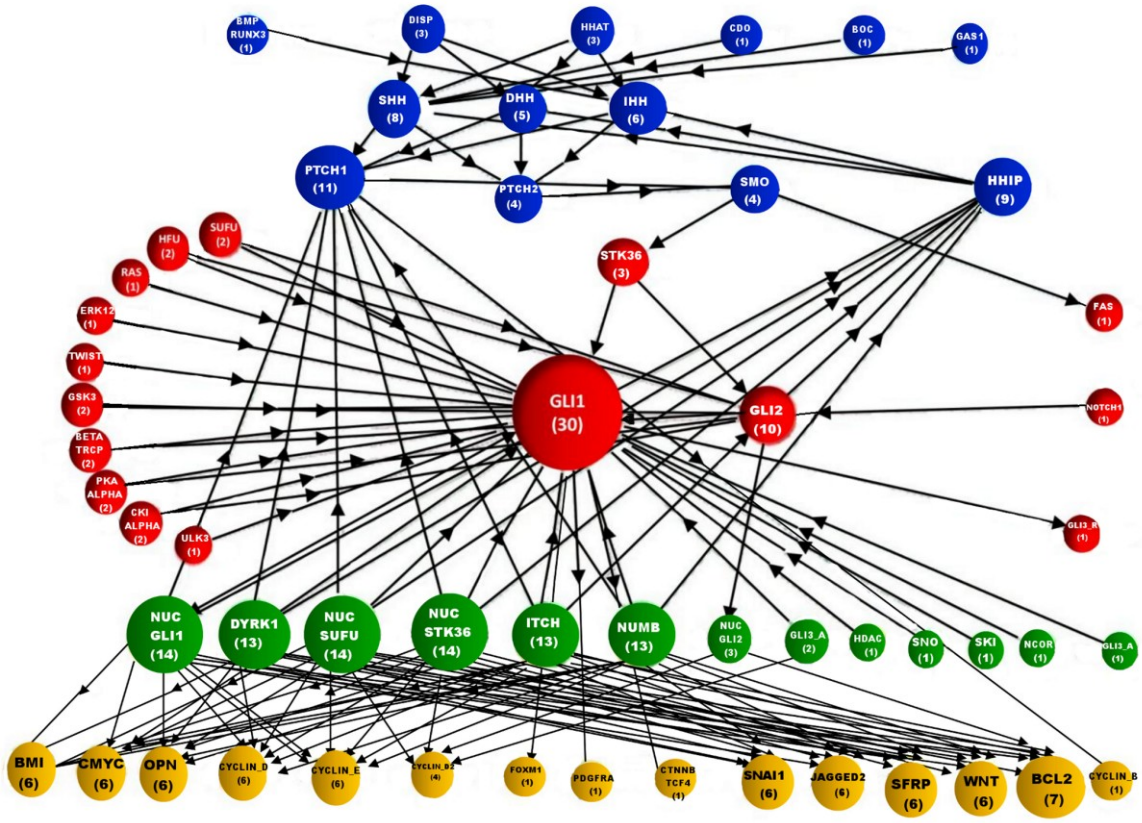

Figure 21: Network picture of Hedgehog signaling pathway.
The nodes are colored according to different groups or layers mentioned in **Figure 19**. These groups are Extracellular and Membrane (Blue), Cytoplasm (Red), Nuclear (Green), and Output proteins (Yellow), respectively. The size of the nodes is assigned according to their total number of connections or degree.

Therefore, we can say that the flow of hedgehog signal from extracellular-membrane region to the downstream target proteins of hedgehog pathway mainly depends on the intermediate cytoplasmic GLI proteins. Due to this reason the canonical hedgehog pathway is also called as 'GLI mediated hedgehog pathway' [362]. In order to understand the topology of the network in a better way, we have analyzed it in three different aspects *viz.* "Connectivity", "Centrality", and "Linear shortest paths".

4.5.2.1 Connectivity Analyses

We have performed this analysis to compute the number of connections of each protein with all other proteins in the network (see **Section 2.4** of **Materials and Methods**). Three types of parameters (IN DEGREE, OUT DEGREE and TOTAL DEGREE) are used in this analysis (See **Section 2.4**). We have calculated and represented these three parameters for each protein of the Hedgehog signaling network in **Figure 22**.

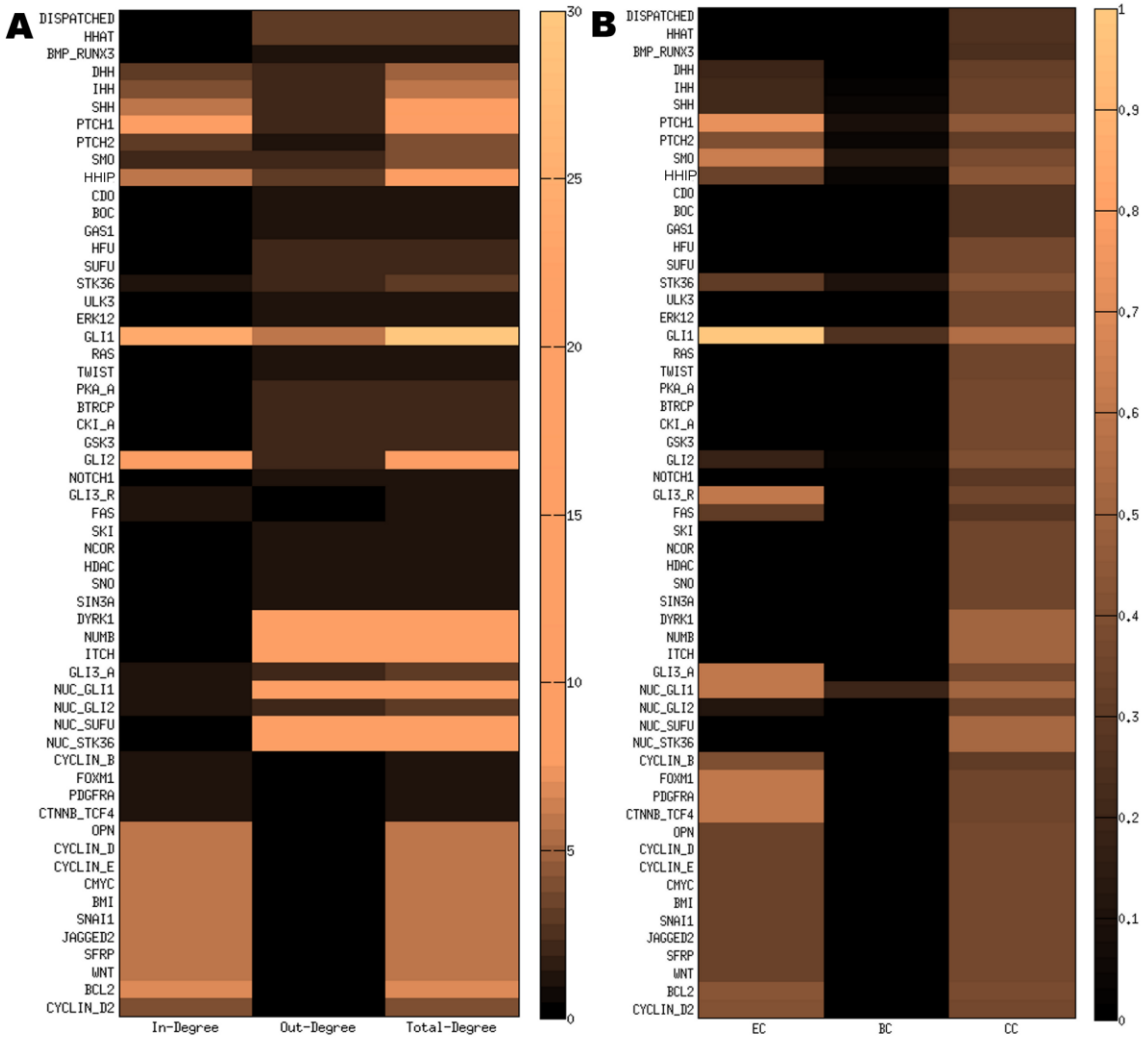

Figure 22: Topological parameters of Hedgehog signaling network.
The parameter values are arranged column wise (X-axis) in the heat map. (**A**) Heat map of the parameter values used in connectivity analysis are IN-DEGREE, OUT-DEGREE and TOTAL DEGREE of each protein. (**B**) Heat map of the individual centrality score of each protein of Hedgehog map. The Centrality measurement parameters used in this analysis were Eigenvector (EC), Betweenness (BC) and Closeness (CC) centrality.

The heat map (**Figure 22A**), representing the values of the parameters (IN-DEGREE, OUT-DEGREE and TOTAL DE- GREE), shows the proteins row wise according to their cellular locations in a cell (top to bottom) and the parameter values column wise. The average IN and OUT DEGREE (all together) of the network was calculated as 2.45 and the average total degree was 4.91. In order to identify the important proteins from this heat plot on the basis of the connectivity parameters, we extracted the proteins which had the parameter values higher than their corresponding average values. All the extracted significant proteins on the basis of this hypothesis were listed in **Table 7**. We have found that there are total 19, 10, and 23 proteins, which have higher values than the average IN-DEGREE, OUT-DEGREE, and TOTAL-DEGREE, respectively.

We have observed that the receptor protein PTCH1 and the transcription factors GLI1, and GLI2 have higher IN-DEGREE values as compared to the other proteins in the entire network due to their higher number of regulations or interactions with other upstream proteins in the hedgehog signaling network. PTCH1 has shown higher IN-DEGREE because most of the extracellular signals pass through this receptor protein to trigger the activation of SMO protein in membrane. On the other hand the cytoplasmic GLI1 and GLI2 have high IN-DEGREE value as these proteins are the most important proteins in the network to activate the pathway.

Also, among the three Hedgehog ligands, Sonic hedgehog (SHH) had the highest IN-DEGREE value as its interaction with PTCH1 and PTCH2 receptors is highly dependent on the proteins DISPATCHED, HHAT, CDO, BOC, and GAS1 at the extracellular region of Hedgehog target cell. The proteins in the nucleus like NUC_GLI1, NUC_GLI2, DYRK1, etc. have highest out-degree value compared to the other proteins in the network. Mainly the output proteins are connected to the outgoing connections or edges of these nuclear proteins in the network structure. Due to the presence of the higher number of outgoing connections from the nuclear proteins to the output proteins, the OUT-DEGREE values of these proteins are

increased in comparison to the other proteins in the whole network. We have also observed that except the nuclear proteins, the proteins from the other sub-cellular locations or regions do not show significant OUT-DEGREE values.

We have also extracted the proteins which have TOTAL- DEGREE values higher than the average total-degree 4.91. **Table 7** shows that in extracellular and ligands region PTCH1, HHIP, SHH, IHH have significant number (greater than the average total degree) of connections or total degree in the network.

Table 7: Significant Proteins Extracted from Connectivity Analysis

Parameters	Extracellular and Ligands	Cytoplasm	Nucleus	Output proteins
In-Degree (> 2.45)	DHH(3),IHH(4), SHH(6),PTCH1(9), PTCH2(3), HIP1(6)	GLI1(24), GLI2(8)	NOT FOUND	**[**OPN, CYCLIN_D, CYCLIN_E, CMYC, BMI,SNAI1, JAGGED2, SFRP,WNT**]**(6) BCL2(7), CYCLIN_D2(4)
Out-Degree (> 2.45)	DISPATCHED(3), HHAT(3), HIP1(3)	GLI1(6)	**[**DYRK1, NUMB, ITCH, NUC_GLI1**]** (13), NUC_STK36 (14), NUC_SUFU (14)	NOT FOUND
Total-Degree (> 4.91)	DHH(5), IHH(6), SHH(8), PTCH1(11), HIP1(9)	GLI1(30), GLI2(10)	**[**DYRK1, NUMB, ITCH**]**(13), **[**NUC_GLI1, NUC_SUFU, NUC_STK36**]** (14)	**[**OPN, CYCLIN_D, CYCLIN_E, CMYC, BMI, SNAI1, JAGGED2, SFRP, WNT**]**(6), BCL2(7)

The numbers mentioned in first parenthesis represent the corresponding parameter values.

It is also clear from **Figure 21**, that GLI1 has the highest TOTAL-DEGREE value among all other proteins in the Hedgehog signaling network. It also signifies that this protein is the most important protein in the Hedgehog signaling network. Out of 57 proteins in the network, it is found to be connected to 30 other proteins in the network. Therefore, in terms of signaling network, it is the biggest 'Hub' in the entire network, which is actually influenced by more number of other proteins as well as influencing more proteins than the other hubs in the entire network. Similarly, in the nuclear region NUC_GLI1, NUC_SUFU, and NUC_STK36 have formed the other larger hubs and thus controlling the synthesis of various target proteins of Hedgehog pathway.

4.5.2.2 Centrality Analyses

We have also measured the 'Centrality score' of each node or protein in the network after identifying the important ''Hub proteins'' from the network. In the connectivity analyses, we have found few important nodes or proteins, which are forming important 'Hub' in the whole network structure. In that case, we have given the highest importance to a node on the basis of its total number of connections or degree value. Albeit in biological as well as any real world network the importance of a node or a protein does not depend only on its number of connections or neighbors [360,363]. Sometimes the importance or significance of a node may increase due to its connections with the other important nodes in the network, though it may have lower number of neighbors or connections or vice-versa. We have used the centrality measurement parameters *viz.* Eigen-vector, Closeness, and Betweenness centrality to determine the relative importance of a node within a network (See **Section 2.4**).

In this analysis, at first, we have calculated 'Eigenvector Centrality' to identify the proteins according to their importance in our newly reconstructed hedgehog signaling network. This parameter is particularly useful to identify the central node,

which is connected to other important nodes (i.e., high degree nodes) in the network. The central node identified by using this parameter may have minimum connections with other nodes. We have calculated the eigenvector centrality scores for each of the protein in the network (**Figure 22B**). We have observed that GLI1and PTCH1 have high Eigenvector centrality score, but GLI2 has comparably lesser score than GLI1, though it has large number of connections or degree in the network. The reason of showing this interesting feature is that in Hedgehog signaling network, GLI2 is connected to NUC_GLI2, FAS, HFU, SUFU, PKA_A, BTRCP etc., which have lowest number of connections in the network and therefore are less important as compare to other nodes. On the other hand, GLI1 is connected with another important protein NUC_GLI1 in the network which regulates the expression of most of the output proteins in the network and comparably has higher number of connections than NUC_GLI2. Therefore, GLI1 is connected to another most important protein NUC_GLI1 in the network and thus the importance of GLI1 increases significantly as compared to GLI2 protein. Also, SMO, GLI3_Repressor (GLI3_R), GLI3_Active (GLI3_A) have higher Eigen-vector centrality values and thus can be considered as important nodes in the network, though they have lower number of connections.

Also, we have calculated the Betweenness centrality and Closeness centrality scores for each protein in the network (**Figure 22B**). We have identified some important proteins of the network based on these parameter values. As expected, we have observed that GLI1 have the highest Betweenness and Closeness centrality scores among all other proteins in the Hedgehog signaling network. Both the centrality scores of GLI1 are high because large numbers of shortest paths between two nodes are passing through it and it is connected to all other proteins with the minimum number of connections in the network. We have also found that besides GLI1, there are some other proteins, such as NUC_GLI1, SMO, STK36, and PTCH1 have high Betweenness centrality score. On the other hand, NUC_SUFU, NUC_STK36, DYRK1, NUMB, and ITCH have shown high Closeness centrality score

after GLI1. Interestingly, all of these proteins are found in the nucleus. As expected, the large number of the proteins, which belong to the "extracellular and ligands" region have lowest closeness scores i.e., they are situated more distantly from all the other nodes in the network and regulating the downstream proteins of Hedgehog signaling network.

4.5.2.3 Linear Shortest Paths

We have calculated the all-pairs shortest paths between each pair of nodes or proteins (directed edges) in the Hedgehog pathway (**Figure 23**).

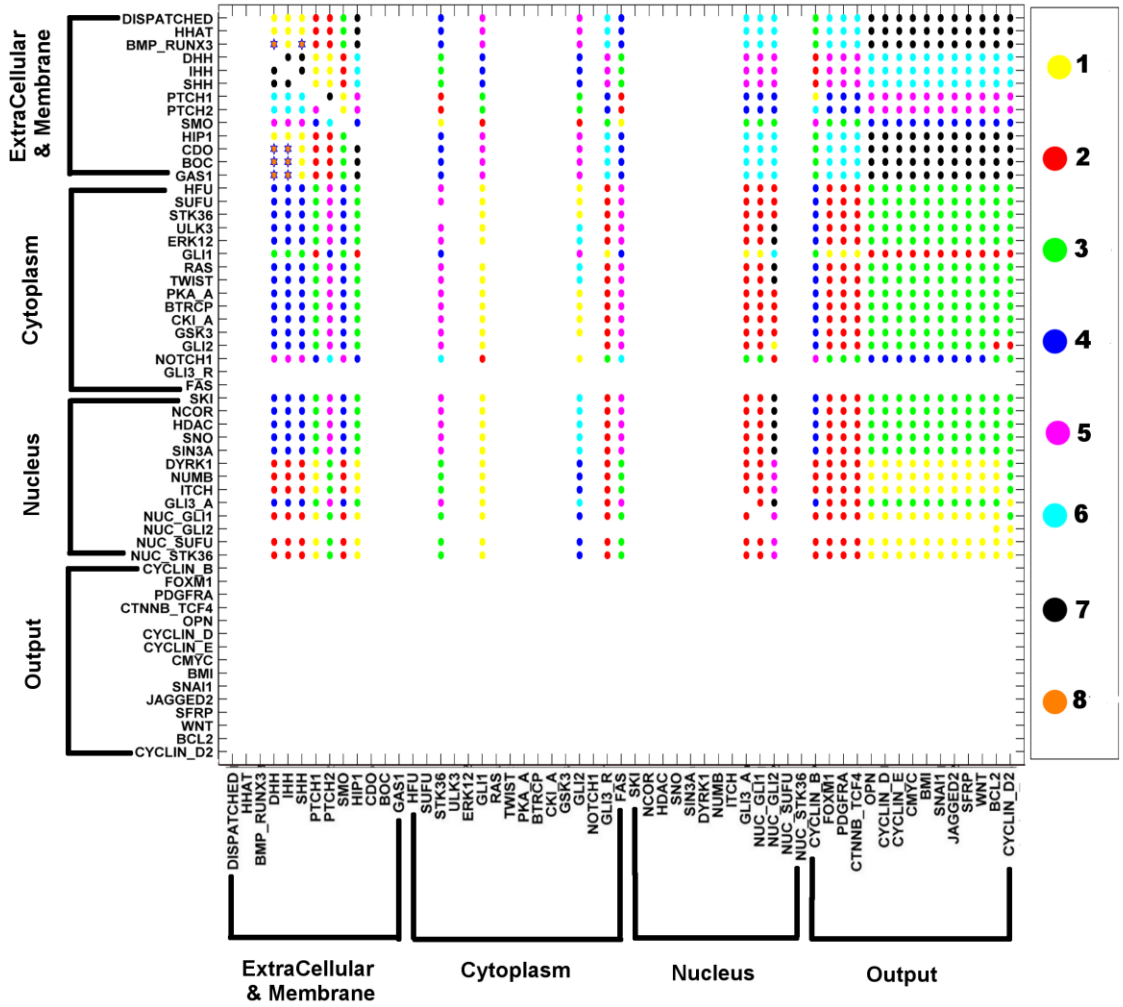

Figure 23: All-pair directed shortest paths of the Hedgehog signaling proteins.
The color codes represent the magnitude of the shortest-path between a pair of proteins in the network.

As obvious, the proteins of the "extracellular and ligand" region are distantly

separated from the Hedgehog target proteins (Outputs/Readouts) with the shortest paths of lengths either 6 or 7. We have also calculated the distribution of all the shortest-paths of the Hedgehog signaling network and the average shortest path is 3.581. This observation signifies that most of the proteins in the Hedgehog signaling network are connected with each other on an average by 3 intermediate steps. For example, the transmission of the signal after ligand-receptor binding takes only 3 (three) intermediate steps or links to activate GLI1 or GLI2 proteins (largest hubs) in the cytoplasm (i.e., PTCH1/PTCH2 --> SMO --> STK36 --> GLI1/GLI2). Similarly, in order to initiate the production of the target output proteins of Hedgehog pathway in the nucleus by transcription factors GLI1, GLI2 or GLI3_ACTIVE, it requires only 2 (two) intermediate steps or links (i.e., GLI1/GLI2 --> NUC_GI1/ NUC_GLI2 --> CYCLIN_D/CMYC/BCL2).

This observation clearly demonstrates that a strong and tightly coupled signal transmission network exists in Hedgehog pathway to rapidly relay the activation signal from the extra-cellular region to the gene transcription network of the cell. Also, we have observed that the identified, important 'hub' proteins (i.e., GLI1, PTCH1, NUC_GLI1, etc.) are connected by each other with shorter number of links, which also signifies the rapid information sharing process between the important hubs in the network. Previous studies of different biological networks have proven that such topological characteristic of the networks (i.e., with a lower value of average shortest-path) is much more robust than the network with higher average shortest-path [364,365]. Researchers have found that removal of the central nodes (or hubs) significantly increases the average shortest-paths of the entire network and decreases the resilience of the system against external perturbation (or mutational attack) [364,365].

Therefore, it is seen that the graph theoretic analyses of the Hedgehog signaling network have helped us to identify the important hubs or proteins in the pathway, understand the topology and resilience of the network, and extract molecular

reaction routes between a pair of protein. The extracted central nodes from our graph theoretic analyses are proven experimentally as the important nodes for initiating the Hedgehog signaling cascades within the cells and their abnormal activations or inhibitions lead to uncontrolled cell proliferation or tumorigenesis in various cells/tissues [337,343-346,357,366]. Hence, the central nodes can be also considered as probable drug targets for inhibiting the growth of the Hedgehog pathway activated cancer cells. It also proves that the topology of network plays very important role in governing the gene transcription mechanisms, and therefore the in-depth analyses are required for understanding the topology of reaction cascades in the network under various pathological conditions. However, one of the major limitations of the graph theoretic analyses is that it just captures the static behavior of the network and is unable to depict any diseased scenario or malfunctions caused by genes/proteins in the pathway. The dynamic interplay of the pathway species after pathway stimulation and their reaction trajectory to reach the steady-states level are also important to understand the mechanisms of pathway activity under different biological conditions. Hence, to overcome this challenge, we have used ''Logical analysis'', where all the interactions of the network are modeled as 'Boolean or Logical equation hyper-graph' using logic-based formalisms (see **Section 2.5**).

4.5.3 Semi-Dynamic or Logical Analyses of Hedgehog Signaling Pathway

A list of all the proteins and other pathway species including phenotypes are provided in **Appendix Table 4**. We have written the logical equations of these pathway species with the help of biological understandings gained from literature sources (**Appendix Table 5**). The developed logic-based equations are considered as the "Master model", which is then used for simulating the pathway activities under normal and different pathological conditions *viz.* Glioblastoma, Colon, and Pancreatic cancers. We have simulated the logical models in CellNetAnalyzer to create the canonical activation process of Hedgehog pathway in normal scenario (NS) and then created the other cancer scenarios by mutating the Hedgehog proteins

found to be responsible for tumorigenesis (see **Section 2.5.2** for further details) [126].

4.5.3.1 Construction of Normal Hedgehog Pathway Scenario (NS)

We have calculated the total number of upstream activators and inhibitors of each protein in the Hedgehog pathway at the steady-state level (see **Section 2.5.2**). Similarly, we have also calculated the total number of downstream activated or inhibited proteins of each protein considered in the logical model of the normal, canonical Hedgehog pathway (see Section 2.5.1). In case of 'Canonical hedgehog pathway' (i.e., NS or Normal Scenario), we have found that at time scale two (2), the ligand Sonic Hedgehog (SHH) has activated overall 26 different proteins in the pathway directly or indirectly (**Figure 24C**). Also, other proteins SMO, STK36 have activated overall 25 and 24 other proteins in the pathway, respectively. Similarly, we have observed that in the normal Hedgehog pathway, the cytoplasmic proteins, such as GLI1 and GLI2 have activated total 21 and 23 proteins, respectively. These results clearly show the importance of these proteins in the activation of Hedgehog pathway in normal human cells [337].

On the other hand, we have found that there are total 20 inhibitor proteins (e.g., PTCH2, SUFU, BTRCP, GSK3, PKA, CKI_A, NCOR, HDAC, SNO, SIN3A, NUMB, and ITCH, etc.) present in the upstream of GLI1 protein in normal scenario (**Figure 24B**). We have observed that such inhibitory effects from the upstream inhibitors are also required for the homeostasis GLI1 protein activation process in Hedgehog pathway under normal condition [355,367-369]. Therefore, it can be said that the interactions between these activators and inhibitors in the canonical Hedgehog pathway control the abnormal activation of GLI protein in the cytoplasm, and thus regulate the excess productions of various target proteins responsible for cell division, migration, and anti-apoptosis processes, etc. We have also observed that our developed model for simulating the canonical Hedgehog pathway activity in normal cells imitates the expression scenarios of various genes or activity profiles

of various proteins under normal conditions (i.e., non-malignant, non-cancerous cells). We have used the simulation outputs of the canonical Hedgehog pathway model as the "Normal Scenario (NS)" of the non-malignant cells and then compared with the three cancer cells models developed in-silico by inserting corresponding oncogenic mutations in the normal model.

4.5.3.2 Construction of Glioblastoma Tumorigenic Scenario (GS)

GBM model development and simulation: In order to simulate the tumorigenic events in Glioblastoma cells caused by abnormal activity of Hedgehog pathway, we have considered the over-expression of Sonic Hedgehog ligand (SHH) in the glial cells, which is a known factor responsible for the over-activation of Hedgehog pathway in difference cells/tissues (see **Section 2.5.2**) [370]. We have done this analyses by constitutively over-activating (i.e., keeping the binary state at '1' or 'TRUE') the state of SHH protein in the normal Hedgehog pathway during pathway simulation. We have also kept the over-activated cytoplasmic kinases HFU and ULK3 in the model and allowed these proteins to hyper-activate the GLI proteins in cytoplasm. Apart from these proteins, we have also considered the cross-talk activation Hedgehog pathway by constitutively over-expressing the cross-talk proteins, such as ERK12, TWIST, and RAS during simulation. Thus, we have made the overall situations (*in-silico*) favorable towards the hyper-activation of canonical and non-canonical Hedgehog pathway (including abnormal activation of GLI1 and GLI2), which are most often observed in Glioblastoma stem and tumor cells [346,371,372]. Therefore, to simulate the Glioblastoma scenario in our study, we have considered the constitutive expressions of SHH, HFU, ULK3, ERK12, RAS, TWIST as ‘‘1’’ or ‘‘ON’’ (**Appendix Table 6**). We have also considered the down regulation or loss of function of few tumor suppressor proteins, such as GAS1, SUFU, NUMB, SNO, etc. in the new Glioblastoma model simulation (**Appendix Table 6**). Down regulation or loss function of these proteins can also cause the up-regulation of GLI proteins in cytoplasm as well as in nucleus [373].

Simulation outcomes: We have analyzed the dependency matrices obtained from both normal and glioblastoma scenario and observed that the total numbers of upstream activators of SHH, STK36, GLI1, GLI2, NUC_GLI1, NUC_GLI2 and all the output proteins (OPN, BMI, SNAI1 etc.) of hedgehog pathway in Glioblastoma scenario (GS) are higher as compared to the normal scenario (NS) (**Figure 24**).

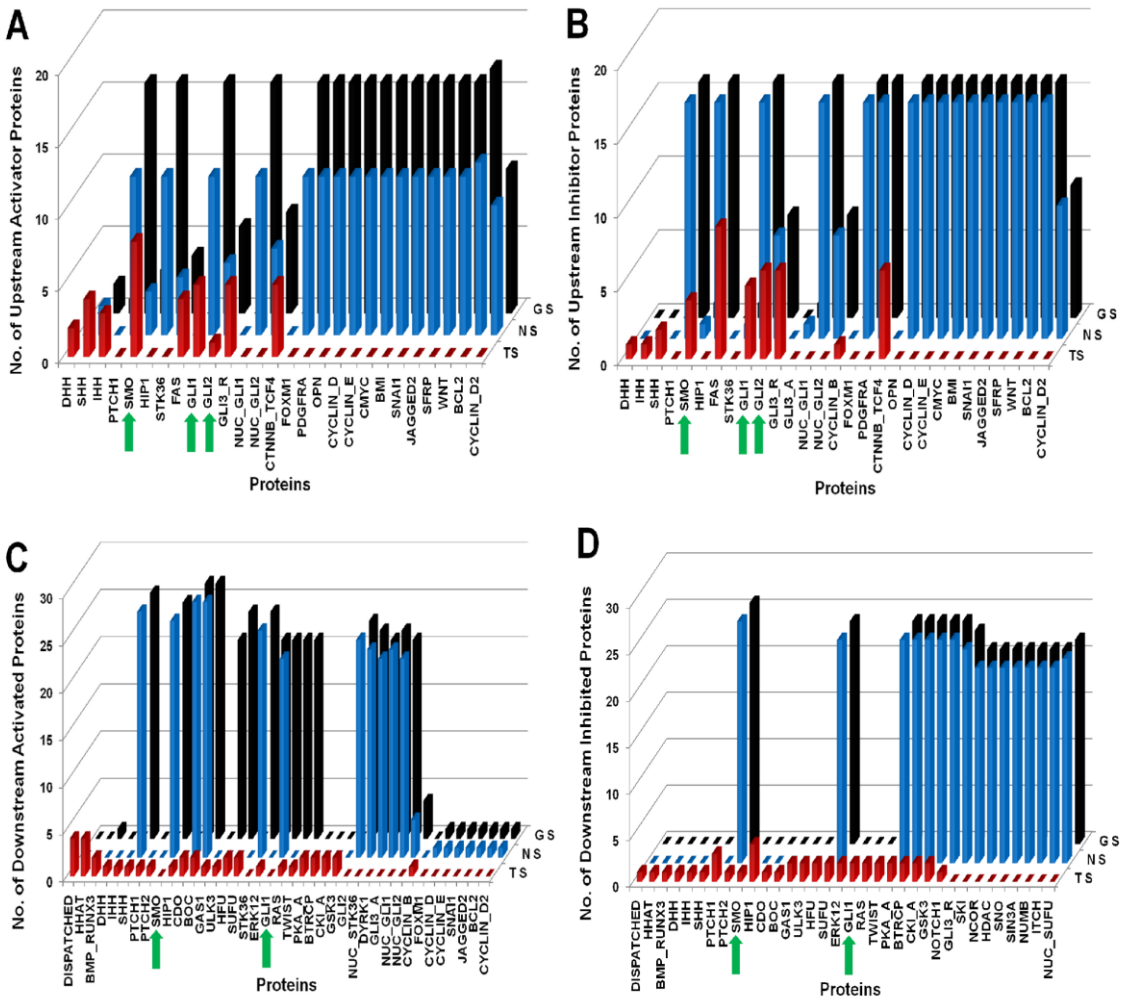

Figure 24: Comparative analyses of Glioblastoma scenarios.
TS: Treated Scenario; NS: Normal Scenario; GS: Glioma Scenario. The green arrow heads are indicating the minimal combination of proteins which was inhibited in the drug treated perturbation analysis. **(A)** Represents number of Upstream activator proteins (Y-axis) activating the proteins (X-axis) representing significant variations. **(B)** Represents number of Upstream inhibitory proteins (Y-axis) inhibiting the proteins (X-axis) representing significant variations. **(C)** Represents number of downstream proteins (Y-axis) activated by the proteins (X-axis) representing significant variations. **(D)** Represents number of downstream proteins (Y-axis) inhibited by the proteins (X-axis) representing significant variations.

In case of Glioblastoma, the total numbers of activators acting on GLI1 and GLI2 proteins are 16 and 6, respectively, whereas in case of normal scenario there are total 11 and 5 proteins activating the expression of GLI1 and GLI2 proteins, respectively. These simulation results clearly show the differences between Normal (NS) and Glioma scenarios (GS), which can be observed from our developed master model (**Appendix Table 5**). We also have found that the total numbers of activator proteins acting on target oncoproteins of Hedgehog pathway, such as FOXM1, PDGFRA, OPN, CYCLIN_D, CYCLIN_E, BMI, SNAI1, JAGGED2, and SFRP are increased in Glioblastoma scenario (GS) as compared to the normal scenario (**Figure 24**) [374-383]. On the other hand the total number of upstream inhibitor proteins on these output proteins remains unchanged while we have compared the Normal versus GBM scenario (Figure 4B). This clearly indicates that the activator proteins of these target oncoproteins in GBM cells are highly active as compared to the inhibitor proteins of the same target oncoproteins.

On the other hand, to identify and compare the potential activators and inhibitor proteins of this pathway, we have calculated the total number of proteins that are activated or inhibited directly or indirectly by each protein of the Hedgehog pathway model (**Figure 24C - D**). We have found that as compared to GLI1 protein, GLI2 protein is more potential activator in GBM scenario. GLI2 is connected with GLI1, which is also an important activator protein in the network (**Figure 24C - D**). Therefore it is worth mentioning that to suppress the GLI1 activation, GLI2 protein should be suppressed simultaneously. We have also observed that HFU, ULK3, RAS, TWIST, ERK12 can act as the activators in GBM scenario (GS), but not in the normal Scenario (NS). This indicates the effect of cross-talk activation process between the core Hedgehog pathway molecules (e.g., GLI1, GLI2, etc.) with the other pathway molecules in GBM cells. On the other hand, we have found that the other core proteins of Hedgehog pathway, such as SHH, SMO, GLI1, GLI2, GLI3_A can activate its downstream proteins in both normal as well as GBM scenarios (**Figure 24C).**

Overall, the simulation results of GBM model have helped us to identify the key regulatory proteins that could be used as potential target proteins to inhibit the growth of tumor cells in GBM treatment. However, at the time of determining the target proteins for our *in-silico* drug treatment strategy (TS) of GBM, we have also taken into the consideration of the biological relevance and the feasibility of targeting that target protein(s) in real therapeutic scenario. We have identified that SHH, SMO, STK36, RAS, TWIST, ERK12, HFU, ULK3 are the proteins responsible for activating most of the target oncoproteins in GBM cells by using GLI transcription factor (**Figure 24C**). It is experimentally proven that SMO plays a very crucial role to activate STK36 as well as GLI1 in cytoplasm [337]. Our graph theoretical study has shown that SMO receptor protein has significant high Betweenness centrality score, which in turn also proves the importance of this protein to transduce the activation signal into the downstream (**Figure 22**). Inhibition of this trans-membrane protein by small molecule inhibitor/drug in tumor cells has shown effective inhibition of Hedgehog signal and the reduction of tumor size [384,385]. Although, SMO is important to mediate the Hedgehog signal in GBM cells by activating GLI transcription factors, but still it is worth to mention that the inhibition of SMO by its inhibitor/drug molecule is not sufficient to completely suppress the Hedgehog pathway activity and reduce the growth of GBM cells. This is due to the activity of the intracellular activator proteins, which are present in the cytoplasm of tumor cells to augment the expression of GLI transcription factors via non-canonical process. Therefore, in this work, we have proposed that to suppress the growth of GBM cells, individual targeting of SMO protein by drug is not sufficient and we also need to take into the account of the intra-cellular, cross-talk molecules present in cytoplasm.

Drug target(s) identification: In order to determine the other factors, we have revisited our graph theoretical analysis and identified the IN-DEGREE neighbors of GLI proteins from cytoplasm. We have found that the upstream activators (IN-DEGREE neighbors) of GLI1 and GLI2 proteins in cytoplasm are HFU, ULK3, ERK12,

RAS, and TWIST. We have individually perturbed the logical states of these activators in GBM scenario (GS), but targeting these activator proteins individually have shown no effect to reduce GLI activity and also the phenotypes, such as cell division, anti-apoptosis, etc. We have tried various combinations of HFU, ULK3, ERK12 to suppress the Hedgehog signal in GBM scenario (GS), but we have observed that only the perturbations of all the activator proteins simultaneously can suppress the GBM growth (i.e., cell division and anti-apoptosis) in the *in-silico* treatment scenarios (result not shown). However, targeting all the activator proteins in GBM cells simultaneously is not biologically feasible due to the possibility of drug induced toxicity. Therefore, to suppress the GLI activation process in cytoplasm, we have directly inhibited the activity of GLI1 and GLI2 proteins in the cytoplasm and SMO protein in cell membrane by constitutively keeping these proteins at "0" or "OFF" states in the GBM scenario (GS). As a result of this perturbation simulation, we have observed that the expression of the output proteins is blocked in GBM scenario (GS). We have used the dependency matrices of the tumorigenic glioblastoma model (GS) and the *in-silico* treatment model (TS) to extract and compare the total number of proteins that are directly activated or inhibited by each protein in both scenarios. We have found that the total number of upstream activator and inhibitor proteins of all the target/output oncoproteins of Hedgehog pathway in the treatment scenario (i.e., GLI1/2 & SMO = 0) has come down to zero. Due to this perturbation, we have also observed that the potential activators and inhibitors of Hedgehog pathway (e.g., RAS, TWIST, GSK3, BTRCP) are activating or inhibiting less number of proteins in the downstream region of the pathway (‘‘Treated Scenarios (TS)’’ of **Figure 24C-D**).

4.5.3.3 Construction of Colon Cancer Scenario (CC)

Model development and simulation: We have found that in case of Colon cancer, canonical Hedgehog pathway mediated by IHH and SHH ligands, and the RAS/RAK mediated signaling pathways simultaneously up-regulates the activity of the GLI proteins in colorectal cancer cell [386]. Experimental evidences have shown that over-

expression of SHH and IHH ligands can cause up-regulation of GLI transcription factors. Also, the over-activation of HRAS protein can also up regulate GLI proteins in Hedgehog pathway and are responsible for Colon cancer [387-391]. In order to capture the effects of Hedgehog ligands IHH and SHH, and RAS in Colon cancer scenario (CC), we have constitutively over-activated these proteins in the master Hedgehog pathway model and then tried to analyze logical states of the GLI transcription factors in the steady-state. To develop the CC mode scenario, we have also considered the down-regulation or loss-of-function states of the tumor suppressor proteins, such as GAS1, SUFU, NUMB, SNO, etc. (**Appendix Table 6**).

Simulation outputs: The simulation outputs of CC model scenario have shown that the total numbers of upstream activator proteins of GLI1, GLI2, and GLI3_A proteins are greater than the normal scenario (NS) (**Figure 25A)**. As a result, the total numbers of upstream activators of the target/output oncoproteins of Hedgehog pathway are also increased in colon cancer scenario as compared to the normal scenario (NS) (**Figure 25A**). Therefore, it is clear that the abnormal activation of GLI proteins in colon cancer cells not only occurred due to the abnormal expression of hedgehog ligands (IHH/SHH), but also due to the over-activation of RAS protein and its subsequent interactions with GLI1 and GLI2 transcription factors. Also, HFU and ULK3 are over-activated as these are the auto-phosphorylated kinase proteins present in cytoplasm to activate GLI proteins [316,343]. The relative potential of activating the other molecules between GLI1 and GLI2 are also measured and we have found that GLI2 have higher number of downstream activated species in the colon cancer scenarios (**Figure 25A** and **Figure 25C**), as it has connection with GLI1.

Drug target identification: These findings have helped us to identify the combination of potential drug targetable proteins in the aberrantly activated Hedgehog pathway for Colon cancer cells. We have perturbed the activation signal transducing from SHH and IHH via PATCHED (PTCH1/2) and SMO proteins to GLI transcription factors. Simultaneously, we have perturbed the interactions between HFU, ULK3,

and RAS proteins with GLI transcription factors in the cytoplasm. In order to stop the SHH/IHH signal transduction, we have perturbed SMO protein by keeping its logical state constitutively at "0" or "OFF" state.

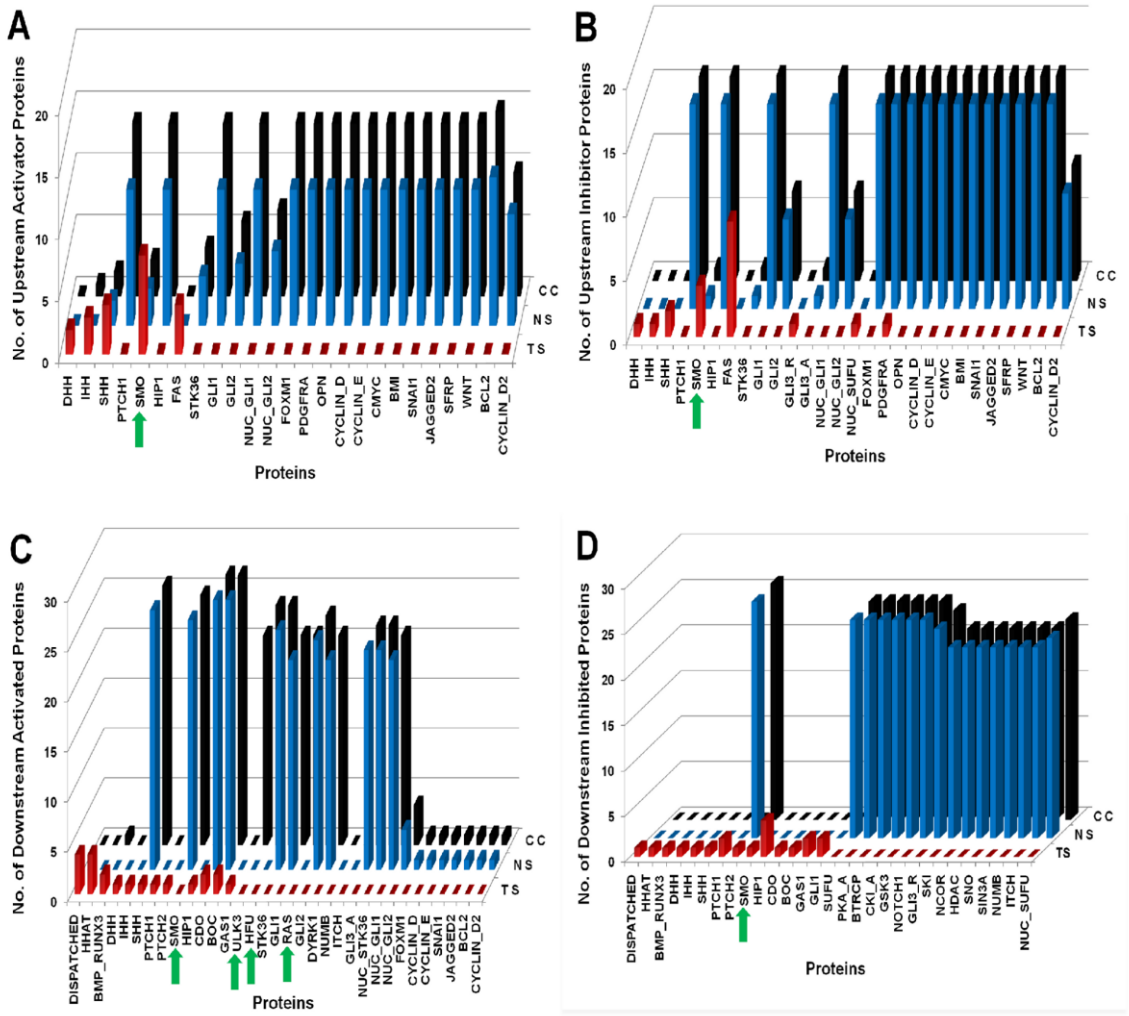

Figure 25: Comparative analyses of Colon Cancer scenarios.
TS: Treated Scenario; NS: Normal Scenario; CC: Colon Cancer Scenario. The green arrow heads are indicating the minimal combination of proteins which was inhibited in the drug treated perturbation analysis. **(A)** Represents number of Upstream activator proteins (Y-axis) activating the proteins (X-axis) representing significant variations. **(B)** Represents number of upstream inhibitory proteins (Y-axis) inhibiting the proteins (X-axis) representing significant variations. **(C)** Represents number of downstream proteins (Y-axis) activated by the proteins (X-axis) representing significant variations. **(D)** Represents number of downstream proteins (Y-axis) inhibited by the proteins (X-axis) representing significant variations.

On the other hand, we have cut the interactions of HFU, ULK3, and RAS proteins with GLI transcription factors by keeping the logical states of these proteins at OFF state and thus created the treatment scenario (TS) on Colon cancer (CC) model. We

have found that the total numbers of the activated proteins by GLI1, GLI2, and GLI3 proteins are decreased as compared to the colon cancer scenario (CC) (**Figure 25C**). Moreover, we have observed that the activation and inhibition level of GLI transcription factors in nucleus and cytoplasm by the upstream proteins are reduced as compared to the CC scenario (**Figure 25A** and **Figure 25B**).

4.5.3.4 Construction of Pancreatic Cancer Scenario (PC)

Model development and simulation: In case of pancreatic cancer, we have found that the expressions of IHH, PTCH1, and SMO proteins in pancreatic cancer cells are highly associated with the tumor cell growth and metastasis [392]. Simultaneously, over-expressions of kinase proteins, such as RAS and ERK12 in the pancreatic cancer cells are also reported [345,393]. Pathological and biopsy reports of pancreatic cancer cells have shown that the mutation or over-expressions of these proteins can trigger tumorigenic events Pancreas [345,393]. Hence, to capture the effect of these proteins in pancreatic cancer cell development, we have constitutively over-activated IHH and RAS in the master Hedgehog pathway model. Apart from these proteins, we have also down-regulated the tumor suppressor or Hedgehog pathway inhibitor proteins, such as GAS1, SUFU, NUMB, SNO, etc. to account their loss-of-function effects in the initiation of pancreatic cancer (PC) scenario (**Appendix Table 6**).

Simulation outcomes: We have observed that in PC scenario, the total numbers of upstream activators of GLI transcription factors and its target oncoproteins are higher than the normal scenario (NS) (**Figure 26**). The increase of total number of activators of the GLI proteins and its target oncoproteins in PC scenario clearly indicates the effect of the over-activations of IHH, RAS and down-regulation of tumor-suppressor proteins in PC.

Drug target(s) identification: We have noticed that the individual perturbations of SMO or PTCH1/PTCH2 receptors are not effective to reduce the activation of GLI1/GLI2/GLI3_A in pancreatic cancer model. Therefore, in order to completely

repress the over-expression of GLI proteins in cytoplasm, we have perturbed the logical states of RAS and ERK12 proteins simultaneously with SMO in pancreatic cancer scenario (PC).

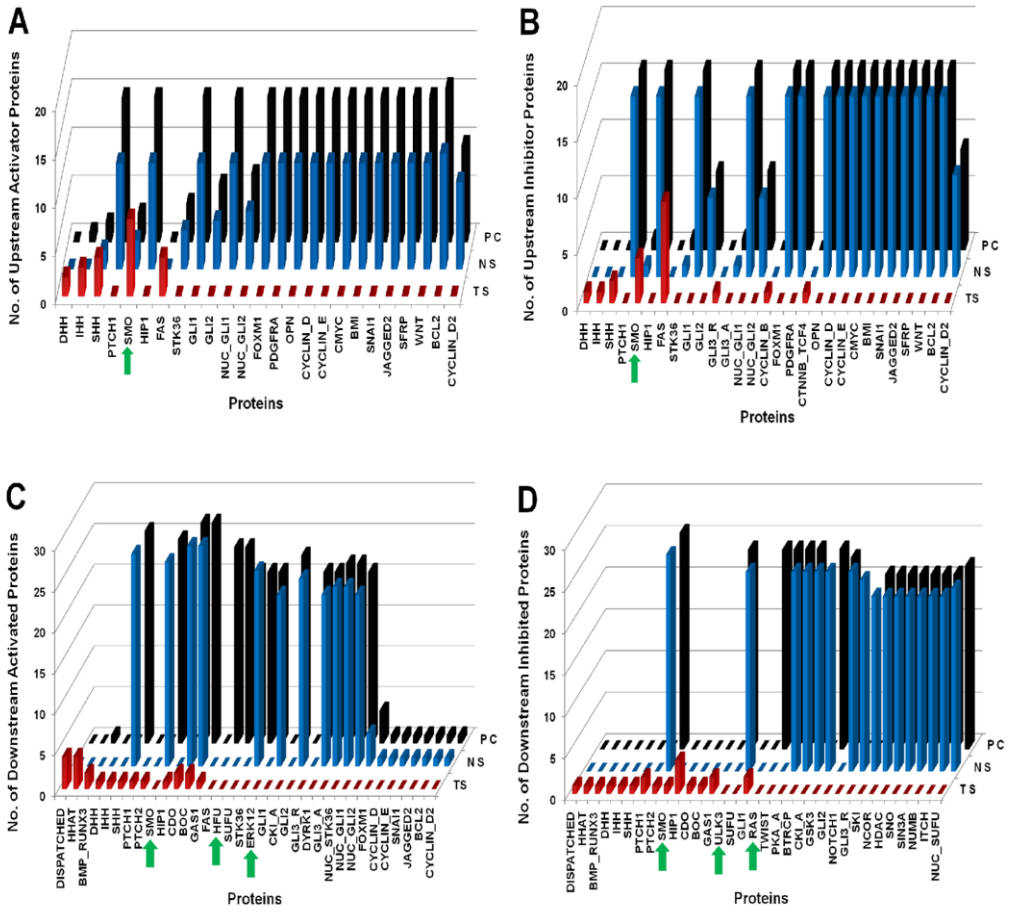

Figure 26: Comparative analyses of Pancreatic Cancer scenarios.
TS: Treated Scenario; NS: Normal Scenario; PC: Pancreatic Cancer Scenario. The green arrow heads are indicating the minimal combination of proteins which was inhibited in the drug treated perturbation analysis. **(A)** Represents number of Upstream activator proteins (Y-axis) activating the proteins (X-axis) representing significant variations. **(B)** Represents number of upstream inhibitory proteins (Y-axis) inhibiting the proteins (X-axis) representing significant variations. **(C)** Represents number of downstream proteins (Y-axis) activated by the proteins (X-axis) representing significant variations. **(D)** Represents number of downstream proteins (Y-axis) inhibited by the proteins (X-axis) representing significant variations. The numbers of downstream proteins inhibited in normal versus pancreatic cancer scenario remain same.

We have also perturbed the logical state of HFU and ULK3 proteins as these proteins are the common and essential auto-phosphorylating kinase proteins responsible for enhancing the activation of GLI1 and GLI2 in cytoplasm [316]. We have observed that the targeting SMO, HFU, ULK3, RAS, and ERK12 proteins

simultaneously are highly effective for suppressing the expressions of GLI transcription factors in cytoplasm. This combinatorial drug treatment scenario (TS) is able to down-regulate GLI transcription factor in cytoplasm and thus caused the down-regulation of the Hedgehog target oncogenes or oncoproteins (e.g., BMI, FOXM1, etc.). The expression of various output proteins like OPN, BMI, SNAI1, JAGGED2, PDGFRA are also not observed in our drug treated perturbation scenario (TS). We have compared the total number of upstream activators and inhibitors of each protein for Normal, Pancreatic and Treatment scenarios, respectively (**Figure 26A** and **Figure 26B**).

4.5.3.5 Validation of the Developed Cancer Models with Experimental Data

We have validated the developed cancer models with the experimental data available in published literature (see **Section 2.5.2**). We have validated the model outcomes by performing two different simulation approaches (SIM1 and SIM2). In Simulation 1 (SIM1), at the initial time step (Time = 0), we have considered the logical states of the Hedgehog proteins for each cancer scenario by collating the expression data of each protein observed in different experimental conditions (**Appendix Table 6**). Hence, the dataset we have used in SIM1 have variations within itself due to the different experimental set-up, but is not prone to the error caused in a single experimental set-up and thereby reduces the chance of getting true negative outputs. In Simulation 2 (SIM2), we have considered the expression of the input proteins specifically observed in a single experimental set-up (microarray experiment or proteomics study; EXP) for each cancer type. In SMI2, we have considered the logical state of a protein as "1" or "ON", if the corresponding gene of that protein or the proteomic expression of that protein is found to be differentially up-regulated (significantly) with respect to the normal cells or vice-versa. The proteins which showed insignificant expression in the microarray or proteomics datasets are considered as "un-determined" protein and their logical states are considered based on literature evidences.

Analyses of glioblastoma scenario (GS): In the microarray experimental data (GEO ID: GSE4290), we have found the up-regulation of IHH, RUNX3, SMO, STK36, TWIST, ERK12, RAS and down-regulation of tumor suppressor protein SUFU in GBM grade-IV cells (**Figure 27A**) [219].

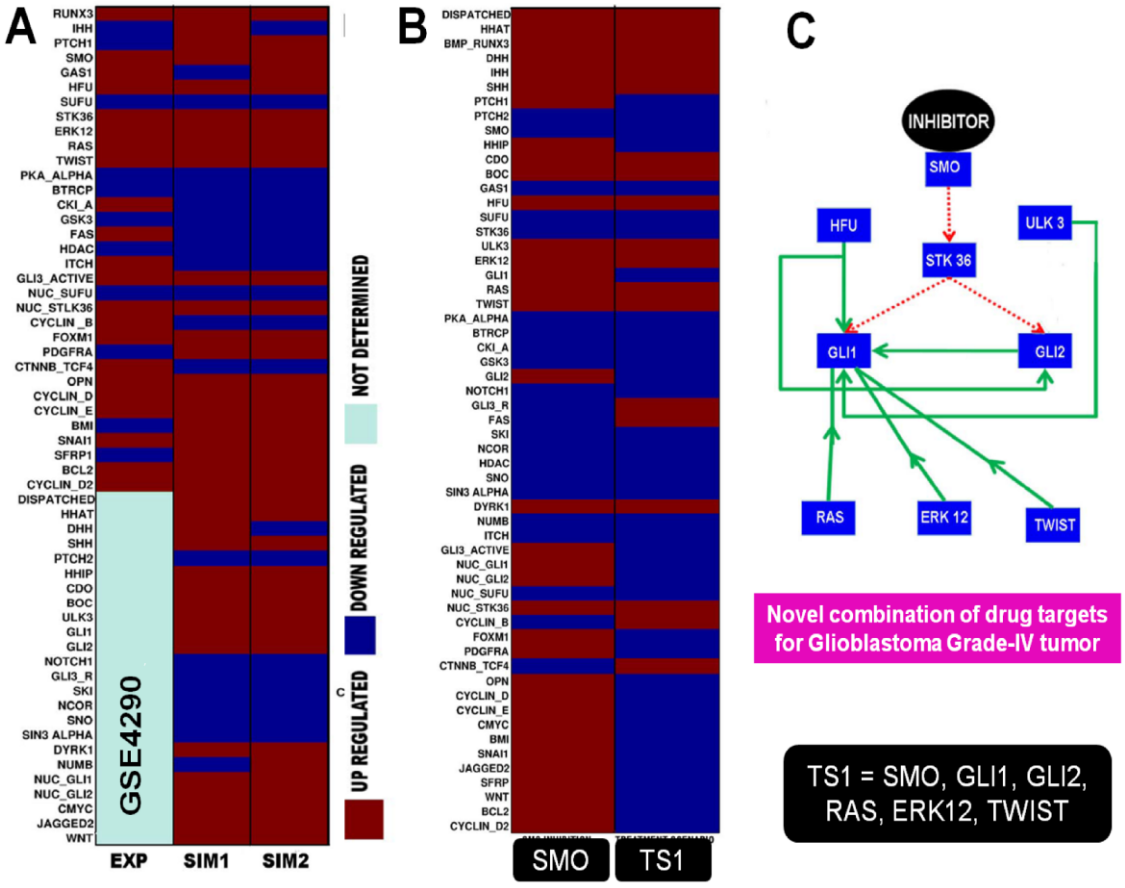

Figure 27: GBM model validation and signaling motif behind anti-SMO drug resistivity. **(A)** Represents the comparisons of the expressions of Hedgehog pathway proteins found in the experimental data (EXP) of GBM Grade-IV cells [219] and in the corresponding simulation (SIM1 and SIM2) data from our logical model. **(B)** First column represents the expressions of the proteins observed after inhibiting only SMO protein in GBM scenario (GS). The second column represents the expressions of the proteins observed in the alternative, combinatorial treatment scenario (TS) by perturbing SMO, GLI1, and GLI2 simultaneously in GS. **(C)** Shows the identified alternative pathways or reaction motif (shown by solid green arrows), which remain active even after the inhibition of SMO (anti-SMO treatment by the drugs Cyclopamine, Vismodegib, etc.) in GBM cells.

We have observed total 33 proteins out of 57 proteins of Hedgehog pathway model are differentially expressed (Up or Down) in Grade-IV GBM tumor cells as compared to the normal astrocytes cells (**Figure 27A**). The rest of the proteins, which

have shown insignificant expression level as compare to normal cells are also grouped together in **Figure 27A**. Within these 33 determined proteins, our simulation (SIM1; **Figure 27A**) has correctly predicted the expression levels of 22 proteins (66.66% accuracy). This result depicts the effect of the over-expressions of the activator proteins *viz.* HFU, RAS, TWIST on Hedgehog pathway in the development of Grade-IV GBM tumor cells. Furthermore, using the experimental expression data, we have performed Simulation number 2 (SIM2; **Figure 27A**) and compared the simulation outcomes with both Experimental (EXP) and (SIM1) (**Table 8**).

Table 8: Comparison of the Percentage of Accuracy between Experimental and Simulation Results.

Disease	Comparison	Not Determined/ Not Available	Correct Predictions	Incorrect Predictions	*Accuracy (in %)
Glioblastoma Grade-IV tumor	SIM1 with EXP	24	22	11	66.66
	SIM2 with SIM1	0	54	3	94.37
	SIM2 with EXP	24	25	8	75.75
Colon Cancer	SIM1 with EXP	52	5	0	100.00
	SIM2 with SIM1	0	57	57	100.00
	SIM2 with EXP	52	5	0	100.00
Pancreatic Cancer	SIM1 with EXP	13	25	19	56.80
	SIM2 with SIM1	0	47	10	82.45
	SIM2 with EXP	13	32	12	72.72

***Accuracy (%) = ((Correct Predictions) / (Correct Predictions) + (Incorrect Predictions)) X 100 (%)**

After comparing the simulation results of Simulation #2 (SIM2) with the Experimental dataset (EXP; **Figure 27A**), we have found that out of 33 experimentally determined proteins, we have correctly predicted the expression levels of 25 proteins

(75.75% accuracy). On the other hand, while comparing the simulation result between Simulation #1 (SIM1) and Simulation #2 (SIM2), we have found correctly predicted the expression levels of 54 proteins out of 57 proteins with the accuracy of 94.37% (**Table 8**). Therefore, in both the cases our GBM model (GS) has shown promising predictions as compared to the experimental data of GBM Grade-IV cells. Hence, the combination of drug targetable proteins identified from our drug treatment scenario (TS1) of GBM mode could be also used as potential drug targets for the treatment of GBM Grade-IV cells.

Moreover, in order to test our hypothesis that the newly identified combinatorial drug treatment therapy will be more effective than the anti-SMO targeted therapy, we have created the *in-silico* treatment scenario by targeting SMO protein in the GBM model. After that we have compared the simulation outputs with our treatment scenario (TS1) of GBM model (**Figure 27B**). We have observed that in the treatment scenario (TS1) of GBM model, the repressor form of GLI3 (i.e., GLI3R), and the pro-apoptotic protein FAS are up-regulated in comparison to the anti-SMO treatment scenario. Previous reports suggest that the up-regulation of these proteins will help to attenuate the active Hedgehog signaling in the rapidly proliferating GBM cells [394,395]. On the other hand, the target onco-proteins, such as FOXM1, PDGFRA, OPN, C-MYC, BMI, SNAI1, SFRP, BCL2 and the cell cycle regulatory proteins, such as CYCLIN-D, CYCLIN-D2, and CYCLIN-E are down-regulated in the treatment scenario (TS1) as compared to anti-SMO treatment (**Figure 27B**). We have also found that anti-SMO treatment scenario does not down-regulate the activity of GLI transcription factors (i.e., GLI1, GLI2, and GLI3_A).

Hence, we have hypothesized that there are some alternative molecular reaction routes present in the Hedgehog pathway, which can activate the GLI transcription factors even after inhibiting SMO protein in the cell membrane of the tumor cells. Hence, to find out such alternative routes, we have calculated and compared the dependency matrices of anti-SMO treatment and combinatorial treatment scenarios

performed on GBM model (GS). We have computed the inter-dependencies of each pathway species with every other species in the two scenarios and found out the proteins whose influence (activation) on the GLI transcription factors have reduced significantly in the combinatorial treatment scenario with respect to anti-SMO treatment scenario. We have also used the structural connectivity map of general Hedgehog signaling network (**Figure 21**) to locate the identified species and thus extracted the molecular reaction network through which the identified proteins activate GLI proteins in the GBM cells. We have observed that although in anti-SMO treatment scenario, the canonical Hedgehog pathway is blocked by SMO inhibitor, but the non-canonical activation via cross-talk molecules can still play the significant role to over-activate GLI transcription factors in the cytoplasm (**Figure 27C**). Cytoplasmic proteins, such as HFU, ULK3, RAS, TWIST, and ERK12 over-activates GLI1 and GLI2 proteins and thus continue the Hedgehog signaling mechanism in the GBM tumor cells.

Analyses of Colon cancer scenario (CC): In order to validate the model developed for simulating the Colon cancer scenario, we have considered the protein expressions profile of colorectal adeno-carcinoma cells, determined by RT-PCR, in situ hybridization, and immunohistochemistry techniques [396]. We have found that in the colon cancer cells, the Hedgehog pathway proteins such as SHH, PTCH1, HHIP, GLI1, GLI3_Active, and PDGFRA are significantly expressed as compared to the normal cells **(EXP; Figure 28A)**. The expression levels of rest of the proteins are considered as ''Not available'' and are grouped separately.

Our first simulation (SIM1) has correctly predicted the expressions of these 5 proteins with 100% accuracy (**SIM1; Figure 28A**). From this simulation, we have also predicted the expression levels of the other undetermined proteins, whose expression levels were not available in the experimental data [396].

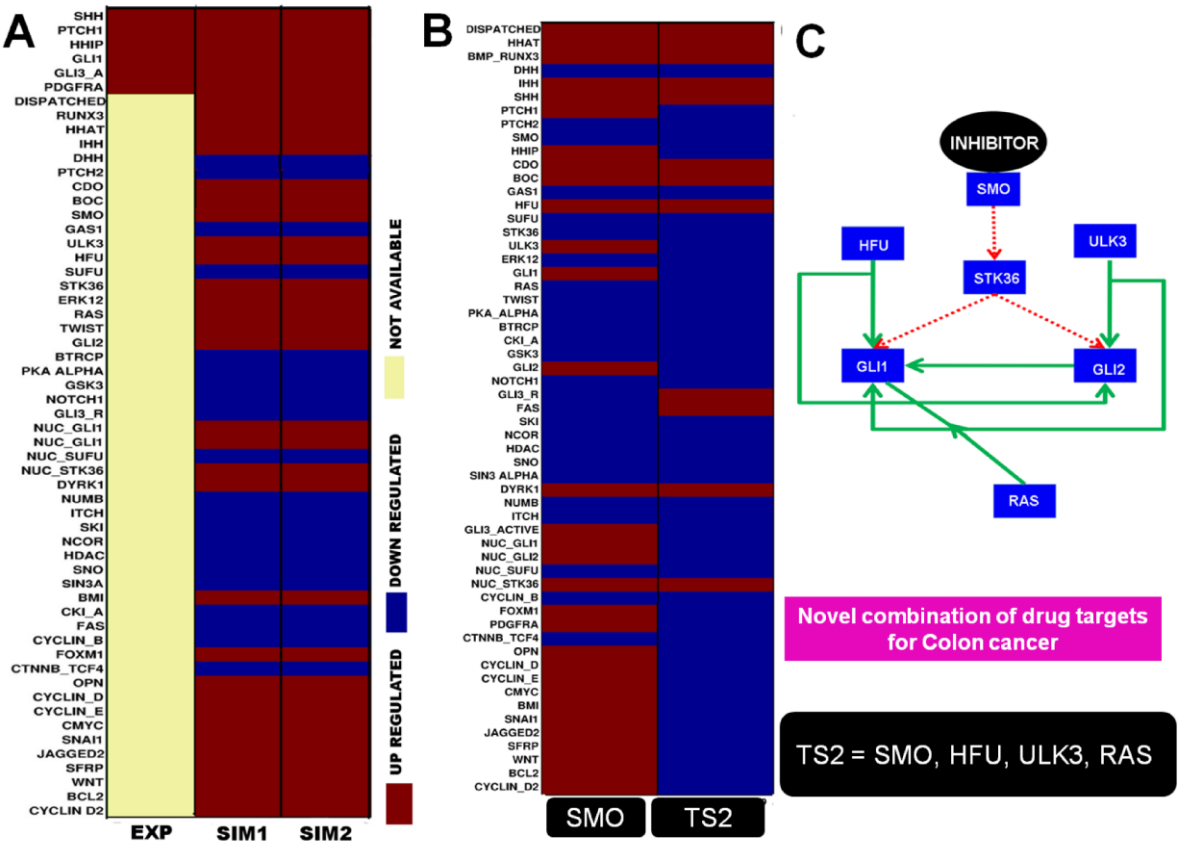

Figure 28: Colon cancer model validation and signaling motif behind anti-SMO drug resistivity.
(A) Represents the comparisons of the expressions of Hedgehog pathway proteins found in the experimental data (EXP) of colorectal adeno-carcinoma cells [396] and in the corresponding simulation (SIM1 and SIM2) data from our logical model. **(B)** First column represents the expressions of the proteins observed after inhibiting only SMO protein in colon cancer scenario (CC). The second column represents the expressions of the proteins observed in the alternative, combinatorial treatment scenario (TS2) by perturbing SMO, HFU, ULK3, and RAS proteins simultaneously in CC. **(C)** Shows the identified alternative pathways or reaction motif (shown by solid green arrows), which remain active even after the inhibition of SMO (anti-SMO treatment by the drugs Cyclopamine, Vismodegib, etc.) in colon cancer cells.

In the next simulation (SIM2; **Figure 28A**), we have computed the expressions of the same 5 identified proteins and compared the outputs with both Experiment (EXP) and Simulation 1 (SIM1) results (**Table 8**). In both the cases, we have observed same expression levels of the proteins (with 100% accuracy), which strongly validates our *in-silico* model of Colon cancer scenario. Hence, we can conclude that the combination of drug targetable proteins identified from our drug treatment scenario (TS) of Colon cancer model could be also used as potential drug targets for the

treatment of colon cancer cells.

Similar to the GBM model validation analyses discussed in the previous sections, here we have also found out that anti-SMO treatment of colon cancer cells is not effective to suppress the expressions of Hedgehog target onco-proteins and thus it is not effective to halt the growth of colon cancer cell progressions (**Figure 28B**). Hence, to identify the alternative molecular reaction routes, which can cause resistivity against anti-SMO treatment, we have studied the anti-SMO and combinatorial drug-targets (TS2) treatment scenarios. We have found out that like GBM cells, there also exists an alternative molecular reaction module, which can cause resistivity against anti-SMO treatment. This molecular route is basically the cross-talk interactions of HFU, ULK3, and RAS proteins with SMO, STK36, GLI1, and GLI2 reaction motifs in the proliferating cancer cells (**Figure 28C**). The reaction motif is very much robust and the inhibition of SMO protein in the cell membrane cannot down-regulate the GLI1 and GLI2 proteins completely and thus expressions of the Hedgehog pathway target onco-proteins becomes difficult to inhibit.

Analyses of Pancreatic cancer scenario (PC): In order to validate our developed pancreatic cancer model, we have used the previously microarray expression datasets of pancreatic cancer cells (GEO ID: GSE16515) [397]. We have extracted the expressions levels of 44 out of 57 proteins of our Hedgehog model from the published microarray expression data. The rest of the proteins (insignificant) are considered as "not determined" category (**Figure 29A**). The up regulation of HFU, ERK12, RAS are observed in the microarray expression data (**EXP; Figure 29A**). Out of the extracted 44 significantly expressed proteins, our first simulation method (SIM1) have correctly predicted the expression levels of 25 proteins with 56.80% accuracy (**Table 8**).

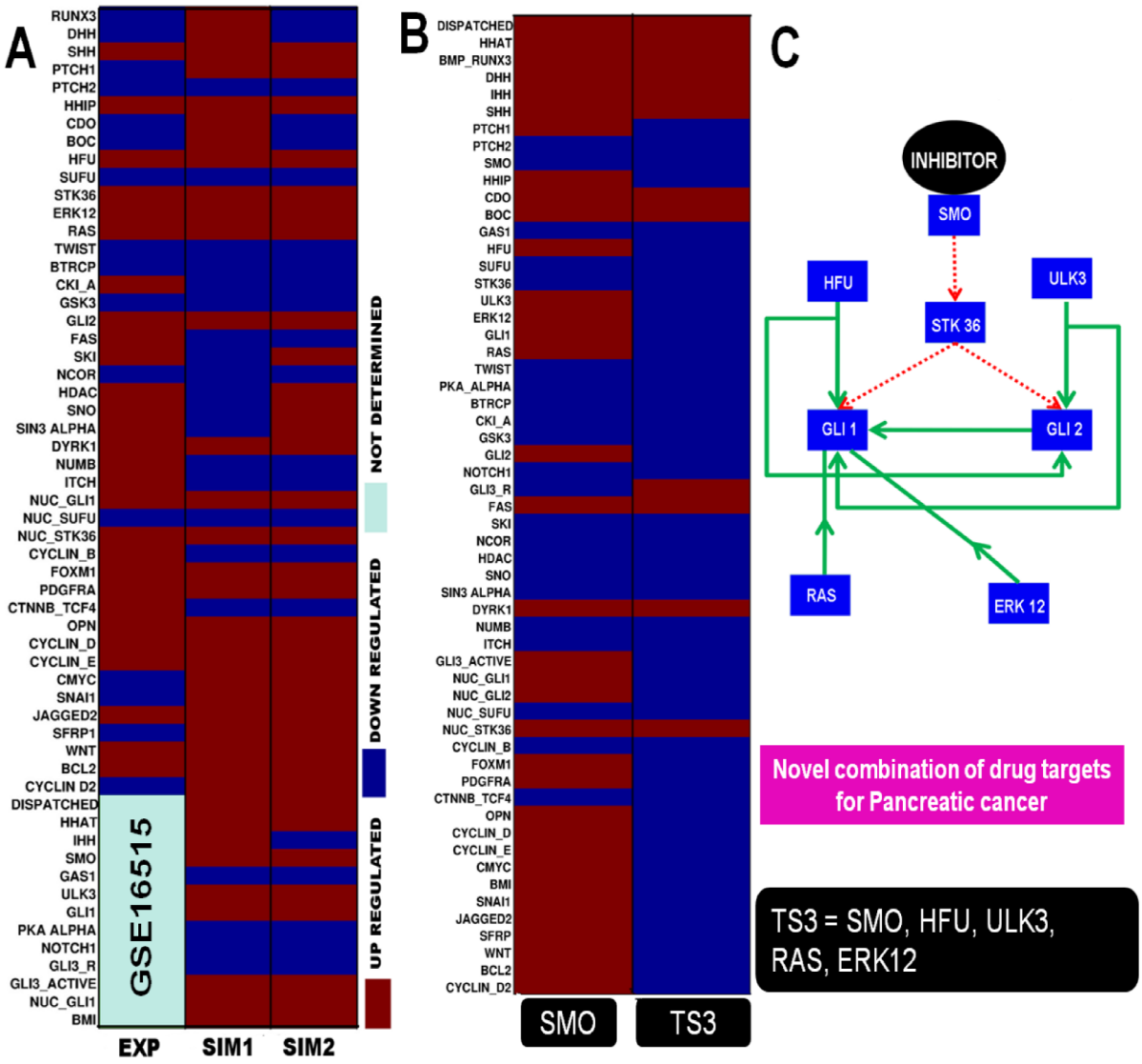

Figure 29: Pancreatic cancer model validation and signaling motif behind anti-SMO drug resistivity.

(A) Represents the comparisons of the expressions of Hedgehog pathway proteins found in the experimental data (EXP) of pancreatic cancer cells with respect to normal cells [397] and in the corresponding simulation (SIM1 and SIM2) data from our logical model. **(B)** First column represents the expressions of the proteins observed after inhibiting only SMO protein in pancreatic cancer scenario (PC). The second column represents the expressions of the proteins observed in the alternative, combinatorial treatment scenario (TS3) by perturbing SMO, HFU, ULK3, RAS, and ERK12 proteins simultaneously in PC. **(C)** Shows the identified alternative pathways or reaction motif (shown by solid green arrows), which remain active even after the inhibition of SMO (anti-SMO treatment by the drugs Cyclopamine, Vismodegib, etc.) in pancreatic cancer cells.

After that we have compared the expressions profiles of the genes from experimental data (EXP) and second simulation method (SIM2) (**Figure 29**) and found that the Simulation 2 have correctly predicted the expressions level of 32

proteins out of 44 significantly expressed proteins with 72.72% accuracy (**Table 8**). On the other hand while comparing the simulation results between Simulation 1 (SMI1) and Simulation 2 (SMI2), we have found that out of 57 proteins, our simulations have correctly predicted the expression levels of total 47 proteins with 82.45% accuracy (**Table 8**). Therefore, it is worth to mention that the developed model of pancreatic cancer (PC) model and the expression pattern are useful and can be used for further analyses, such as drug target identification, new hypotheses testing, etc.

Similar to the analyses of glioblastoma (GS) and colon cancer (CC) scenarios, here we have also observed the effect of anti-SMO treatment in pancreatic cancer cell model (PC) by perturbing SMO expression. We have observed that the inhibition of SMO by small molecule inhibitor can reduce the canonical Hedgehog pathway, but is unable to stop the activation signal transduces via cross-talks molecules, which eventually activate the GLI transcription factors in the cytoplasm. As a result, the anti-SMO treatment performed in our pancreatic cancer (PC) model has still shown the active expression pattern of the Hedgehog target onco-proteins (**Figure 29B**). On the other hand, our combination of drug targets (TS3: SMO, HFU, ULK3, RAS, ERK12) have successfully reduced the activity of the target onco-proteins (e.g., FOXM1, PDGFRA, BCL2, etc.) in the pancreatic cancer (PC) model.

We have also identified the alternative molecular reaction paths in the pancreatic cancer cells, which have caused the anti-SMO resistivity (**Figure 29C**). We have observed that the activation signal on GLI1 and GLI2 in cancer cells can come from the cytoplasmic molecules HFU, ULK3, RAS, and ERK12, even after blocking the activation signal from SMO protein to GLI1 or GLI2. The alternative reaction paths for GLI protein activation depict the intra-cellular molecular mechanisms through which the pancreatic cancer cells become resistance of anti-SMO treatment therapy.

4.6 DISCUSSION

Several attempts have been made so far to study the Hedgehog signaling pathway to understand its role in organ development and cancer pathogenesis using experimental and theoretical approaches [83,167,398,399]. Dillon et al. proposed reaction-diffusion kinetic models of Hedgehog signaling pathway to study Patched-Smoothened interaction and function of SHH as a long range morphogen [317]. A dynamic model based approach to analyze the signal transduction and transport mechanism of Sonic Hedgehog to study tissue patterning has also been done successfully [167]. All of these studies are based on quantitative modeling of coupled Ordinary Differential Equations (ODEs) or Partial Differential Equations (PDEs), the success of which immensely depend on rate equations, kinetic rate constants, and initial concentrations of the pathway species, and hence requires substantial amount of data from the biochemical studies. In this context, qualitative modeling approach is much more amenable to model larger biochemical reactions network than the quantitative modeling approaches based on coupled ODEs/PDEs.

Qualitative modeling approach, based on semi-dynamic, logic based equations using universal logic gates (AND, OR, NOT) can be particularly useful for this purpose, in which the model formulation does not depend on finer details of reaction kinetics and a large number of pathway species and their interactions/ reactions/ connections are modeled by using their logical relationships [400,401]. This also enables our understanding that how a large scale biochemical reactions network behaves in different conditions at steady-state. Previous attempts have used Boolean logic to model developmental pathways for the topological study of interactions that enable prediction of patterning in *Drosophila melanogaster* [402], and exploration of the effect of transient perturbations on development of wild type pattern for the segment polarity network [403]. However, the implication of logic based modeling approach to study the Hedgehog pathway dynamics in cancer pathogenesis is not performed earlier. We have observed that the successful implications of logic based

approach require the comprehensive knowledge of the network topology and the prior knowledge of reaction mechanisms and interdependency of all the associated pathway species, starting from extra-cellular ligands to gene transcription regulators. We noticed that there was no single database or literature evidence available from which we could get such comprehensive information of the pathway. The pathway information was highly dispersed in various literature and pathway databases, and hence it required manual curation with the expertise of domain knowledge.

In this study, at first, we reconstructed the Hedgehog signaling network using the information available from various sources and tried to include as many as possible target output proteins of this pathway (**Figure 19)**. We included the connections of the output proteins JAGGED2, WNT, SFRP, CYCLIN_B, CYCLIN_D, CYCLIN_D2, CYCLIN_E, OPN, SNAI1, CMYC, BMI, BCL2, FOXM1, and PDGFRA with the phenotypic outcomes or cellular responses (like Cell proliferation, Cell cycle progression and Endothelial to Mesenchymal Transition etc.) and also with three other important pathways, *viz.* WNT, NOTCH, and Anti-Apoptosis [86,337,376,377,379-381,404-407]. The inclusion of these output proteins in the reconstructed pathway map helped us to understand that how Hedgehog signaling pathway controls the major developmental procedures of a cell, such as cell division, cell proliferation and also maintains the cross-talks with other pathways. Also, it is important to note that most of the output proteins presented in the reconstructed map are oncoproteins and thus the developed model is correlated with various types of cancers. Also, we included the cross-talk proteins, such as ERK12, RAS, TWIST, FAS, NOTCH1, which are not the core hedgehog pathway proteins. The inclusion of these proteins in our reconstructed map helped us to depict the regulation or cross-talks of Hedgehog pathway with other molecules associated with different signaling pathways, *viz.* WNT, NOTCH, and MAPK, etc. Including these non-core proteins, as far as the literature and database are concerned, this reconstructed map of Hedgehog signaling pathway represents the highest number of molecules and interactions, and

is considered for further computational analysis.

After the pathway reconstruction, we performed graph theoretical and logical analysis in the newly reconstructed signaling network. We observed that GLI1 protein had most number of connections with other proteins in the entire network (**Figure 21**). Hence, we denoted GLI1 as the "hub" protein in the network, which if targeted by inhibitor, would trigger maximum damage in the flow of signaling cascades in the entire network. Our observation also matches with biological evidences, in which it is already identified that certain perturbation (e.g., mutation, malfunction, high or low expression, etc.) of this protein can affect the normal functioning of Hedgehog pathway and may cause several disease pathologies, such as cancers, birth defects, improper organ development, etc. [408-411]. This proteins is also found to be associated with the etiology of Glioblastoma, Colon, and Pancreatic cancers, in which cancer cells have shown aberrant activation of GLI1 [388,412,413]. Moreover, we found that PTCH1 had higher importance in the network (**Figure 22**), which supports numerous experimental studies where mutation of PTCH1 protein was shown to affect the flow of normal Hedgehog signal and cause pancreatic and colon cancer [414]. Higher concentration (i.e., active or ON state) of PTCH1 protein in membrane is helpful to regulate the activated SMO protein in membrane, so that it cannot further activate GLI proteins in the cytoplasm [415]. Also, the high IN-DEGREE value showed by SHH also implies that this ligand is mostly regulated by some other extracellular proteins at the time of binding with PTCH1/2 (**Table 7)**. It is experimentally proven that the proteins CDO and BOC both help to process (or activate) the SHH ligand in the Hedgehog ligand secreting cells, whereas GAS1 inhibits SHH to bind with Patched proteins (PTCH1/2) in the membrane [416]. SHH is one of the main activator of this pathway in several cancer cells, hence from our result it is clear that targeting CDO or BOC by small molecule inhibitor in the ligand secreting cells could be helpful to attenuate the pathway activity in the receiving cancer cells.

On the other hand, the higher OUT DEGREE value of the nuclear proteins signifies the role of these proteins to process the hedgehog input signal into the output products (**Table 7**). Therefore, mutation or malfunction of these proteins in nucleus would cause over production of various target oncoproteins of the Hedgehog pathway. In the case of Eigenvector centrality score, GLI1, SMO, PTCH1, GLI3_A had shown significant higher scores as compared to other proteins in the network. It implies that these proteins not only have higher number of connections, but are also connected with other highly prestigious nodes in the network. We also observed that although GLI1 and GLI2 had higher number of connections in the network, but the eigenvector centrality of GLI2 was lower as compared to GLI1. We analyzed the reason behind this phenomenon, and found that the downstream proteins of GLI1 and GLI2 were making the main difference. The downstream node of GLI1 is NUC_GLI1, which is also a central node in the network and is connected to many other proteins. The downstream of GLI2 protein is NUC_GLI2, which is connected with comparably lesser number of proteins in the network as compared to NUC_GLI1. This result was also reflected in our Logical analyses, where we observed that in case of Glioblastoma, Colon and Pancreatic cancer scenarios the number of upstream activator/inhibitor species and downstream activated species of NUC_GLI1 was high compared to the NUC_GLI2.

Also, GLI1, NUC_GLI1, SMO, STK36, and PTCH1 had shown high Betweenness centrality score and GLI1, NUC_SUFU, NUC_STK36, DYRK1, NUMB, and ITCH had shown high Closeness Centrality score in our analysis (**Figure 22**). It was clearly seen that GLI1 had significant scores in all three kinds of centrality parameters and thus we can say that GLI1 was the most centrally situated protein in our reconstructed Hedgehog signaling network. Therefore, knock-out or mutation of this protein from the Hedgehog signaling network would cause most significant effect in the normal cell and would lead the cells towards the cancerous stage [406]. Apart from GLI1 protein, SMO was also found as another important protein which featured high

Eigenvector and Betweenness centrality scores (**Table 7**). The Closeness centrality score showed by most of the nuclear proteins implied that these were the proteins that were connected with lower number connections to the other proteins in the network and thus regulated by maximum number of proteins in the network. Therefore, certain disturbance in the other proteins can perturb the normal activity of these proteins in the network.

In order to analyze the importance of this signaling pathway as well as the individual proteins involved in Glioblastoma, Colon, and Pancreatic cancer cell lines, we developed Boolean or Logical models or scenarios. We found that, if we perturbed the logical states of SMO, GLI1 and GLI2 proteins from 1 to 0 in Glioblastoma model scenario (GS), it will be possible to suppress the expression of various output proteins (e.g., JAGGED2, WNT, SFRP, CYCLIN_B, CYCLIN_D, CYCLIN_D2, CYCLIN_E, OPN, SNAI1, CMYC, BMI, BCL2, FOXM1, PDGFRA, etc.) as well as the phenotypic expressions of the Glioblastoma affected cell (**Figure 24**). Therefore we propose that inhibition of these proteins would be helpful for the therapeutic treatment of Glioblastoma. We observed that there were several proteins which were activating the GLI transcription factors in cytoplasm and these proteins were connected with other signalling pathways. Therefore inhibiting those proteins could affect the normal functioning of other pathways. In order to prevent such side-effects, we propose that, in Glioblastoma cancer cells, selectively targeting SMO in cell membrane, and inhibiting the nuclear translocation of activated GLI1 and GLI2 within cytoplasm would be more effective to completely shut down the Hedgehog pathway by suppressing the activity of different proteins responsible for uncontrolled cellular proliferation. Also, inhibition of the activity of GLI2 protein was necessary to prevent its positive feedback loop to GLI1 activation.

In case of Colon cancer scenario, we found that inhibition of SMO, HFU, ULK3 and RAS was useful to suppress the expressions of various responsive proteins of Hedgehog pathway (**Figure 25**). Several experimental studies have already proven

the correlation of the mutation of ras gene with the Colon cancer [417,418]. As this protein was one of the activator of GLI1 in cytoplasm, therefore inhibition of RAS may also help to shut down the Hedgehog pathway. Hence, targeting RAS in cytoplasm would be effective to reduce the over activation of GLI proteins in cytoplasm and consequently several oncoproteins like PDGFRA, BMI, SNAI1, etc. of Hedgehog signaling. Few studies have already reported that mutated RAS family proteins could be the suitable drug targets for treating various types of cancer [419,420] [90,91], which also supports our computational findings. On the other hand, our analysis also revealed that suppression of SMO, HFU and ULK3 by external drug would be required to completely shut down the Hedgehog signaling in the colon cancer cells. Therefore, we propose that in order to shut down the effect of Hedgehog signaling in colon cancer cell line for in-vitro or in-vivo analysis, one could also think a combination of drugs that will suppress the activity of SMO, HFU, ULK3 and RAS proteins altogether.

From our analysis, we also suggested a minimal combination of proteins, SMO, HFU, ULK3, RAS and ERK12, the expressions of which needed to be inhibited so as to control the effect of mutated hedgehog signaling pathway in pancreatic cancer by suppressing the activity of different proteins responsible for uncontrolled cellular proliferation (**Figure 26**). Also, suppression of HFU and ULK3 proteins in cytoplasm would be helpful to block the enhanced activation of GLI proteins. Experimental studies have already proven their role to activate and enhance the production of phosphorylated GLI transcription factors in the Hedgehog signaling pathway [343]. This perturbation would decrease the concentration of NUC_GLI1, NUC_GLI2 and GLI3_A in nucleus and consequently bring down the production of various output proteins of Hedgehog signaling network.

All the developed cancer models were validated with the existing microarray data for these three types of cancer and also simulated using the experimental data (**Figure 27A, Figure 28A,** and **Figure 29A**). In both the cases our model predictions

showed promising validations with the experimental results (**Table 8**). Moreover, to determine the alternative pathways which are still active under drug treatment, we refined the model simulation by only inhibiting SMO in three types of cancer scenarios and compared the results with the corresponding treatment scenarios (**Figure 27B, Figure 28B,** and **Figure 29B**). Using this analysis we found the alternative pathways in each cancer scenarios, which have the ability to up-regulate the GLI proteins without the help of hedgehog ligands (**Figure 27C, Figure 28C,** and **Figure 29C**).

4.7 CONCLUSION

Hence, we have proven our proposed hypotheses of the existence of active signaling modules in Hedgehog pathway, which are responsible for creating drug-resistivity during anti-SMO treatment. Also, we have proven that the combinations of potential drug targets (TS1, TS2, and TS3) can be more effective than targeting the single receptor protein SMO in target-based anti-cancer therapeutics. However, our logical model had also some limitations, as it was unable to present the quantitative measurements of the expression levels of the genes/proteins, the rate of inhibition required in combinatorial therapy, or the dose dependent scenarios in the treatment of cancer cells. Despite its limitations, our model helped us to identify the expression levels of different proteins, which were ‘‘not determined’’ in the experiments. The proteins identified as probable drug targets from these simulations were not novel targets, and individually their efficacy as drug targets was tested experimentally. But the optimal combinations of drug targets used in these simulations are new and to the best of our knowledge, the effectiveness of targeting these proteins in combinatorial drug target therapy is not tested yet. Therefore, our *in-silico* simulations identified the "novel combinations" of drug targets from the large scale network of Hedgehog pathway, which can be helpful to the experimental biologists as well as pharmacologists to develop optimal targeted therapy.

Chapter 5

5 UNDERSTANDING THE ROLE OF NOTCH SIGNALING PATHWAY IN THE GROWTH OF GLIOBLASTOMA CELLS UNDER STEADY STATE

5.1 INTRODUCTION[4]

Notch signaling pathway is a conserved developmental intra-cellular pathway, which is involved in cell fate determination, stem cells renewal at the stage of embryonic development and also plays essential roles in tissue and organ development, hematopoiesis, vascular growth, etc. in the adult stage of human [421-424]. The canonical Notch pathway is activated when the membrane-associated ligands (DELTA or JAGGED family proteins) of the transducer cells interact on the membrane-bound NOTCH receptor proteins of the juxtaposed receiver cells [425]. After the ligand-receptor binding, the complex undergoes several enzymatic reactions, which involve the enzymes, such as ADAM, TACE, and the GAMMA-SECRETASE enzyme complex formed by PRESENILIN, NICASTRIN, APH-1, and PEN-2 proteins in the membrane of the receiver cells. The ankyrin domains present in the intracellular portion of NOTCH family receptor proteins (NOTCH1, NOTCH2, NOTCH3, and NOTCH4) are catalytically cleaved with the help of the enzyme complex GAMMA-SECRETASE. The cleaved portion of is a truncated protein and generally known as the Notch intracellular domain (or NICD). There are four types of NICDs *viz.* NICD1, NICD2, NICD3, NICD4 can be present in the cells, which are mainly produced from the four homologues of NOTCH proteins. NICDs then translocates into the nucleus, where it binds with the transcription factor complex

[4] The materials of this chapter has been taken verbatim from our previously published articles (a) Chowdhury and Sarkar, *Clin Exp Pharmacol*., 2013

CSL (CBF1/Su(H)/Lag-1) and starts transcription of its target genes, such as HES/HEY family of genes and various others genes related to cell cycle progression, anti-apoptosis, and so on [426-429]. Thus, Notch pathway operates various cellular functions including cell division, cell cycle progressions, cell growth, anti-apoptosis, differentiation, etc. through transcription of various target genes and activation of other signaling and metabolic pathways in the cell [430-434].

However, sometimes this entire concerted process diverges from its main objective and triggers the development of various types of cancers, such as glioblastoma, breast cancer, osteosarcoma, prostate cancer, melanoma, T-cell acute lymphoblastic lymphoma/leukemia, and so on [435-442]. In the previous reports, we have found that Notch pathway plays significant roles in the maturations of astrocytes/oligodendrocytes cells from the neural stem cells in the sub-ventricular zone (SVZ) of the adult human brain [443-445]. Previous reports have also asserted that during tumorigenesis in the human brain, specific somatic or genetic mutations in the matured astrocytes/oligodendrocytes (commonly known as Glial cells) cause abnormal growth and eventually convert the cells towards tumor formation [446,447]. Hence, it is expected to observe that the suppressor pathways of tumorigenesis, such as Notch pathway also gets hijacked by the mutations in the glial cells, and is redirected towards the development of astrocytomas or oligodendrogliomas (or Glioblastoma) in SVZ [448,449].

5.2 CURRENT CHALLENGES

The identification of drug-targetable proteins in the Notch signaling pathway to reduce the growth of Glioblastoma tumor cells is always a major area of research in target-based tumor therapeutics [450]. The membrane-bound enzyme protein, GAMMA-SECRETASE is widely used as a potential drug target to reduce the Notch pathway activity in Glioblastoma treatment therapy [451]. Although, the GAMMA-SECRETASE inhibition (GSI) therapy had proven succesfull for reducing the growth

of GBM cells, but later studies have found that GBM cells may generate resistivity against GSI therapy [452]. Also, during the clinical trial (Phase III trial of Alzheimer's disease) of the small molecule inhibitor of GAMMA-SECRETASE protein Semagacestat (LY450139), the molecule has failed to meet the desired objective and is found to be associated with several risk factors including skin cancer [453]. GSI is also associated with several others long term side effects, such as gastrointestinal toxicity and diarrhea [454]. The principal reason behind the appearance of such toxicity in GSI is because of its wide array of substrates in the cells (including NOTCH) and its involvement in the enzymatic cleavage of different proteins in the cell and mitochondrial membrane [455-458]. An *in-vivo* study performed on transgenic mice model to observe the consequence of the inhibition of GAMMA-SECRETASE protein had shown that GSI therapy had impaired the development of lymphocyte and altered the tissue morphology in the intestine [459]. Due to these reasons, although GSI is effective to inhibit Notch pathway activity in cancer cells, still the target protein GAMMA-SECRETASE cannot be used for the treatment of GBM and other cancer patients. The other drug targets of this pathway, such as NOTCH1, NOTCH4, DLL4, NRARP etc. have also failed to show desirable outcomes due to toxicity and side effects [31].

On the other hand, identification of suitable and alternative drug targets for the inhibition of aberrantly activated Notch pathway in Glioblasotma (GBM) cells is also a major challenge. It requires the understanding of the complex mechanisms that are governing the normal functioning of Notch signaling pathway in normal cells, and the processes through which it gets deregulated in the cancer cells. The complex circuitry of this pathway, which includes several regulations, feedback loops, cross-talks, etc. have made the drug targets identification process a more challenging task. Moreover, the comprehensive information of this pathway is also limited in the cell signaling databases, and the information is mostly scattered in literature. Integrations of the scattered experimental findings are also not done yet. Moreover, to study

effects of new drug targets through *in-vitro* and *in-vivo* analyses is also a challenging task.

5.3 HYPOTHESES AND OBJECTIVES

In order to address the previously mentioned challenges, here we have hypothesized that the alternative of GAMMA-SECRETASE protein as drug target can be used for the treatment of GBM cells which are non-responsive to GSI therapy. Our conjecture is that by suppressing the Notch pathway activity at the steady-state level, there is a chance of reducing the expressions of the target oncogenes of Notch pathway and thus we can reduce the growth of Glioblastoma. Hence, to test our hypotheses, we need to understand the mechanistic regulations of Notch signaling cascades under normal and GBM pathogenic scenario and the effects of pathway regulations in the expressions of the target genes/proteins in these two scenarios. We require the absolute understandings of the topological properties of the network by finding out the largest "hub" and central nodes, diversities of the flow of reaction cascades from the extra-cellular region to the nucleus, influence of several feedbacks and cross-talk reactions, dependencies of various pathway species with cellular phenotypes, and so on in the outcomes of tumorigenesis process.

Our main objective in this work is to identify the alternative molecular reaction routes other than the GAMMA-SECRETASE mediated pathway activation process in the comprehensive network of Notch pathway, which can be further used for suppressing the pathway activities in GBM cells. To fulfill this objective, we will at first curate a most up to date and comprehensive map of human cell specific Notch signaling pathway by assembling the interactions data found in literature and signaling pathway databases (**Appendix Table 7**). The reconstructed map will be then analyzed using graph theoretical analyses with the aim of identifying central nodes in the network. The reconstructed map will be also analyzed using logical equations and a master model will be created to simulate the normal and GBM

scenarios. The expressions/activities of the proteins at the steady-states will be measured and compared between these scenarios. The models outcomes will be validated with the experimental data and if the master model shows promising results to simulate the normal and GBM scenarios, then it will be further used for simulating the effects of non-responsive GAMMA-SECRETASE inhibition (GSI) scenario in the GBM model. The objective of developing the GSI scenario is to analyze and compare the expression patterns of target oncogenes/proteins in GSI resistant cells versus GBM scenarios, and subsequently identify the alternative molecular reaction routes (or drug targets) to suppress the Notch activity without using GSI. Finally, by combining the outcomes of graph theoretical results for identifying the hub proteins and the outcomes of logical simulations for identifying the molecular reaction routes, our objective will be to perform *in-silico* perturbation analyses of an individual or a set of proteins on GBM scenario with the aim of finding out alternative drug target of GAMMA-SECRETASE enzyme complex.

5.4 RESULTS

5.4.1 Reconstructed Map of Notch Signaling Network

5.4.1.1 Pathway Statistics

In this work, we have reconstructed a comprehensive, most up to date and largest human cell specific Notch signaling pathway to the best of our knowledge. The entire Notch pathway (**Figure 30**) was annotated and reconstructed manually by collating the data from various literature, experimental findings, and biological databases (**Appendix Table 1**). Although the basic core pathway is same as the pathway map available in the existing signaling databases, but to the best of our knowledge the newly reconstructed pathway map consists up to date information of interaction data which is not available in any freely available major academic databases.

In this reconstructed pathway map, we included 115 molecules (96 core and 19 cross talking pathway molecules including proteins and organic compounds) and

231 molecular interactions.

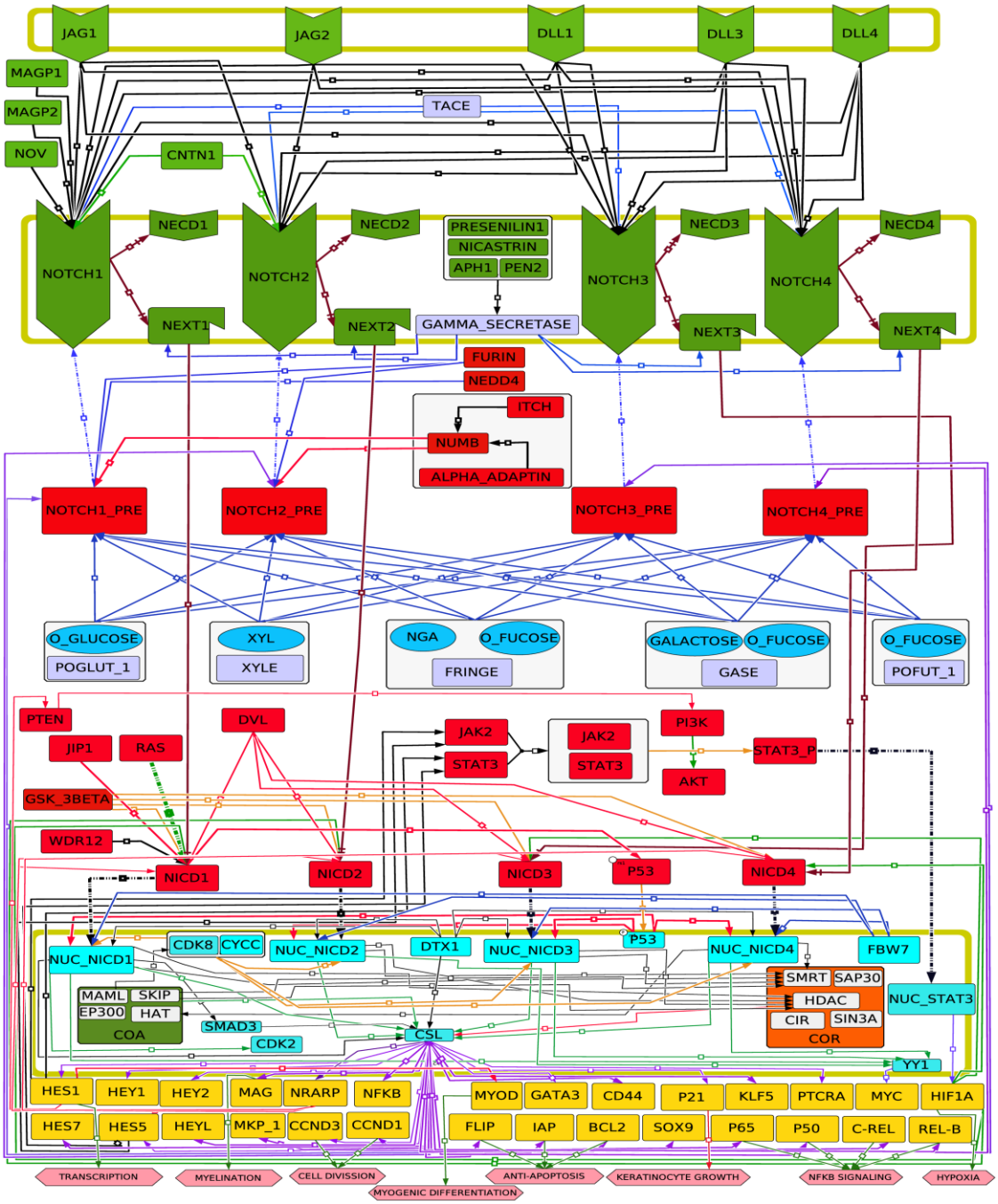

Figure 30: Reconstructed map of the human Notch signaling pathway.
All the pathway species (proteins, metabolites, truncated proteins, co-activators, co-repressors, and phenotypes) in the diagram are color coded as per the locations and hierarchy. Physical interactions are shown by black arrow, whereas blue colored arrows are enzyme mediated reactions. Protein truncation is shown by brown arrow. Gene transcription, followed by protein synthesis is depicted by magenta colored arrows, nuclear translocation is by dotted black arrows, activation is by green arrows, and phosphorylation is by orange arrows. Red arrows depict inhibitions. All the target genes/proteins (yellow colored) of Notch pathway are mapped with phenotypes at the bottom level of the diagram.

Different types of molecular reactions such as physical interaction, enzymatic reactions, phosphorylation, protein synthesis, activation, inhibition, and nuclear translocation etc. were also considered to construct the pathway map. A comparison between this reconstructed notch pathway data (i.e., total number of molecules and interactions) with the pathway information from other major biochemical signaling databases (e.g., KEGG, Biocarta, Netpath, etc.) is presented in **Table 9**.

Table 9: Comparative Statistics of the Reconstructed Notch Pathway Data

Database Name	Molecules*	Interactions*
Our Reconstructed Pathway	115	231
KEGG	24	15
Biocarta	7	10
NetPath	85	138
Pathway Central	16	11
Cell Signaling Technology	22	18
Protein Lounge	13	12

5.4.1.2 Description of the Reconstructed Notch Pathway

Reactions in the extracellular & membrane region: Previous experimental findings have identified that all four homologues of NOTCH receptor protein can express in the cell membrane of the Notch "Signal receiver" cells, whereas the membrane bound ligands JAG/DLL class proteins are expressed in the membrane of notch signal "Transducer cells" [460]. In the reconstructed pathway map (**Figure 30**), we have shown these proteins in the membrane regions of both the cells. Moreover, Notch pathway can also be activated by the interactions between microfibrillar proteins MAGP1 and MAGP2, CONTACTIN/F3 (CNTN1), or Nephroblastoma over-expressed (NOV) proteins with Notch receptors [461-464]. We have shown these proteins separately in the extracellular regions in the pathway map. However, all ligand-receptors interactions (i.e., JAG/DLL mediated or MAGP1/ MAGP2/ CNTN1/ NOV mediated) follow the complex common proteolytic cleavage of NOTCH receptors and produces Notch Extracellular Domain (NECD) and Notch Extracellular Truncated Protein (NEXT) from the four Notch receptor proteins [460].

These proteolytic cleavage reactions are shown by brown arrows in the pathway map. We have found that the metalloprotease enzyme TACE helps at this step to cleave the Notch receptors [465]. Subsequent to these reactions, all the truncated proteins NEXT1/2/3/4 are again cleaved by another proteolytic enzyme GAMMA SECRETASE enzyme complex present in the cell membrane (**Figure 30**).

Reactions in the cytoplasm: Followed by the GAMMA-SECRETASE mediated reactions, four NEXT proteins produce four homologues of Notch Intracellular Domains (NICD1/2/3/4) which are then translocated into the nucleus [460]. During this translocation time, NICD proteins encounter other activators (RAS, GSK_3BETA, and WDR12) and inhibitors (DVL, JIP1) proteins in the cytoplasm [466-469]. The post translation modification of Notch precursor proteins (NOTCH1_PRE, NOTCH2_PRE, NOTCH3_PRE and NOTCH4_PRE) are also shown included in the reconstructed map. In these reactions, notch precursors undergo several glycosylation or fucosylation reactions by Glucose, Galactose, Fucose and the enzymes POGLUT_1, FRINGE, GASE, POFUT_1, etc. in the cytoplasm or ER [470-473]. These post translational modifications of notch precursors increase the specificity of ligand receptors interactions in the membrane. On the other hand, Xylosylatin by Xylose with the help of the enzyme Xylosyltransferase (XYLE) is also observed in several cases which in turn reduce the specificity of notch ligand bindings [470].

Reactions in the nucleus: The truncated NICD1/2/3/4 proteins enter into the nucleus and interact with another transcription factor complex CSL [460]. In the absence of NICDs, CSL forms a transcription repressor complex with the transcription co-repressor (COR) proteins SMRT, SAP30, HDAC, CIR, and SIN3A in the nucleus [474-477]. On the other hand, in presence of NICDs, when CSL makes transcription factor complex, there are other set of proteins, *viz.* MAML, SKIP, HAT, and EP300 act as transcription co-activator (COA). The transcription factor complex of NICD-CSL-COA transcribe the Notch target genes/proteins, such as HES1, HES5,

HEY1, HEY2, HEYL, BCL2, P65, and NOTCH1/2/3/4, etc. [33,478-482] Besides this CSL/NICD/COR/COA mediated transcription, it is also reported that Notch pathway can be activated by CONTACTIN/F3 (CNTN1) mediated interaction, which requires the involvements of DTX1 as a transcription co-activator to produce the protein MAG. This protein is involved in the oligodendrocyte maturation and myelination [463].

Cross-talks with other pathways: Notch pathway has cross connections with different signaling pathways such as, JAK/STAT, PTEN/PI3K/AKT, RAS/MAPK, TGFB/SMAD3, CYCLIN/CDK, HYPOXIA/HIF1A, BCL2/IAP/ANTI-APOPTOSIS, P65/P50/NFKB etc., [445,473,479,483-491]. In our reconstructed Notch pathway map, we have tried to include more number of cross-talks reactions with core proteins of Notch pathway. We have included only those cross-talk molecules of other pathways that had direct interaction/influence on the core proteins of Notch pathway. For example, we have added the cross-talk of PTEN/ PI3K/AKT pathway with HES protein in the pathway map. We have found in the literature that the output proteins HES1 or HES5 directly inhibit the activity of PTEN protein, and thus activate the PI3K/AKT pathway [492,493]. Please see **Section 2.2** of **Materials and Methods** for more details of pathway reconstruction procedures.

5.4.2 Topology of Notch signaling network

5.4.2.1 Identification of Feedback Loops

The reconstructed map of Notch signaling pathway is further analyzed by graph or network theory. We have used MCODE algorithm to find out the sub-graphs (or cluster of nodes) in the entire network [494]. Using this algorithm (implemented as a plug-in package in Cytoscape [209]), we have identified few feedback loops in the reconstructed map of Notch pathway. We have found a cyclic feedback loop HIF1A --> NICD1 --> HES1/5 --> JAK/STAT --> HIF1A in the network. We have observed that HIF1A (a core protein of Hypoxia pathway) can activate NICD1/2/3/4 in the

cytoplasm, which in turn helps to produce HES1/5 as well as the other Notch pathway target proteins [495]. We have found that HES1/5 proteins help to stabilize the JAK2/STAT3 complex formation. The stable JAK3/STAT3 complex is required to produce phosphorylated form of STAT3 protein (STAT3_P), which binds in the promoter region of *HIF1A* gene and synthesize its transcript [491].

We have also found a double negative feedback loop between P53 and Notch signaling proteins NICDs. We have found that the phosphorylated P53 protein inhibits NICDs in the nucleus (NUC_NICD1/2/3/4) and terminates its transcription mechanisms. On the other hand, the phosphorylation reaction of P53 is blocked by NICD1/2/3/4 in the cytoplasm [488,496]. Besides, these two sub-graphs or clusters, we have also observed a comparatively larger feedback loop within the pathway, which is formed by the the precursor of NOTCH proteins, Notch receptors, and NICDs (NOTCH_Precussors --> NOTCH --> NICD --> NOTCH_Precussors). Synthesis of Notch precursor proteins contributes a strong positive feedback effect in the entire network. Besides, we have also found another strong negative feedback loop, which is formed between Notch-regulated ankyrin repeat-containing protein (NRARP) and NICDs, which was one of the Notch targeted output proteins. NUC_NICD1/2/3/4 produces this protein as a target, but this protein inhibits the translocation of NICD1/2/3/4 proteins to the nucleus [497].

5.4.2.2 Identification of Central Nodes in the Network

In order to identify the important "Hubs" or central nodes or proteins in the network, we have measured the connectivity and centrality parameters of the network, such as "Degree", "Closeness", "Betweenness", and "Eigenvector" centrality (See **Section 2.4**). We have computed the numerical values of these parameters for each node and measured the average values of each of the parameters. After that, we have extracted the significant nodes/proteins, if the protein/node has higher value of the parameter than the corresponding average value of the parameter (**Table 10**).

We have observed from In-Degree calculation that almost all the four homologues of Notch receptors (NOTCH1/2/3/4), the Notch precursors and the Notch Intracellular domains (NICD1/2/3/4) proteins have high In-Degree values as compared to the other proteins in the network (average value of In-Degree of the entire network: 1.97).

Table 10: Significant Proteins of Notch Signaling Pathway Identified through Topological Analyses

Network Parameters	Average Value	Name of molecules
In-Degree	1.97	NOTCH1, NOTCH2, NOTCH3, NOTCH4, GAMMA_SECRETASE, NICD1, NICD2, NICD3, NICD4, NOTCH1_PRE, NOTCH2_PRE, NOTCH3_PRE, NOTCH4_PRE, NUC_NICD1, NUC_NICD2, NUC_NICD3, NUC_NICD4, CSL, COA, SMRT, COR, HDAC, YY1, STAT3
Out-Degree	1.97	JAG1,JAG2,DLL1,DLL4,DLL3,TACE,GAMMA_SECRETASE,DVL, POGLUT1,O_GLUCOSE,POFUT_1,O_FUCOSE,XYLE,XYL,GASE, GALACTOSE,FRINGE,NGA,GSK_3BETA,P53_P,NUC_NICD1,N UC_NICD2,NUC_NICD3,NUC_NICD4,CSL,DTX1,FBW7,SKIP,C DK8,HIF1A,HES1,NRARP
Total-Degree	3.94	NOTCH1, NOTCH2, NOTCH3, NOTCH4, GAMMA_SECRETASE, NICD1, NICD2, NICD3, NICD4, NOTCH1_PRE, NOTCH2_PRE, POGLUT1, O_GLUCOSE, NOTCH3_PRE, NOTCH4_PRE, POFUT_1, O_FUCOSE, XYLE, XYL, GASE, GALACTOSE, FRINGE, NGA, P53_P, NUC_NICD1, NUC_NICD2, NUC_NICD3, NUC_NICD4, CSL, , YY1, HIF1A, NRARP
Eigenvector Centrality	0.20	NICD1,NICD2,NICD3,NICD4,NOTCH1_PRE,NOTCH2_PRE,NO TCH3_PRE,NOTCH4_PRE,NUC_NICD1,COA,HAT,SMRT,COR, HDAC,YY1,HES1,STAT3,HES5,JAK2,HEY1,HEY2,MAG,NRARP, NFKB,MYOD,GATA3,CD44,P21,KLF5,PTCRA,REL_B,C_REL,P50, P65,SOX9,BCL2,IAP,FLIP,CCND1,CCND3,MKP_1,HEYL,HES7
Closeness Centrality	0.002	JAG1,NOTCH1,JAG2,DLL1,DLL4,DLL3,NOTCH2,NOTCH3,NOT CH4,MAGP1,MAGP2,TACE,NOV,CNTN1,PRESENILIN1,GAMM A_SECRETASE,APH1,NICASTRIN,PEN2,NEXT1,NEXT2,NEXT3, NEXT4,NICD1,NICD2,NICD3,NICD4,DVL,WDR12,GSK_3BETA,J IP1,RAS,P53,P53_P,NUC_NICD1,NUC_NICD2,NUC_NICD3,NU C_NICD4,SMAD3,CSL,DTX1,FBW7,EP300,COA,SKIP,HAT,MAM L,SMRT,COR,SAP30,HDAC,CIR,SIN3A,CDK8,STAT3_P,NUC_ST AT3,HIF1A,HES1,STAT3,HES5,JAK2,NRARP
Betweenness Centrality	107.94	NOTCH1,NOTCH2,GAMMA_SECRETASE,NEXT1,NEXT2,NEXT 3,NEXT4,NICD1,NICD2,NICD3,NICD4,NUC_NICD1,NUC_NICD 2,NUC_NICD3,NUC_NICD4,CSL,STAT3_P,NUC_STAT3,HIF1A, HES1,STAT3,HES5,NRARP,PTEN

NOTCH1 has high values compared to the other homologues NOTCH2/3/4.

Similarly, NOTCH1_PRECURSOR and NICD1 have shown higher as compared to the other homologues present in the network. This result clearly signifies the importance of NOTCH1 receptor protein, its precursor, and its truncated portion (NICD1) in compare to all other homologues in the entire network. As expected, in case of Out-Degree parameter calculations data (average Out-Degree of the network: 1.97), we have observed that the nuclear protein CSL has highest number of Out-Degree value in the network. CSL is mostly connected with the target genes of Notch pathway the network, and therefore its importance in terms of Out-Degree calculation is very high as compared to all other nodes. Apart from CSL, the NOTCH ligands, GAMMA_SECRETASE enzyme complex, and the enzymes, such as POGLUT_1, POFUT_1, Galactosyltransferase (GASE), have higher maginitudes of Out-Degree values in the network. On the other hand there are also some inhibitor molecules or complex like Co-repressor complex (COR), HDAC, SMRT and the phosphorylated form of P53 (inhibitors of NUC_NICD1/2/3/4) which also have significant number of Out-Degree values in the network. This result signifies that the stimulation of Notch pathway activity is mostly triggered by these molecules in the network and targeting these proteins during pathway activation will create profound effects in the signal transductions mechanisms (**Table 10**).

The average Eigen vector centrality of the whole network is 0.20 (**Table 10**). We have found that the nuclear transcription factor CSL has the highest value in the network. Interestingly, we have also found that the cross-talk protein STAT3 has significant Eigenvector centrality in the network, although it has lower number of connectivity (Degree centrality) in the network. The reason behind of STAT3 showing higher Eigenvector centrality is that it is connected (in-degree connections) with the Notch target proteins HES1 and HES5. These two proteins are again connected with the hub proteins, such as NUC_NICD1/2/4 and CSL in the network. Due to this reason, STAT3 gets higher Eigenvector centrality as its connections HES1 and HES5 are connected with the most influential nodes in the network. Similarly,

we have observed the higher Eigenvector centrality of the transcription co-repressor proteins HADC and SMRT, which in turn implies the importance of these proteins in the transcriptional mechanisms and the activities of the whole network.

The Closeness centrality (average 0.002) values of all individual molecules in Notch pathway have revealed that CSL is the closely places node in the entire network with lowest number of shortest paths or connections (**Table 10**). This signifies that to influence a node in the network, CSL protein will be the highest priority protein to start with. Simultaneously, NRARP, HIF1A, STAT3 have also shown higher Closeness centrality values in the network. Similar to the Eigenvector centrality, herein we have also identified the feedback effects of other important proteins, such as NICD1/2/3/4 or HES1/5 in the network, which in turn has increased the closeness of HIF1A, STAT3, NRARP proteins in the network.

We have also measured another important centrality parameter called Betweenness centrality (**Table 10**). As expected, we have found that CSL has highest Betweenness centrality value in the network as the synthesis of all the output proteins are mediated by this protein the nucleus. We have also found that NICD1 has higher Betweenness centrality value as compared to its other homologues (i.e., NICD2/3/4). We have checked its connectivity profile and found that unlike the other three homologues of these proteins, NICD1 has extra three upstream regulators proteins: RAS, JIP1 and WDR12 as well as P53 protein in its downstream. It is also connected with its nuclear counterpart NUC_NICD1, which has additional downstream target genes (e.g., BCL2, FLIP, IAP, P21, P65, P50, C_REL, REL_B) for transcription than its counterparts NUC_NICD2/3/4. Hence, more number of shortest paths intersects this protein and thus enhances its Betweenness centrality score.

Although the topological analysis is useful to identify the important proteins from such a large complex network and have helped us to identify the probable drug targets to suppress the activity of maximum Notch target proteins, but using this

technique we are unable to identify the exact proteins that are directly or indirectly influenced by these identified proteins. This is one of the limitations of the graph theoretical analyses. Hence, to overcome this drawback and to test the effect of mutation or deregulation of important proteins in the network under certain circumstances as well as to identify the new biomarkers of Notch pathway, we have performed logical or semi-dynamic analysis of the reconstructed network.

5.4.3 Logical Analyses of Notch Signaling Network

The entire reconstructed map of Notch pathway is modeled using Boolean algebra (See **Section 2.5.3**). We have also extracted the input nodes (source) and the rest intermediate and output nodes (sink) from the logical hyper-graphs. The logical equations (master model) are at first used for constructing the normal Notch pathway scenario (NNS) (**Appendix Table 7**), in which we have considered the normal expression pattern (**Appendix Table 8**) of the input nodes as binary values (Up or Down). We have simulated the model in CellNetAnalyzer to observe the steady-state expression patterns of the intermediates and target proteins (outputs). Followed this, to create the GBM tumor scenario (GBS), we have at first performed the differential expression analysis of primary Grade-IV GBM tumor versus normal cells (GBE) [219] and extracted the differentially expressed transcripts of the input, intermediate, and output proteins of the master model (**Appendix Table 8**). The initial states of the input proteins are considered as "1" or "Up", if the transcript of the protein in the differential expressions analyses has shown significant up regulation and vice-versa (**Appendix Table 8**).

However, in few cases, the transcripts of the input proteins have shown insignificant expression pattern as compared to the normal cells. In that scenario, we have consulted the literature references for the expected expressions or activities (up or down) of those insignificant proteins in the GBM cells and considered their initial expressions accordingly. For example, in the differential expression analysis, we have

observed that the expression of JAK2 protein in GBM cells is not significantly expressed as compared to the normal cells, because JAK2 is a common protein and it can express normally in both GBM and normal cells. However, in another experiments (from literature references), we have found that JAK2 protein can be also over-expressed in the GBM cells [498,499]. Hence, we have considered initial expression of JAK2 protein as 1 or Up in the GBS model (**Appendix Table 8**).

5.4.4 Validations of GBS and NNS Models

We have extracted the logical steady-states of all the nodes in the developed NNS and GBS models using the initial states (See **Section 2.5.3**) (**Figure 31A**). The logical steady-states of the intermediate and the output nodes observed in NNS and GBS models are shown in **Figure 31B.** We have found that in the differential expression analyses of the microarray data of GBM cells versus normal cells (GEO: GSE4290), 35 proteins out of 54 proteins in the input list have shown significant expressions (either Up or Down) in the GBM cells in contrast to the normal glial cells (**Figure 31A**) [219]. Similarly, in the intermediate and output list, 25 out of 62 proteins have shown significant expressions in the GBM cells (**Figure 31**B). Our model simulation of GMB model (GBS) has correctly predicted the expressions of 21 proteins with accuracy of 84% out of the 25 significantly expressed proteins in the output list (**Figure 31**B). Our model simulation (GBS) has predicted the expressions of NOTCH1, NOTCH2, NOTCH3, HES7, and HEYL proteins (up-regulation) exact to the expressions observed in the differential expression analyses of GBM cells (GBE). Also, it has correctly predicted the oncoproteins related to GBM growth, such as HIF1A, MKP-1, FLIP, SOX9, etc. The tumor suppressor protein PTEN, which is down-regulated in the GBM cells, is also found at the down regulated state in our logical steady-state simulation of GBS. Apart from this, our simulation has also predicted expressions of the remaining 38 proteins, which can be validated with other experimental (microarray, RNA-Seq) data. The accuracy level of GBS model confirms the useful of the GBS model for further GSI analyses and new drug targets identification process.

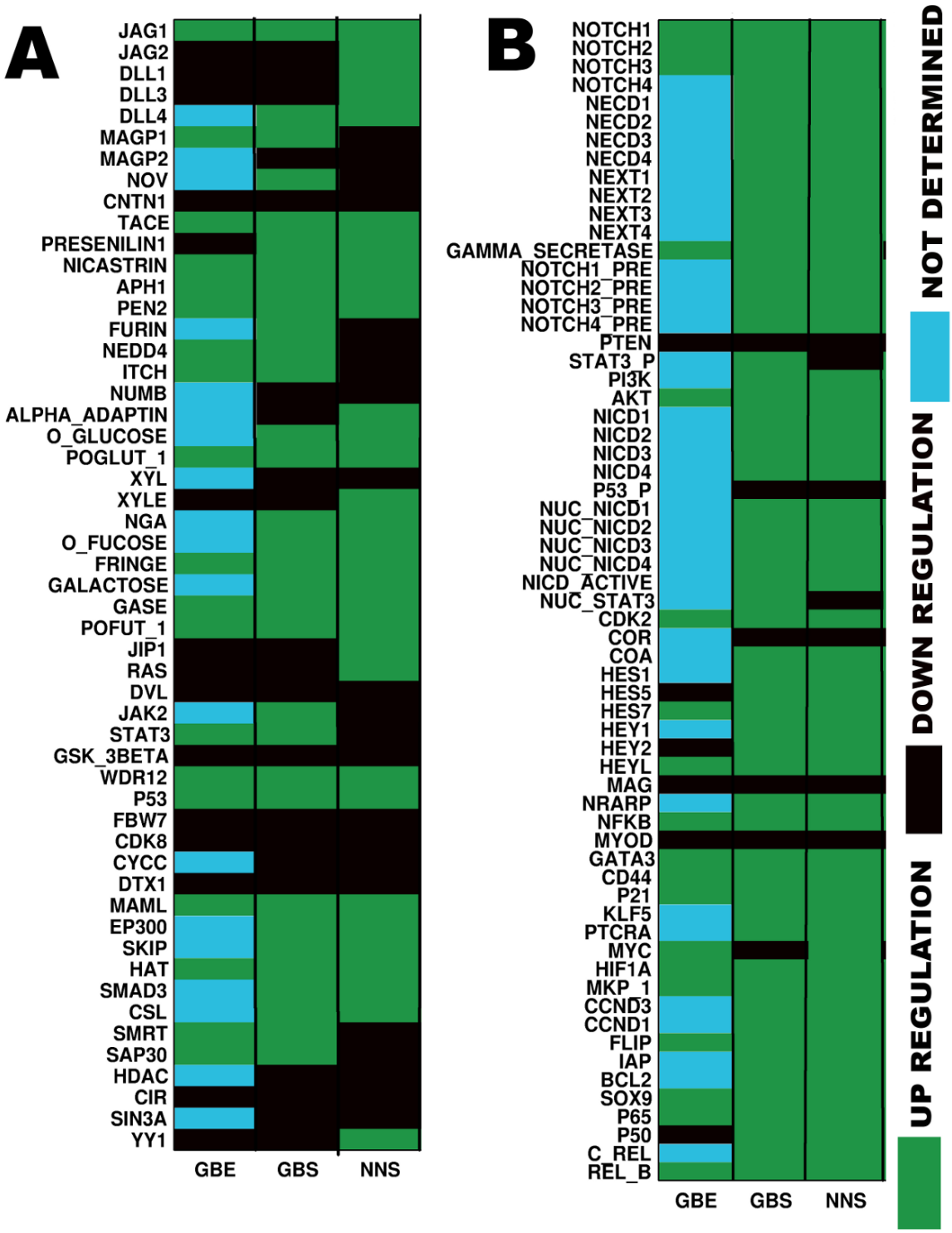

Figure 31: Logical states of the nodes observed in differential expression analyses and simulations of NNS and GBS models.

(A) Represents the logical states of the input nodes observed in experimental data of Glioblastoma cells (GEO: GSE4290) and in simulations of GBS and NNS scenarios. **(B)** Represents the differential expression patterns observed for the output proteins in the same microarray experiment and the logical states under steady-states of NNS and GBS simulations. The transcripts of few proteins from both the lists have shown insignificant expression in the differential expression studies of normal and GBM cells.

5.4.5 Simulation of GSI Non-Responsive Scenario

We have targeted the GAMMA-SECRETASE enzyme complex in our master model to create the non-responsive GSI scenario (See **Section 2.5.3**). We have constitutively kept the expression of this node in the model simulation as '0" or 'OFF' in the GBM model scenario (GBS) and simulated the expressions of the intermediate and output proteins. The expressions of Notch pathway proteins in GSI are collected from the previous performed experiment in GBM cells [500]. The known GAMMA-SECRETASE inhibitors, such as DAPT, BMS-708163, and RO4929097 are used in GBM cells and the expressions of Notch pathway proteins (**Appendix Table 8**) are measured. In the experimental outcomes, it is observed that in the GSI responder (sensitive) cells, the Notch pathway proteins, such as *NOTCH1, NOTCH3, HES1, MAML, DLL3,* and *JAG2,* etc. are up-regulated before treamtent and down-regulated after performing GSI using DAPT, BMS-708163, and RO4929097.

We have found similar results (data not shown) in our GBS scenario, if we consider only the canonical Notch pathway activity in GBM cells. However, if we consider the non-canonical activation and cross-talk effects of Notch signaling in GBM cells, we have found that GSI is not effective for the treatment of GBM cells. Almost, every Notch target genes (except MYC, MYOD, and MAG) are still active in after suppressing GAMMA-SECRETASE enzyme complex in GBS (**Figure 32B**). Thus, we have created the GSI non-responsive (resistive) GBM scenario, in which GSI treatment will not be effective.

5.4.6 Variations in the number of activator and inhibitor nodes in the NNS, GBS, and GSI scenarios

We have also computed the total number of upstream activator and inhibitor, and the downstream activated and inhibited nodes for each of the node in the NNS, GBS, and GSI scenarios. We have observed that mostly all the Notch target proteins (including the oncogenic proteins, such as BCL2, FLIP, SOX9, etc.) of GBS has higher number of upstream activator nodes in GBS as compared to the NNS (**Figure 33A**).

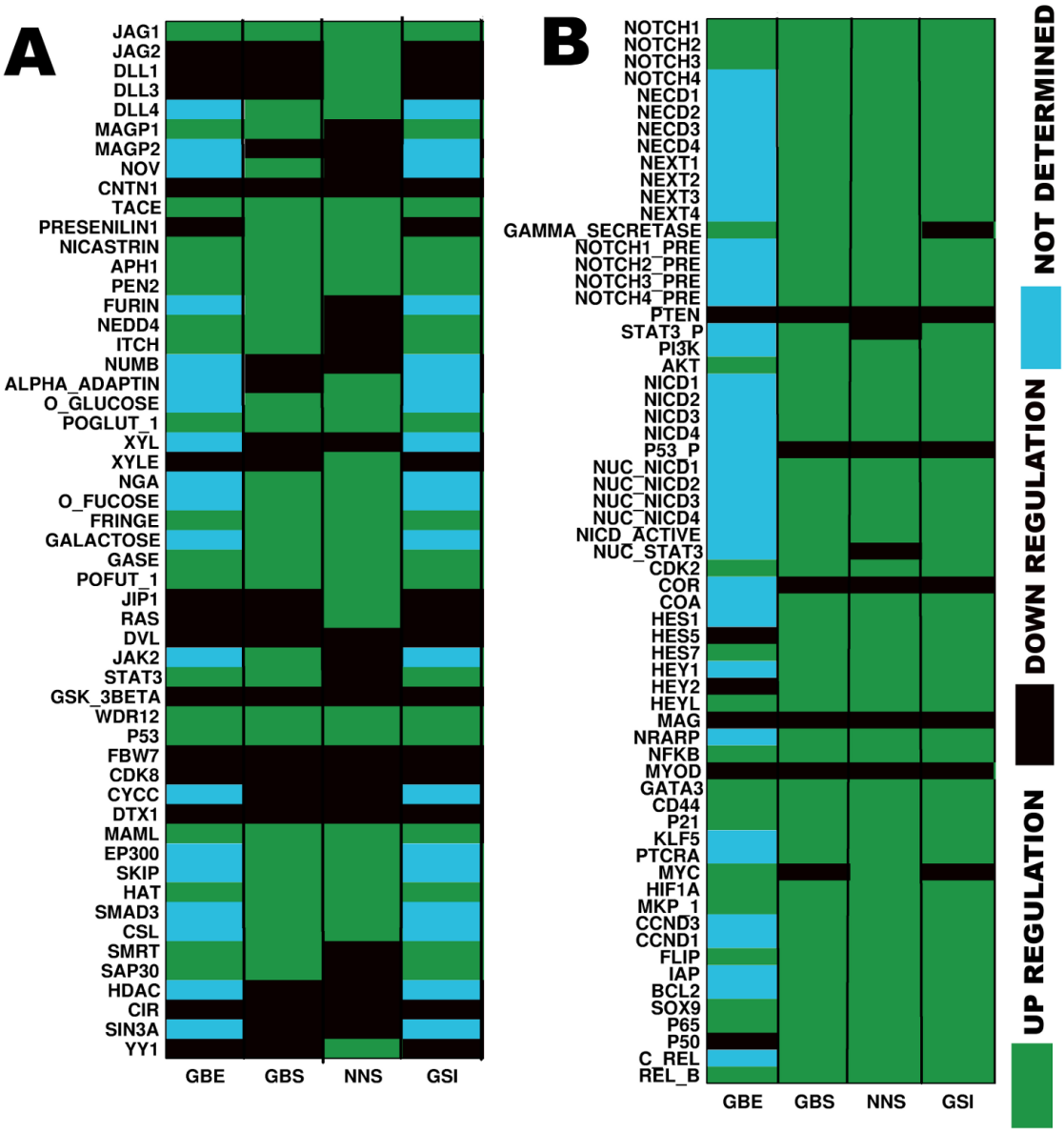

Figure 32: Logical states of the nodes observed in experimental analyses and simulations of NNS, GBS, and GSI models.
(A) Represents the logical states of the input nodes observed in experimental data of GBS, GSI, and the logical states considered in GBS, NNS, and GSI non-responder (resistant) models. **(B)** Represents the differential expression patterns observed for the output proteins in the experimental conditions and the logical steady-states observed in NNS, GBS, and GSI scenarios.

On the other hand the total number of inhibitor molecules (Upstream inhibitor) acting on the Notch target proteins of NNS scenario are comparably higher than the GBS scenario. This result clearly indicates that the in the normal scenario (or cells), the Notch pathway is highly controlled by other inhibitor proteins or molecules in the reaction cascade.

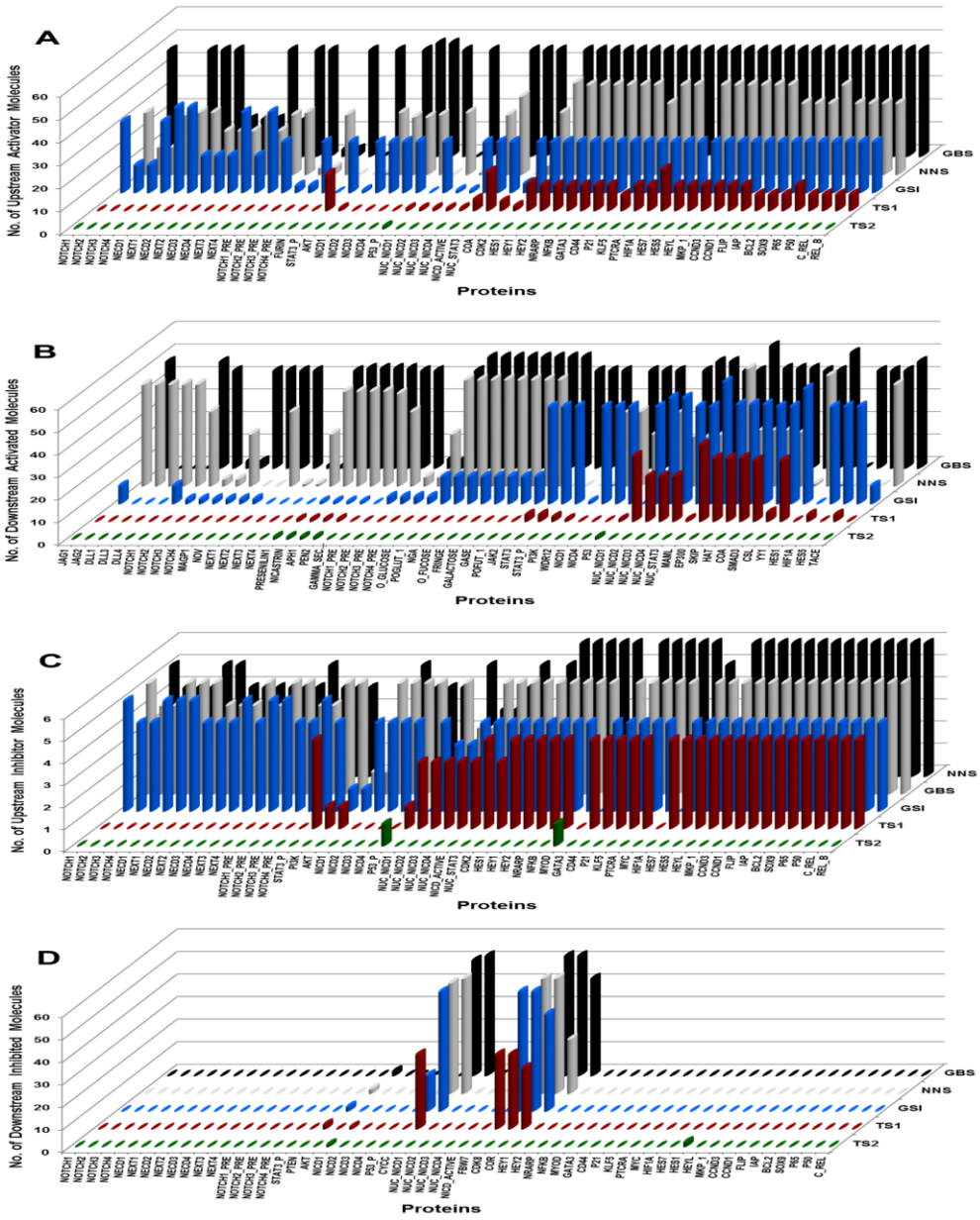

Figure 33: Comparison between Normal, Glioblastoma, Gamma Secretase inhibition and the two proposed drug target scenarios.
NNS: Normal Notch Pathway; GBS: Glioblastoma Scenario; GSI: Gamma Secretase inhibition; TS1: NICD1 and HIF1A combinatorial inhibition; TS2: NICD1 and MAML combinatorial inhibition. **(A)** Represents number of Upstream activator molecules (Y-axis) activating the molecules (X-axis) representing significant variations. **(B)** Represents number of downstream activated molecules (Y-axis) activated by the molecules (X-axis) representing significant variations. **(C)** Represents number of upstream inhibitor molecules (Y-axis) inhibiting the molecules (X-axis) representing significant variations. **(D)** Represents number of downstream inhibited molecules (Y-axis) inhibited by the molecules (X-axis) representing significant variations.

Thus, the pathway is able to regulate the phenotypic expressions of cells, such as cell proliferation, anti-apoptosis, hypoxia, etc. in a homeostatic state. Most often, this

homeostatic regulation of the Notch pathway gets perturbed by over-activations of several oncogenic factors and down-regulations of tumor-suppressor proteins (e.g., P53, PTEN) in Glioblastoma cells and thus the controlled regulations of the output proteins are perturbed.

On the other hand, the total numbers of downstream activated proteins of GAMMA_SECRETASE, WDR12, NICD1, NICD4, EP300, MAML, SKIP, HAT, etc. have also shown significant variations in GBS as compared to the NNS (**Figure 33B**). We have observed that downstream activated proteins of DLL1, DLL3, DLL4, HES1, MAML, HAT, and HIF1A, etc. are increased around 2 folds to 50 folds in GBS as compared to NNS. The increase in total number of downstream activated proteins by these proteins indicates that these proteins play significant roles in the up-regulations of other target oncoproteins and down-regulations of tumor-suppressor proteins in Notch pathway under GBS model simulation. Over-activations of these proteins and their downstream effects in the acceleration of GBM cell growth are also experimentally verified in previous experiments [501,502]. Hence, we are now able to distinguish the proteins which are highly influenced by other upstream proteins and also the proteins, which influence higher numbers of other proteins in the downstream of Notch pathway during GBM cell growth. However, the total numbers of downstream inhibited proteins by most of the proteins in this pathway, (except for NUC_NCD1, FBW7, CDK8, COR and HEY1), have not shown any significant variations in both GBS and NNS scenarios.

On the other hand, in the simulation of GSI non-responder GBM cells, we have found that the total numbers of upstream activator nodes of NOTCH1 receptor protein is significantly reduced as compared to GBS scenario, but not with respect to NNS (**Figure 33A**). This result proves that although the higher rate of activation of NOTCH protein is suppressed by GSI in GBM cells, but the normal rate of activation of this protein still persists. As a result, the non-responder of GSI tumor cells may still have active Notch pathway, which in turn provides the tumor cells a scope to

escape the GSI induced break on active cell cycle. Apart from this protein, the other proteins in this pathway, such as NICD1, STAT3 (in Nucleus), HIF1A, and the Notch target genes/proteins (e.g., MYC, SOX9, BCL2, HES1, GATA3, etc.) also have reduced numbers of upstream activators in non-responder GSI scenario as compared to GBS, but not significantly lesser than NNS simulation. This result also proves the existence of active Notch pathway in the non-responder GSI tumor cells. Simultaneously, we have observed that the downstream activated proteins by the transcription factors, such as NICD1/2/3/4 (in Nucleus), CSL, YY1, and including the co-activators, such as MAML, EP300, HAT are not significantly reduced in non-responder GSI tumor cells as compared to GBS (**Figure 33B**). The activities of these transcription factors and co-activator proteins indicate that GSI resistant GBM cells are still able to transcribe Notch targets oncogenes and thus activate the cell cycle progressions.

We have not observed any significant variations in the total numbers of upstream inhibitor molecules in GBS, NNS, and GSI scenarios (**Figure 33C**). This signifies that in all the three cases, *viz.* NNS, GBS, and GSI, the inhibitors of Notch pathway are mostly unchanged and thus the activities of this pathway do not get impede. On the other hand, we have observed that in GSI scenario, the total number of downstream inhibited proteins by P53 protein is increases as compared to the NNS and GBS scenarios (**Figure 33D**). This result proves the synergistic effects of P53 and GAMMA-SECRETASE inhibition in canonical Notch pathway to suppress cell cycle progressions.

5.4.7 Identification of Alternative Drug Targets

In order to find out the novel drug targets as the alternative of GAMMA-SECRETASE enzyme to impede the growth of GBM tumor cells, at first we have analyzed the topological parameters (i.e., connectivity and centrality) and network structure (e.g., feedbacks) of the entire Notch signaling network. We have seen that the proteins, such as ADAM/TACE, CSL, NICD1, MAML, HI1A, NRARP, HES1,

HES5, etc. have high centrality values (i.e., Degree, Eigenvector, Closeness, etc.) in the network (**Table 10**). Out of these proteins, we have filtered out the proteins, which are biologically feasible to target, such as ADAM/TACE, NICD1, MAML, HIF1A, and DLL4. We have selected these proteins as the probable drug targets for our next *in-silico* perturbation analyses on GSI resistant GBS model.

However, after performing the single knock-out of the selected target proteins, we have not found any potential protein as the alternative of GSI, which can suppress the expressions of target oncoproteins of Notch pathway in the GSI non-responsive Grade-IV GBM cells (results not shown). Next, we have performed the *in-silico* dual knock-out of the selected proteins in the GSI non-responsive GBM scenario. After performing several trials, we have found that targeting NICD1 with either HIF1A or MAML give optimal results of suppressing target oncoproteins. Hence, we have selected two target combinations, *viz.* NICD1 and HIF1A as TS1, and NICD1 and MAML as TS2 as the alternative of drug targets.

We have observed that, in case of dual knock-outs of NICD1 and HIF1 (TS1) in GSI resistant GBM model cells, the total numbers of upstream activator proteins and the total numbers of downstream activated proteins are significantly reduced in TS1 simulation as compared to GBM and GSI scenarios (**Figure 33A-B**). However, the steady-states of the target oncoproteins in TS1 scenario are not down-regulated as compared to GBS and GSI scenario, which makes our hypothesis that this combination (NICD1 and HIF1A) can cause partial suppression of the growth of GBM cells (**Figure 34B**).

On the other hand, in case of dual knock-outs of NICD1 and MAML proteins (TS2), we have observed that the expressions of all target onco-proteins are completely suppressed (**Figure 34B**). Moreover, the decrease in fold changes of total numbers of upstream activator proteins and the total numbers of downstream activated proteins in TS2 are very much lower as compared to GBM and GSI

scenarios (**Figure 33A - B**). Thus, the complete inhibition of the growth of GBM cells and the GSI resistive cells is possible by using this target combination.

5.5 DISCUSSION

We analyzed the reconstructed Notch signaling network using graph theoretical techniques and identified the centrally located proteins or "Hubs" (e.g., CSL, NICD1, NICD2, MAML, HIF1A, HES1, HES5, NRARP) in the network.

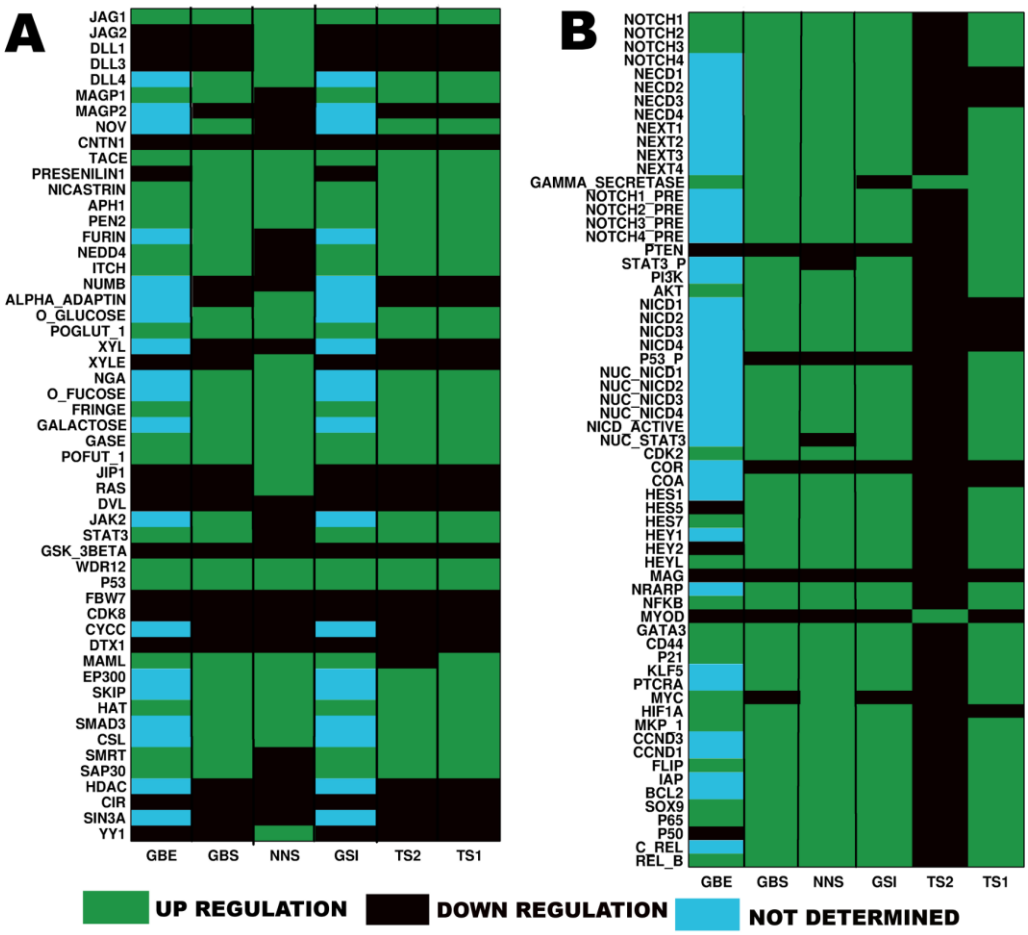

Figure 34: Logical states of the nodes observed in experimental analyses and simulations of NNS, GBS, GSI, TS1, and TS2 models.
(A) Represents the logical states of the input nodes observed in experimental data of GBS, GSI, and the logical states considered in GBS, NNS, GSI non-responder (resistant), TS1 and TS2 models. **(B)** Represents the differential expression patterns observed for the output proteins in the experimental conditions and the logical steady-states observed in GBS, NNS, GSI non-responder (resistant), TS1 and TS2 models.

We found that the nuclear transcription factor CSL, STAT3 had the higher

eigenvector centrality value in the network along with HES1 and HES5. Similarly the Eigenvector centrality of NICD1/2/3/4 was also increased due to its connections with the output proteins NRARP, implying that these proteins not only have high number of connections but also are connected with other highly prestigious nodes that possess higher number of connections in the network. The molecules connected in feedback regulations with the output proteins had higher Eigenvector centrality and showed strong influence in the network, even though they had lower number of connections in the network. This finding helped us to identify the unknown feedback interactions of a particular protein in the network. We were also able to identify the connection of several cross-talk molecules, such as, HIF1A from Hypoxia, PTEN from PI3K/AKT pathway, P53, RAS etc., with the core molecules of Notch pathway, which create either positive or negative feedback loops in the network. Moreover, CSL, NRARP, HIF1A, STAT3 were also showing high Closeness centrality score, where the feedback connections of these proteins with the other important proteins, such as NICD1/2/3/4 or HES1/5 in the network gave the access of these proteins to regulate more number of other proteins in the network. As a result the closeness centrality values of these proteins were also increased. This result also signifies that certain perturbations or mutations of these proteins will cause worst effect than the other proteins having lower closeness centrality value. The higher Betweenness centrality value found for NICD1 compare to its other homologues (i.e. NICD2/3/4) shows that unlike the other three homologues of these proteins, NICD1 had extra three upstream regulators proteins: RAS, JIP1, and WDR12 as well as P53 protein in downstream. We found that this protein was connected with its nuclear counterpart NUC_NICD1, which had additional downstream target genes (e.g., BCL2, FLIP, IAP, P21, P65, P50, C_REL, REL_B) for transcription than its counterparts NUC_NICD2/3/4, implying enhancement of Betweenness centrality value as more numbers of shortest paths intersect this protein. Therefore, this structural and topological analysis helped us to identify the important proteins from a large complex network and indicated the probable drug targets to suppress the activity of

maximum Notch target proteins.

We also created the logical models for normal (NNS) and GBM cancer (GBS) cells to test the effect of mutation or deregulation of important proteins in the network under certain circumstances as well as to identify the new biomarkers of Notch pathway. In the logical analysis, we modeled the entire pathway reactions by logical equations and created five scenarios: Normal (NNS), Glioblastoma (GBS), GAMMA SECRETASE inhibition (GSI), Treatment scenarios by inhibiting NICD1 and HIF1A (TS1) and by inhibiting NICD1 and MAML (TS2). We also validated our model with the mRNA expression profiles of human Glioblastoma cells [219]. Further we were also able to verify the expression or activation of proteins in the non-responsive GAMMA-SECRETASE inhibited GBM cells (GSI) scenario. From our *in-silico* simulation of GAMMA-SECRETASE inhibition scenario (GSI) by comparing the number of upstream activators genes/proteins in GBS and GSI scenarios, we observed that the downstream activated proteins of several Notch pathway activator proteins (e.g., JAG1/2, DLL1/3/4, MAGP1, NICD1, etc.) were getting reduced by administering the GAMMA-SECRETASE inhibition in GSI resistive GBS cells.

Moreover, by comparing the Normal Notch and Glioblastoma scenarios, we were also able to identify the proteins which were abnormally getting activated or inhibited in GBM cells as compared to the normal scenario. This findings and the results obtained in our network analysis, gave us the opportunity to extract the potential drug targets from a large numbers of pathway molecules (total 115 species) of the reconstructed Notch pathway. We identified several probable targets through single and double knock-outs of few selected proteins by perturbing the logical states in the GBS model. We found that as compared to single knock-out, the double or multiple knock-outs show promising results. Among these possible double knock-out combinations, we observed that the combination of NICD1 and HIF1A (TS1) is suitable for the partial suppression of Notch pathway activity in the GSI resistant GBM cells. On the other hand, the inhibition of the proteins NICD1 and MAML (TS2)

as a combination is useful to completely suppress the Notch pathway activity, which targets most of the oncoproteins in the GBM cells. Therefore, it can be concluded that depending on the choice of suppressions of Glioblastoma cells, one can use any of the above drug target combinations (TS1 or TS2) to suppress the growth of Glioblastoma cells.

5.6 CONCLUSION

Our Notch pathway reconstruction and computational analyses is able to provide the insights of the mechanistic view of the signal transduction processes in primary GBM and GSI resistant GBM cells. The entire analysis is helpful to identify the alternative drug targets for Glioblastoma tumor cells. The computational strategy and alternative drug targets predicted from our analysis may help to achieve more accurate therapeutic strategy, not only for Glioblastoma cells, but also for other diseased conditions in which Notch pathway is found to be aberrantly activated. The strategy used for identifying the minimal combination of drug targets for GBM cells is easy to implement. Our findings will be useful for the experimental and clinical biologists to select the biomarkers for cancer diagnosis, treatment response and will show new direction in human cell specific signaling pathway study.

Chapter 6

---

6 IMPLICATION OF NOTCH PATHWAY ANALYSES IN THE EVOLUTIONS OF INTRA-TUMOR HETEROGENEITY AND RISK PREDICTION OF GLIOBLASTOMA

6.1 INTRODUCTION[5]

In the previous chapter, we have shown that how the developmental cell signaling pathway mediated by Notch receptors regulate the growth of proliferating GBM cells in the advanced stage (i.e., Grade-IV) [185]. We have also shown the methodologies using state-of-the-art graph or network theory and Boolean logic to dissect the network properties of the molecular reactions behind the growth of GBM and how one can use the network topology and its logical inter-connectivity to identify novel drug targets [503]. However there are certain aspects of GBM tumor growth, which we did not cover there. For example, the activity of Notch pathway in the maintenance of GBM stem cells was not highlighted there [141,504]. We did not observe the somatic mutations in the Notch pathway, which trigger the normal neural stem cells (NSCs) to transform into Glioma initiating stem cells (GSCs) [141]. Moreover, our approach had the limitations to simulate the occurrence of intra-tumor heterogeneity within the GBM tumor cells, [505] and most importantly it was unable to classify and predict the risk of GBM tumor in the early stage by analyzing the omics data. We have understood that to answer these unsolved questions and decipher the complex pattern of the evolutionary mechanisms, a predictive model with the aim of performing mechanistic simulation of the emergence of genetically different sub-clones of tumor cells with the normal neuronal and astroglial cells in

[5] The materials of this chapter have been taken verbatim from our communicated article Chowdhury & Sarkar, 2018 (manuscript under review).

the tumor microenvironment is very much important.

However, the tumor microenvironment created by GBM is highly heterogeneous and complex in nature, which also accompanies with multiple non-tumorigenic cells, e.g., quiescence, apoptotic (or necrotic), normal or healthy neurons, astrocytes, oligodendrocytes [139]. The sub-ventricular zone (SVZ) of the human brain, recognized as the primary location for nurturing neural stem cells (NSCs), also provides a sustainable ground for the maintenance of GSCs and the development of matured Glioblastoma (GBM) tumor cells [141]. Stemness properties (tumorigenic/non-tumorigenic) of few cells within the tumor ecosystems are also observed and are proven to be the underlying causes of the emergence of low and high-grade astrocytomas during tumor development [506]. The single-cell RNA-Seq analyses performed by Patel et al. on single 420 primary GBM tumor cells collected from 5 individual patients has depicted the cell-to-cell variability of the expressions of surface receptor genes (e.g., EGFR, PDGFRA, FGFR1, ERBB2, ERBB3, KIT, and NOTCH2, etc.) [139]. This observation has demonstrated the role of most of the RTKs and its downstream signaling cascades in the emergence of intra-tumor heterogeneities within a tumor. This study has also confirmed that these surface receptor molecules, especially the RTKs are mutually expressed in different tumor cells and trigger uncontrolled cell divisions. The mosaic expression pattern of NOTCH2 receptor gene (a non-RTK surface gene/protein) observed across the tumor cells is also interesting to realize not only the development of genetic heterogeneities between the tumor cells but also the evolutions of various cell-lineages co-existing within the GBM ecosystems. Hence, this work is aimed to explore the ability of Notch pathway in the regulation of intra-tumor heterogeneity as well as the developmental dynamics of the evolutions of variable cell lineages in the GBM ecosystems.

Notch signaling pathway, which is commonly known for the maintenance and proliferation of the adult neural stem cells (aNSCs) in the neurogenic niche of SVZ, is

also implicated in the proliferation of GSCs[141]. It suppresses the neurogenesis and gliogenesis by targeting the expression of bHLH transcription repressor proteins HES1-7 and HEY1, 2, L, which help to maintain the self-renewal of aNSCs/GSCs by suppressing the transcriptions of the differentiation genes [507]. On the contrary, the other component proteins of Notch pathway (e.g., NOTCH1, RBPJ/CSL, NICD1, γ-Secretase complex, and so forth) are observed to be over-expressed in the differentiated astrocytes and GBM tumor cells [508,509]. However, how the Notch pathway reprograms its functional mechanisms to trigger the differentiation processes in aNSCs/GSCs, despite being the gate-keeper of the maintenance of stem-like cells, is unknown hitherto. Previous reports suggest that several cross-talk pathways of Notch signaling can manipulate the self-renewing processes of aNSCs/GSCs and trigger the differentiation of astrocytomas [510,511]. However, the mechanistic regulations through which the cross-talk molecules influence the dynamics of Notch pathway and help to sustain its dichotomous nature between stem cells renewal and differentiation (i.e., neurogenesis and gliogenesis) processes are also not clearly understood yet. Hence, identification of the molecular switches and reaction motifs in Notch signaling balancing the stemness and differentiation properties of aNSCs/GSCs will be helpful to decipher the complexities of cellular dynamics involved in GBM tumorigenesis.

6.2 CURRENT CHALLENGES AND OBJECTIVES UNDERTAKEN

Various experimental analyses have corroborated the existence of the driver genes/proteins (e.g., P53) causing tumorigenesis, but a follow-up study is essential for studying their functional regulations with Notch signaling which help to regulate the self-renewal or differentiation of aNSCs to GSCs and drive the self-renewing GSCs or Astrocytes towards GBM [512]. Simultaneously, the molecular processes triggering the formation of neurosphere from aNSCs or the isolated GSCs collected from GBM tumor are also required to be studied [507]. Moreover, the underlying mechanisms of the evolutions of different grades (Grade I-IV) of GBM tumor cells

with different clinical responses (e.g., drug sensitivity and resistivity) and its relationship with Notch signal transductions are also not explored yet. Previous reports suggest that evolution of such phenotypically distinct sub-types of GBM cells originated from GSCs or low-grade astrocytomas along with other cell lineages (e.g., neurons, astrocytes, and quiescence cells) in the GBM tumor cells niche, impose significant hurdles in GBM therapeutics[234,437,513,514]. Moreover, the appearances of genetically distinct, sub-clones of tumor cells in the molecular level within the same grade of GBM tumor create another layer of challenges as these sub-clones most often evolve against the selection pressure created by the chemotherapeutic drugs during therapy [234,513]. Hence, understanding the mechanisms of Notch signaling cascades and its implications in the gene regulatory network on the developmental pathway of the evolutions of different sub-types and genotypically distinct sub-clones of GBM tumor cells is the main objective of this work [515].

6.3 FORMULATIONS OF THE PROBLEM AND WORK FLOW

Predictive mechanistic models, considering the dynamic activities of the reconstructed Notch and its cross-talks pathways (see **Section 2.5.4**), are developed here for analyzing the underlying mechanisms of neurogenesis and GBM tumorigenesis. The underlying molecular processes of Notch and its cross-talks pathways are explored by using this computational approach, which in turn helps to assess the governing principles working behind the self-renewal, differentiation, apoptosis, and cell growth arrest (i.e., quiescent state) of aNSCs/GSCs or GBM tumor cells. Also, by assessing the variances in the genetic makeup of the individual tumor cells, this work is competent to dissect the underlying mechanisms working on the emergence of different sub-types and sub-clonal heterogeneities of GBM tumor. Based on these basic understandings, these models are further implicated to perform futuristic predictions of the transformation of low-grade astrocytoma to high-grade GBM, the chances of occurrences of heterogeneous populations of GSCs, and the developmental tracks of the evolutions of intra-tumor molecular sub-clones. Apart

from these, these models are used to explore other fundamental aspects of neurogenesis and tumorigenesis, for example, what are the molecular mechanisms driving the excess rate of apoptosis of aNSCs observed during neurogenesis? [516]. Also, is there any role played by the Notch pathway during apoptosis and cell growth arrest (or quiescent state) of aNSCs/GSCs? Does P53 mutation hinder the neurogenesis or other cellular differentiation of GSCs? What are the main component proteins of the Notch pathway, which form the regulatory switch to regulate the differentiation of different cell lineages from aNSCs/GSCs?

The models are developed to predict the probability of the emergence of GSCs within the normal neurogenic niche and identify the driver proteins responsible for the development of different tumor sub-types. The chances of developing different sub-types of GBM tumor cells from GSCs or astrocytes cells are also quantified and further applied as a personalized approach to detect the risk of early onsets of GBM tumor with appropriate grades (or sub-types). A novel scoring technique is introduced here for quantifying the risk of the tumor development by considering the genetic variances of the individual patients as inputs. Moreover, a novel approach is developed for *in-silico* perturbations analyses to perform the assessments of targeting multiple combinations of proteins on the developed models. The implication of this approach would help to screen and identify potential molecules for target-based GBM therapeutics.

6.4 RESULTS

6.4.1 Mechanistic Models of aNSC, GSC, and GBM Tumor Development Illustrate Cell Proliferation and Differentiation Dynamics

6.4.1.1 Simulation of aNSC Model

Simulation outputs of the developed aNSC model with highest Shannon entropy score (1.456 Shannon, **Figure 35A**) had produced multiple cellular states (i.e., cell lineages, **Figure 35B**) of normal brain cells, e.g., quiescent, neural stem, matured neurons and astrocytes in the steady state or attractor space (see **Section 2.5.4**). It was

also observed that a large number of cellular states were apoptotic in the simulation outcomes. This result justifies the previous experimental outcomes, where it is shown that most of the stem cells undergo apoptosis during the normal developmental stages of adult NSCs [516]. Activity ratio (AR) scores of each input protein of aNSC model showed that by decreasing the activities of the Notch pathway activators and in the presence of wild type active tumor-suppressor P53 (TP53) protein, most of the aNSCs were directed towards the apoptotic state (**Figure 36A**). Hence, from this observation, it was understood that during aNSCs proliferation, a large numbers of newborn stem cells did not get ample stimulations from Notch pathway to proliferate, but due to the increased activity of P53, a significant fraction of stem cells went to apoptotic state. This process is proven useful to maintain sustainable proportions of aNSCs in SVZ, which further divide and differentiate to mature neurons and glial cells [516].

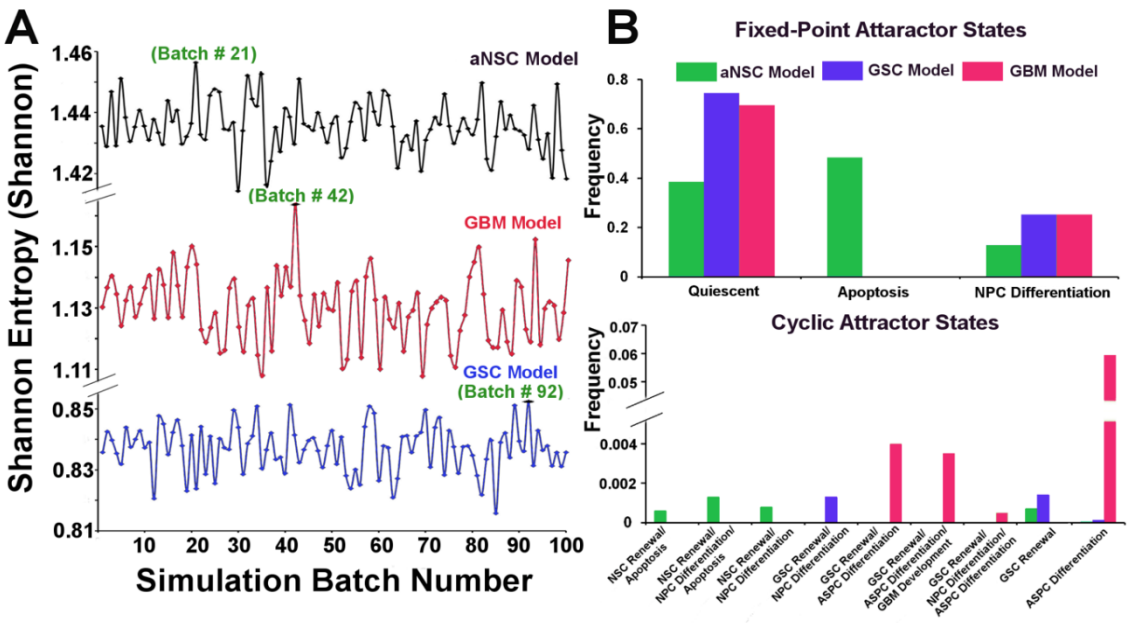

Figure 35: Simulation results of the aNSC, GSC, and GBM models.
(A) Shannon entropy scores observed, and **(B)** Normalized frequencies of cellular states (or cell lineages) observed in aNSC, GSC, and GBM model simulations. Batch numbers 21, 92 and 42 in aNSC, GSC, and GBM models with highest Shannon entropy scores in each model simulations, respectively, are chosen for calculating the normalized frequencies of each attractor states of different models.

The AR-scores (-1.44 to +1.44) of each input protein in the aNSC, GSC, and GBM

models were used to identify the driver proteins responsible for the development of a particular cellular state during the developmental stages of neurogenesis and tumorigenesis (see **Section 2.5.4**). For example, the driver proteins, stimulating the aNSCs towards the development of a few numbers of Glioblastoma stem cells (GSCs), were easily identified by observing the AR-scores of the core component proteins of Notch signaling and the functional state of the P53 protein. It was observed that core Notch pathway proteins (e.g., APH1, RBPJ, MAML1, etc.) were required to be in the active state (AR score = +1.44) and P53 in the inactive (or mutated) state in the self-renewing GSCs (**Figure 36B**). However, the probability of this combination (i.e., protein activities) was infrequent in proliferating aNSCs, and thus the normalized frequency obtained for proliferating GSCs was lowest in the simulation outcomes (**Figure 35B**).

6.4.1.2 Simulation of GSC model

However, it was observed that P53 mutation in GSCs did not influence the developmental dynamics of the stem cells and the cells still retained its self-renewal and differentiation properties like aNSCs. Simulation outputs of GSC model (highest Shannon entropy 0.852 Shannon) were able to exhibit all sub-types of cells observed in the neurogenic niche (e.g., neurons, astrocytes, quiescent cells, etc.), except the apoptotic cells, which was not present due to P53 mutation (**Figure 35B**).

6.4.1.3 Simulation of general GBM model

On the other hand, while simulating the general GBM model (in which P53 is inactive and JAK2, STAT3, and RBPJ are constitutively over-expressed), another three new cyclic attractors (or cellular states viz. i) GSC Renewal/ASPC Differentiation, ii) GSC Renewal/ASPC Differentiation/GBM Development, and iii) GSC Renewal/ASPC Differentiation/NPC Differentiation) were found in the attractor space with varying normalized frequency distributions (**Figure 35**). Fixed-point attractor states *viz.* Quiescent and NPC differentiation and the cyclic attractor states

viz. ASPC differentiation were also observed in this model simulation as observed in aNSC and GSC models. The cellular state ASPC Differentiation (with $P53^{-/-}$, $JAK2^{+/+}$, $STAT3^{+/+}$, $RBPJ^{+/+}$ mutations) observed in this model simulation is an oscillating attractor state with the expression of GFAP protein, uncontrolled proliferation or division of mutated astrocyte cells and no apoptotic break. Hence, the phenotypic characteristic of this cellular state was tumorigenic in nature and thus it was denoted as "Low Grade Glioblastoma (LGG-I)" state (**Figure 37A-B**).

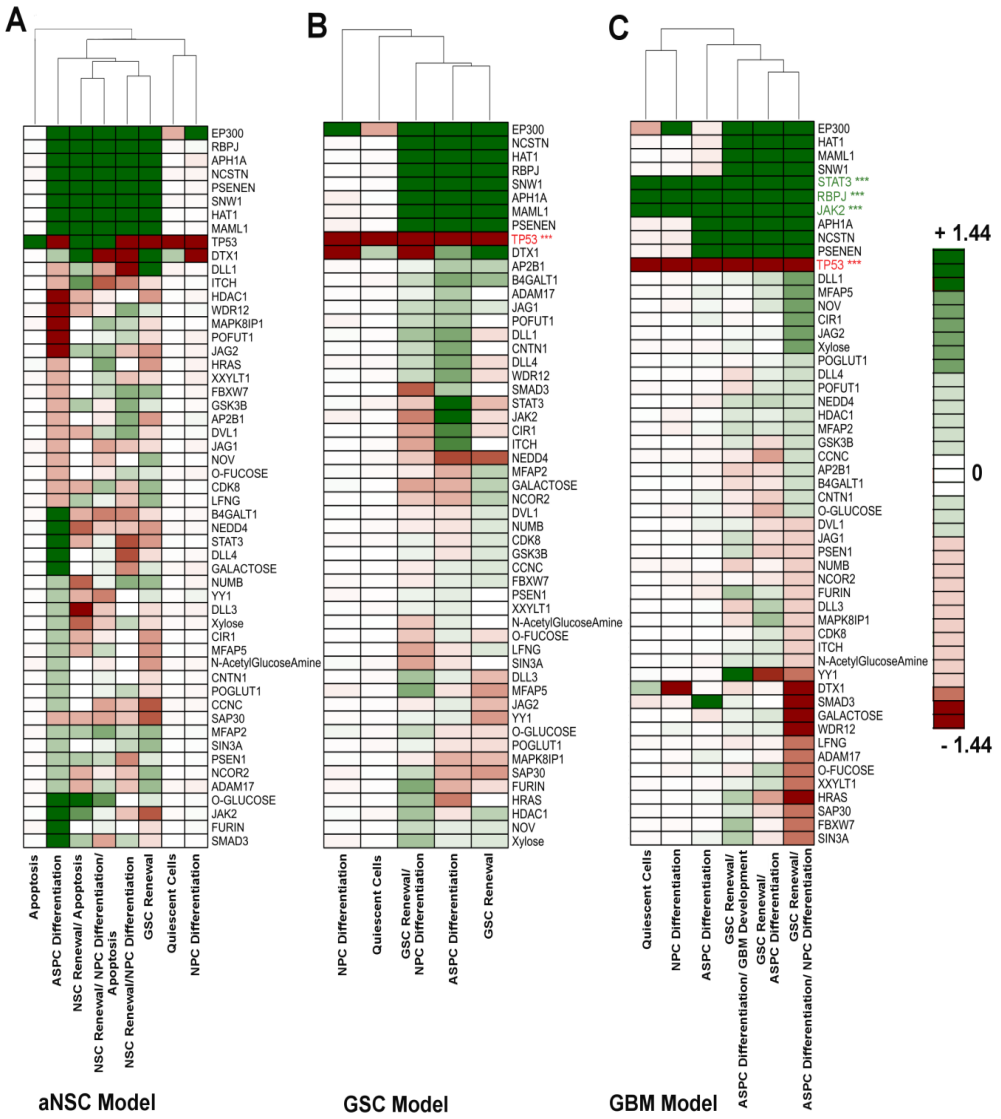

Figure 36: Activity Ratio (AR) scores of the input molecules.
(A) aNSC **(B)** GSC and **(C)** GBM Model simulations**.** The AR-scores of each input molecules were calculated for every cellular state. TP53 protein was found as mutated in the "GSC Model" simulation. On the other hand, while simulating "GBM Model", TP53 protein was found mutated along with increased activities of RBPJ+/+, JAK2+/+ and STAT3+/+ proteins.

Moreover, the protein activity profile observed in another cellular state: "GSC Renewal/ASPC Differentiation" revealed that the expression of the astrocytes differentiation marker GFAP is oscillating along with the stem cell markers (tumor) and the anti-apoptotic markers are constitutively expressed, which also make this cellular state tumorigenic in nature (**Figure 37C**).

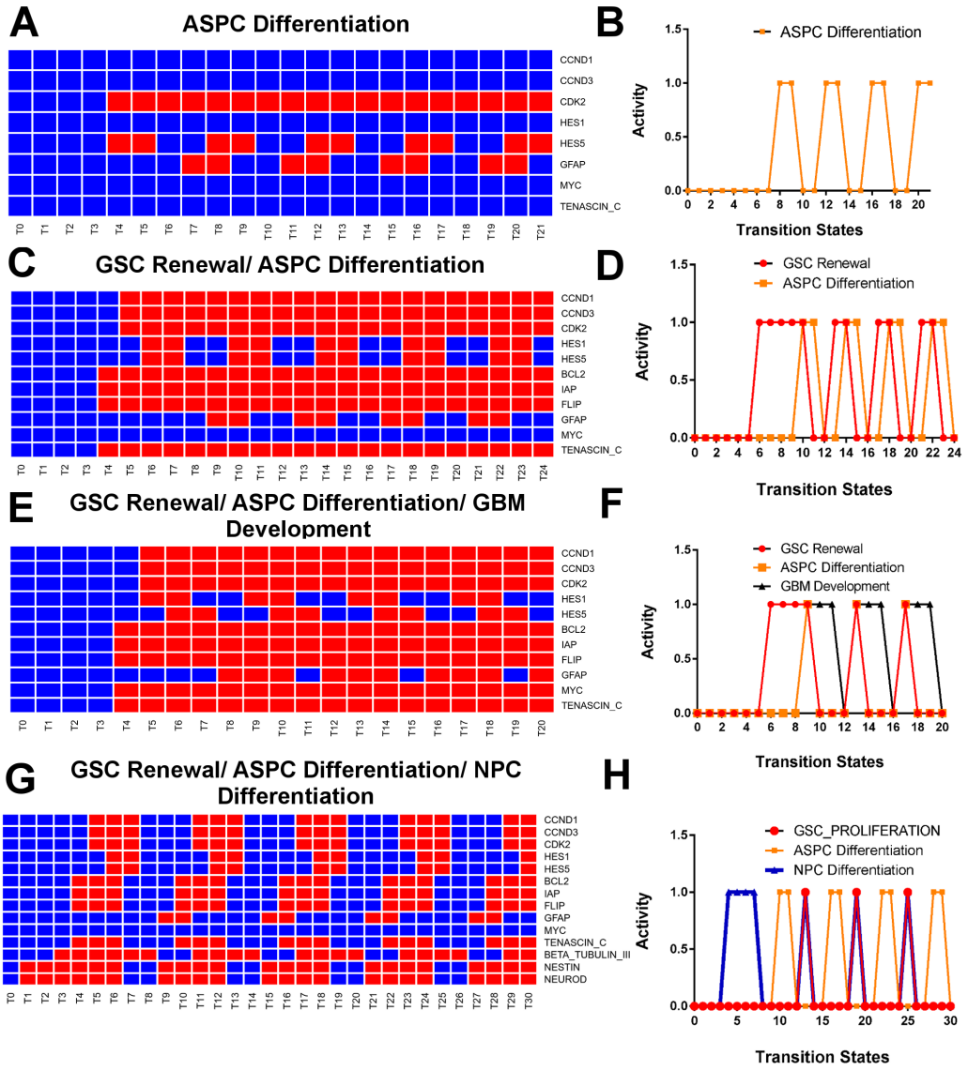

Figure 37: Activity profiles and the temporal dynamics of the marker proteins observed in general GBM model simulation.
(A) & (B) ASPC Differentiation; **(C) & (D)** GSC Renewal/ASPC Differentiation; **(E) & (F)** GSC Renewal/ASPC Differentiation/ GBM Development; **(G) & (H)** GSC Renewal/ASPC Differentiation/NPC Differentiation.

It was observed that both the phenotypes GSC renewal and ASPC differentiation processes are oscillating with fixed time periods throughout the transition time

points around their respective steady state levels (**Figure 37D**). Both the phenotypes had same periodicities (i.e., 4 units of transition states) and out of these two cellular states, GSC renewal process was moving ahead from the ASPC differentiation process. This observation clearly indicates the real biological scenario, in which self-renewal/proliferations of GSCs are initiated in the brain tissue followed by its differentiation and maturation into ASPCs (**Figure 37D**). It also proves the role of Notch pathway to maintain the inertia of GSCs at self-renewal state compare to differentiation and maturation into glial cells and also demonstrates the slow rate of differentiation of mutated and proliferating astrocytes/glial cells from GSCs in the tumor niche. It should be noted that the ASPCs generated through this process were purely tumorigenic in nature with multiple oncogenic mutations and uncontrolled rate of proliferation without apoptosis. However, the genotypic signature of ASPCs found in this cellular state had down-regulated C-MYC (a marker for high-grade GBM), which in turn made this cellular state less tumorigenic (i.e., low grade) as compared to grade-IV glioblastoma tumor cells. As this cellular state contains both tumorigenic stem cells (GSCs) and mutated astrocytes (ASPCs), hence this state is defined here as the "LGG-II" state.

Apart from these two major cellular states, the state transition dynamics of the model also lead towards the two other complex attractor states, which were phenotypically mapped with the "GSC Renewal/ASPC Differentiation/GBM Development" and "GSC Renewal/ASPC Differentiation/NPC Differentiation" cellular states. In the first one, it had shown the cyclic expression patterns of all the GSCs self-proliferation, ASPC maturation or differentiation and simultaneously the constitutive over expressions of important high-grade glioblastoma marker proteins C-MYC and TENASCIN-C (**Figure 37E**). Although both the processes, involved in GSC renewal and ASPC differentiation, were showing oscillatory dynamics around the steady state time points, but the observed periodicities of both the cellular states (i.e. 3 units of transition states) were lesser than the previously mentioned LGG/GSC

cellular state. More importantly, in this cellular state both the phenotypes were oscillating synchronously and dividing rapidly in the tumor niche, which will eventually accelerate the entire population of the mutated astrocytes to grow exponentially without any apoptosis (**Figure 37F**). Interestingly, it was observed that if the malignant GBM markers C-MYC and TENASCIN-C, both were expressed in the rapidly proliferating ASPCs, then the cumulative dynamics of the entire cellular state would transform more aggressively into the self-proliferating stage. Hence, based on these phenotypic and genotypic characteristics, this particular cellular state observed in this simulation was further denoted as "High-Grade GBM" state. The characteristic dynamics of GBM developmental process was also periodic in nature, but the periodicity and the amplitude were comparably higher than the GSC renewal and ASPC differentiation processes.

The Shannon entropy landscape showed that the differentiating GSCs were also directed towards the development of high-grade or "Grade-IV" tumorigenic cellular state in the general GBM model simulation (highest Shannon entropy score, 1.164 Shannon, **Figure 35A**). After analyzing the AR-scores of the input molecules driving towards this cellular state, it was observed that YY1 transcription factor was one of the driver proteins associated with the development of Grade-IV tumorigenic state in the simulation and formed a nexus of the transcriptions regulatory genes associated with tumorigenesis and metastasis (**Figure 35B**).

6.4.2 Network Motifs of Notch Pathway in the Regulation of Cellular States

These overall simulation studies on aNSC, GSC, and GBM models had further revealed the underlying molecular reaction mechanisms (sub-networks) in Notch signaling pathway through which different cellular states occurred during the developmental stages (**Figure 38**).It was observed that, in the normal situation, when the core Notch pathway is active and the P53 protein was not mutated in the aNSCs, the target proteins (*HES1, HES5* etc.) produced at the end of this pathway were

normally expressed and follow cyclic temporal expression pattern (results not shown) [517]. It was experimentally proven and also shown in the identified network motif that the transcription complex (NICD:CSL:MAML) of Notch pathway produces the precursor of NOTCH receptor protein [185], which was then turns to post-processing (most often glycosylated) and transferred to the cell membrane, where it further binds to the Notch pathway activation ligands Delta/Jagged (**Figure 38A**). Thus, this reaction pathway makes a positive feedback loop in the reaction motif and triggers oscillatory dynamics of the Notch target genes *HES1, HES5, CYCLIN-D1, CYCLIN-D3* etc. The entire reaction motif works in such a way that it enforces the transcription regulatory network of notch pathway to produce the cell cycle regulatory proteins CYCLIN-D1, CYCLIN-D3, etc. in periodic pattern. In the previous experiments and these current simulation studies, oscillations in the dynamic expressions of HES1, HES5 genes/proteins were also observed [518].

However, the expressions of HES1/HES5 proteins negatively influence the transcriptions of their own expressions and thus create a negative feedback loop in the motif [519]. The periodic expressions of CYCLINS and CDK proteins, a well known indicator of cell proliferation process, were also observed due to the presence of these feedback loops in the simulations [377]. It was also observed from the simulations that the controlled regulation and fine-tuning of this positive and negative feedback loops were the main regulators behind the maintenance of stem cell renewal and proliferation process. Further, perturbation studies performed by eliminating or targeting any of this feedback loops would shift the cellular dynamics towards different cellular states (neurogenesis or gliogenesis).

In another reaction motif, it was observed that when the expressed HES1/HES5 proteins, which are also known as bHLH transcription repressor protein, interact with the transcription regulatory network of MASH1 gene, then the entire reaction motif was redirected to suppress the expressions of neuronal cell markers e.g., NGN1, β-TUBULIN-III (**Figure 38B**). In this case, HES1/5 proteins act as a

transcription repressor of MASH1 gene, which is the activator of the production of NGN1 and β-TUBULIN-III proteins.

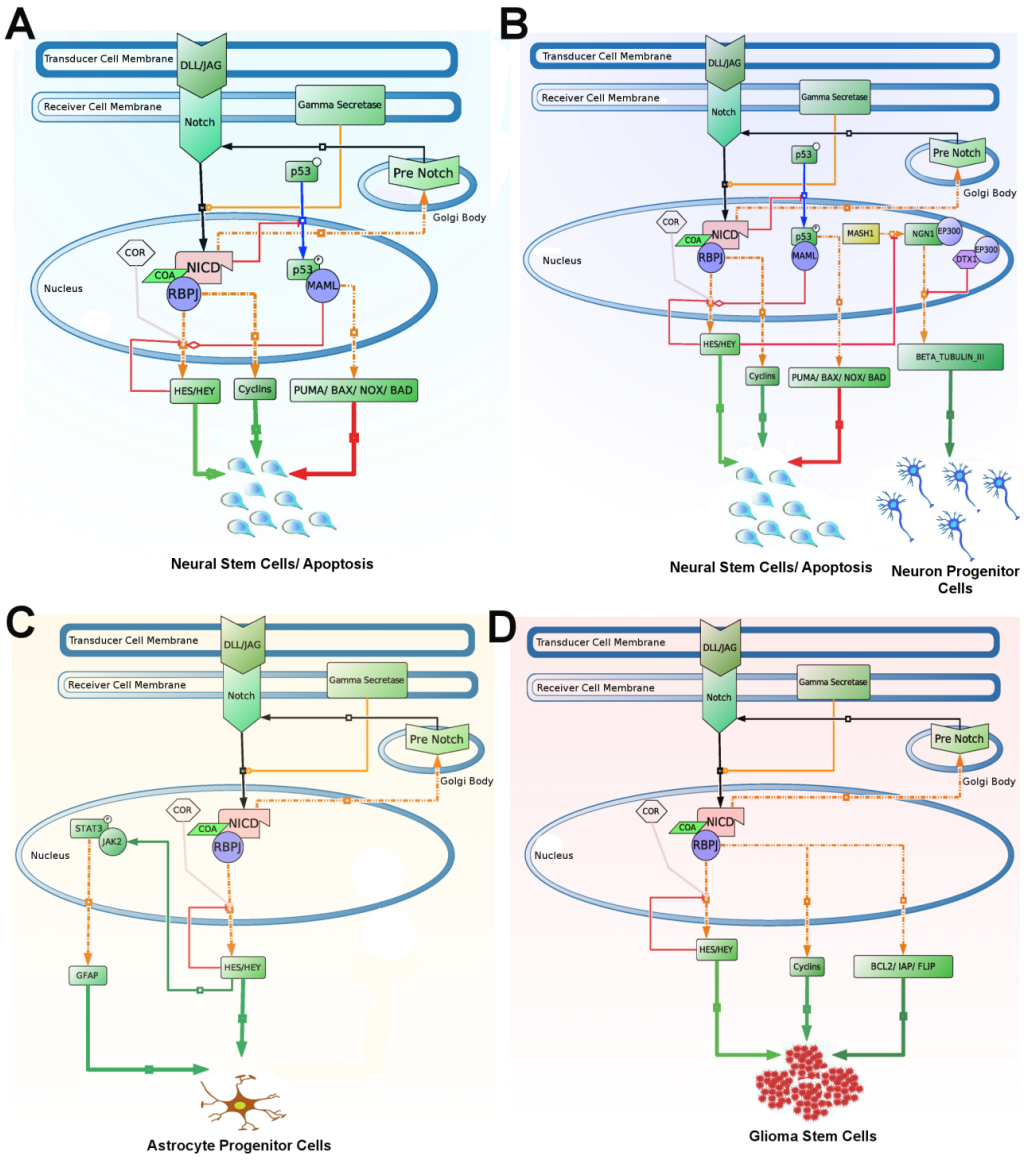

Figure 38: Active network modules in Notch pathway.
(A) NSC Renewal/ Apoptosis; **(B)** NSC Renewal/NPC Differentiation/Apoptosis; **(C)** ASPC Differentiation; and **(D)** GSC Renewal. COA and COR represent co-activator and co-repressor complexes.

It was also observed that during the periodic intervals, at which the HES1/HES5 protein expressions were at the lower activity level, the expressions of neuronal marker proteins were found at the higher level and differentiation of neuron progenitor cells appeared (i.e., neurogenesis). On the other hand, it was observed that when the expression dynamics of HES1/5 protein were at the up-regulated state,

and JAK2/STAT3 proteins were also simultaneously expressed, then the expression of the marker protein GFAP, responsible for astrocytes differentiation (ASPC) was at ON state. In this case, the entire reaction motif of Notch signaling network was redirected towards gliogenesis process (**Figure 38C**).

These entire feedback reaction mechanisms governing neurogenesis and gliogenesis processes can sometimes can go awry, if the tumor suppressor protein P53 is absent in the neural stem cells. In that case, anti-apoptotic proteins such as BCL2, IAP, and FLIP will be expressed at higher quantities and the stem cells become immortal. However, due to the active regulations of both the positive and negative feedback loops, Notch signaling network will produce HES1/5, CYCLINS at regular interval of time (**Figure 38D**). This will eventually lead the rapid proliferations of the stem cells without any apoptosis and the cells will be converted into Glioblastoma stem cells (GSCs).

6.4.3 Increased Activities of JAK2/STAT3, RBPJ, YY1, γ-Secretase Complex and P53 Mutation Promoted Low-Grade GBM to High-Grade

In the high-grade (Grade-IV) GBM tumor states, increased activities of JAK2/STAT3, RBPJ, MAML1, YY1, γ-Secretase complex (APH1, NCSTN, PSENEN, PEN2) and mutation of P53 were observed as a minimal number of deregulations (total mutations, $\mu = 10$) in the Notch pathway derived high-grade GBM cells (**Figure 36A**). Hence, a new model was built by constitutively expressing Notch pathway molecules (RBPJ, MAML1, γ-Secretase complex), JAK2/STAT3 and YY1 proteins and by mutating P53 protein to simulate the evolution of high-grade (Grade-IV) GBM cells. This model was named as high-grade or the Grade-IV GBM model. In this new model, the mean normalized frequency of Grade-IV tumor state was found comparatively high (Mean 0.094 ± 0.003 S.D.) as compared to the normalized frequency (Mean 0.002 ± 0.0005 S.D.) observed in general GBM model and thus this new model was named as high-grade (Grade-IV) GBM model. A comparative study, performed to analyze the changes of the mean normalized frequency differences of

different cellular states observed in general GBM and Grade-IV model, depicted significant changes in all scenarios (**Figure 39**).

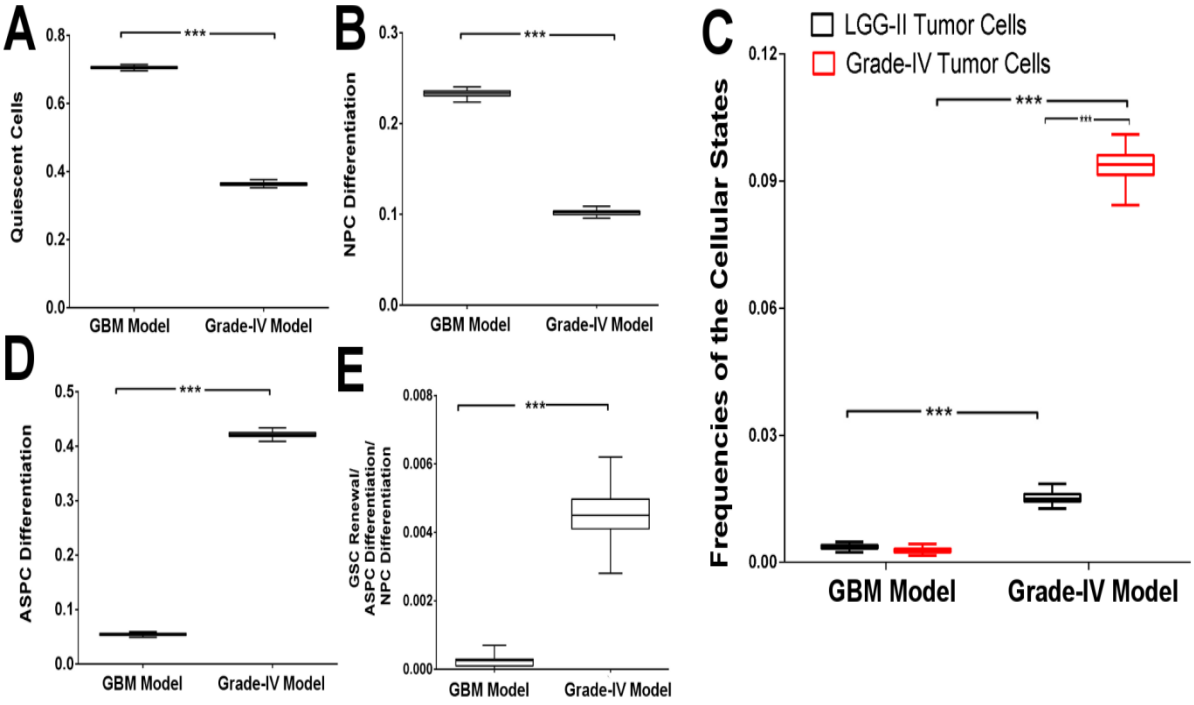

Figure 39: Comparative analyses of the normalized frequencies of cellular states observed in general GBM and Grade-IV models.
Normalized frequencies of the cellular states representing **(A)** Quiescent cells; **(B)** NPC Differentiation; **(C)** LGG-II vs. Grade-IV tumor cells; **(D)** ASPC Differentiation or LGG-I; and **(E)** GSC Renewal/ASPC/NPC Differentiation states observed between general GBM and Grade-IV GBM models are compared here. The simulations of all the models were replicated 100 times (N = 100). Data represents the means ± S.D.; *** P-value < 0.001.

In the Grade-IV GBM model, the normalized frequencies of quiescent cells and the differentiating NPCs were found to be significantly reduced in numbers (**Figure 39A-B**). In contrast, the normalized frequencies of differentiated ASPCs (LGG-I state) and the cellular state "GSC Renewal/ASPC/NPC Differentiation" were significantly increased in the Grade-IV GBM model as compared to the general GBM model (**Figure 39D-E**). A significant increase of LGG-I cells depicted the possibilities of the presence of spatiotemporally distributed, genetically distinct, heterogeneous populations of differentiated astrocytes cells within the high-grade (Grade-IV) GBM tumor niche. Most often these cells (LGG-I) progress to form high-grade (Grade-IV) glioblastoma tumor cells and cause poor prognosis of the patients [520]. It was

observed that in the general GBM model, although a significant difference between LGG-I and Grade-IV cells (**Figure 35B**) exists, no difference was found between LGG-II and Grade-IV cells (**Figure 39C**). In contrasts, the simulation outcome of high-grade (Grade-IV) GBM model showed the significant differences between the frequencies of the two states: LGG-II and Grade-IV. A higher number of Grade-IV tumorigenic cells were observed, which in turn justified the redirection of low-grade GBM cells towards the high-grade state in this high-grade GBM model (**Figure 39C**).

6.4.4 Origin of the Evolution of Intra-Tumor Heterogeneity and Distinct Sub-Types of GBM Tumor

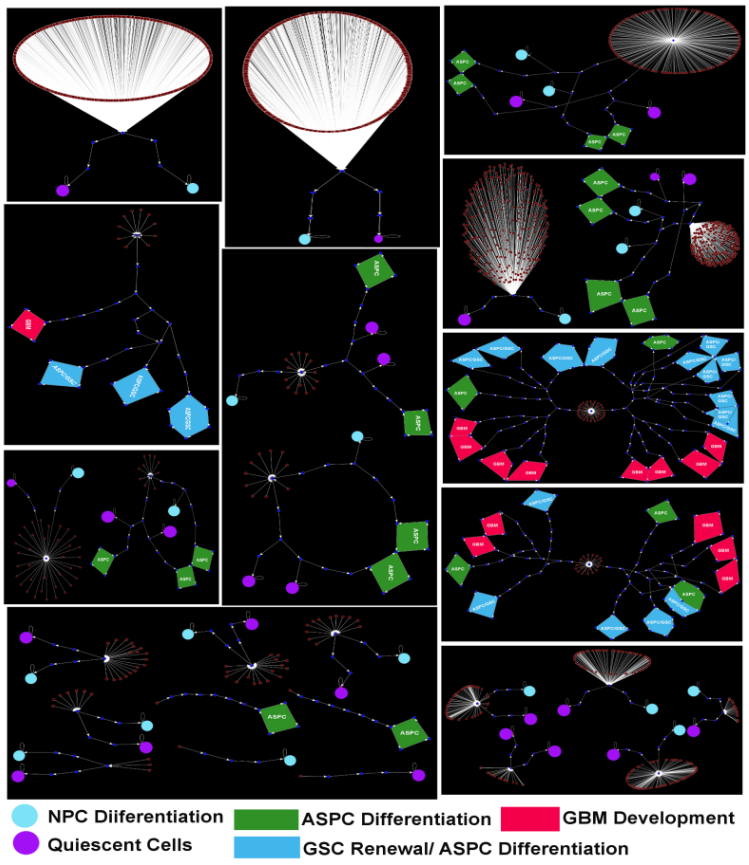

Figure 40: Full State transition graph (STG) of general GBM model.
NPC Differentiation and Quiescent cells are the fixed-point attractors and the others represent the cyclic attractor states corresponding to different sub-types of GBM tumor cells.

The state transition graph (STG) of the logical model developed for general glioblastoma development (GBM) was used to identify the initial and transient states (or cells), which were eventually redirected at either the singleton (i.e., fixed-point) or

periodic (i.e., cyclic) attractor states. Here, STG was used for representing the phylogeny and the hierarchical arrangements of variant tumor cells within the tumor microenvironment. In the steady state level of general GBM model simulation, the constructed STG contained 26 distinct groups of clusters, each of which was associated with either stable or cyclic attractor states or sometimes both (**Figure 40**).

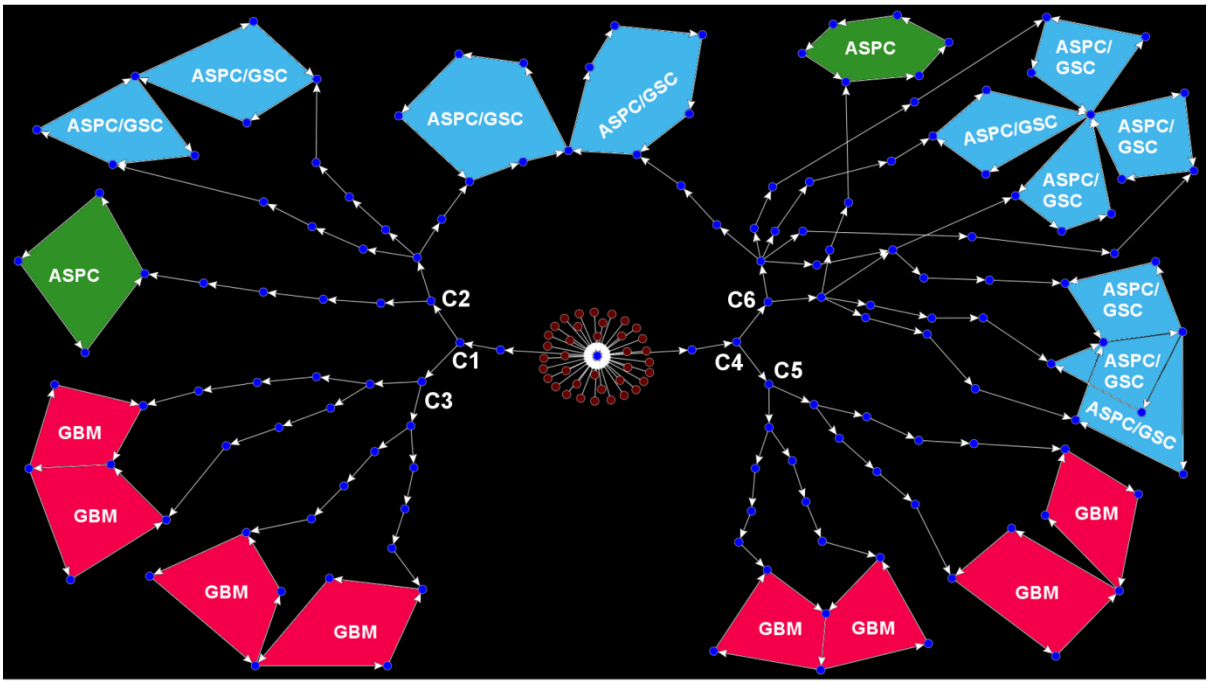

Figure 41: Partial stage transition graph (STG) of the general GBM model.
The Red and blue nodes represent the initial and transient cellular states of the general GBM developmental model. It shows that starting from a group of common tumor-initiating cells (red colored nodes), the tumorigenic developmental process leads towards distinct grades (sub-types) and sub-clones of GBM tumor. ASPC and ASPC/GSC refer the cyclic attractor states "ASPC Differentiation or LGG-I" and "GSC Renewal/ASPC Differentiation or LGG-II" respectively, whereas GBM refers the cellular state corresponding to "GSC Renewal/ASPC Differentiation/GBM Development or Grade-IV" cellular state.

Analysis of one such cluster extracted from the STG showed that the developmental dynamics of the tumor-initiating mutated GSCs were able to develop any of the three different sub-types of GBM *viz.* LGG-I, LGG-II and Grade-IV [234,521] (**Figure 41**). In this STG (**Figure 41**), the basin of attractors contained the critical nodes (e.g., C1, C2, C3, C4, C5, C6, etc.) at which the transient states were separated and proceed towards different cellular states (i.e., LGG-I, LGG-II, and Grade-IV). The critical nodes in STG, such as C1, C2, C4, and C6 are the nodes, at

which the stimulated GSCs undergo asymmetric cell division and produce distinct sub-types of GBM cells. On the other hand, C3 and C5 represent the critical junctions during the development at which the transient tumorigenic stem cells perform symmetric cell divisions and produce sub-clones of each different grade of GBM tumor (**Figure 41**). For example, there were total 14 distinct sub-clones (periodic attractor states) obtained in the simulation of general GBM model, each of which represented Grade-IV tumor cells and had different protein activity patterns at the steady state (periodic) levels (**Figure 42**).

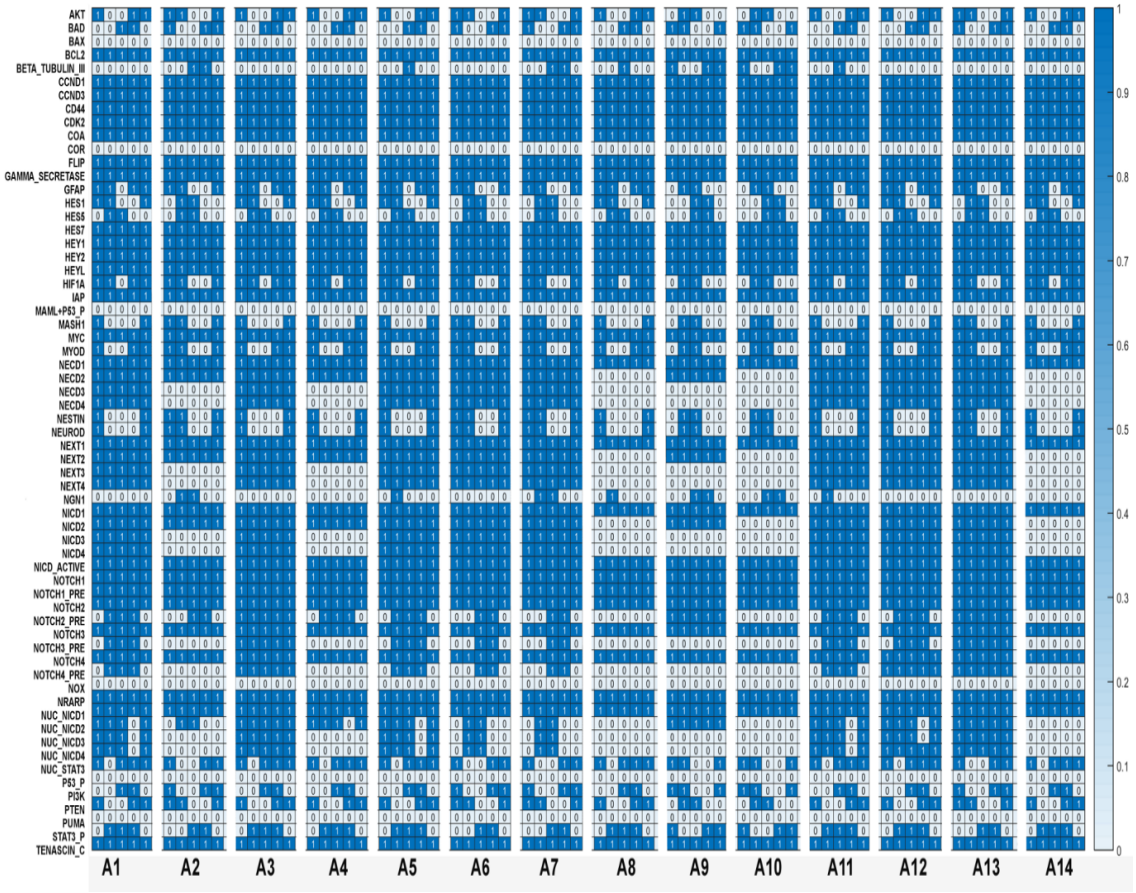

Figure 42: Activity patterns of the pathway molecules observed in all the periodic attractor states corresponding to Grade-IV tumor state in the general GBM model simulation.
The temporal expression/activity profiles (ON/Up-regulation OR OFF/Down-regulation states) of the pathway molecules in each of the distinct attractor states (total 14) clearly depict the intra-molecular heterogeneities, which exist within the sub-clones of Grade-IV tumor cells found in GBM tumor.

Comparative analyses of the activity patterns of the pathway molecules were performed successively to extract essential marker proteins whose variations of

expressions were creating heterogeneities within the sub-clones of Grade-IV tumor cells. This STG also helped to track the small but effective fluctuations within the tumorigenic stem cells, which significantly contributed to the emergence of different cell types, grades and the sub-clonal heterogeneity previously observed in the GBM tumor cells niche [139]. For example, after the critical junction at C1 in STG, it is visible that the developmental paths lead to either ASPC (LGG-I) or GSC/ASPC (LGG-II) cellular states via the transient node C2. Otherwise, it can also be attracted towards high-grade GBM (Grade-IV) states via C3 junction (**Figure 41**). Hence, by comparing the protein activity profiles of the transient nodes C2 and C3, it was revealed that transitions from low-grade GBMs (LGG-I or LGG-II) to high-grade GBM was possible if the cells had increased activities of the proteins YY1, C-MYC, CYCLIN-D1, CYCLIN-D3, etc. Comparing C5 and C6 nodes, the similar expression pattern of the proteins behind the origin of high-grade tumor cells was also observed. Moreover, to assess the molecular heterogeneities of the sub-clones of GBM (high-grade) cells, all the different 14 attractors states observed in Grade-IV tumor state were compared, which helped to find the origin of the heterogeneous sub-clones (**Figure 42**).

6.4.5 Understanding the Bias in the Outcomes of Different Cellular States under Different Conditions

The probabilities of occurrences of different cellular states were not homogenous in the attractor space (**Figure 35**). For example, in Grade-IV GBM model, the normalized frequency of Grade-IV tumor cells was comparatively higher than LGG-II cellular state, whereas, in the general GBM model, these two cellular states did not have much difference (**Figure 39C**). Here, it is hypothesized that the intra-cellular network, through which the activation signal flows from the ligand and receptors to the tumor marker proteins via the cytoplasmic and nuclear molecules, plays crucial roles to determine the bias towards a particular cellular state during tumor development. During the flow of such signal, which is imposed on a cell at the time

of cellular development, a set of signaling molecules (or proteins) alter their activity/abundance (ON/OFF states) patterns rapidly and relay the signal to the transcription factors in the nucleus. Hence, it could be considered that the activities of different ligand molecules, the complicated topology of the intra-cellular network, and the presence of several intrinsic and extrinsic fluctuations (e.g., mutations, over-activation, phosphorylation, etc.) were the critical factors behind the emergence of cellular heterogeneities in the attractor space. It is evident that if higher numbers of alteration of the molecular activities are required for reaching at a particular cellular state or phenotype, then a cell has to perform multiple chemical reactions. It will eventually increase the overall costs for the cells during the developmental stage and thus decrease the probability to reach that particular cellular state. Hence, to quantify the effects of diverse molecular events (e.g., chemical reaction, physical interaction, translocation, etc.) associated with the development of different cellular states, a novel phenotype cost function (Ψ_C) was defined in this present work. It was defined as the summation of the average rate of changes of the activity patterns of all the pathway molecules (i.e., the signaling cost) and the mutational cost associated with that cellular system (see **Section 2.5.8**).

Analyses of the phenotype cost function were found useful to understand the bias towards the determination of particular cell fate during neural stem cell maintenance and differentiation processes. The violin plots depict the comparative analyses of the distributions of the average cost values associated with the cellular states "NPC Differentiation", "ASPC Differentiation (LGG-I)", "GSC Renewal/ASPC Differentiation (LGG-II)", and "GBM Development (Grade-IV)" cells observed in aNSC (non-tumorigenic), general GBM and high-grade GBM (tumorigenic) models (**Figure 43**).

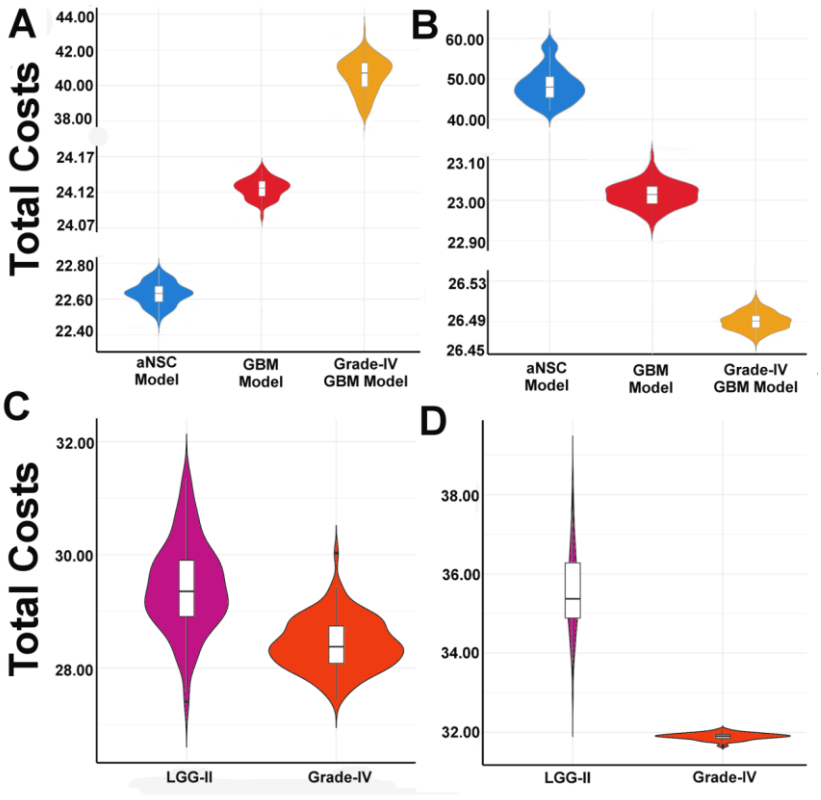

Figure 43: Violin plots of the total phenotype cost function or total costs calculated for different cellular states.
Comparisons of the median and the probability density functions of the cellular states **(A)** NPC Differentiation and **(B)** ASPC Differentiation observed in aNSC, general GBM, and high-grade GBM models respectively. In high-grade (Grade-IV) GBM model, the median costs for reaching the differentiated NPC and ASPC cells are opposite to each other. In adult NSC model, the median costs for reaching the differentiated NPC cellular state is lowest compared to the others tumorigenic models. In contrasts, the required total cost is significantly lower in the tumorigenic GBM model as compared to the non-tumorigenic aNSC model. Comparative analyses are also performed between the total costs required for reaching the LGG-II and Grade-IV tumor cell states in **(C)** general GBM and **(D)** High-grade GBM models.

It was observed that in the non-tumorigenic, adult NSCs model, the overall median costs to produce the differentiated neurons (NPCs) was significantly lower than the costs required for producing differentiated astrocytes cells (ASPCs) (**Figure 43A-B**). This result was validated with the previous observation, which suggests the higher amount of neurogenesis happens in the adult neural stem cell niche in SVZ [522]. Whereas in the general and high-grade GBM models, the total costs for reaching the ASPC differentiation state at the steady state level found to be reduced significantly, which in turn proved the previous experimental finding that the higher

number of matured astrocytes cells also exist in GBM tumor niche [523]. Hence, it is proven that the induced mutations e.g., P53 knock-out and increased activities of JAK2/STAT3, RBPJ, YY1 proteins, etc. can divert the normal functioning of Notch pathway in the aNSCs from undergoing the stem cell renewal or neuronal differentiation processes to the development of mutated, differentiated astrocytes cells [523].

On the other hand, while performing the intra-model comparisons of the total costs required for LGG-II and Grade-IV tumor cells, it was observed that in both the tumorigenic (i.e., general GBM and Grade-IV GBM) models, the median total costs required for developing Grade-IV tumor cells were comparably lower than the LGG-II cells. Although, the median difference was not very much high in general GBM model (**Figure 43C**), but it was found to be significantly higher in high-grade (Grade-VI) GBM model. Hence, it proved that to reach the Grade-IV tumor state in comparison to the LGG-II cellular state at the steady state level, the high-grade (Grade-IV) GBM model requires less cost than LGG-II cellular state. This result also explained the previous observation that why the Grade-IV tumor cells were observed with higher normalized frequency in the high-grade (Grade-IV) GBM model (**Figure 43C**).

6.4.6 Applications of Phenotype Predictor Scores to Predict the Appearances of Cellular States

The normalized frequency (P_C) and the total phenotype cost function or total cost required for reaching a particular cellular state were the two main parameters, which showed to regulate the outcome of a specific cellular state during the cell fate determination process. It was observed that the probability of occurrences of a particular cellular state was directly proportional to its total frequency and inversely proportional to the phenotype cost function. A novel scoring function "Phenotype Predictor Score (Θ_c) " was introduced here by defining this function as the ratio of the total frequency to total phenotype cost function associated with the cellular states.

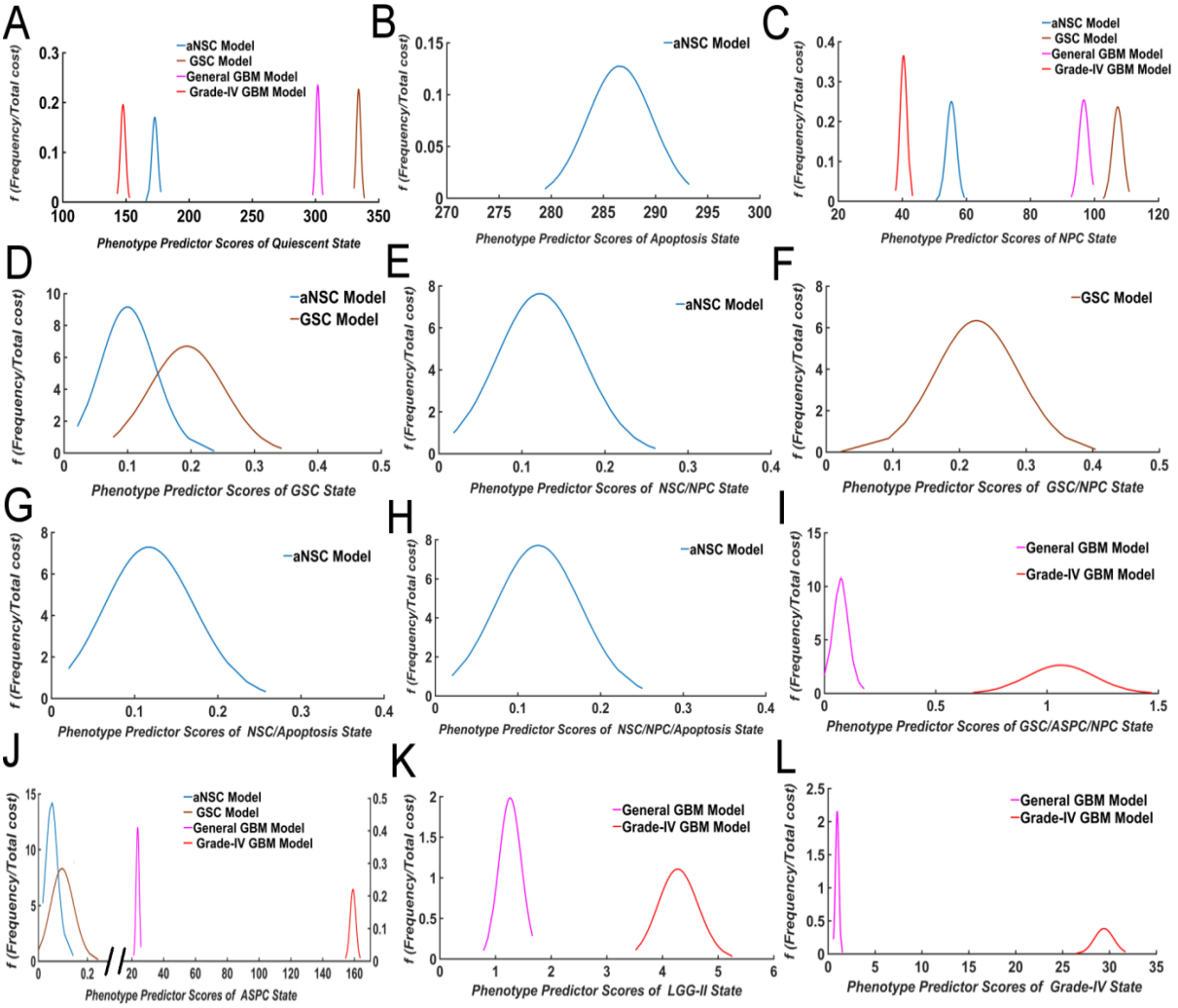

Figure 44: Distributions of the Phenotype Predictor Scores.
Distributions are plotted for all distinctly observed cellular states (total 12) *viz.* **(A)** Quiescent state; **(B)** Apoptosis; **(C)** NPC Differentiation; **(D)** GSC Renewal; **(E)** NSC Renewal/NPC Differentiation; **(F)** GSC Renewal/NPC Differentiation; **(G)** NSC Renewal/Apoptosis; **(H)** NSC Renewal/ NPC Differentiation/Apoptosis; **(I)** GSC Renewal/ASPC Differentiation/ NPC Differentiation; **(J)** ASPC Differentiation (LGG-I); **(K)** GSC Renewal/ASPC Differentiation (LGG-II); **(L)** GSC Renewal/ASPC Differentiation/GBM Development (Grade-IV) state.

The distributions of the frequencies (P_C) and total phenotype cost function (Ψ_C) of all the cellular states in the attractor space followed normal distributions with different variances. Hence, the probability density function of this newly defined score (Θ_C) is the joint probability distribution $(\Theta_C = \frac{P_C}{\Psi_C})$ of the ratio of two multivariate normally distributed variables and the shape of the resultant distribution is a fat-tailed Cauchy-like distribution (**Figure 44**) [237]. A detailed

description of this distribution including its mathematical expression is provided in the **Section 2.5.8**.

A higher value of phenotype predictor score of an arbitrary cellular state signifies the greater chance of that cellular state to appear within the attractor space (see **Section 2.5.8**). The mean values with 95% CI (calculated using Fieller's theorem [524]) of all the cellular states (total 12) observed in all the four aNSC, GSC, general GBM and Grade-IV models are provided in **Table 11.** It was observed that the mean phenotype predictor score (mean 147.688, 95% CI [146.630 148.747]) of the "Quiescent" cellular state was lowest in the highly mutated Grade-IV GBM as compared to the other models. This result was also in accord with the previous observation (**Figure 39A**), in which lower normalized frequency of quiescent state was detected in Grade-IV GBM model as compared to general GBM model (**Figure 39A**). Therefore, it can be state that the percentage of quiescent cells will be less in the heavily mutated tumorigenic niche (i.e., Grade-IV GBM cells) as compared to normal neurogenic niche (adult NSCs). A similar result also observed for "NPC differentiation" state, which showed that in the same tumorigenic niche, the heavily mutated tumor cells were least influenced (mean 40.372, 95% CI [39.804 40.940]) to differentiate into matured neurons as compared to normal aNSCs in the neurogenic niche. The score observed for GSC renewal (mean 0.193, 95% CI [0.163 0.224]) state was also higher in the GSC model as compared to the aNSC model (mean 0.100, 95% CI [0.078 0.124]).

A comparative analysis of LGG-I, LGG-II, and Grade-IV cellular states observed in general GBM, and Grade-IV GBM models revealed that the phenotype predictor scores for all of these cellular states were comparably higher in Grade-IV GBM model as compared to other model simulations. For example, the score observed for Grade-IV cellular state in general GBM model was less (mean 0.970, 95% CI [0.873 1.066]) as compared to Grade-IV GBM model (mean 29.414, 95% CI [28.875 29.955]) (**Table 11**).

Table 11: Mean Values with 95% CI of the Phenotype Predictor Scores of all Cellular States.

Observed Cellular States or Phenotypes	Models			
	aNSC Mean [95 % CI]	GSC Mean [95 % CI]	General GBM Mean [95 % CI]	Grade-IV GBM Mean [95 % CI]
Quiescent	172.813 [171.596 174.032]	334.160 [333.246 335.074]	301.827 [300.945 302.709]	147.688 [146.630 148.747]
Apoptosis	286.581 [284.957 288.206]	NA	NA	NA
NPC Differentiation	55.326 [54.496 56.156]	107.264 [106.388 108.140]	96.677 [95.862 97.492]	40.372 [39.804 40.940]
GSC Renewal	0.100 [0.078 0.124]	0.193 [0.163 0.224]	NA	NA
NSC Renewal/NPC Differentiation	0.123 [0.095 0.150]	NA	NA	NA
GSC Renewal/NPC Differentiation	NA	0.226 [0.194 0.259]	NA	NA
NSC Renewal/Apoptosis	0.118 [0.090 0.147]	NA	NA	NA
NSC Renewal/NPC Differentiation/Apoptosis	0.124 [0.098 0.151]	NA	NA	NA
GSC Renewal/ASPC Differentiation/ NPC Differentiation	NA	NA	0.072 [0.053 0.0917]	1.063 [0.984 1.141]
ASPC Differentiation (LGG-I)	0.054 [0.040 0.069]	0.098 [0.073 0.123]	23.655 [23.150 24.160]	159.059 [158.118 160.000]
GSC Renewal/ASPC Differentiation (LGG-II)	NA	NA	1.265 [1.161 1.370]	4.282 [4.096 4.470]
GSC Renewal/ASPC Differentiation/GBM Development (Grade-IV)	NA	NA	0.970 [0.873 1.066]	29.414 [28.875 29.955]

Thus, the induced alterations of the protein expressions in Grade-IV GBM models accelerated the growth of LGG-I, LGG-II and high grade (Grade-IV) glioblastoma

cells, which in turn helped to predict the chances of occurrences of the oncogenic cells in the tumor ecosystem.

Based on these observations, it was ascertained that the phenotype predictor score could be used to highlight the possible outcomes of the specific grade of tumor cells in the tumorigenic niche and could be a good estimator to quantify and assess the heterogeneity of glioblastoma tumor cells. Hence, after its successful executions on the master tumorigenic models (i.e. general GBM and Grade-IV GBM), it was interesting to analyze the performance of this novel scoring function to detect and quantify different grades and molecular sub-clones of tumor cells present in the tissue samples, collected from the individual glioblastoma patients. Here, the general hypothesis was that given the molecular expressions (e.g., transcriptomics, proteomics, etc.) data from the individual glioblastoma tumor patient or cohort, the phenotype predictor scores calculated for the tumorigenic (e.g., LGG-I, LGG-II, Grade-IV, etc.) and non-tumorigenic (e.g., Quiescent, NPC differentiation, apoptosis etc.) cellular states from the tumorigenic model would be able to predict the chances of occurrences of tumor cells and the corresponding grades.

6.4.7 Retrospective Studies of the Low and High- Grade GBM Patient's Cohort

Retrospective case studies using the RNA-Seq transcriptomics data, available in "The Cancer Genome Atlas (TCGA)" data portal of the P53 mutated low (TCGA-LGG) and high grade (TCGA-GBM) tumor samples from the cohorts of glioblastoma patients (**Table 12**) were performed to assess the potential of the developed models for identifying tumor grades by using the newly introduced estimator, "phenotype predictor score". A detailed description of the entire methodology used here is provided in the **Section 2.5.11** of *Chapter 2* (*"Materials and Methods")* of the thesis. The transcriptomics profile of the transcripts of 9 out of 53 input molecules was found (**Table 13**) in the differential expression analyses performed on the RNA-Seq data extracted from the cohort of P53 mutated, low-grade glioblastoma (TCGA-LGG)

tumor samples available in TCGA data portal. Using this TCGA-LGG transcriptomics data as inputs in the master aNSC model, a new simulation was performed to calculate the normalized frequencies of the observed cellular states (see **Section 2.5.11**).

Table 12: Statistics of the TCGA Glioblastoma Patient Cohorts.

Cohorts	# of Cases (Patients)				# of Samples (RNA-Seq Raw Counts Data)			
	M	F	U	Total	# Normal Tumor Samples	# Primary Tumor Samples	# Recurrent Tumor Samples	Total
TCGA-LGG (General)	282	228	1	511	0	513	16	529
TCGA-LGG (TP53 Mutated)	141	100	1	242	0	240	15	255
TCGA-GBM (General)	366	230	21	617	5	13	156	174
TCGA-GBM (TP53 Mutated)	32	21	0	53	0	57	0	57

M = Male, F = Female, U = Undefined/Not mentioned

At first, the new values were compared with the normalized frequencies obtained for each cellular state in the master aNSC model. After that, Chi-square goodness-of-fit test was performed between the normalized frequencies observed for each cellular state in the new simulation (observed data) versus normalized frequencies observed for each cellular state in the master aNSC model (expected data). This statistical test showed significant differences between the expected and observed data, which in turn proved that the transcriptomics profile extracted from the TCGA-LGG cohort did not indicate the development of normal neurogenesis of adult NSCs (**Figure 45A**). For example, due to the imposed induction of the proteins (such as P53 mutation, and increased activities of DLL1, DLL3, etc.) in the new model, cellular states "Apoptosis" and "NSC/NPC/Apoptosis" were not found in the attractor space. On the other hand, the normalized frequencies of the cellular states

"ASPC differentiation" and "GSC renewal" were found in higher numbers in the TCGA-LGG transcriptomics data model of aNSC. Hence, it proved that the neural stem cells having transcriptomics profile of TCGA-LGG cohort were more inclined to the development of glioblastoma stem-like (GSC) and mutated astrocytes or tumor cells (ASPC).

Table 13: Differential Expressions Genes in Glioblastoma Patient Cohorts.

TCGA-GBM (TP53 Mutation) Vs. Normal solid tumor		TCGA-LGG (TP53 Mutation) Vs. Normal solid tumor	
Up Regulation	**Down Regulation**	**Up Regulation**	**Down Regulation**
APH1,DLL3, FRINGE,GASE, HAT,HDAC,JAG1, MAGP1,MAGP2, NEDD4,POFUT1, SAP30	CNTN1,DVL,DTX 1, FBW7, JAG2, JIP1	DLL1, DLL3, FRINGE, NEDD4, POFUT1	FBW7, JAG2, NOV

Further, the present study was aimed to assess the same transcriptomics profile (TCGA-LGG) on the development of GBM tumor cells (**Table 13**). It was hypothesized here that if the same profile was taken as input in the master general GBM model, then the outputs of the model simulation will be able to capture the developmental dynamics of the tumorigenesis of low-grade glioblastoma starting from the mutated GSCs.

It was observed that all the cellular states, which were observed in the master general GBM model, were also appeared in the attractor space of this new model simulation. Furthermore, Chi-square goodness-of-fit test found that the values of the observed normalized frequencies of each cellular state fitted well with the expected normalized frequencies of each cellular state of the master general GBM model (**Figure 45B**). In this simulation, the mean values of the phenotype predictor scores of the tumorigenic cellular states viz. LGG-I (mean 19.00, 95% CI [18.60 19.41]), LGG-II (mean 1.10, 95% CI [0.98 1.22]), and Grade-IV (mean 0.83, 95% CI [0.73 0.93]) were similar to the values observed in master general GBM model (**Table 11**).

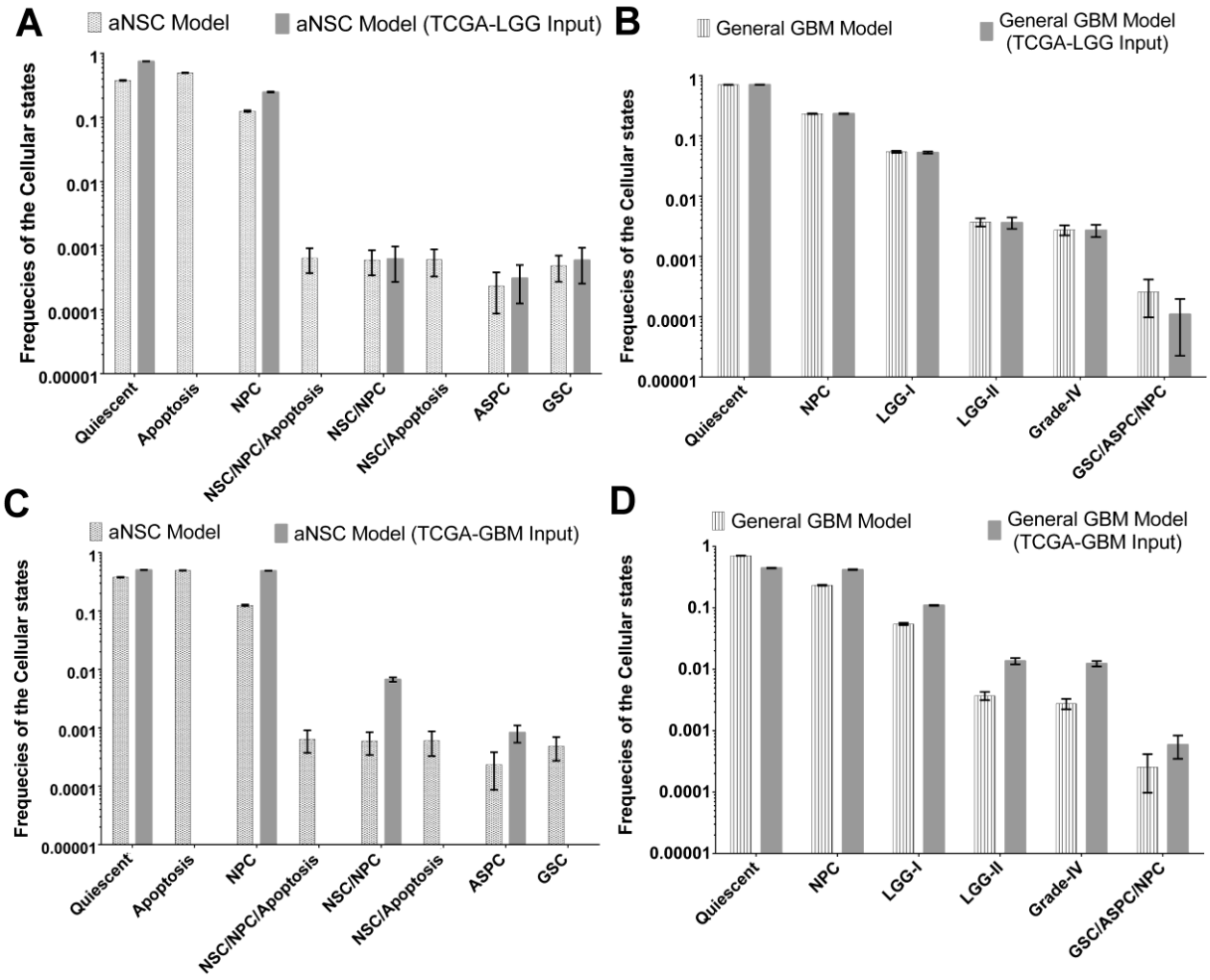

Figure 45: Comparative statistics of the normalized frequency distributions of different cellular states in master aNSC and general GBM models.
(A) A significant difference (Chi-Square goodness-of-fit test, P-Value < 0.001) is observed in the normalized frequency distributions while comparing the effect of TCGA-LGG expression profile in aNSC model simulation. Cellular states such as Apoptosis and NSC Renewal/Apoptosis are found to be absent in the aNSC model with TCGA-LGG expression profile. Normalized frequencies of ASPC Differentiation and GSC Renewal are also slightly increased in this scenario as compared to the master aNSC model simulation. **(B)** The normalized frequency distribution of the cellular states of general GBM model is well fitted with the simulation outcomes of general GBM model with transcriptomics profile of TCGA-LGG patient cohort. Similar to these analyses, mRNA expressions profile of the TCGA-GBM patient cohort is also given as inputs in **(C)** aNSC and **(D)** general GBM models. In both the scenarios, significant differences are observed between the expected and observed (outcomes from TCGA-GBM) normalized frequency distributions of the cellular states. These results prove that the developed aNSC and general GBM models are capable of differentiating the mRNA expression profiles of LGG and GBM patient cohorts correctly.

Hence, from this retrospective analysis, it was concluded that the transcriptomics profile of TCGA-LGG tumor samples was perfectly matching with the developed master general GBM model and therefore, this model could be further used for

testing other unknown transcriptomics profile to predict the existence of low-grade glioblastoma. This new simulation was also able to indicate the early risk of the development of low-grade GBM in an individual by analyzing the transcriptomics profile of that individual.

Another retrospective analysis was also performed by using the transcriptomic profile of the input molecules, which were differentially expressed in the P53 mutated TCGA-GBM tumor samples collected from the high-grade Glioblastoma patients' cohort (**Table 12**). Total 19 transcripts out of the total 53 transcripts of the input molecules were found significantly expressed in the high-grade GBM tumor samples with respect to the normal (control) solid tumor samples (**Table 13**). This transcriptomics profile was then provided as inputs in the master aNSC model to assess whether the differentially expressed genes/proteins of the transcripts of the input molecules were able to trigger normal neuronal cell development or not. It was observed that the transcriptomics profile while taken as inputs in the aNSC model were able to shift the NSCs developmental dynamics from the stem cell renewal process to the cell differentiation processes (i.e., NPC and ASPC). The normalized frequencies of cellular states observed in this new simulation were not found consistent with the cellular states observed in the master aNSC model (Chi-square goodness-of-fit test) (**Figure 45C**). It was observed that the cellular state ASPC differentiation had a higher normalized frequency in the attractor distribution in the new simulation, which signified that the transcriptomics profile observed in TCGA-GBM sample cohort were biased towards the development of astrocytes cells in the neurogenic niche. These astrocytes cells had a higher number of mutations, abundant expressions of Notch and JAK/STAT pathway molecules with rapid proliferation rate, which in turn helped these cells to differentiate and transform into tumorigenic states. Differentiation of NPC was also observed in higher numbers, which indicated that the tumorigenic niche also contained matured (mutated) neuron-like cells. Cellular state Apoptosis was also absent in the new simulation, which was an

indicator of the development of tumor cells. Overall, this study depicted that the transcriptomics profile of TCGA-GBM cohort was not helpful to the normal development of NSCs and could trigger tumorigenesis in future time-course.

To check the probabilities for developing high-grade GBM cells from the mutated GSCs a new simulation was again performed on this transcriptomics profile (extracted from TCGA-GBM cohort) by considering it as an input in the master general GBM model (see **Section 2.5.11**). The simulation outcomes showed that although there was no other cellular state appeared in the simulation as compared to the master general GBM model, but the normalized frequencies of the cellular states were not consistent (Chi-square goodness-of-fit test, rejecting the null hypothesis at 99.99% significance level). It was found that the tumorigenic cells viz. LGG-I, LGG-II, Grade-IV and GSC/ASPC/NPC were significantly increased in the new simulation (**Figure 45D**). Quiescent cells were significantly reduced, and simultaneously the differentiated NPC cells were increased. These results proved that the transcriptomics profile of TCGA-GBM tumor samples cohort could trigger stem cell differentiation as well as tumorigenesis. The inconsistency appeared between the normalized frequency values were due to the higher number of differentiation of neuronal and tumorigenic cells in the new model simulation.

The reasons behind the inconsistency of the normalized frequencies of cellular states were further analyzed by assessing the phenotype predictor scores. It was observed that the phenotype predictor scores of the tumorigenic states viz. LGG-I, LGG-II, GSC/ASPC/NPC and most importantly the Grade-IV tumor cells were significantly increased in the new simulation as compared to the scores observed in the general master GBM model (**Table 11**). For example, the mean phenotype predictor score of Grade-IV tumor cells was calculated in the new simulation was 3.22 (95% CI [3.05 3.39]), which was approximately 3 times greater than the mean score (mean 0.970, 95% CI [0.873 1.066]) of Grade-IV cells observed in master general GBM model. The mean scores of other two tumorigenic cellular states LGG-I and

LGG-II were 32.72 (95% CI [32.46 32.97]) and 3.51 (95% CI [3.30 3.71]) respectively, which were also slightly increased in the new simulation as compared to the general master GBM model simulation. On the other hand, the mean phenotype predictor score 133.04 (95% CI [132.21 133.87]) of "Quiescent" state in the new simulation was drastically reduced as compared to the mean score 301.827 (95% CI [300.945 302.709]) observed in the master general GBM model. Hence, this result indicated that the transcriptomics profile of TCGA-GBM tumor sample cohort was not only triggering the differentiation of mutated GSCs but also lead to the development of Grade-IV (high-grade) GBM cells. Therefore, it was concluded that the general master GBM model could predict the risk of the development of high-grade GBM cells, and simultaneously, the prospective study of the determination of the probable tumor grades could be performed by using its transcriptomics profile.

6.4.8 A Case Study of Screening and Ranking of Potential Drug Targets in High-Grade GBM Cells

Potential drug targets screening and its ranking by assessing their ability to suppress the Grade-IV tumorigenic cellular state were also performed on the developed Grade-IV (high-grade) GBM model. The objective of this study was to check the efficacy of an individual or a combination of protein(s) to suppress the Grade-IV GBM tumor cells in the tumorigenic niche. A detailed description of the methodologies used for the drug target screening is discussed in **Section 2.5.12**. The significant correlation and delay between the temporal dynamics of the Grade-IV cellular state observed in the Grade-IV GBM model (i.e., target signal) with the intermediate signaling molecules (i.e., query signal) was calculated in their corresponding normalized frequency domains (**Table 14**). The critical proteins identified from this analysis (**Table 14**) were targeted individually or in combinations in the high-grade GBM model, and the normalized frequencies of the LGG-I, LGG-II, and Grade-IV cellular states were compared with the master Grade-IV GBM model simulation results.

Table 14: Delay Difference and Significant Correlation Observed Between Grade-IV Trajectory and Pathway Molecules.

Original Signal	Query Signal	Delay 3	Correlation >= 0.6	P-Value
Grade-IV	**GFAP**	1	1	0
Grade-IV	**HIF1A***	1	1	0
Grade-IV	**NUC_STAT3***	2	1	0
Grade-IV	**MASH1***	3	-1	0
Grade-IV	**NESTIN**	3	-1	0
Grade-IV	**NEUROD**	3	-1	0
Grade-IV	**PTEN***	3	-1	0
Grade-IV	**STAT3_P***	3	1	0
Grade-IV	**BAD**	0	-0.7698004	0.005588
Grade-IV	**AKT***	1	0.7637626	0.010131
Grade-IV	**NGN1***	2	-0.7559289	0.018452
Grade-IV	**PI3K***	2	0.7559289	0.018452
Grade-IV	**HES1**	0	0.6236096	0.040347
Grade-IV	**HES5**	0	0.6236096	0.040347

*** Proteins selected for drug target screening as they do not belong to the class of marker proteins responsible for defining different cellular states.**

It was observed that inhibition of STAT3 protein strongly affected the development of all grades of Glioblastoma tumor cells. Besides, increased activity of MASH1 (or NGN1) protein (TC4) or its higher expression with the inhibition of PI3K protein as a combination (TC5) were able to suppress the LGG-I glioblastoma cells completely (**Figure 46A)** but were unable to suppress LGG-II and Grade-IV cells significantly (**Figure 46B-C**). On the other hand, inhibitions of PI3K/AKT and HIF1A simultaneously (TC6) was necessary to suppress the LGG-II tumor cells (**Figure 46B**), but not the LGG-I (**Figure 46A**) and Grade-IV cells (**Figure 46C**) in the tumorigenic niche. Significant partial suppression of Grade-IV tumor cells was also observed while activating NGN1 (or MASH1) protein in the Grade-IV tumor cells (**Figure 46C**). Therefore, the activation of MASH1/NGN1 was proven to be significantly useful for Glioblastoma treatment as it was able to suppress both LGG-I and Grade-

IV tumor cells simultaneously. However, it should be noted that LGG-II tumor cells which were mostly unaffected by this treatment, had a probability of the recurrence of Grade-IV cells from the remaining LGG-II cells in the future.

Table 15: Calculated ranks of the effective drug targets for suppressing Grade-IV tumor cells.

Target Cellular State	Query Proteins	Delay ≤ 3	|Correlation| ≥ 0.75	P-Value	Rank
GSC/ASPC/GBM	**NUC_STAT3***	2	1	~ 0	1st
GSC/ASPC/GBM	**NGN1***	2	-0.7559289	0.018452	2nd
GSC/ASPC/GBM	**STAT3_P***	3	1	~ 0	1st
GSC/ASPC/GBM	**MASH1***	3	-1	~ 0	2nd

Hence, from this analysis, it was possible to rank the target proteins, which had a significant effect on perturbing the LGG-I, LGG-II, and Grade-IV tumor cells in the high-grade GBM tumor niche. The ranking for each protein to target the tumorigenic cells is provided in **Table 15**.

In the previous experiments, inhibition of STAT3 protein is found highly effective to suppress Glioblastoma tumor cells [525,526]. However, the inhibition of STAT3 protein in cancer therapy also triggers various toxic side effects as this protein is involved in many other signaling pathways responsible for the development of normal neural cell lineages [527]. It has been observed that targeting MASH1 by over-expressing this protein in the tumor cells is sufficient for the treatment of high-grade GBM [528]. MASH1 is a neurogenic gene and over-expression of this protein will trigger the differentiation process of neurons from the GSCs present in the tumor niche and would be helpful to redirect the mutated stem cells from further astrocytes differentiation process [529].

Applications of such targeted, cell-based differentiation therapy would be very much practical as it has a less chance of developing tumorigenic (or differentiated, mutated astrocytomas) cells after therapy and thus may reduce the probability of the tumor relapse.

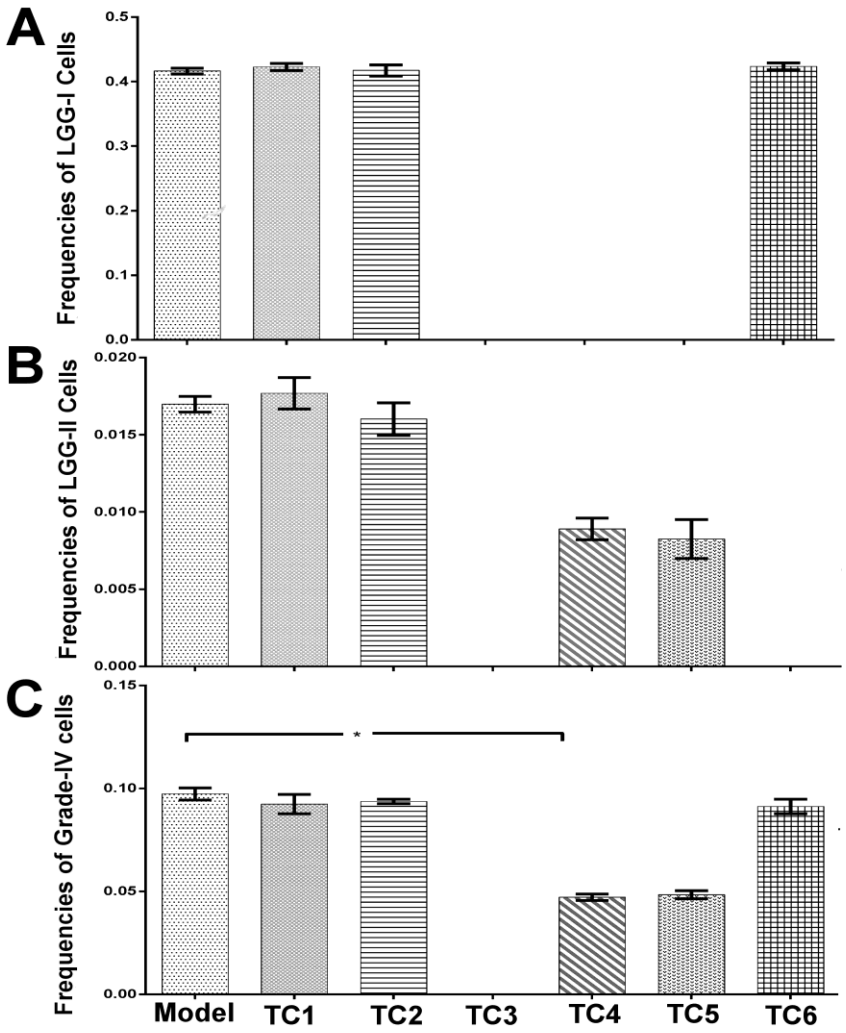

Figure 46: Simulation outcomes of drug targets screening analyses.
Normalized frequencies observed for (A) LGG-I (B) LGG-II, and (C) Grade-IV tumor cells in High-grade GBM model and different target screening scenarios are shown here. TC1: HIF1A Inhibition; TC2: PI3K & AKT Inhibition; TC3: STAT3 Inhibition; TC4: MASH1 & NGN1 Activation; TC5: PI3K Inhibition & MASH1 & NGN1 Activation; TC6: PI3K & AKT & HIF1A Inhibition. (* P-Value ≤ 0.05).

6.5 DISCUSSION

Population based study on Glioblastoma tumor by TCGA consortium has demonstrated the existence of four different tumor subtypes within the high-grade GBM patient cohort [530]. Later, Patel et al. have explored the possible cellular states with diverse transcriptional programs within the same tumor cells using single-cell RNA-Seq analysis and provided a substantial evidences of the existence of intra-tumor heterogeneity in GBM [531]. These studies have triggered several intriguing questions related to the mechanisms of the development of tumor sub-types and the

evolutions of genetically distinct sub-clones in GBM tumor. The molecular signatures between adult NSCs and tumorigenic GSCs are observed to be similar in various aspects, and it is understood that both of these cells share a common origin of evolution during development [506]. However, the dynamic regulations behind the emergence of tumorigenic lineages in the SVZ of the human brain are unknown hitherto.

In this work, a novel approach is proposed to analyze the developmental dynamics (i.e., self-renewal, differentiation processes, etc.) of adult neural stem cells in the neurogenic niche and the molecular mechanisms of the emergence of GSCs are attempted to explore. The role of juxtacrine Notch signaling and its cross-talk reaction mechanisms are considered for this purpose as the involvement of Notch target genes HES1, HES5, etc. are proven to be strongly correlated with the regulation of neurogenesis as well as gliogenesis in brain tissue [532]. While studying the normal neurogenic and tumorigenic niches, the simulation studies performed on the aNSC, GSC, general GBM and Grade-IV GBM models were also able to show the emergence of different cell lineages (**Figure 47)** which are similar to the observations found in the previous experiments [533,534].

The performance of these developed models was measured by calculating the maximum number of appropriate cell types appeared in a simulation with highest normalized frequency distributions (and Shannon entropy scores) [535]. Further, by introducing a novel scoring parameter "Activity Ratio (AR)" score, the active regulatory motifs which played the critical roles to maintain the dynamic balance of this entire developmental process of aNSCs were also identified (**Figure 38**).

It was observed that the three interconnected gene regulatory networks and the P53 dependent apoptotic network, which were responsible for the transcription of marker genes involved in stem cell maintenance, apoptosis, and the differentiation of neurons and astrocytes, were greatly influenced by the activities of Notch signaling

network during GBM development (**Figure 38**).

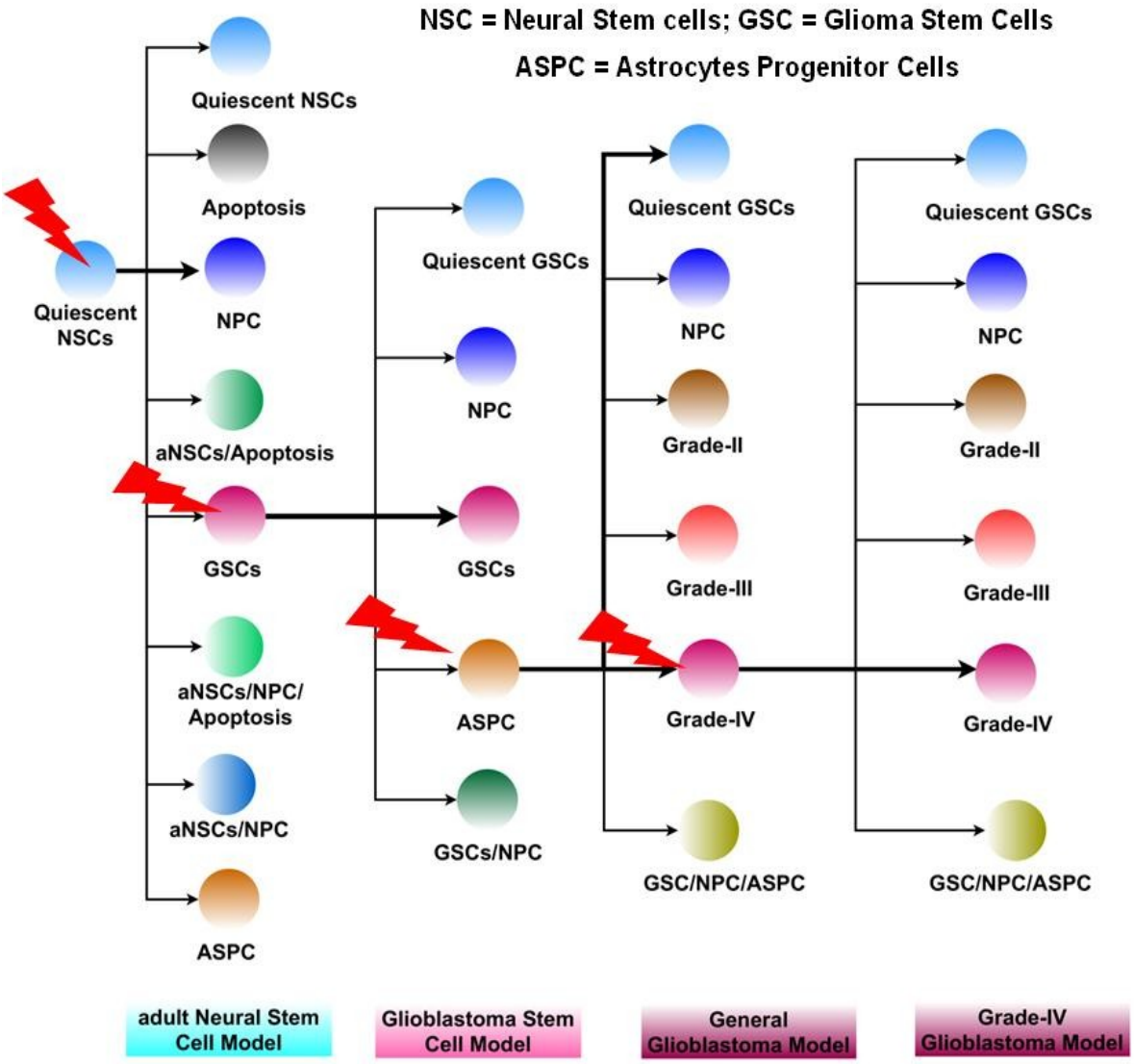

Figure 47: Simulation outcomes of Neural and Glioma stem cells development and the emergence of different grades of the tumor cells.
This figure describes the simulation outputs of the developmental model of aNSCs in the neurogenic niche and how different neuronal and astrocytes cell lineages are emerged. In the first step, aNSCs produce 8 different cell types, out of which 3 cell types (e.g. aNSC/NPC) are bipotent or multipotent (i.e. carrying multiple marker genes). Apoptotic cells are only observed in the aNSC model simulation, and are absent in the later tumorigenic simulations. This result clearly shows the capability of the developed models to correctly predict the differentiation and renewal of various cell types observed in the neurogenic niche of the SVZ of human brain.

On the other hand, Notch signaling network was stimulated by several extrinsic (such as ligands) and intrinsic (e.g., kinase, transcription co-activator, repressors,

crosstalk reactions, etc.) factors, which profoundly differed based on the microenvironment of the cancer cells as well as its genomic/proteomic profiles. However, it was observed that Notch pathway alone could not regulate or initiate all the cellular states. For example, to trigger the gliogenesis process, Notch target proteins HES1/5 should be highly abundant in the cells to stimulate the kinase activity of the JAK2 protein and the successive phosphorylation of the transcription factor STAT3 protein. STAT3 is the transcription factor of the gene responsible for the expression of GFAP protein, which is a known marker of astrocytes cells [443]. Hence, from this motif identification analyses, it was clearly understood that how the intracellular Notch signaling network performed multiple developmental processes in adult NSCs and governed the neurogenesis, gliogenesis as well as tumorigenesis. This mechanistic understanding further helped to assess the effects of the perturbations within Notch signaling network through which normal aNSCs develop proliferating GSCs and further trigger the development of different grades of Glioblastoma tumor in the human brain.

Finally, it was observed that the core Notch pathway is primarily used in the adult NSCs as a rheostat, which can be tuned by regulating (i.e., mutation, activation, inhibition, etc.) various proteins to achieve desired phenotypic outputs (**Figure 48**). Simulation outcomes showed that if the activities of the core component molecules of Notch pathway (e.g., Notch receptors and ligands, γ-secretase enzyme complex, MAML, RBPJ or CSL, HES/HEY, etc.) were very less, then the stem cells would stay at its quiescent state (qNSCs). On the other hand, the cells would lose its stemness or quiescent properties, if EP300 protein had increased activity in the aNSCs/aNSCs. In this case, the stem cells will switch to the differentiation process and develop into matured neurons.

The neural stem cells would retain its self-renewing state if the core components of this pathway were abundant in the cells to produce HES/HEY, CYCLIN-D1, etc. proteins periodically and simultaneously possessed the wild-type active P53 protein.

Furthermore, if the P53 protein was mutated but Notch pathway was active, then the stem cell dynamics would alter and redirect towards the development of self-renewing Glioblastoma stem cells (GSCs).

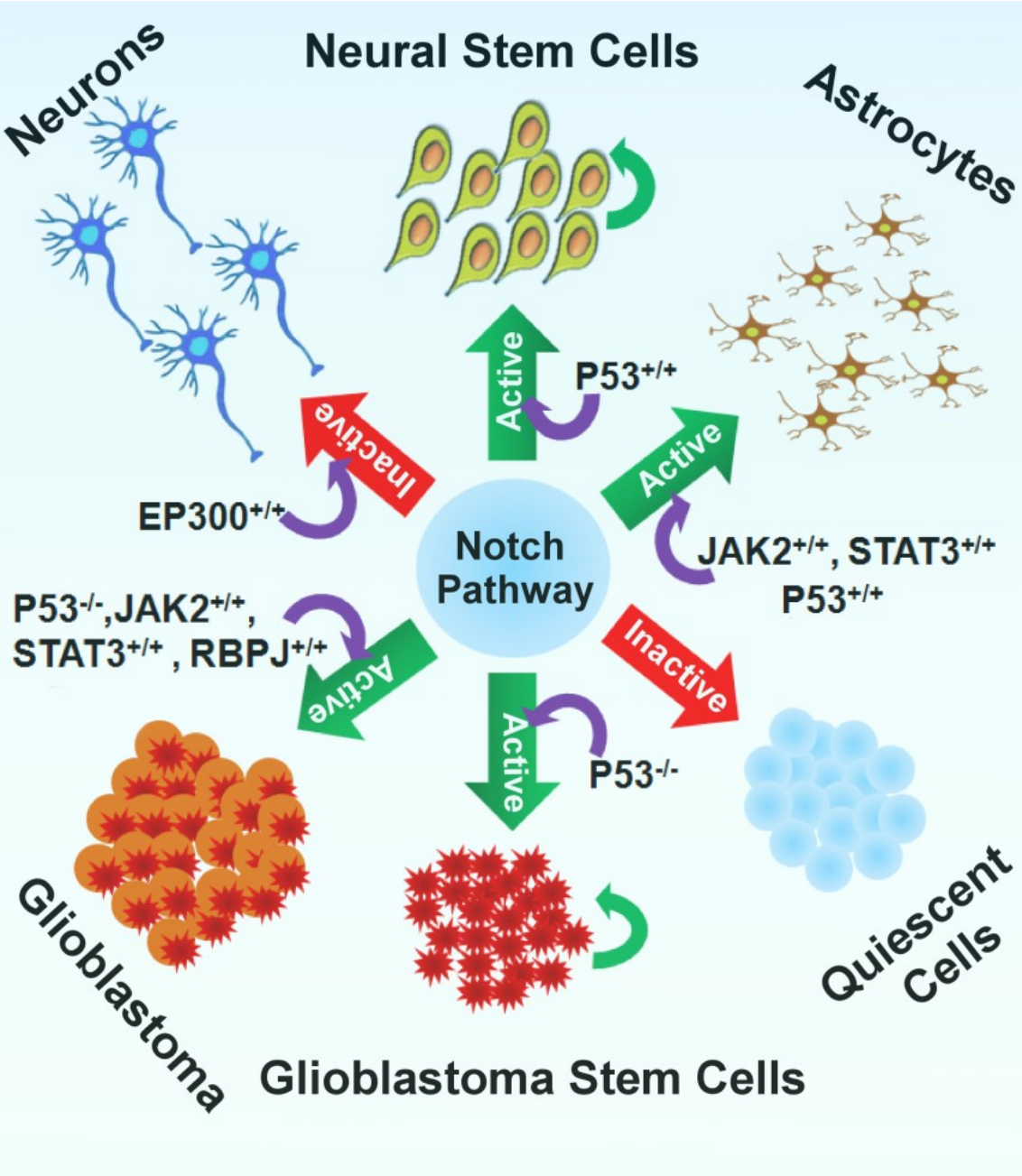

Figure 48: The rheostat model proposed for Notch signaling network.
This model shows the mechanistic regulations of core Notch pathway during the evolutions of different sub-types of normal and tumorigenic cells from adult neural cells.

It was also observed that in the normal, wild-type adult NSCs (with wild-type P53), if the targets proteins (i.e., HES1, HES5) of active notch pathway interact with

the JAK2/STAT3 protein, then the developmental dynamics of adult NSCs would redirect towards the development of gliogenesis (i.e., astrocytes) process. It was also observed that the same gliogenesis process could be malfunctioned and lead to the development of Glioblastoma tumor formation if the differentiating cells had mutant P53 protein.

The master general GBM model, developed for simulating the tumorigenesis process, was also able to predict the developmental routes through which the intra-tumor heterogeneities of tumor cells of different grades emerged in the tumorigenic niche of Glioblastoma (**Figure 40, Figure 41,** and **Figure 42**). The critical time points, at which the different grades (sub-types) of GBM tumors and its sub-clones occur, were also possible to extract for analyzing the differences in the molecular profiles of the tumor cells at these critical junctions. Moreover, using the newly introduced phenotype prediction score, it was possible to quantify the probabilities of occurrences of different cellular states as well as tumor grades during the time of tumorigenesis.

This scoring technique correctly predicted the bias towards the development of low-grade astrocytoma from mutated aNSCs/GSCs or the bias towards the development of high-grade GBM cells from low-grade GBM. To validate the existence of such bias, a case study was performed considering the TCGA RNA-Seq data from low and high-grade GBM tumor samples cohort as the inputs for master aNSC and general GBM models, respectively. The simulation results correctly predicted the higher probabilities of low grade and high-grade (Grade-IV) GBM tumor cells in the general GBM model (**Figure 45**). After this prediction study, the proposed Grade-IV GBM model was used for screening and ranking of potential drug targets for the suppression of the growth of Grade-IV tumor cells. A novel methodology was developed for this purpose in which Fast Fourier Transformation (FFT) technique was used to find out the correlation and lag between the temporal dynamics of Grade-IV tumor state with the temporal expression pattern observed for

each pathway molecule. It was observed that the protein molecules correlating higher than 0.6 and lag less than or equal to 3 time-steps with the Grade-IV tumor cell were the suitable drug targets to perturb the Grade-IV tumorigenic signal. *In-silico* perturbations were performed on the Grade-IV GBM model to explore the effects of targeting these identified drug targets individually or in combinations. The drug target (such as MASH1), which showed huge reductions of the normalized frequencies of LGG-I, LGG-II, and most importantly Grade-IV tumor cells, but did not affect the growth of non-tumorigenic cells (NPCs), were considered as the highly preferable target(s) (**Table 15**).

6.6 CONCLUSIONS

This work has provided new concepts for better understanding of the biological signal transduction mechanisms in normal and diseased scenarios. The overall methodologies provided here can be applicable to multiple biological processes (such as stem cell development, neurogenesis, tumorigenesis, etc.) and can be implicated for a diverse range of investigations in future. If integrated with robust experimental data, this work would be helpful to substantially advance the understanding of the mechanisms underlying cellular responses to external or internal cues, and refine the current views of the signaling processes.

Chapter 7

---

7 IMPLICATIONS OF THE RECONSTRUCTED INTEGRATED CELL SIGNALING NETWORKS TO UNDERSTAND THE GROWTH AND METASTASIS OF GLIOBLASTOMA

7.1 INTRODUCTION[6]

In the previous chapters, we described the importance of Hedgehog and Notch pathways, and their individual cross-talks with other pathway molecules (e.g., JAK/STAT, PI3K/AKT, HIF1A, P53, ERK, RAS, etc.) in Glioblastoma tumor cells. We performed literature-based pathway reconstructions followed by topological analyses and binary logic-based dynamic simulation of the pathway species to predict their expression levels in normal and tumorigenic cells (e.g., Glioblastoma, Colon, Pancreatic cancers) [185]. We have shown our approach effective as well as efficient to model the temporal dynamics of multiple pathway species (i.e. genes, proteins, complexes, metabolites, etc.) and their reaction dynamics using Boolean logic-based formalism without the use of kinetic rate equations and quantitative dynamic modeling. The semi-dynamic binary logic (BL)-based approach was useful to predict the expressions or activity profiles of the genes/proteins of the modeled pathways in different initial conditions and thus we were successful to create different cellular systems (e.g. normal, glioblastoma, drug treatment). The steady-state expressions of the target genes of both the pathways in GBM scenarios were compared with the experimental evidences (e.g., microarray, proteomics data) of Grade-IV GBM cells and thus the performances of the developed models were calculated. During the comparative studies of the pathway models, we observed

[6] The materials of this chapter have been taken verbatim from our communicated article Chowdhury & Sarkar, 2018 (manuscript under review).

higher level of accuracies while predicting the expression levels of target genes/proteins in cancer cells with respect to normal cells or controls.

We were also able to analyze the state-transition dynamics of the overall system (cellular state) under different input conditions, and successively analyzed the course of the appearances of multiple attractor states correspond to different cellular fates. The attractor analyses in the state-transition graph (STG) of the logical model developed for Notch signaling pathway was specifically helpful to understand the fate of adult neural stem cells (aNSCs) under different pathological conditions and the predictions of the trajectories to the development of glioblastoma stem cells (GSCs) from aNSCs. The follow up analyses of the results obtained from the attractors of the STGs were specifically useful for determining the steady-state expression patterns of the important marker genes, particularly expressed in the self-renewing stems cells (aNSCs, GSCs), matured and differentiated cells (e.g., astrocytes, neurons, and different grades of GBM tumor cells), and the apoptotic cells. Perturbation studies of important genes/proteins in the network were also performed in the logical model by making the target node(s) time invariant during the Boolean updates. The resultant attractors under the perturbation analyses were compared with the normal and tumorigenic scenarios, and subsequently the efficacies of the drug target(s) were measured.

Applications of logic based (qualitative) modeling technique have been implicated in the analyses of other biological systems (most often gene regulatory and signaling pathways), in which prediction of the qualitative expression patterns of the nodes (i.e., genes/proteins) at equilibrium condition are of particular interest to the model developers [536-538]. Also, while developing mathematical model of a large biochemical reaction network, in which the quantitative information of the reaction parameters are largely unknown, binary logic-based logical modeling approach is proven to be effective in such instances [539]. Various attempts of modeling the signaling pathways, such as T-cell signaling [244,309,540], MAPK

pathway [541], JAK/STAT [542], Apoptosis [543], etc., and several gene regulatory networks of yeast [544], mammal [545], and plat cells [546] are successfully executed by using binary logic-based Boolean modeling approach. The resultant attractor states obtained from the STGs of the model simulations are used for explaining the cell fate decision mechanism during development and identification of the molecular pathways leading to the expressions of various phenotypic states (e.g., cell division, differentiation, apoptosis, metastasis, etc.) in the organisms [537,547,548]. Logic-based computational models are also developed for studying the effects of different extra-cellular and environmental factors (e.g., radiation, carcinogens) on cell signaling and gene regulatory networks [549].

The expression/activity of the nodes considered in the logical model is mapped discretely with only two binary states i.e. active/over-expressed/present (or binary state = 1/TRUE) and inactive/under-expressed/absent (or binary state = 0/FALSE). To run the simulation, the initial binary states (TRUE/FALSE) of all the nodes are required to provide at the start of the simulation, which are mainly considered by discretizing the continuous expression values, obtained from different biological assays. Most often, the discretization of the continuous expression values are performed by differential expression analyses of the samples with respect to the controls, and thus different cellular conditions are created for modeling and simulation. The binary logic-based analyses solely rely on the assumptions of only two possible states of the nodes based on their "significant changes" in the expression levels observed in the biological samples versus controls [82,185,542]. Most often, a "threshold" (e.g., P-value, FDR, log fold change) is determined and applied on the expression values obtained in array-based experiments based or from other experimental assays to discretize (up/true/active/present/1 or down/false/inactive/absent/0) the expressions or activities or abundances of the nodes (pathway species such as genes, proteins, microRNA, etc.) in a given biological condition [82,126,185,550]. Besides, perturbations analyses in the logical model can be

performed by freezing the update rules of the select nodes during the execution of the update rules. This is also an advantage of the application of logical model, in which the temporal dynamics of the nodes in a perturbed condition can be predicted and thus the expressions of the target nodes (or outputs) of the perturbed model can be compared with the unperturbed model at the steady-state. Such perturbation analysis is very much effective to examine the sensitivity of the nodes in a model without having quantitative information of the perturbation on the select nodes [551]. The most sensitive nodes identified through the perturbation analyses in the logical models of cell signaling or gene regulatory networks are considered as "target nodes" of the network [552,553].

7.1.1 Challenges in the Selection of Binary States of the Nodes

Despite its successful implications in the analyses of gene circuits and cell signaling networks, binary logic-based model formalism has few inherent limitations [554]. However, in some cases assumption of the binary states to represent the possible outcomes of a node in a biological sample can be limited to the over-simplification of the target problems. For instance, certain transcription factors, receptors, ligands, etc. are found to be expressed in almost equal quantities in both the normal and cancer cells. In that case, the differential expressions of those transcription factors and other molecules in both the cells will be computed as "insignificant" after performing the differential expression studies. For example, the presence of equal amount of transcription factors responsible for the transcription of MYC oncogenes in both normal and cancer cells (e.g., Burkitt lymphoma) may trigger different level of transcriptions (weak or string). It is observed that MYC oncogenes in Burkitt lymphoma cells translocates into the enhancer regions of immunoglobulin heavy chain (IGH) locus, which in turn causes strong transcription of MYC [555]. Logical modeling of such molecular events $(TFs \rightarrow Gene)$ to distinguish the magnitude of gene expressions in normal and cancer cells will remain elusive, as the transcription factors (TFs) of the genes (activators) are present in both

the cells (i.e. at ON state) and thus the logical state of the gene in both scenarios will be at active (or ON) state. On the other hand, if we consider the enhancer elements with the TFs during gene transcription in the logical equation i.e., $(TFs\ AND\ Enhancers \rightarrow Gene)$, then the absence (OFF state) of enhancers in normal cells will output down regulation or inactive (OFF) state of the gene, even though the TFs of the gene is still available in the normal cells. This is one of the serious challenges of the use of binary number-based logical modeling approach in gene transcription network, in which multi-valued states of the expression levels (absent, low, normal, strong) of the genes/proteins are difficult to model. On the other hand, there are some genes/proteins in the cells, which are always expressed in lower quantities but play effective roles in the functioning of both the normal and cancer cells. In such situation, the differential expression studies will also compute those genes as insignificant, and thus determining the initial binary states will be again difficult.

Similar problem may also arise while modeling the receptor mediated pathways (e.g., ERBB pathway), in which the genes (*EGFR*) responsible for encoding receptor proteins have higher copy number gains in the target samples (e.g., cancer cells) in comparison to the normal controls [556,557]. Binary logic-based modeling approach is not suitable in this case to model the effects of the copy number gains of the receptor proteins and its downstream regulations in both normal and cancer cells, and subsequently predict and compare the gene expression patterns at the steady-state or equilibrium conditions. Also, the homozygous ($P53^{-/-}$) and heterozygous ($P53^{+/-}$) deletions of protein coding genes can have different effects in the downstream of signaling cascades and gene regulatory networks in the target biological cells/tissues [558,559]. In such instance, developing different computational models, which consist of homozygous deletions, heterozygous deletions, and wild-type expressions of the select gene can be challenging as both the deletion scenarios will have to be considered in the logical model as down-regulated or inactive (i.e., FALSE

or 0). Hence, we can understand that binary logic-based algebra will mostly fail to develop and compare the outputs of the logical models of signaling and gene regulatory networks, which consist of the genes/proteins with copy number gains, homozygous and heterozygous deletions, and wild-type expressions.

Apart from considering the multi-valued variations of the expression of genes/proteins in the logical models, binary logic-based method is also unable to capture the differences in the signal propagation under strong and weak stimulations of the receptor proteins (e.g., T-cell receptor and co-receptors) in signaling pathways or the presence of very high abundance and regular concentrations of transcription factors in the promoter regions during the simulation of gene transcription regulatory networks. There are various examples, in which the bi- or multi-phasic responses of target genes and phenotypic expressions observed after strong and weak stimulations of the receptor proteins in signaling pathways [560-563]. Different multitudes of signal intensities or the abundance of transcription factors, co-repressors, and co-activator molecules in the nucleus are proven significant in the pathology of various diseases including cancers and immunological disorders [561,563,564]. Due to the presence of such variances, alternative molecular reaction pathways or gene regulatory motifs can be activated in the cellular systems, which sometimes may change the course of expression dynamics of the target genes or phenotypes of cells or tissues [564,565]. However, comparative analyses of such scenarios are not possible via Boolean modeling due to its limitations of the scope of considering only TRUE (or active state) for both the strong and weak stimulations or very high and modest concentrations of the transcription factors and co-factors in the logical model simulations.

7.1.2 Methodological Challenges of the Constructions of Logical Equations

We have also observed few methodological challenges while writing the logical equations based on Boolean algebra to model biochemical reaction pathways. For

instance, when a single protein molecule or any other species (say ***A***) in the signaling or gene regulatory networks is negatively regulated by another single inhibitor (say ***B***), the logical state of the negatively regulated species will have complementary state of its upstream inhibitor. Such type of reaction is modeled in Boolean logic by using negation $(B \neg A)$ or universal NOT gate $\{A(t+1) = !B(t)\}$. Hence, if we consider $A(0) = 0$ and $B(0) = 0$, then based on the condition of negation we can say that $A(1) = 1$, even though the expression of ***A*** at 0th time-point was null. Now, let us consider the node ***A*** is a protein, which gets phosphorylated by another protein ***B*** and then the protein ***A*** goes to proteosomal degradation. Most often, such type of chemical reactions are written in Boolean logic by using the above mentioned logic of negation rule. The output of the logic equation in such biochemical reaction holds true in terms of the biochemistry of the event, if only the protein ***A*** is present or TRUE in the initial time-point or 0th state. Whereas, if the protein ***A*** is not at all present in the initial time-point, then the presence (TRUE) or absence (FALSE) states of ***B*** at the initial time-point has no consequence in the downstream protein ***A***. But, we have observed that the binary output state of ***A*** at ***(t+1)th*** time step will be "TRUE", if ***B*** is "FALSE" at the ***t th*** time step. Indeed, this output will propagate false chemical signal in the downstream of the network. Also, the logical output state of a node ***Z*** at ***(t+1)th*** time-point regulated by multiple co-operative activators (***X1, X2, X3,...***) and inhibitors (***Y1, Y2, Y3,...***) considered as conjunction (AND gate) in the Boolean equation $\{f(Z) = AND(X_1, X_2, X_3, ..., !Y_1, !Y_2, !Y_3, ...)\}$ will be always at FALSE state if any of the upstream inhibitor of that node is at TRUE state in the previous time-point (***t***). As a consequence, the output of the overall system could depict strong canalizing effect in the attractor space.

7.1.3 Alternative Solutions to Overcome the Methodological Challenges of Boolean Logic

In this work, we have proposed an extended quaternary logic (ExQuLogic)-based method as the alternative solution to overcome the challenges or limitations

discussed in the previous sections. In order to consider multiple qualitative states of logical nodes of the Boolean equations, previous research works have already shown the use of multi-valued logic in the update scheme [539,566-569]. Several software, such as CellNetAnalyzer (CNA), GINsim, etc. are available freely to model the signaling and gene regulatory networks using binary and multi-valued logical update schemes [218,570]. In CNA, the logical equations of the nodes are integrated with stoichiometry of the equations e.g. $2A \wedge B = 3C$ to consider the multi-valued logic. From this equation, CNA assumes that the logical states of the node ***A*** is when 2 and the node ***B*** is 1, then the activation state of C will reach at 3. Hence, using this approach, one can easily model the strong and weak stimulation of the receptors, higher and regular abundances of transcription factors and other molecules in the nucleus and its effects in the gene transcription network. The effect of copy number gain or homozygous and heterozygous deletions of the genes/proteins can be also performed by using CNA. Moreover, to account the effects of multiple inhibitors in the regulations of a downstream node in the model, *Guebel et al.* have also proposed an extension of multi-valued quaternary logic, named as *graded-inhibition* theory [569]. Here, the authors have argued the use of four logical ordinal states (**0**, **1**, **2**, and **3**) $S = \{fully\ inactive, lowly\ inactive, high\ activity, maximal\ activity\}$ to represent the quaternary states. The authors have also shown the procedure to integrate the *graded-inhibition* method with CNA, so that the model developers can realize the full potential of the use of multi-valued logic implemented in CNA software environment. The other popular software GINsim can also take multi-valued states of the nodes, whose update rule can be defined by user specified threshold value. Using these software, researchers have successfully modeled the gene regulatory system of *lac* operon found in prokaryotic system [554], E2F1 transcription network [539], apoptosis pathway [566], functional model of hypothalamic-pituitary-adrenal (HPA) axis during stress response [567], the mathematical model of anti-inflammatory effects of reactive oxygen species (ROS) in immune signaling network [568] and so on [571].

However it should be noted that, although both CNA and GINsim can take multi-valued logic states of the nodes to successfully simulate the multi-valued logic operations, the very basic theory or logic behind the update rules of the nodes used in this software is still based on general Boolean algebra. In the previous models of quaternary-logic based simulation of biochemical pathways, the update rule of the quaternary states of the logical nodes is dependent on either stoichometric constraint or the canalization rules of upstream nodes. In these cases, the threshold values, canalization rules or the stoichiometry selected in the logical state update schemes are completely relied on prior domain knowledge of the model developers. However, for a large signaling and gene regulatory network, providing the threshold values or imposing the constraints on the update rule using stoichiometry may not be always the feasible options to the model developers. Apart from the quaternary state update rules, binary logic-based models are not capable to discretize the insignificant genes/proteins in appropriate logical state, except "Up (1)" and "Down (0)". In few cases, researchers have also used fuzzy-logic-based modeling approach to discretize the continuous expressions of the genes/proteins/metabolites of the pathway [572-574]. However, in spite of its superiority over conventional binary logic-based Boolean modeling approach, the performance of Fuzzy-logic is largely depends on the appropriate selection of membership functions to activate the nodes and most often it requires large amount of training datasets to come up with a best fuzzy-network to predict gene expression pattern in a given biological scenario [575,576].

7.1.4 Proposed Approach

Hence, to overcome the above mentioned issues, in this work we propose to implement the previously developed truth table of quaternary number system [243] and its corresponding algebra to perform the basic logic operations of negation (NOT operator), conjunction (AND operator), and disjunction (OR operator) rules (see **Section 2.6.1**). Like the truth table used in conventional Boolean algebra, the update

scheme of multi-valued quaternary logic will be also defined by the already established truth table of quaternary number system **0**, **1**, **2**, and **3** [243]. In the logical model of biochemical pathway using our proposed ExQuLogic-based operation, the quaternary states are considered as ordinal values, in which **0** refers the fully inactive or absent or lowest possible active state of the nodes. The quaternary state **1** represents the lowly active or mild expressions of the pathway species, but higher level of expression than the state **0**. Another state **2** represents the regular or normal or modest expression or activity of nodes, which is higher than the state **1** (lowly active), but lower than the quaternary state **3**. Here, the quaternary state **3** represents the maximal active or highly over-expressed or strongly stimulated nodes in the network. Unlike the conventional logic, herein we can model the weak or mild stimulation or responses of genes/proteins by considering the quaternary state either as **1** or **2** with respect to strong stimulation/response denoted by the quaternary state **3**. Likewise, we can also consider the state **3** to define the copy number gain, state **2** to define wild-type expression, state **1** to define heterozygous deletion, and the state **0** to represent homozygous deletion or knock-out of the genes/proteins or any other pathway species in the logical model. Similarly, the quaternary state of the significantly under-expressed genes having very high negative fold change ratio (threshold) in the target samples in comparison with control observed after differential expression studies can be considered as "**0**" and "**1**" for the condition of *vice-versa*. On the other hand, the quaternary states of insignificant genes can be defined as either **2** or **1** based on the statistics observed in differential expression studies. The elaborate descriptions of the quaternary state selection rule used for considering the initial states of the input nodes from the statistics obtained from the differential expression data are provided in the ***Materials and methods*** section (see **Section 2.6.4**).

However, we have found that there are few scenarios in the biochemical reaction events, which cannot not be captured appropriately while implementing the logic

rules described in the quaternary logic truth table. Hence, we argue to use the concept of extended quaternary (multi-valued) logic (ExQuLogic), in which the update rule of a node is not defined by any pre-specified subjective threshold values or canalization rules. It also does not depend on the specification of the stoichiometry of the reactants and the products of biochemical reactions to perform the multi-valued logic operation. In ExQuLogic-based operation, we have introduced a new extended quaternary state upstate scheme (ExQSUS) to model and simulate the functioning mechanisms of biochemical reactions, which involve multiple activators and inhibitors in more realistic way.

7.1.4.1 Rationale of Extended Quaternary States Update Scheme

It is known that the biochemical reactions inside the cells are highly intricate in nature and there exist multiple routes for signal propagations, gene regulations, or metabolite conversions. A single pathway species in a biochemical network thus have multiple activators and inhibitors molecules or external factors (stress, radiation, drug molecules), which play important roles to either activate or inhibit the chemical and physical states of that node in the consecutive time-step. The core objective of logical modeling is to observe the dynamical changes of the chemical or physical states of all the nodes in biochemical network, based on the logical states of its upstream activators and inhibitors at the previous time-step and subsequently compute the equilibrium state of the overall system. We have observed that during the logical update of the quaternary states of a node, sometimes there may exist a strong canalization effect of its single or multiple upstream activator or inhibitor node(s), which can always drive the quaternary state of that node towards a biologically, chemically or physically inappropriate state. As a result, the system may always trap into a specific phenotypic state even there are multiple variations existed in the logical states of the input nodes. Following are the examples of few biochemical reaction events the quaternary algebra defined in the truth table fails to model the reaction propagation dynamics in the downstream of reaction networks.

Let us consider a set of activators $\{A_1, A_2, A_3, A_4, ..., A_m\}$ and inhibitors $\{I_1, I_2, I_3, I_4, ..., I_n\}$ are regulating a node ***D***. Let us also assume that the activators are cooperative in nature, which can be modeled using conjunctive rule or AND operator as follows.

$$D(t+1) = A_1 \wedge A_2 \wedge A_3 \wedge A_4 \wedge !I_1 \wedge !I_2 \wedge !I_3$$
$$D(t+1) = A_1 \wedge A_2 \wedge A_3 \wedge A_4 \wedge I_1' \wedge I_2' \wedge I_3' \quad Where, I' = I$$

(Eq. 34)

Example 1: If $I_1' = 0, A_1 = A_2 = A_3 = A_4 = I_2' = I_3' = 2/3$, then the output of ***D*** from the logical equation ***(Eq. 34)*** is the quaternary state **0**. Let us also assume that I_1 and I_2 are the co-repressor proteins which act with the transcription repressor I_3. Hence, the individual expression of either I_1 or I_2 have no role in the regulation of ***D***, if I_3 is absent in the reaction. But, the logical output of ***D*** in the quaternary logic simulation is **0**, if we use the general rule of AND gate operation, which is clearly not capturing the real biological scenario. In such scenario, we should consider the quaternary states of other activators (A_1, A_2, A_3, and A_4) and inhibitor molecules (I_2 and I_3), and can consider the particular quaternary state at which maximal number of upstream nodes are present in the previous time-step.

Example 2: If $A_1 = A_2 = A_3 = 3$ & $A_4 = 2$ & $I_1' = I_2' = I_3' = 3$, then the output of ***D*** from the logical equation ***(Eq. 34)*** is the quaternary state **2**. In this case, it is seen that the there are total three activators (A_1, A_2, and A_3) of ***D*** which are at their maximal activation state (**3**) and all the inhibitors (I_1, I_2, and I_3) are at their lowest activation state or absent (**0**). Despite this the logical output or the quaternary state of ***D*** is not reaching at the maximal activation state (**3**). The activation state of ***D*** is not reaching at the maximal activation state because of the quaternary state of another activator A_4 in the equation, which has the quaternary state **2**. For example, if we consider the activator node A_4 as the transcription co-activator molecule of the main activator, which can cooperatively binds with another transcription factor and activates the transcription

of a gene (say ***D***). In such scenarios, it is clear that although A_4 is an important factor (activator) of the node ***D***, but it cannot solely bring down the state of ***D*** in the chemical reaction, especially when other activators of ***D*** are present at the maximal active quaternary state (**3**). Herein, we propose to use the extended rule of quaternary state update scheme (ExQSUS) instead of the general rule of conjunction defined in the truth table (see **Section 2.6.3**).

Example 3: If $A_1 = A_2 = 2$ & $A_3 = A_4 = I_1' = I_2' = I_3' = 3$, then the output of ***D*** from the logical equation ***(Eq. 34)*** is the quaternary state **2** as per the rule mentioned for AND logic in quaternary truth table. However, it can be seen that in this equation, all the inhibitors are at lowest inactive state or absent (**0**), and there are two activators A_3 and A_4 at maximal active state (**3**). In terms of biochemical reaction, there is a possibility that the output of such reaction will lead to the maximal activation of ***D***, even though the state of A_1 and A_2 are at regular activation state (**2**). In such scenario, we can also apply our ExQSUS algorithm and choose the state at which most of the upstream activators and inhibitors are present at the previous time-step (see **Section 2.6.3**).

Example 4: If $A_1 = A_2 = 2$ & $A_3 = A_4 = 3$ & $I_1' = I_2' = I_3' = 0$, then the output of ***D*** from the logical equation ***(Eq. 34)*** is the quaternary state **0** as per the rule mentioned for AND logic in quaternary truth table. Here, it can be seen that all the three inhibitor molecules are at active state and out of four activator molecules, there are two activators which are at normal active states (**2**) and other two are at highest active state (**3**). Clearly the dominating factors in this scenario are the inhibitor molecules, which will play most influential role to determine the state transition of ***D*** at the next time-step. Hence, the quaternary state of ***D*** at the consecutive step will be **0** as per proposed algorithm designed for the extended quaternary state update scheme (ExQSUS). It should be noted that in few cases, the rule proposed by ExQSUS may give exactly the same output as it is supposed to compute through general

quaternary truth table.

Apart from these examples, depending on the distribution of the quaternary states of upstream activator and inhibitor nodes in a logical equation, numerous other situations may also appear during the signal propagation or gene regulation. In all such scenarios, our newly proposed EXQSUS algorithm at first tries to compute the quaternary state at which most of the upstream nodes are present at the previous time-step (for more details see **Section 2.6.3**). The most dominating state at which most of the upstream molecules are present is always chosen as the output state of the downstream node at the subsequent time-step. If same numbers of upstream molecules are found to be present at multiple states, then the ExQSUS algorithm will realize the output from the general quaternary truth table. It should be noted that the algorithm is only applied for the nodes which are cooperatively dependent on each other and are modeled via conjunction rule (AND operator). The upstream nodes acting independently and modeled by disjunction rule (OR operator) follow the output of general quaternary truth table of OR gate.

7.1.5 Hypothesis and Objectives

In this work, we have hypothesized that the use of quaternary logic-based operations can be successfully implemented to test our different conjectures of large scale biochemical pathways and its relationships with different human diseases mainly oncogenesis, which we could not explained via conventional logic-based operations. Herein, we aim to develop a general computational framework, which will be useful for automatic update of quaternary states of the nodes with respect to time without using any constrained-based approach or pre-defined thresholds during execution. We have divided the entire workflow of the current work in three modules. In the first module, we aim to validate our proposed methodology in the existing logical models of biochemical pathways and seek the answers of the unexplained phenomena which were remained unanswered due to the limitations of

the conventional logical model. The next module is directed towards the application of quaternary logic to examine and compare the potential of our quaternary state selection algorithm with the conventional binary logic-based model simulations while predicting the steady-states of the genes/proteins in diseased (Glioblastoma) cells. The overall objectives of performing these two modules are to validate and measure the potential of our quaternary logic-based operation in the simulation of biochemical pathways and the accuracies of the predictions of genes/proteins expression patterns in biological samples. In the last or third module, we implement our quaternary logic-based operations to model a completely new pathway model consists of multiple signaling pathways found to be responsible in the development of Glioblastoma cells. Hence, the overall approaches of the first two modules of this work are retrospective analyses and validations of the developed models using quaternary logic-based operation, whereas the last module is performed for prospective analyses of integrated pathway model to predict the target gene expression patterns of all the pathways in different GBM tumor cells. Subsequently, the integrated model with quaternary logic-based simulation technique is used for our new hypotheses testing and tasks for the validation of previously identified or novel drug-target identification process of GBM tumor.

7.2 RESULTS

7.2.1 Case Studies of the Previously Published Logical Models

7.2.1.1 Activation of JNK Protein in CD28 Stimulated T-Cells

In the previous experiments, it was observed that the stimulation of only CD28 co-receptor protein in T-cells could not activate JNK protein at higher the expression level or sometimes it can induce a very weak activation of JNK [577-579]. A seminal paper published by *Rodriguez et al*. had tried to simulate this molecular event by developing an integrated logical model of T-cell receptor and co-receptors mediated intracellular signaling pathways [244]. The authors had considered this molecular event in T-cells as a binary response, in which CD28 stimulation alone induced the

JNK activation (i.e., binary state 1 or UP) and vice-versa in the developed *in-silico* model. They had also performed an *in-vitro* experiment on mouse primary T-cells to observe the effect of CD28 stimulation on JNK and found that stimulation of only CD28 protein by non-superagonistic CD28 antibody induced a sustained expression of JNK protein. Thus, verified the binary response phenomena of JNK activation by CD28 stimulation with the new *in-vitro* experiment. Although their simulation result perfectly matched with their newly performed experimental evidence, the author mentioned that the discrepancy of the JNK activation level in comparison to the literature could be due to the different cellular systems (e.g., primary T-cells or T-cell lines) used in the previous [577-579] and current experiments [244].

Hence, considering all the previous results, we can argue that CD28 stimulation does not trigger the activation of JNK or weakly activates its expressions or in some cellular system, it can trigger the sustained expression of JNK. The binary assumption (either ON or OFF) of the logical states of the pathway species (nodes) is not sufficient to simulate all of these three possible outcomes of JNK activation level. We have hypothesized that the strength of CD28 stimulation and the initial states of the pathway species of T-cell signaling network are the main factors behind these three possible (i.e., inactive, weakly active, and sustained active) states or outcomes of JNK protein. Here, we propose to implement our modified quaternary logic based method and algorithm to simulate the previous model of T-cell signaling pathway developed by *Rodriguez et al*. [244]. In this logical model, we have only changed the mathematical function of those nodes, which have only inhibitor proteins in their upstream region (see **Section 2.6.2**). Apart from these nodes, we have not altered the mathematical equations of any other nodes in the select model. However, before trying to prove our proposed hypothesis, we have attempted to reproduce the simulation results obtained by *Rodriguez et al*. in the Boolean logic based simulation of T-cell network [244]. The authors had simulated the logic model in three different time scales (t_0, t_1, t_2), in which t_2 time scale contains few feedback reactions. The first

time scale t_0 refers the simulation outcomes of the logical model without any stimulation of the activators/receptor proteins CD28, CD4, and TCRLIG. In the t_1 and t_2 timescales, the logical steady states of t_0 and t_1 timescales are considered as the inputs, respectively. In this work, at first, we have simulated the t_0 timescale by considering the initial states of the input nodes cd28, cd4, and tcrlig as "0" or "Absent". To simulate the next timescale t_1, we have considered the initial states of each input nodes cd28, cd4, and tcrlig at three different expression levels, *viz.* "1" or "low expression"; "2" or "normal expression"; and "3" or "over-expression". Here, the low expression (i.e., quaternary state = 1) of the input nodes refers to mild stimulation, whereas the normal and overexpressions states are referred to as the normal and strong stimulations, respectively.

We have followed the similar simulation steps executed by *Rodriguez et al*. to measure the expression profiles of the read-out proteins, *viz.* pag, tcrp, cblb, ccblp1, lckp1, zap70, lat, pkb, and erk at t_0, t_1, and t_2 timescales by stimulating the input nodes cd28, cd4, and tcrlig [244]. We have successfully reproduced the similar expression profiles of these proteins (except cblb and pkb) under strong stimulation of cd28, cd4, and tcrlig input proteins (**Figure 49**). Hence, it is proven that our newly developed quaternary logic method is capable of reproducing (qualitatively) the previous outcomes of Boolean logic based T-cell pathway model. Additionally, we have observed that under the mild and normal stimulations of all the three input proteins, the expressions profiles of the select proteins have shown different expression patterns. The varied responses of the read-out proteins under different initial conditions, which could not be retrieved in the previous model due to the limitations of considering only the binary states of the nodes, are now can be observed by implementing our newly developed quaternary logic based analyses. Hence, altogether it can be concluded that the application of quaternary logic based method can be best suitable for studying the effects of the strength of stimulation of the receptor proteins by the ligand molecules as well as assessing the magnitude of

the activation/inactivation of the input proteins in signal transduction pathway. The current approach can be reduced to simulate only the binary responses or multi-valued responses depending on the objective of the work.

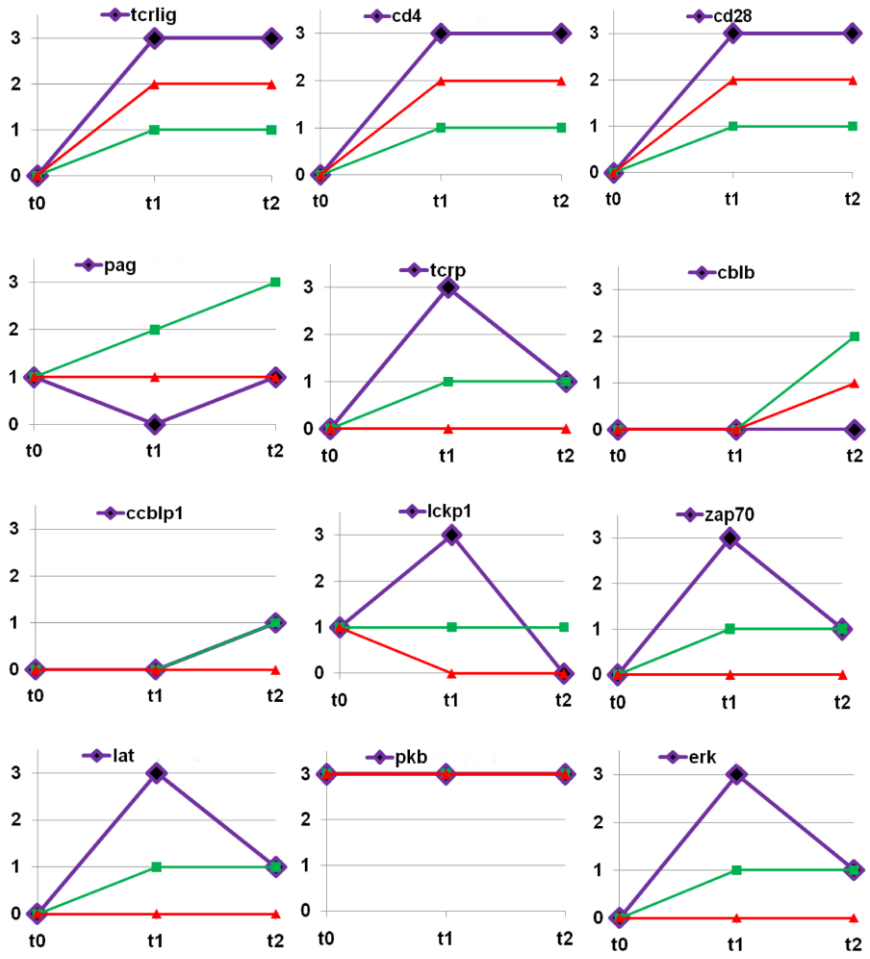

Figure 49: Activity profiles of the select proteins of T-cell pathway model.
At timescales, t_1 and t_2, the quaternary states of the input nodes cd4, cd28, and tcrlig are considered as "**1**", "**2**", and "**3**" to simulate the effect of mild, normal, and hyperstimulation of these input proteins in the select downstream proteins of T-cell signaling network. The observed activity profiles of the select downstream proteins, simulated with our newly developed quaternary logic based method, show the quaternary states ("**0**", "**1**", "**2**", and "**3**") of the select proteins. Here, the quaternary state "0" refers the absence or null expression of the proteins, whereas "**1**", "**2**", and "**3**" refers the low expression, normal or basal, and over expressions of the proteins, respectively. The activity profiles of the select proteins under hyper-stimulation of input proteins qualitatively match (except cblb and pkb proteins) with the activity profiles observed by Rodriguez et al. in their developed Boolean logic-based model of T-cell signaling pathway [244].

Next, we have tried to prove our hypothesis by simulating the dichotomous or binary response pattern of JNK activation upon stimulation of CD28 protein,

observed by *Rodriguez et al.* [244] as well as the multi-level responses (null expression, weak activation, and sustained expressions) of JNK protein by CD28 stimulation observed in all experiments (including Rodriguez et al.) [577-579].

We have traced back the JNK activation pathway in the model diagram and found that a node named as rac1r is situated at the upstream of JNK and downstream of CD28. The pathway species (node) rac1r was considered as the reservoir node of rac1 and rac2 proteins. Here, we have assumed that the initial concentration of rac1r species in the T-cell pathway is heterogeneous in different cellular systems (primary T-cells and T-cell lines), which can regulate the inital states of rac1 and rac2 proteins and finally the activation of JNK. We have observed that the effect of varied strengths of stimulation and its consequence in the expressions of JNK protein can be only observed if we consider different initial expressions (quaternary states) of rac1r in the logical model. We have found that the cellular system with null expression of rac1r shows no activation of JNK protein under all three possible stimulations of CD28. The similar observation was also found in the previous experiment (**Figure 50A**) [577,578]. In case of a cellular system with a shallow concentration level of RAC1R protein (quaternary state = 1), we have observed a very weak activation of JNK (quaternary state = 1) under CD28 stimulation (**Figure 50B**). The weak JNK activation state upon CD28 stimulation was also observed in the previous experiment [579]. Simultaneously, we have also observed the binary response pattern of the JNK activation level upon CD28 stimulation in the cellular system having normal or basal expression level of RAC1R protein (**Figure 50C**). The sustained expression pattern of JNK protein was observed by Rodriguez et al. in their experiment on primary mouse T-cells after stimulating CD28 co-receptor with non-superagonistic antibody [244]. Apart from these observations, we have also predicted the graded, multi-valued response patterns of JNK activation levels under the gradually increasing strength of CD28 stimulation in the cell types having overexpressed RAC1R (**Figure 50D**). Hence, we can conclude

that in previous experiments the initial concentrations of RAC1 and RAC2 proteins before CD28 stimulation may have played significant roles to trigger multi-valued expression patterns of JNK protein in the T-cells. Simulation of such multi-valued outcomes of the target genes/proteins in large-scale signaling pathway is difficult to perform by Boolean modeling approach, but it can be much easier with the help of our newly developed quaternary logic-based approach.

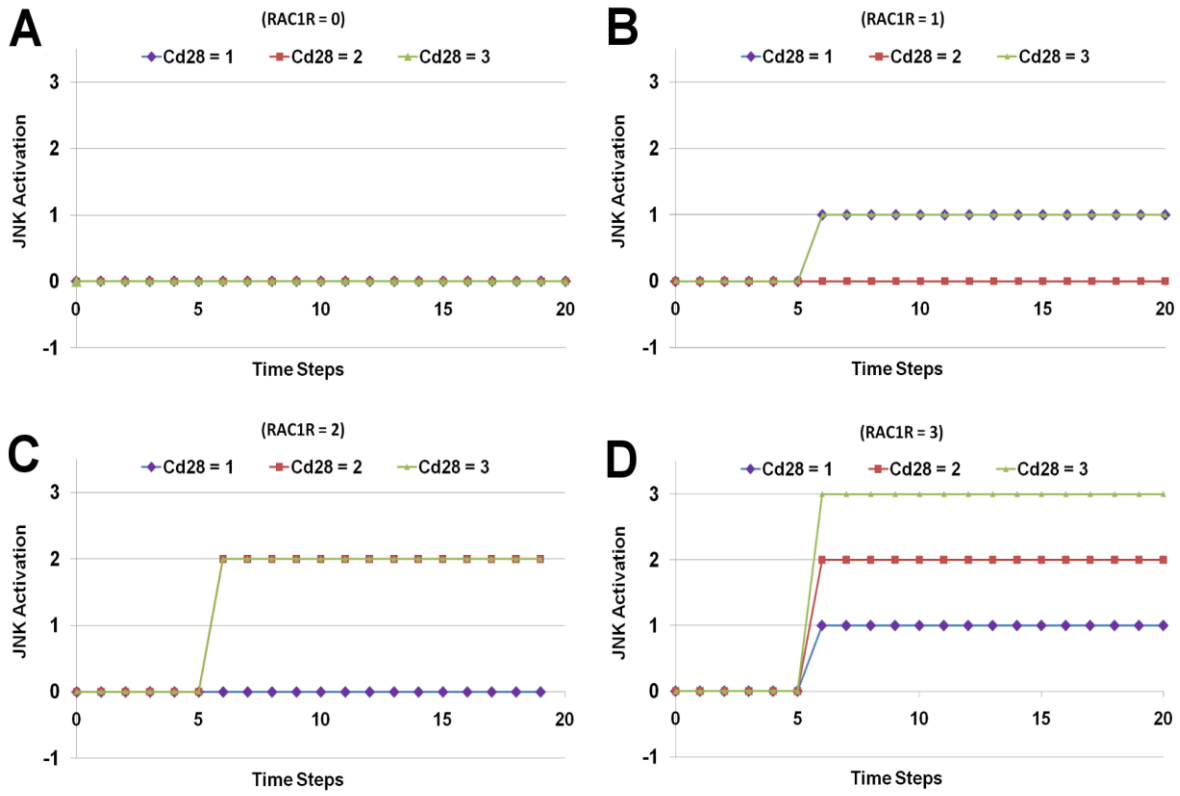

Figure 50: JNK activation profile in the different cellular systems.
Total four different cellular systems **(A)** null, **(B)** low, **(C)** normal or basal level and **(D)** over expressions of RAC1R are considered in our simulation to assess the effects of CD28 stimulation in JNK activation level. The JNK activation profile is simulated in these cellular systems under gradually increasing strengths of only CD28 stimulation. The quaternary states of CD28 equal to 1, 2, and 3 refer to the mild, normal, and hyper stimulations (*in-silico*) of CD28 co-receptor protein of T-cells, respectively.

7.2.1.2 Effects of Copy Number Gain or Gene Amplification in Signaling Pathways

Apart from the effects of varying signal transduction intensities exerted by ligand molecules on the receptor proteins, genetic aberrations *viz.* gain or loss of copy

numbers of the vital signal transducing genes are also found responsible for manipulating the expressions of target genes/proteins during disease pathogenesis. For example, epidermal growth factor receptor or EGFR (also known as ERBB1 or HER1) is regularly expressed in the membrane of epithelial, mesenchymal, and neuronal cells, which most often binds with the growth factor ligands (EGF, TGF, etc.) and triggers the downstream kinase signaling pathways for normal mitogenic (e.g., G1/S transition) events [580]. One of the target proteins at the downstream of EGFR pathway is the tumor suppressor protein retinoblastoma (pRB). In the early G1 stage of cell cycle, the hypo-phosphorylated pRB binds with E2F family transcription factors and represses the transcription of the genes (e.g., Cyclin A and E, CDC2, CDC25C, p21, etc.) responsible for cell division [581]. In order to accelerate the cell cycle progression, the mitogenic signaling cascades mediated by growth factor receptor proteins, such as EGFR are specifically directed towards the phosphorylation of pRB in the late G1 phase and subsequently inhibit the interaction between pRB and E2F [245,581]. Hyper-phosphorylation of pRB is found to be strongly correlated with the irregular cell divisions and the pathogenesis of various types of cancers [582]. Similarly, hypo-phosphorylation of pRB is associated with the G1 cell cycle arrest and the inhibition of cellular growth [583]. Hence, the modest expression of the phosphorylated pRB protein after EGFR pathway activation in the late G1 phase followed by its hypo-phosphorylation is always desirable to maintain the cell cycle progression dynamics at a regular rhythm [581]. The EGFR receptors family mediated pathway play important role in the maintenance of cell cycle progression and cell proliferation pattern by tightly regulating the phosphorylation of pRB and other CDKs. Therefore, the imbalances caused due to the irregular expressions of the ligands, such as EGF, over-expression or amplification and mutation of the receptors in the membrane are found to be strongly associated with the hyper-phosphorylation (or inactivation) of pRB, irregular cell divisions and oncogenesis [584].

Amplification of the *EGFR* gene with higher expression of EGFR protein is detected in various cancer cells including primary Glioblastoma (57.4% of GBM patients) [556]. Different types of EGFR variants in GBM cells, such as *EGFR* deletion (EGFRvIII, deletion of exons 2–7) or point mutations (R108K, A289V/D/T, G598D) are also co-expressed with the amplified $EGFR^{WT}$ in the cell membrane [556]. Also, the previous studies have found that these variants are present in the GBM cells at constitutively active state and trigger the aberrant mitogenic signal to the intra-cellular kinase proteins, such as CDKs [556]. The constant activations of the kinase proteins augment the phosphorylation of pRB. Therefore, it can be said that the activation levels of various proteins (e.g., AKT1, MEK1) or phosphorylation (inactivation) of pRB in the downstream of EGFR pathway will not be in equal magnitudes in the cells expressing only $EGFR^{WT}$ with modest expression level and the cells which express $EGFR^{WT}$ with amplification or higher copy number, or the cells which have both the amplified $EGFR^{WT}$ and EGFRvIII variants co-expressed [245] (**Figure 51A**). As a result, the expressions of cell cycle regulatory genes in these cell types will be highly varied and thus the rhythm of cell division process. The comparative analyses of such downstream proteins (e.g., CDK2, CDK4, CDK6, CYCLIN D1, and CYCLIN E1, etc.) after EGFR pathway activation in different cells expressing the variants of EGFR receptor will be helpful to understand the regulatory mechanisms behind the irregular cell divisions and tumor formation. Previous dynamic logical models (based on binary Boolean algebra) of ERBB pathway have successfully simulated the dynamics of the activation pattern of cell cycle regulatory genes/proteins during cell division process under various pathological conditions [585-587]. However, a general logical model developed for the simulations of all such molecular variants of EGFR protein simultaneously and subsequently the comparison of the expressions of downstream genes/proteins of this pathway in different cellular systems is hitherto unavailable. Using conventional binary logic-based approach, a comparative study of the different cellular systems with regular, amplified, and variant EGFR in the cell membrane is not possible to

perform due to the limitation of considering the initial logical state of the EGFR protein in all the cellular systems at only the up-regulated (or Active/True) state. On the other hand, the quaternary logic-base approach can overcome this limitation by considering the initial state of the regularly expressed $EGFR^{WT}$ as "2" and the amplified or copy number gain as quaternary state "3" (without EGFRvIII expressed).

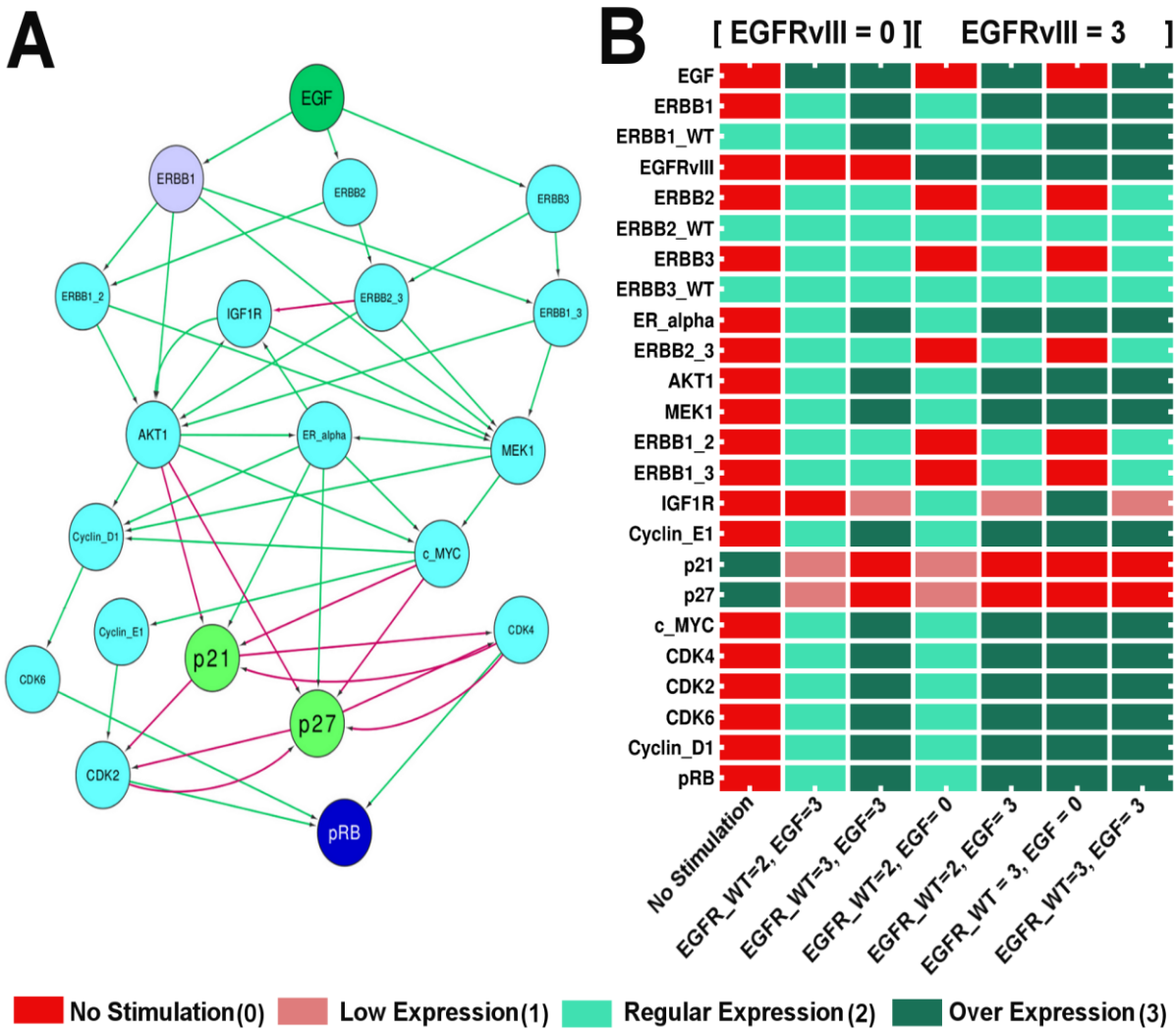

Figure 51: Simulation output of ERBB family receptor proteins mediated pathway in Glioblastoma cells.

(A) Network diagram of ERBB pathway. The green and red arrows in the network are activation and inhibition, respectively. **(B)** Predicted quaternary states of the pathway species in various experimental situations. Two types of EGFR (or ERBB1) *viz.* EGFR_WT (wild type) and EGFRvIII (deletion of exon of 2-7) are considered in the simulations to create different cell types. EGFR_WT = 2 and EGFR_WT = 3 represent the regular and amplified EGFR expressions, respectively. Here, the logical state of pRB represents the phosphorylation status of the protein.

We can also create four different possible scenarios of the cellular systems which co-express $EGFR^{WT}$ and EGFRvIII *viz.* (i) regular $EGFR^{WT}$ expression (= 2) and no stimulation (EGF = 0), (ii) regular $EGFR^{WT}$ expression (= 2) with stimulation (EGF = 3), (iii) $EGFR^{WT}$ amplification (= 3) and no stimulation (EGF = 0), and (iv) $EGFR^{WT}$ amplification (= 3) and with stimulation (EGF = 3). In order to analyze the effects of these quaternary states of EGFR protein in all the cellular systems, we have used a previously published logical model of ERBB pathway [245] and have modified the formulation of the logical equations (**Appendix Table 11**) by following our proposed quaternary logical equation formulation rule (see **See 2.6.2**).

We have observed that when there is no external stimulation and no EGFRvIII variant is present in the cellular system, our model simulation perfectly predicts the expressions or activities of the pathway species at the logical steady states (**Figure 51B**). The downstream proteins of ERBB pathway responsible for cell cycle progression (e.g., phosphorylated pRB, Cyclin D1, Cyclin E1) are found inactive in this scenario. The expressions of these proteins are increased and reached at the regular expression level (= 2) when EGF (= 3) is present in the extracellular region as ligand or stimulant for the wild-type receptors *viz.* ERBB1 ($EGFR^{WT}$), ERBB2, and ERBB3. The regular or normal expression patterns achieved in our simulation can be represented as the normal functioning of the cells. On the other hand, we have found that the downstream proteins are over-expressed (as compared to normal cells) when $EGFR^{WT}$ (ERBB1) is amplified in the cells and received appropriate signal from EGF ligand. The over-expression of the downstream proteins, such as phosphorylated pRB, is observed in the GBM cells having amplified or high copy number of EGFR (ERBB1) protein [245]. Over-expressions of phosphorylated form of pRB, and Cyclin D1 are linked with rapid cell cycle progressions and malfunctions in the normal development of the cells, and thus cause cancer or tumor [245,588]. Hence, we can assert that our developed quaternary logic-based model is able to simulate both the normal (with or without stimulation) and cancer (high copy gain of EGFR) scenarios

with correct expression levels of the downstream proteins.

Simulation of the cellular system co-expressing amplified EFGR and EGFRvIII variant can be also performed to compare the protein activity profiles of ERBB pathway with various other cellular systems. We have used our modified, quaternary logic-based model (**Appendix Table 11**) in four different cellular systems consist of EGFRvIII variant and computed the activity profiles of the downstream proteins of ERBB pathway (**Figure 51B**). We have observed that EGFRvIII variant cells without any extracellular stimulation and co-expressing with regular $EGFR^{WT}$ transduces weaker activation signal to the downstream proteins as compared to the amplified EGFR and/or EGF stimulated cellular systems. This result validates the previous experimental findings in which it was shown that the constitutively active EGFRvIII variant has weaker kinase activity in the downstream proteins of ERBBB pathway than the EGFR, but it is reported to be sufficient for cell cycle progression in the absence of extracellular stimulation in GBM and other tumorigenic cells [556,584]. The augmentation in the signal intensity in the downstream proteins was observed when EGFRvIII variant is co-expressed with the amplified EGFR protein in the GBM cells [556]. We have also observed the similar expression profiles of the downstream proteins in our simulation studies, which proves that the EGFRvIII variant promotes the oncogenic transformation of the cells by producing dimer with EGFR protein (**Figure 51B**). Hence, it can be seen that all the cellular situations consisting of the molecular variants of EGFR protein can be possible to simulate with the help of our newly developed quaternary logic-based modeling approach, and can better explain the activity profiles of pathway species as compared to the binary logic-based approach.

7.2.1.3 A Comparative Study of the Previous Models of Hedgehog and Notch Pathways

We have also used our newly developed quaternary logic-based modeling technique to simulate the previously developed logic-based models of Hedgehog and

Notch pathways [185]. In our previous models of Hedgehog and Notch pathways, we tried to predict the logical steady-states of the target genes/proteins of these pathways in grade-IV GBM cells and matched the expression profiles with the microarray data (GSE42902) [219]. To simulate the models, we considered the initial states (binary 1 and 0) of all the input genes/proteins of the pathways based on the expression profile obtained from the microarray data. While simulating the GBM cellular system, we considered the initial states of the input proteins/genes in the logical models as "1" or "True" if the corresponding transcripts are differentially over-expressed (significant or adjusted P-Value ≤ a defined cut-off) in the GBM cells as compared to normal cells. Similarly, the initial states were considered as "0" or "False" if the transcripts of the input genes/proteins were differentially down-regulated (significant) in the GBM cells. We used this data as inputs of our Hedgehog and Notch pathway models, and achieved 75% and 84% accuracies, respectively while we compared the activity profiles of all proteins obtained from the simulations of the pathway models using Boolean logic with the expression profiles of the corresponding transcripts obtained from microarray data.

However, we could not determine the binary initial states of a few input proteins/genes in both the model simulations as the differentiation expressions of the respective transcripts of those genes/proteins in the GBM cells were not statistically significant [185]. We have observed that the transcripts whose expressions were insignificant (adjusted P-Value > a user defined cut-off), sometimes had fold change value more or less than ± 2 (i.e. log2 [FC] ≥ 1 or log2 [FC] ≤ -1) in the GBM cells as compared to the normal cells. On the other hand, few transcripts had significant expression level (adjusted P-Value ≤ a defined cut-off) in the GBM cells, but had fold change ratio lies between -2 to +2. Such scenarios signify that these transcripts are either modestly active (but not hyperactive) or hypoactive (but not absent) in the GBM cells as compare to the normal cells. Indeed, such expression scenarios we cannot capture in Boolean logic-based model, in which we can either use binary 1

(active or over-expressed) or 0 (inactive or under-expressed) states to represent the logical states of the genes/proteins. Hence, we need a better approach to perfectly quantize (multiple states) the logical states of the nodes by using the statistics we obtain from the differential expression analyses. Our proposed technique, to select multi-valued, quaternary logical states of the nodes can be specifically useful for this purpose (see **Section 2.6.4**).

Here, we have used the quaternary state selection criteria on GSE4290 microarray datasets to select the quaternary logical states of the proteins considered in Hedgehog and Notch pathway models for simulating the primary GBM scenario. We have chosen the adjusted P-value cut-off (i.e. θ_1) at 0.05 and log2 [FC] (i.e. θ_2) equals to ±1 for determining the quaternary states of all the transcripts to implement the quaternary state selection rules. We have also modified the logical equations of both the models as per our proposed rules for the constructions of logical equations and subsequently computed the logical-steady states of the models using our modified quaternary logic (ExQuLogic)-based simulation rules. We have observed that both the models have reached the fixed-point attractor states during simulations. Followed this, we have extracted the quaternary logical states of all the genes/proteins of the models at the steady-states and compared with the logical states obtained just after applying quaternary selection rule. We have achieved 77.55% and 87.65% accuracies from both the model simulations, respectively. Although, the accuracy levels have not improved significantly in Hedgehog pathway model, but the prediction of the expression levels of the statistically insignificant genes/proteins, which we could not explicitly computed in the binary logic based simulation, can be now possible to predict by using our ExQuLogic-based method.

In case of binary logic (BL)-based simulation of Hedgehog (HH) pathway model, we observed that the important activator proteins of this pathway, such as CDO, BOC, DISPATCHED, etc. had insignificant expression level in GBM cells as compared to the normal cells. However, to activate the HH pathway, we know that

these proteins play significant role to pre-process the HH pathway ligands *viz.* Sonic (SHH), Desert (DHH), and Indian (IHH) Hedgehog. Due to the limitation of considering only two possible states either up-regulation (or True) or down-regulation (or False) in BL-based simulation, these proteins were required to be considered as up-regulated or over-expressed at the initial time point of the simulation to activate the HH pathway in the *in-silico* model of Glioblastoma tumorigenic scenario (GBS). On the other hand, after applying the quaternary state selection rules on the transcripts of these proteins of the same microarray dataset of GBM tumor cells (GEO ID: GSE4290), we have seen that these proteins are modestly active (but not hyper-active or over-expressed) but insignificantly expressed in GBM cells (**Figure 52**). Hence, in our ExQuLogic-based simulation of GBM cells, we do not need to consider the initial states of these insignificantly expressed proteins as up-regulated (or TRUE) state. Instead, we can consider the activity of these proteins as modestly active (quaternary state = 2) at the initial state of the simulation (**Figure 52**). Also, On the other hand, the biological expressions of the co-repressor proteins of HH pathway, such as SKI, SNO, SIN3α, and NCOR found insignificant in the same expression dataset of GBM cells (GSE4290), but they were also considered as inactive/down-regulated to activate the HH pathway during the BL-based simulation [82]. After applying our quaternary state selection rules on the same datasets, we have observed that SNO and SIN3α are modestly active (quaternary state = 2), and SKI and NCOR are hypo-active (quaternary state = 1) in the dataset. The simulation outputs of our modified quaternary logic (ExQuLogic)-based simulation method on HH pathway depict that the expressions of the target oncoproteins (e.g., FOXM1 [589], PDGFRα [590]) of HH pathway will be still hyper-active (over-expressed) at the steady state (**Figure 52**). This result clearly shows that the modest activation of CDO, BOC, DISPATCHED, SNO, and SIN3α and hypo-activation of SKI and NCOR proteins of HH signaling network are sufficient to trigger the over-activation of tumorigenic proteins.

However, it should be noted that the predicted quaternary states of HH pathway activators *viz.* CDO, BOC, and DISPATCHED and the co-repressor proteins *viz.* SKI, SNO, NCOR, and SIN3α are sufficient but not necessary for promoting the expressions of the target oncoproteins.

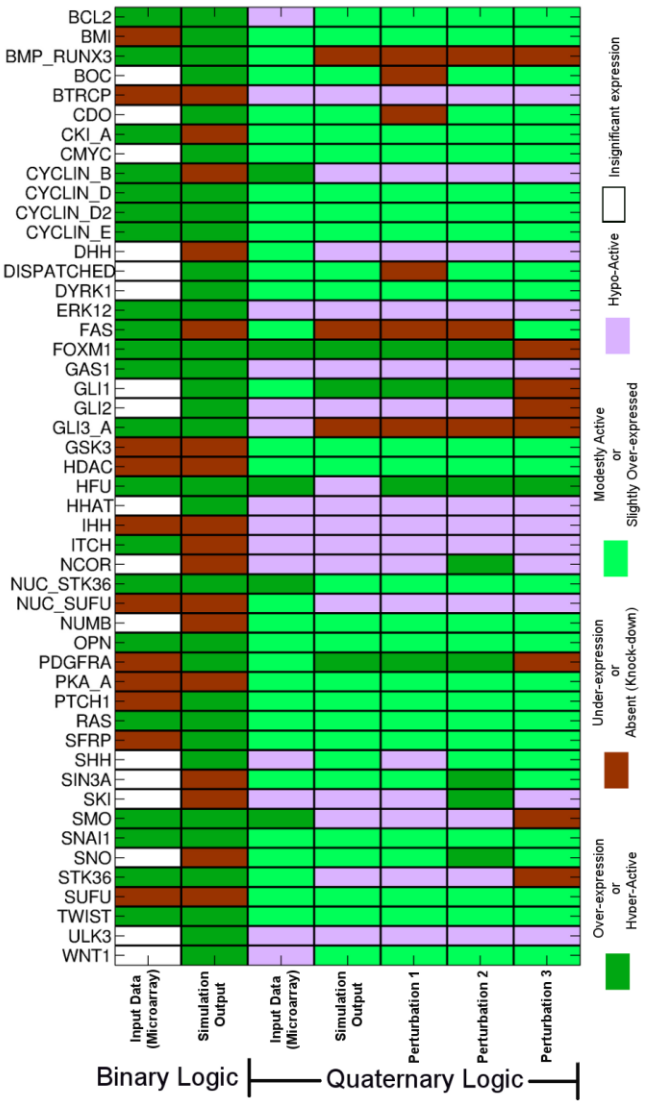

Figure 52: Logical steady-states of the proteins in Hedgehog pathway model.
The logical states of the proteins shown under binary logic-based simulation (first two columns) are obtained from the previously performed logical-steady states analyses of Hedgehog pathway model. The transcripts of the Hedgehog pathway proteins are quantized using the "quaternary states selection rules" and the logical steady-states of these proteins are computed by using the modified quaternary logic (ExQuLogic)-based simulation technique. Perturbation 1 refers the *in-silico* knock-out of CDO, BOC, DISPATCHED and Perturbation 2 indicates the over-expression of SKI, SNO, SNI3A, and NCOR proteins in the GBM cells. Perturbation 3 shows the simulation outputs of the knock-out experiments on SMO, GLI1, and GLI2 proteins in the GBM cells as the proposed, combinatorial drug target.

Perturbations studies of these two groups of proteins (activators and co-repressors) separately (perturbation 1 and 2) have proven that the over-expressions of the target oncoproteins (e.g., FOXM1, and PDGFRα) in GBM scenario do not get affected even after perturbing the quaternary states of these proteins (**Figure 52**). Hence, it can be said that the suppression of CDO, BOC, and DISPATCHED (positive effectors) or over-activation of the co-repressors proteins SKI, SNO, NCOR, and SIN3α have no potential to reduce the activities of the oncogenic signaling network in GBM cells. Indeed, certain cross-talks from other pathway molecules and/or the positive feedback loops can still keep the signal at ON state even after suppressing its activators or over-activating its suppressor proteins. Similar to the previous binary logic-based, perturbation analysis of the activated Hedgehog signaling network on GBM cells have revealed that by targeting SMO, GLI1, and GLI2 proteins simultaneously, down-regulation of the target oncoproteins FOXM1 and PDGFRα is possible to accomplish [82].

In addition, we have also predicted that perturbation of SMO, GLI1, and GLI2 proteins will not completely down-regulate the cell cycle regulatory proteins Cyclin D, Cylcin D2, and Cyclin E, which will in turn reduce the chance of severe toxicity induce on the normal cells during target-based therapy (**Figure 52**). Therefore, we have established that using our new quaternary logic-based simulation of Hedgehog pathway, we can now test the precision of the proposed drug-target in the GBM or other tumorigenic cells, which was not made possible in the binary logic-based simulation.

7.2.2 Application of ExQuLogic-based Simulation Techniques on Other Datasets

We have shown that by taking the inputs from the microarray gene expression datasets (GSE4290) of GBM tumor cells, how our newly developed ExQuLogic-based simulation technique is helpful to predict the expression profiles of each of the genes/proteins involved in Hedgehog pathway with much higher level of resolution

than the BL-based method [82] (**Table 16**). On the other hand, in case of Notch pathway model, the ExQuLogic-based logical simulation have shown a slightly higher level of accuracy (91.35%) than the accuracy level (84%) obtained in BL-based method [185] (**Table 16**).

Table 16: Percentage of Accuracy Obtained in BL- and ExQuLogic methods

Models / Datasets	Hedgehog Pathway		Notch Pathway	
	Binary Logic (BL)	Extended Quaternary Logic (ExQuLogic)	Binary Logic (BL)	Extended Quaternary Logic (ExQuLogic)
Glioblastoma tumor cells vs. Normal cells (GSE50161)	**80.77%*** **69.23%****	**81.25%****	**85.71%*** **97.62%****	**91.36%****
NICD over-expressed GBM cells vs. Neural stem cells (GSE44561)	**NA**	**75%**	**NA**	**89.87%**

*** FDR = 0.05 & log2 [FC] = 0; ** FDR = 0.05 & log2 [FC] = 1**

However, in both the analyses, we have used same gene expression datasets obtained from GBM tumor cells. Hence, to examine the efficiency of ExQuLogic-based method for the prediction of the gene expression or protein activity profiles, we have also used other experimental datasets (microarray, proteomics) as benchmark for measuring its performance with BL-based method. It should be noted that similar to the previous approach used in BL-based simulation [185], herein the entire analyses are also performed in retrospective mode to compute the percentage of accuracy of the predicted activity profile or expression pattern of the associated genes/proteins of the signaling pathways (see **Section 2.6.7**).

We have observed that in both the pathway models, while predicting the percentage of accuracy of the gene/protein expression patterns with respect to the

experimental datasets (GSE50161 and GSE44561), the ExQuLogic-based approach slightly outperforms the BL-based approach (**Table 16**). We have also observed that the BL-based method outputs different magnitudes of the percentage of accuracies in both the pathway models after applying different level of fold change (log2 [FC]) values as thresholds to determine the up or down regulated transcripts. Indeed the change of fold change values in constant FDR have altered the expression profiles of the transcripts significantly and thus have changed the magnitudes of the percentage of accuracy of all the pathway species in GBM cells.

However, the main advantage of the use of ExQuLogic-based method is not limited to the calculation of the percentage of accuracies of the expressions of genes/proteins in the GBM cells under steady-state. This new method has the capability to consider and correctly compute the expression patterns of all the genes/proteins even though the transcripts are not significantly expressed in the cellular system (i.e., GBM cells) to be considered as the binary up ("1") or down-regulated ("**0**") states. For example, in the GSE44561 microarray assay performed by *Guichet et al.*, the GBM cells were transfected with NICD (Notch Intracellular domain) and the expressions of the transcripts were compared with the normal neural stem cells (with NICD transfection) for further differential expression analysis [591]. Interestingly, the authors observed that even after strong activation of Notch signaling pathway in the GBM cells, the main target genes of this pathway, such as HES5, HEY1, and HEY2 were not significantly expressed or sometimes modestly expressed (FDR > 0.05). Hence, the application of BL-based method to retrospectively predict the gene/proteins expression pattern of the NICD transfected GBM cells is not helpful in this case. After reanalyzing this microarray dataset, we have also observed that most of the genes/proteins involved in Hedgehog and Notch pathways are not significantly expressed in the NICD transfected GBM cells. Hence, we cannot apply a suitable threshold of significance level on the expression patterns of the transcripts to determine the binary states (Up or Down) of the involved genes/proteins considered

in our logical models.

We have overcome this problem by implementing our quaternary state selection rule on the output of the differential expression analyses of the microarray dataset, and thus quantized the expression pattern as per our defined quaternary states $\{0, 1, 2, 3\}$. Subsequently, we have extracted the quaternary states of all the genes/proteins considered in both the pathway models as inputs for further logical steady-state analyses. We have obtained 75% and 89.87% accuracies from Hedgehog and Notch pathway models, respectively by using the ExQuLogic-based method for pathway simulation. This result clearly depicts the prospective potential of the developed ExQuLogic-based method for the prediction and analyses of the gene expression pattern in a cellular system. Also, we have observed that in the Notch pathway model of GBM cells, the logical steady states of the target proteins of the pathway, such as HES5, HEY1, and HEY2 are modestly active (quaternary state = 2) (result not shown). This observation completely agrees with the experimental evidences observed by *Guichet et al.* [591]. Hence, it is observed that the BL-based method, which specifically depends on the significant expression levels of the genes/proteins of the pathway to decide the binary logical states of the inputs, fails to capture the expression pattern of the genes/proteins of the cellular systems (or disease scenario) in which the transcripts/genes/proteins are insignificantly or modestly activated and inhibited.

7.2.3 Activity Profiles of the Genes/Proteins in the Integrated Logical Model of Hedgehog, Notch, WNT, and EGFR Pathways

7.2.3.1 Study of Integrated Pathway Model using ExQuLogic

In order to examine the applicability of our proposed ExQuLogic-based method on another pathway model, we have reconstructed a new logical model of the reaction networks consisting of Hedgehog, Notch, WNT, and EGFR pathways (see **Section 2.6.6**) and subsequently used it for further *in-silico* analyses (**Appendix Table 12**). These pathways and their cross-talks are proven to be significantly associated

with the development and maintenance of the matured and tumor initiating stem cells [314,405,584,592,593]. These pathways are also involved in the angiogenesis, tumor cell invasion, Epithelial-Mesenchymal transitions (EMT) during cancer pathogenesis [594-598]. Hence, a predictive mathematical model is very much useful, which is specifically developed for computing the expressions of genes/proteins associated with these pathways during cancer cell development and maintenance. In this work, we have used our proposed ExQuLogic-based method on the newly developed, dynamic logical model of the integrated pathways and subsequently compared the qualitative expression levels of the genes/proteins observed in the experimental microarray data of GBM tumor cells with the predicted quaternary expression/activity levels of the respective genes/proteins at the steady-state of the dynamics. We have used same logical model, but different initial conditions for the inputs of the integrated pathway. The initial conditions of the input proteins are considered from the three microarray datasets, which we have used earlier for the individual Hedgehog and Notch pathway models of GBM development to computer the logical steady-states (LSS) of the genes/proteins of these pathways.

7.2.3.2 Initial States of the Input Proteins and the Expected Expressions of the Target Genes of the Model are Selected from Microarray Data

In order to select the quaternary states of the input proteins for each model simulation, at first we have used the differential expression statistics of each microarray experiments (i.e., GSE4290, GSE50161, and GSE44561) to extract the log2 [FC] and FDR values of each of the transcripts correspond to the input proteins of the model. After that, we have applied the quaternary state selection rules on the extracted transcripts and assigned the quaternary states to the input proteins. We have also used the differential expression statistics to assign the quaternary states of the output nodes or the target genes of the integrated pathway model. The quaternary states of the target genes obtained from three microarray datasets are the "expected values", which will be compared with the computed quaternary states (i.e.

"predicted values") of the target genes to predict the "level of accuracy" of the corresponding model simulation. During logical simulation of a single microarray dataset, we have shuffled the quaternary states of the inputs and generated random 1000 samples. However, while shuffling the quaternary states of the inputs, we have made sure that the activity status of the input proteins remains conserve as per the quantization level achieved from the differential expression study. That means, if the quaternary state of a transcript is obtained from the differential expression study is "2 (or modestly expressed)", then the corresponding random initial state of the input protein will be either "2 (or modestly active)" or "3 (or highly-active)" during shuffling. On the other hand, if the quaternary state of the transcript is determined as "3 (i.e., Over-expressed)", then random initial state of the corresponding input protein during shuffling will be either "2 (or modestly active)" or "3 (or highly-active)". Similar rule is applied for the quaternary states "1 (or lowly expressed)" or "0 (or absent)" of the transcripts. The quaternary states of the intermediate nodes (e.g. signal transducer proteins, protein complexes, metabolites, ions, etc.) are randomly assigned at the initial time point of the simulation and no activity conservation rule is applied.

7.2.3.3 Percentage of Accuracy of the Model Simulations are Calculated Based on the Predicted Expression Levels of the Target Output Genes

Here, we have assumed that if the transcript of an input protein is found modestly expressed (quaternary state "2"), then the activity of that protein after post-translational modification will be either remained modestly ("2") or highly ("3") active and thus the activity status of the input nodes will be remained conserved as per the observed experimental gene expression pattern. By this way, we have created 1,000 random initial states including the randomized states of the intermediate nodes and successively computed the LSS of the target genes or output nodes of each 1,000 models of a single microarray dataset. Simultaneously, we have computed the active and inactive target genes from the LSS of each model and compared with the

expected activity profile of the target genes obtained from the corresponding microarray data of GBM cells. All the target genes, which have the quaternary the states "2 or 3" and "0 or 1" in the expected and predicted datasets, are hot encoded as "active" and "inactive" states, respectively. The percentage of accuracy of the activity profile of a model is thus measured by using the hot encoded activity profiles of the expected versus predicted dataset of the target genes.

In our integrated pathway model, we have total 158 input, 220 intermediate, and 50 output (or target genes) nodes. The accuracy of the model prediction is performed on these 50 target genes of the developed pathway model. After running the LSS simulation, we have achieved maximum 72%, 68%, and 67.34% level of accuracies in each 1,000 total set of simulations (random initial states) of the logical model developed for all the three experimental microarray datasets *viz.* GSE4290, GSE50161, and GSE44561, respectively. The activity profile of the target genes of the model obtained after taking the initial states of the input proteins from the three microarray datasets are shown in **Figure 53**.

7.2.3.4 Predicted Quaternary States of the Target Genes Depict Better Resolutions during Comparison of the Expected Expression Levels in GBM Cells

We have observed that our developed logical model, based on the input dataset obtained from the GSE 4290 microarray experiment, is correctly able to predict the expressions of various target genes (e.g., *CD44, MYOD, BAX, CYCLIN D2*, etc.) observed in the same assay of GBM cells. Important cell cycle regulatory, apoptosis, and E-to-M transition related genes (also responsible in GBM cell division and metastasis), such as *CD44, FOXM1, SOX9, CYCLIN-D2, CYCLIN-E, MMP7, OPN*, etc. have "active" expressions (either regular- or over-expressed) in the corresponding GBM microarray assay (GSE 4290) as compare to the normal cells (see EXP, **Figure 53A**). In the simulation of GBM model using ExQuLogic method, we have also observed similar expression pattern of these proteins in the computed logical steady states (LSS) (see SIM, **Figure 53A**). Similar results are also observed for the other

assays i.e. GSE 50161 (**Figure 53B**) and GSE 44561 (**Figure 53C**). It should be noted that for each logical models, the quaternary states of the extracted output or target genes are different, which actually indicates the context specific expression patterns of the select target genes in different GBM cells. Hence, it can be said that the simulation outputs of our developed model (ExQuLogic-based) are also able to predict such context specific variations of the gene expression pattern, which are most frequently observed in different GBM tumor cells.

7.2.3.5 Phenotypic response profile of cell cycle progression and EMT show different response patterns in the simulated GBM cell scenarios

Moreover, in the developed logical model, we have also considered few phenotypic responses, such as cell cycle progressions, pro-apoptosis, anti-apoptosis, Epithelial to Mesenchymal transition (EMT), etc. to predict the future effects of the expressions of the target genes in the tumor cell divisions, metastasis, and cell death. The temporal phenotype response profiles of cell cycle progression and E-to-M transition for each cancer model are plotted in **Figure 53D-H.**

These temporal profiles clearly depict that both the phenotypes reach homeostasis (sustained level of activation observed in GSE 4290 and GSE 50161 models) or revolves around a fixed-point attractor state (cyclic or periodic activation observed for E-to-M transition in GSE 44561 models) in all the three developed logical models of GBM cell system. It should be also noted that the response profile of cell cycle progression state in the steady-state of GSE 4290 and GSE 50161 assay-based GBM models are comparably higher (quaternary state = 3) and sustainable than the GSE 44561 assay-based model (quaternary state = 2). The occurrence of such variation mainly arises due to the differences in the expression levels of the dependent genes (e.g., *CYCLIN-D1*, *CYCLIN-D2*, *CYCLIN-E1*, *CYCLIN-E2*, *CYCLIN-B*, *CYCLIN-C*, etc.) in the GBM cells.

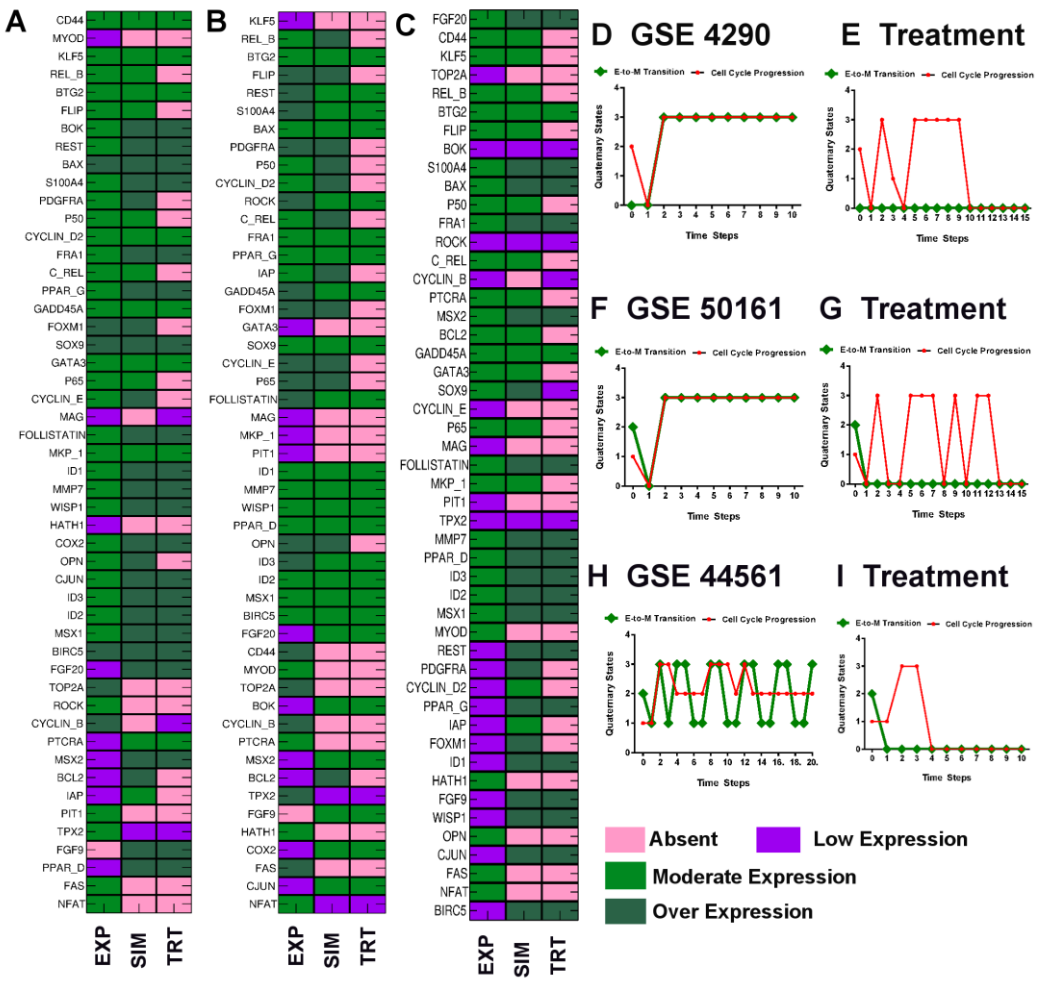

Figure 53: Quaternary states of the target genes and temporal dynamics of the corresponding phenotypes of integrated pathway model.
The logical steady states of the target genes of integrated pathway model are calculated for three models based on the input dataset obtained from microarray data *viz.* **(A)** GSE 4290, **(B)** GSE 50161, and **(C)** GSE 44561. EXP: Experimental data obtained from the microarray data (expected values) of GBM cells, SIM: Simulation outputs (LSS) of the corresponding model simulation, and TRT: Treatment refers the combinatorial inhibition scenarios (perturbations of multiple proteins on the corresponding model). **(D)**, **(F)**, **(H)** The temporal dynamics (pattern of activations) of two important phenotypes *viz.* "Cell Cycle Progression" and "E-to-M Transition or Epithelial to Mesenchymal transition" obtained after considering the initial states of the input proteins from the three microarray datasets of GBM cells. **(E)**, **(G)**, **(I)** The temporal dynamics of the same phenotypes observed after inhibiting multiple proteins on the corresponding GBM models. GLI1, NICD, HEY1, NRARP, and SNAI1 are targeted in the developed GBM models based on GSE 4290 and GSE 44561 microarray datasets **(E)** & **(I)**. Whereas, in GSE 50161 dataset based developed GBM model, another protein Estrogen receptor-alpha (ER-α) is required to inhibit in combination with the previously mentioned target proteins.

Appearance of such variation in the response profile is analogous to the unequal rate of cell cycle progression dynamics associated with GBM tumor cells. The

response profiles of cell cycle progression observed in the three logical models of GBM cells describe the strong and moderate activation of cell division process of the tumor cells obtained from three different sources of GBM tumor. The variable rate of cell cycle progression mainly depends on the expression levels of cyclin genes/proteins after the initiation of signal transduction. Indeed, the variations in the transcription of cyclin genes depend on the initial states of the input proteins (or signal intensities) and the concentrations of the component proteins involved in the transcription machinery of the signal transduction pathways activated in GBM tumor cells. Hence, we have seen that using ExQuLogic-based logical modeling approach we can readily create such cellular environment and successively predict the phenotypic outcomes in higher resolution. Obviously, the binary-logic based approach will fail in this scenario to create such cellular system and successively predict the ordinal expression/activation pattern of the cellular phenotypes depending on variable signal intensities.

7.2.3.6 SCCs are Important in the Reaction Module of Cell Cycle Progression and EMT of GBM

We have also applied the ExQuLogic-based logical modeling method to predict the important drug targets (or crucial nodes) in the whole signaling network used for the development of the three independent models of GBM tumor cells. Here, the objective is to identify the signal transduction nodes in all the three logical models of GBM cells, whose perturbations will subvert the cell cycle progression and E-to-M transition dynamics of the rapidly proliferating and metastatic tumor cells. However, the existing cross-connections between different signaling pathways, numerous positive and negative feedback loops, sub-cellular compartmentalization of the reactions, alternative shortest paths connecting the ligands/receptors to the components of transcription machineries within tumor cells make the entire system highly convoluted for the identification of drug-targetable crucial nodes [599-601]. Previous research works have shown the possible way out to overcome such

problems by splitting the entire network in different functional and regulatory modules [602-604]. In this work, at first we have extracted all possible existing reaction modules from the whole reconstructed signaling network (**Figure 54A)**. The existing modules and the few large of cluster of nodes can be visualized in this network diagram, which in turn proves the existence of highly connected nodes (hub) or proteins and the preferential attachment of the other nodes to that hub proteins in the whole signaling transduction network. The SCCs of the whole network are mainly situated in between the clusters and helping to maintain the flow of signal across the network. Due to this topological importance, the SCCs of any large signaling network are proven the best targets to alter the functional properties of the signaling network under a specific pathological condition [235,537,605].

However, all the SCCs, which we have identified in the whole network, are not suitable for perturbing the cell cycle progression and EMT of the tumor cells. Most importantly, there are total 62 SCCs in the network, and therefore systemic perturbations (or knock-outs) of all the SCCs on the logical model are difficult to perform. Hence, to reduce the overall complexities, we have analyzed the extracted modules of the whole network and identify a specific module which mainly contains the genes/proteins directly or indirectly influencing the cell cycle progression and EMT (**Figure 54B)**. We have observed that in this extracted module (**Figure 54B),** there are total eight strongly connected components or nodes (Inactive GLI1, activated GLI1, Nuclear localized GLI1, repressor of GLI3, activated NICDs, HEY1, NRARP, and SNAI1) present, which connect the other genes/proteins/complexes in the network involve in the regulation of the phenotypic nodes: cell cycle progression and E-to-M transition (EMT). GLI1 protein (transcription factor activated in Hedgehog pathway), Notch intracellular domains or NICDs (produced during Notch signaling), NRARP (Notch-regulated ankyrin repeat-containing protein), and SNAI1 (the nodes present in the outside of the circular layout of the regulatory module) (**Figure 54B)** are the important genes/proteins responsible for regulating the two

phenotypic responses. The peripheral nodes (e.g., SUFU, PKAα, CKIα, MAML, PTCRA, etc.) present in the circular layout are mainly the intermediate, but important signaling proteins in the entire network module, which help to process and transduce the incoming activation signal from the input to trigger the select phenotypes. However, it should be noted that the identified SCCs are also the intermediate nodes in the network, which mainly maintain the intra-module connectivity to smoothly relay the signal to the downstream for activating the phenotype responses at the end of the signaling cascade.

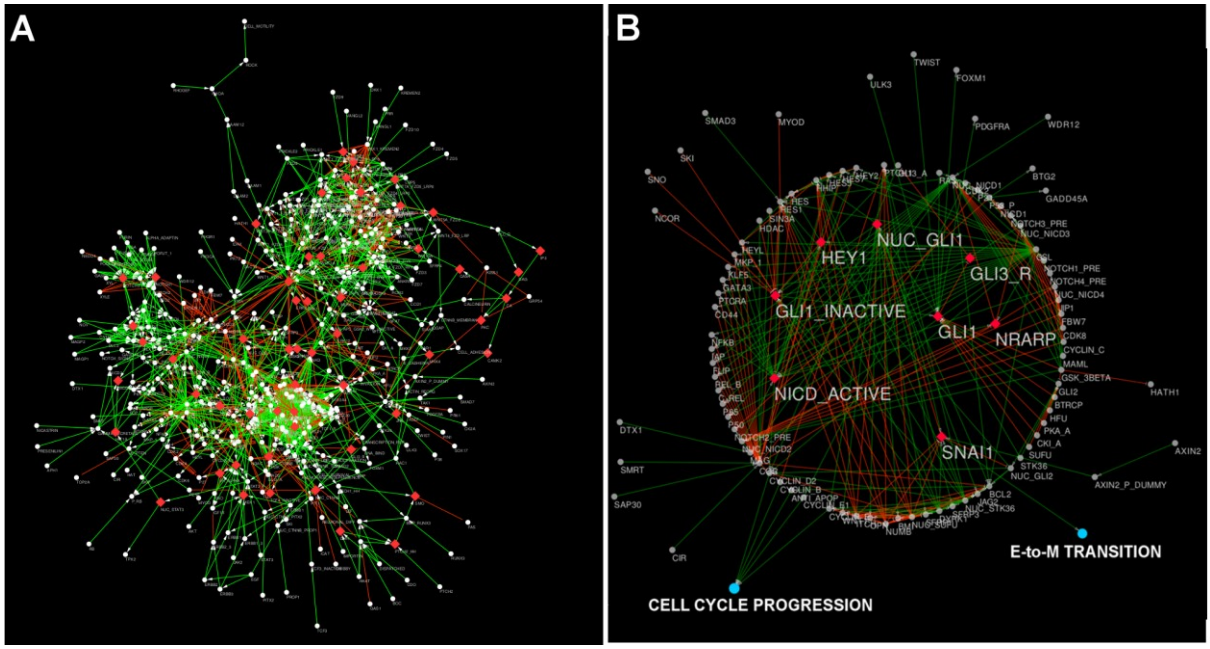

Figure 54: Network view of the whole reconstructed signaling network and the identified cell cycle progression and EMT regulatory module.
The green and red colored edges in the **(A)** whole reconstructed signaling network and **(B)** cell cycle progression and EMT regulatory module represent the activation and inhibition processes, respectively. The red colored, diamond shaped nodes in both the network diagrams represent the strongly connected components (SCC) of the respective networks. The topological positions of the SCCs in the extracted regulatory module describe the importance of the nodes connecting the outside (mainly inputs) and peripheral (intermediates) nodes to the select phenotypes.

7.2.3.7 The Identified SCCs are the Minimal Target Proteins to Deregulate Cell Cycle Progression and EMT in the GBM Cells

We have further studied the importance of the identified SCCs in the response patterns of cell cycle progression and E-to-M transition (EMT), observed in the

logical model simulations of the GBM cells (**Figure 53D-H**) (see **Section 2.4.2** of **Materials and Methods**). In order to do that, we need to systematically perturb the identified SCCs in each of the developed models and subsequently computed the steady-state expression pattern of the select target genes of the integrated pathway model. Interestingly, after individually knocking-out the identified SCCs in each logical model of GBM cells, we have not observed any significant differences in the expression patterns (quaternary states) of the select target genes at equilibrium conditions (results not shown). This result proves that the single knock-out of any of the identified SCCs is not capable enough to down-regulate the cell cycle progression genes (e.g., *CYCLIN D1*, *CYLCIN-D2*, and *CYLCIN-E1*, etc.) and simultaneously inhibit the hyper-phosphorylation of *pRB* protein in the GBM cells. The single inhibition of the SCCs also does not have any effect on the response of E-to-M transition phenotype.

The possible reason behind this outcome can be the existence of the alternative pathway and feedback loops (i.e. conserved signaling motifs) in the module, which can help to sustain as well as relay the response signal from the input nodes to the target genes (**Figure 55**). Follow up studies have revealed various feedback loops, cross-connections between the signaling proteins of all the pathways, which in turn create conserve signaling motif in the module.

Existence of such reaction motifs are the main reason for which suppression or single knock-out of the important SCCs are not effective to suppress the response of cell cycle progression and E-to-M transition dynamics. For example, while targeting the SCC, such as GLI1 or NUC_GLI1 in the GBM model, we have not observed complete suppression of cell cycle progression and EMT. A closer inspection of a signaling motif in the module has shown us the alternative route via GLI2 protein in the signaling network can still activate *CYCLIN-D2* gene expression and thus continues the cell cycle progression (**Figure 55A**).

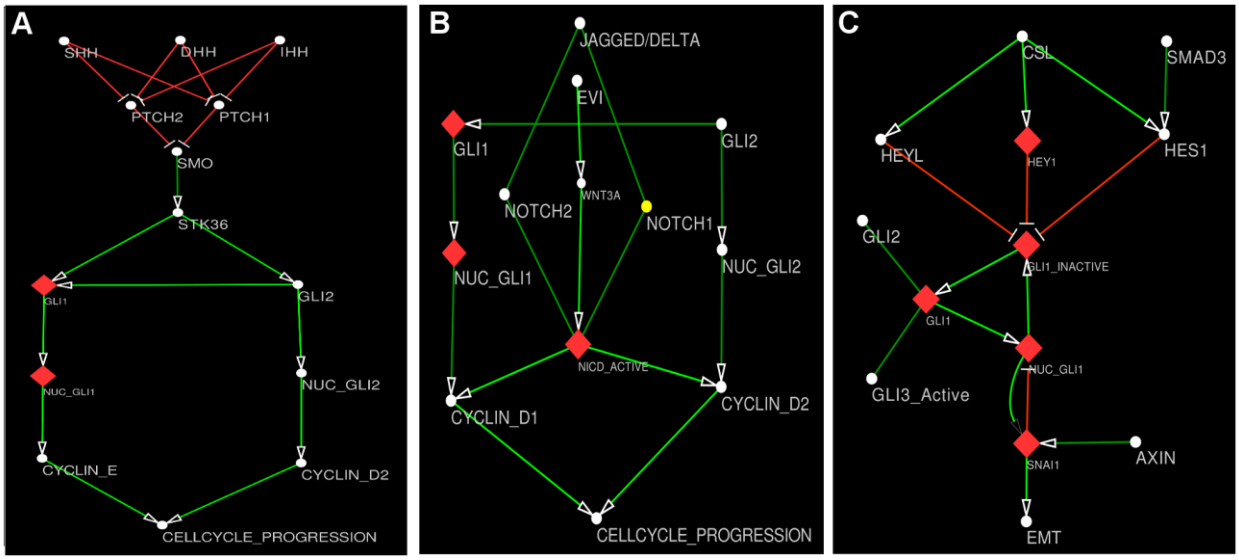

Figure 55: Examples of the signaling motifs identified in the cell cycle progression and EMT regulatory module.
(A) The alternate pathways available for activating cell cycle progression in Hedgehog pathway via GLI1 or GLI2 proteins. **(B)** The cross-talks of WNT, Notch, and Hedgehog pathways in the activation of cell cycle progression genes (*CYCLIN-D1* and *CYCLIN-D2*). **(C)** Cross-talks and feedback reactions between multiple signaling pathways leading towards EMT activation.

Hence, the better solution in this case will be to suppress the upstream activator STK36 protein including GLI1 or target both GLI1 and GLI2 proteins simultaneously to suppress the cell cycle progression. However, suppression of these proteins in combination will be ineffective if the other pathways, such as Notch and/or WNT are active in the GBM cells. We have also identified another conserved reaction motif in the integrated network, which are tightly coupled with each other and provide alternative escape route to the GBM cells to sustain its cell cycle progression, even after targeting a SCC of any of the pathway (**Figure 55B**). In this reaction motif, we have found that even after targeting GLI1 and GLI2 proteins in the GBM cells, the alternate activation route via DELTA/JAGGED --> NOTCH 1/2 --> NICDs can still activate cyclin genes. Also, the cross-talks between WNT pathway and Notch pathway can indirectly activate the Notch intracellular domain (NICDs) and thus sustain the cell cycle progression. On other hand, we have also observed that different signaling pathways can also cross-talk and create a conserve and robust signaling motif for the activation of SNAI1 protein, which is important for the

initiation of Epithelial to Mesenchymal transition (EMT) of the GBM tumor cells followed by metastasis (**Figure 55C**). Hence, it is obvious that single perturbation of the SCCs may not be always effective to subvert the cell cycle progression and EMT due to the presence of such conserve and robust reaction motifs in the rapidly proliferating and metastatic GBM cells.

In this work, we have hypothesized that to break such robust and synergistic network motifs in the GBM cells multiple knock-outs of SCCs including other signaling proteins are utmost required. We have tested this hypothesis on our developed logical models of GBM cells by knocking-out (considering the quaternary state = 0/absent) all the identified SCCs and STK36 proteins simultaneously. Inhibition of STK36 can be made possible by inhibiting its upstream activator protein Smoothened (SMO) in the cell membrane [606]. In case of GBM models based on GSE 4290 and GSE 44561 microarray datasets, important target genes (e.g., *REL-B, BCL2, PDGFRA, CYCLIN-E, OPN, FOXM1, CD44, KLF5, CYCLIN-B*, etc.) responsible for tumorigenesis, uncontrolled cell divisions, and metastasis are either lowly expressed or completely down-regulated (or absent) after targeting all the SCCs and STK36. On the other hand, the pro-apoptotic genes, such as BOK and BAX are over-expressed in the multiple knock-out (or treatment) scenario (**Figure 53A** and **Figure 53C**). Due to this, the activation or response profile of the cell cycle progression and EMT also get declined (quaternary state = 0) in the steady-state or equilibrium condition in the treatment scenario of the logical models developed for GSE 4290 and GSE 44561 microarray datasets (**Figure 53E** and **Figure 53I**). Interestingly, we have also observed that the GBM model based on GSE 50161 microarray dataset does not show any promising results while performing multiple knock-outs of SCCs and STK36 proteins. Further inspection of the experimental datasets and predicted outcomes of the target genes revealed that in parallel with Hedgehog, Notch, WNT pathway, the activation signal coming from EGFR pathway to activate ERα protein (**Figure 51A**) can also trigger the cell cycle progression dynamics in the GBM cells. In this case, the

EGFR signaling cascades act synergistically with other pathways to compensate the damages induced by targeting SCCs in the GBM cells. Hence, after targeting ERα and other SCCs including STK36, we have achieved the complete suppression of cell cycle progression in the respective GBM model (**Figure 53B** and **Figure 53G**).

This result evidently indicates the existence of intra-tumor heterogeneity and the appearance of drug resistant GBM cells against known therapeutic drugs. The overall results of perturbation analyses also depict the potential of our newly developed extended quaternary logic-based approach in the simulation and prediction of the expression pattern of the target genes under tumorigenic and perturbation scenarios. Using this approach, we can more precisely predict the level of expressions (i.e., over-expressed, regular expression, low expression, and down-regulation) of the genes or proteins of the signaling pathways as compare to the conventional binary logic based approach, limited to predict only up or down-regulated states. Also, by connecting the target genes with the phenotypic responses (e.g. cell cycle progression, EMT, apoptosis, etc.), we can perform the comparative analyses of the magnitude of the response pattern (strong, modest, low, absent) of the phenotypes under tumorigenic and treatment cases.

7.3 DISCUSSION

In this work, we have shown the existing limitations of binary logic-based modeling method while implicated in the simulation of biochemical networks, such as signaling pathways, gene regulatory network, etc. We have shown the possible alternative of the use of quaternary logic-based algebra, which assumes four possible quaternary states **0**, **1**, **2**, and **3** of all the nodes of the developed logical model, instead of binary states **0** and **1**. We have also proposed quaternary state selection rule to discretize the genes/proteins of a logical model by utilizing their fold-change ratio and FDR values in a biological sample versus control, obtained in differential expression studies. We have introduced a novel algorithm to perform quaternary

algebra, which we have named as the extended quaternary state update scheme (ExQSUS) algorithm. We have proposed to consider the quaternary state of the downstream node of previous time-step while computing the influence of only the single or multiple inhibitor node(s) on that downstream node. The entire methodology is the extended version of quaternary logic (ExQuLogic)-based operation, but is more suitable and appropriate for modeling and simulating the real biochemical reaction events inside the cells. Hence, in order to assess its efficacy in the simulation of real biochemical networks, we have performed retrospective analyses of the previously published, binary logic-based models of T-cell receptor mediated signaling pathway, ERBB pathway, Hedgehog pathway, and Notch signaling pathway using ExQuLogic. We have compared the results obtained from ExQuLogic and binary logic-based operations, and observed marked improvement in the prediction of biological states (qualitative) of the nodes at steady-state. Using ExQuLogic, we have successfully simulated various biological scenarios, which we could not able to simulate using conventional binary-logic based simulation.

Using ExQuLogic-based operation, one can create and simulate the effects of strong, weak, and the absence of stimulation of the receptor proteins in signaling pathway, consider the expression of the nodes in a network which are neither at significantly up-regulated (active) nor down-regulated states, and distinguish the logical states of the target genes/proteins in the cellular systems, which consist of homozygous, heterozygous, wild-type expression, and copy number gains of important signal transduction and gene regulatory nodes (genes/proteins), etc. We have also applied our proposed method in a newly reconstructed logical model of integrated cell signaling pathways of Hedgehog, Notch, EGFR, and WNT signal transduction networks to perform the prospective analysis for predicting the gene expression pattern in Glioblastoma cells at the equilibrium condition. We have matched the predicted gene expression data obtained from ExQuLogic-based operation with three microarray datasets of GBM cells. The simulation outputs

matches very well with all the three experimental datasets, and correctly predicted the activity profiles of the target oncogenes causing irregular cell cycle progression (cell division) and Epithelial to Mesenchymal transition or EMT (metastasis) of the GBM cells. Further topological analyses of the integrated network have revealed that there exist multiple strongly connected components (SCC) in the entire system, which can act synergistically to establish a robust fail-safe mode in the GBM cells upon any external perturbations. However, extracting the multiple shortest-paths which connect the SCCs in the network with the genes responsible for cell cycle progression and EMT, we have identified that targeting SCCs are the best possible solution if they are targeted simultaneously in the proliferating and migrating GBM cells. We have performed the perturbation analyses using ExQuLogic-based method, which has efficiently depicted the possible required alteration for the perturbed nodes in the network.

7.4 CONCLUSIONS

Overall, our developed method of the extended quaternary logic-based operation has depicted its potential to simulate the biochemical networks in cancerous and immunological cells. However, it should be mentioned that the algorithm proposed in the logic-based operation is not limited to the simulation of cancer or immunological cells, but also useful for modeling and simulating other types of cellular system. Using the ExQuLogic-based modeling approach, one can test various hypotheses related to cell signaling, gene regulatory, and the metabolic networks, etc. The outputs obtained from ExQuLogic have higher resolutions of the qualitative expressions of genes, proteins, and metabolites, etc. in any cellular system at equilibrium condition, which can be compared with the quantitative expression values obtained from the real biological and biochemical experiments. We believe that our newly proposed algorithm will open up new opportunities to model and simulate the biochemical networks in different cellular systems with much more accuracy and resolutions as well as break the imposed barriers causing by the

existing limitations of conventional binary and multi-valued logic-based modeling approaches. Our proposed algorithm (ExQSUS) can be easily implemented in the standard computing environment without using the compilers for multi-valued logical operations.

Chapter 8

8 CONCLUDING REMARKS AND FUTURE WORKS

In the targeted therapy of Glioblastoma tumor and other cancer cells, most often the preferred targets are the proteins or enzymes involved in different signal transduction pathways [607]. Hence, identification of suitable drug targets has always been a major challenge to the researchers working in this field of study [607].

The investigations performed in the thesis have been specifically directed towards the understandings of the regulatory mechanisms in the signaling pathways involved in oncogenesis. The identified regulatory modules in the cancer cells have further helped to identify novel drug target(s) for the treatment of GBM tumor cells. The regulatory motifs have also helped to understand the mechanisms through which the cancer cells get resistivity against drug molecules. A phenotype prediction sore have been also proposed through which the risk of the development of different grades of GBM tumor cells have been computed. This phenotype prediction score is further validated with the TCGA-LGG and TCGA-GBM patient cohorts datasets.

In this thesis, different oncogenic cell signaling networks, such as Hedgehog, Notch, WNT, EGFR, RAS, PI3K/AKT, HIF1A, etc. have been reconstructed for *in-silico* analyses. Also, the reconstructed pathways have been made available to the common users via a newly developed database BIOPYDB. Furthermore, the reconstructed pathways have been analyzed using graph theoretic analyses and successively simulated using semi-dynamic logical models. The topological and semi-dynamic logical modeling approaches have helped us to identify the flow of reaction signals in the normal and GBM tumor cells, extract important modules

responsible for pathogenesis and drug-resistivity, and subsequently discover novel combination of drug-targetable proteins. The drug targets, which have been identified in this study is summarized in the **Appendix Table 13**.

8.1 APPLICATIONS AND FUTURE DIRECTIONS

8.1.1 Applications of the Drug-targets in Cancer Therapy

In this thesis, we have been successfully able to identify the combination of druggable proteins, which can be useful for the targeted therapy of GBM and other cancer cells (i.e. pancreatic and colon). Our *in-silico* perturbation studies have proven the identified targets (**Appendix Table 14**) as highly effective for this purpose. However, to implement these targets in the targeted therapeutics, we need to perform further *in-vitro* and *in-vivo* studies. The efficacy of suppressing the targets, its side-effects, toxicity, pharmacodynamics, etc. should be rigorously measured and fine tuned with the current clinical guidelines of anti-cancer therapy.

8.1.2 Implementation of Predictive Model for Tumor Risk Prediction

The newly introduced phenotype prediction scores can be successfully implemented to measure the tumor risk of an individual model with the help of the logical models discussed in *Chapter 6*. The patient's omics data (e.g., differentially expressed transcripts, proteins, metabolites, etc.) will be considered as inputs in the developed models and the chances or risk of developing GBM tumor will be assessed, followed by drug-targets screening using FFT. Also, the novel AR score calculation technique could be useful to back-trace the probable stimulations in the GBM tumor cells responsible for the emergence of different tumor grade and intra-tumor heterogeneity. The entire protocols to use the developed mathematical models in the prediction of tumor grades are provided in the flow-chart (**Figure 56**).

8.1.3 Applications of ExQuLogic

Our newly developed algorithm Extended Quaternary states Update Scheme

(ExQSUS), implemented in Extended Quaternary Logic (ExQuLogic)-based computational analyses framework has been shown highly useful to capture the various aspects of signal transduction networks, which we could not analyze by using conventional binary logic-based approach.

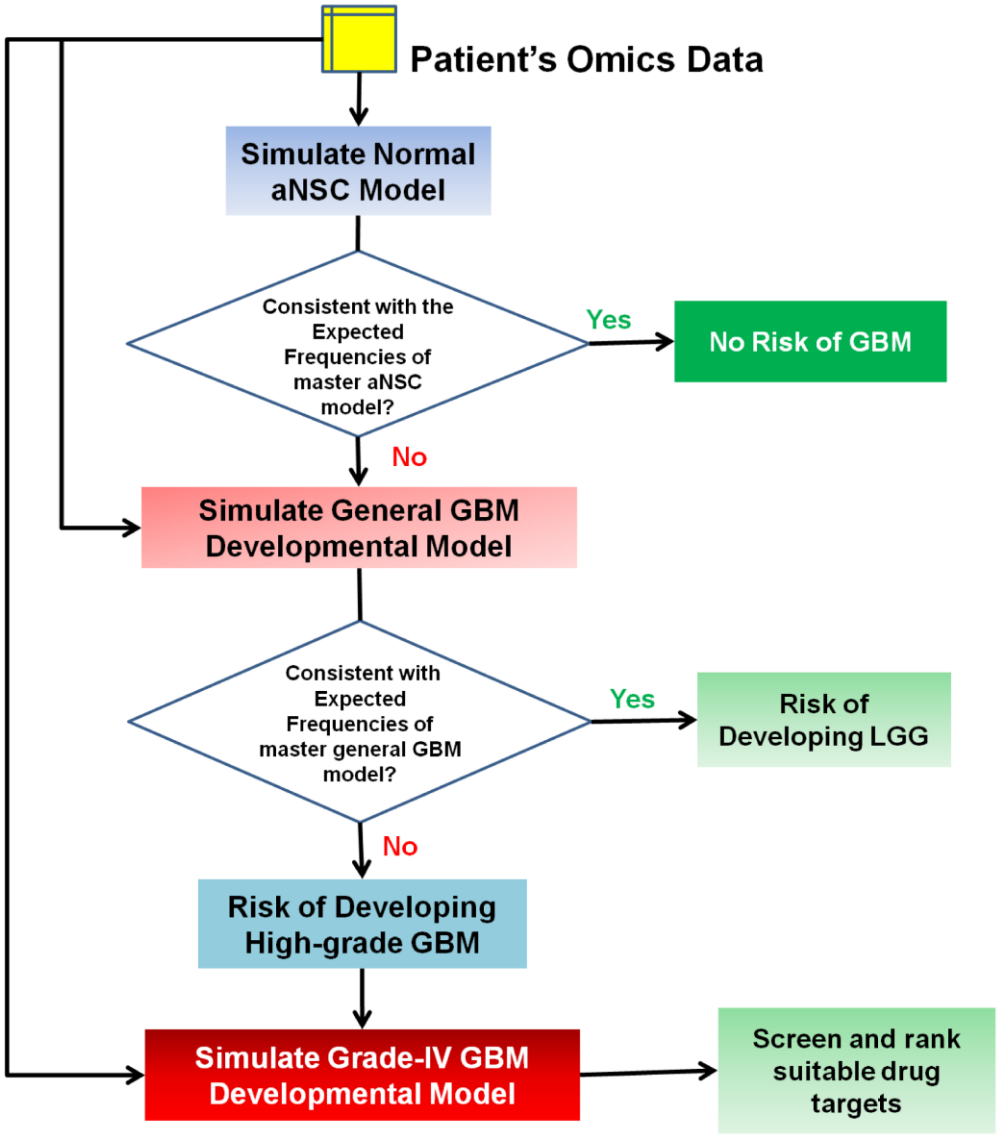

Figure 56: Flow-chart of the decision-making steps to determine risk of GBM.

These algorithm and framework can be implemented to model any cell signaling network and compute the steady-states of the nodes. In the near future, we believe the application of ExQuLogic will help to overcome the theoretical and computational limitations of the conventional binary logic-based modeling and simulation techniques. The source codes will be shared in open platform (e.g.

GitHub), which will help the researchers to gain better access and implementation of the developed method.

8.1.4 Third-party use of BIOPYDB Database

The newly developed, open source, database-cum-web server "BIOPYDB" can be also used by the third-party developers to launch new software or services for pathway data reconstructions, analyses, modeling, and simulations. We believe the third party software developers will find our free RESTful web services as useful API in their projects. Our pathway reconstruction process is still ongoing and in the near future, we believe that the available contents (pathways, reactions, tools) of our database will attract more users and software developers. We will try to update the information of the database as per the latest information available in the literature. The general users can also join the pathway reconstruction initiatives by providing their feedbacks, corrections, or new pathway data. The primary concept of our web service is not only to provide the pathway data, but also create the open source ecosystem in which the general users can model, simulate, and analyze the pathway data using network, logical, and dynamic theories without having advanced knowledge of computer programming.

The investigations performed in the thesis have been mainly restricted to the understanding of major developmental and mitogenic signaling pathways in the pathogenesis of Glioblastoma tumor. However, the computational methodologies proposed in the thesis can be extrapolated to the understandings of the pathogenesis of other types of cancer and tumor cells. The drug-target identification techniques used in this thesis can be also applied to other human diseases, such as autoimmune diseases, heart diseases, and infectious diseases, etc. Also, to get better understanding of oncogenesis, other molecular datasets can be used, which will further help to understand the complex regulatory mechanisms in cancer cells and identify better drug-targets for the future cancer therapy.

9 APPENDICES

All appendix tables cited in the main document have been provided in this section.

Appendix Table 1: Available Resources used for Pathway Data Curation, Modeling, and Simulation

Name	HTTP Link
1A. Pathway Databases (Commercial Developers)	
1. GENEGO: PATHWAY MAPS	http://pathwaymaps.com/maps/
2. BIOCARTA	http://www.biocarta.com/
3. PROTEIN LOUNGE	http://www.proteinlounge.com/
4. CELL SIGNALING TECHNOLOGY	http://www.cellsignal.com/index.jsp
5. MILLIPORE	http://www.millipore.com/pathways/pw/pathways
6. APPLIED BIOSYSTEM	http://www5.appliedbiosystems.com/tools/pathway/
7. INVITROGEN	http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cell-Analysis/Signaling-Pathways.html
1B. Pathway Databases (Academic Developers)	
1. KEGG	http://www.genome.jp/kegg/
2. SIGNALING PATHWAY DATABASE (SPAD)	http://www.grt.kyushu-u.ac.jp/spad/
3. DOQCS	http://doqcs.ncbs.res.in/
4. REACTOME	http://www.reactome.org/ReactomeGWT/entrypoint.html
5. PATHWAY INTERACTION DATABASE (PID)	http://pid.nci.nih.gov/
6. CPDB	http://cpdb.molgen.mpg.de/
7. NETPATH	http://www.netpath.org/
8. PATHWAY COMMONS	http://www.pathwaycommons.org/about/
9. HIPATHDB	http://hipathdb.kobic.re.kr/browse.php?dbType=1
10. SIGNALINK	http://signalink.org/
11. SPIKE	http://www.cs.tau.ac.il/~spike/
12. WIKIPATHWAYS	http://wikipathways.org/index.php/WikiPathways
13. INNATEDB	http://www.innatedb.com/
14. INOH	http://www.inoh.org/
15. BIOMODELS	http://www.ebi.ac.uk/biomodels-main/
16. GOLD.DB	https://gold.tugraz.at/
17. PANTHER	http://www.pantherdb.org/
2. Protein-Protein Interaction (PPI) Databases	
1. MINT	*https://mint.bio.uniroma2.it/*
2. APID	*http://bioinfow.dep.usal.es/apid/index.htm*
3. STRING 9.05	*http://string-db.org/*
4. PIPS	*http://www.compbio.dundee.ac.uk/www-pips/*
5. DIP	*http://dip.doe-mbi.ucla.edu/dip/Main.cgi*
6. BIOGRID 3.2	*http://thebiogrid.org/*
3. Microarray Expression Databases	
1. EBI-ARRAYEXPRESS	*http://www.ebi.ac.uk/arrayexpress/*
2. GENE EXPRESSION OMNIBUS	*http://www.ncbi.nlm.nih.gov/geo/*
3. THE CANCER GENOME ATLAS (TCGA)	*https://portal.gdc.cancer.gov/legacy-archive/*
4. Disease Databases	
1. NCG 5.0	*http://bio.ieo.eu/ncg/*
2. CANCER RESOURCE	*http://data-analysis.charite.de/care/*
3. CANCER CELL MAP	*http://cancer.cellmap.org/*

Appendix Table 2: Detailed Description of the Signaling Pathway Database and their Computational Tools.

Tools	Name of the Database	Remarks
Pathway Data Uploading Tool	BIOCARTA, WikiPathways, REACTOME	These databases have the tools for uploading pathway data in the specific database formats. Users are required to Log in WikiPathways to upload or edit any pathway information.
Pathway Drawing Tool	KEGG	**KegDraw:** It is a Java based application for drawing compound and glycan structures.
	CPDB	It can create the signaling network or map by uploading the pathway interactions in its own format.
	PROTEIN LOUNGE	3D pathway can be created using its ePATH 3D tool. Although it is not free.
	Pathway Commons	It has Cytoscape Plug in to view, edit and analyze the pathway data. It also has a pathway viewer and editor ChiBE which is linked to it. Another pathway visualization application PCVIZ is alos available in this database, which takes a list of genes and by finding its neighbors it generates the pathway diagram.
	SPIKE	It is also a tool for pathway drawing and visualization.
Pathway Analysis Tool	BioModels	It has online ODE simulation tool to simulate the pathway models.
	KEGG	**KegArray:** JAVA based application for microarray data analysis.
	SignaLink	**PathwayLinker:** Identifies and visualizes the first neighbor interaction network of the queried proteins, analyzes the signaling pathway memberships of the proteins in this subset, and provides links to further online resources. **Signalog:** It can predict novel signaling pathway components on a genomic scale based on the signaling pathway membership(s) of its ortholog(s) in 8 signaling pathways of 3 intensively investigated species: *C. elegans*, *D. melanogaster*, and human.
	SPIKE, CPDB	These databases have the tools for pathway enrichment tools using micro array or protein expression data.
	PANTHER	**Gene List Analysis:** This tool is provided in the database to analyze a set of user defined gene lists, and their expression data with PANTHER data. It maps the gene lists to PANTHER ontology and subsequently grouped into various biological process categories, as well as to the biological pathways. It also overlay the analysis results on PANTHER pathway diagrams to visualize the probable functional relationships between genes/proteins in known pathways. **PANTHER scoring:** This tool is useful for scoring a user defined protein sequence against the entire PANTHER library of over 38,000 statistical models based on Hidden Markov Models (HMMs) to obtain PANTHER classifications and alignments. Moreover, PANTHER has also built various in-house tools

		for visualizing, downloading and computing the pathway data. It has also put a significant effort to provide the pathway maps using SBGN standard notation on JAVA based application platform.
	REACTOME	**Pathway Analysis Tools:** This tool merges various pathway analysis related tasks to a single portal, through which one can perform the identifier mapping, overrepresentation and expression analysis. Users can provide Uniprot accession list, NCBI/Entrez list, Small molecule (ChEBI or KEGG) list, microarray, and metabolomics data for the mapping and expression study against REACTOME database. REACTOME has also in-built web based tools to visualize the module based/functional hierarchy based pathway components and reactions in web interface.

Appendix Table 3: Short Descriptions of the Features Available in BIOPYDB

Features		Description
Search	**Text search**	Convenient way of searching proteins, pathways and Diseases in the search box provided in "Home" page.
	Advance Search	Designed for searching multiple queries in BIOPYDB as well as in 17 Similar pathway databases.
Browse Pathway	**Browse Pathway By Name**	Helps to select/browse the pathways by their names.
	Browse Pathway By Category	To browse the pathways by their functional categories: Disease, Immunological and Developmental.
Tools	**Pathway Drawing**	PHP and Graphviz based application to draw the pathway diagram.
	Network Analyses	Built in PHP, Graphviz and igraph based application, which allows users to perform Graph theoretical or Network analysis of signaling network.
	Logical Analyses	Developed by in-house codes written in PHP and the open source software BooleanNet for performing Boolean or logical equation simulation [309].
	Dynamic Analyses	Developed by PHP, C, Gnuplot scripts to perform Ordinary Differential Equation Model on the signaling pathways present in the database or on the user defined new pathway model.
	Find Interaction	PHP and Graphviz based application to "Find interaction of a Protein" with other proteins or to "Find Interaction Between Two Proteins". For each instance reference is also available.
	Find Disease	PHP and Graphviz based application to “Find Mutated Proteins of a Disease of Interest”; “Find Diseases of a Protein of Interest”; or to “Find Connection Between a Protein and a Disease”. For each instance reference is also available.
Download	**Download Pathway Image**	Developed to download "Pathway Image" in SVG, PNG, JPEG or PDF format. For each instance, image is first generated by Graphviz software and then serves for download. The idea is to provide up to date pathway image by using the data from database.
	Download Pathway Data	Developed to provide Pathway related Data in SBML and simple Text (TAB or CSV) formats. Protein-Disease related data is also available. Here, the data is also first generated by PHP based application and then serve for download.
	RESTful API Service	A tutorial to access the RESTful API services of BIOPYDB is provided in this section. The API service is useful to extract pathway list, their BIOPYDB IDs, pathway species and their reactions, protein vs. disease information etc. The pathway images can also be downloaded through this API, which are provided in encoded JPEG, PNG and SVG formats.
FAQ	Frequently Asked Questions (FAQ) is helpful for guiding the users in Database Searching, how to use the tools and data downloading. It is also open for asking new questions or suggestions.	
Upload Pathway	**New Pathway Upload**	It helps to upload and populate a new pathway data from user's side. A very handy application of pathway curators without having any previous knowledge of database maintenance. Also, users can instantly view their uploaded data, pathway image and the graph theoretical analysis of the uploaded signaling network.
	Edit Uploaded Pathway	Helpful to update the user's pre-uploaded data in the database. Users have to provide the accession number entered in the first time of uploading the data from "New Pathway Upload" section.
Contact Us	**Feedback**	This Option is available to provide Comments /Remarks to the database developers for further improvement/modifications.
	Contact Info	Detail contact information of the Database management team.

Appendix Table 4: Abbreviations and Detail Information of the Proteins and Cellular Responses used in the Master Model of Hedgehog Pathway

SHORT NAME USED IN MODEL	FULL NAME	DOCUMENTATION
I. Extracellular and Membrane Proteins		
BMP_RUNX3	Bone Morphogenetic Protein and Runt related transcription Factor 3	BMP-RUNX3 signaling induces expression of IHH in surface differentiated epithelial cells of stomach and intestine.
DISPATCHED	Dispatched	Dispatched regulates the release and extracellular accumulation of cholesterol-modified hedgehog proteins and is hence required for effective production of the Hedgehog signal.
HHAT	Hedgehog Acyltransferase	HHAT is a hedgehog modifier which induces lipid modification to generate mature peptides. Hedgehog proteins with lipid modification are then released from producing cells by Dispatched homologues.
DHH	Desert Hedgehog	Three Hedgehog ligands (homologues proteins) of Hedgehog pathway considered as Input Proteins in this model.
IHH	Indian Hedgehog	
SHH	Sonic Hedgehog	
PTCH1	Patched1	Two homologue of receptor protein Patched. In the absence of hedgehog ligands these proteins inhibit another trans membrane protein Smoothened (SMO).
PTCH2	Patched2	
SMO	Smoothened	G-protein coupled receptor that is normally suppressed by Patched receptors but is activated in the presence of Hedgehog ligands (SHH, DHH, IHH).
HHIP	Hedgehog Interacting Protein1	Regulates the amount of Hedgehog ligand that can bind to Patched receptors
CDO	Belong to Immunoglobin super family.	CDO and BOC represent a subfamily within the Ig super-family, consisting of an

BOC	Brother of CDO.	ectodomain comprised of four (BOC) or five (CDO) Ig repeats, followed by three fibronectin type III (FNIII) repeats and a long, divergent intracellular domain.
GAS1	Growth arrest specific gene	Regulates the amount of Hedgehog ligand that can bind to Patched receptors along with HHIP.
II. Cytoplasmic Proteins		
HFU	Human Fused	Stimulates GLI1 and GLI2 transcription factors.
SUFU	Suppressor of fused homolog	Sequesters GLI proteins in the cytoplasm and prevents tarnscription of target genes.
STK36	Serine/threonine-protein kinase 36	Up-regulation of GLI transcription activity.
ERK12*	Extracellular signal-regulated kinase	Up-regulation of GLI transcription activity.
GLI1	Transcriptional activator Gli1	Mediates target gene expression.
GLI2	Transcriptional activator Gli2	Mediates target gene expression.
GLI3_R	Transcriptional repressor Gli3	Antagonises target gene expression byother Gli factors.
RAS*	Ras protein (GTPase activity)	RAS and TWIST activate GLI1 regulatory sequences.
TWIST*	Twist-related protein	Is known to activate GLI1.
PKA_A	Protein Kinase alpha	Phosphorylates and activates SMO.
BTRCP	Beta-transducin repeat-containing protein	Involoved in ubiquitination of Gli1 resulting in the formation of a transcriptional repressor.
CKI_A	Casein Kinase I isoform alpha	Known to elicit negative effects on GLI
GSK3	Glycogen synthase Kinase 3	Known to elicit negative effects on GLI
NOTCH1*	Notch1 protein	Known to elicit negative effects on GLI
FAS*	Apoptosis-mediating surface antigen FAS	Mediates apoptosis

ULK3	Unc-51-like kinase 3	Serine Threonine kinase present in addition to STK36 that functions in up-regulation of GLI transcriptional activity
III. Nuclear Proteins		
NUC_GLI1	Nuclear GLI1	Represents the nuclear form of GLI1 protein.
NUC_GLI2	Nuclear GLI2	Represents the nuclear form of GLI2 protein.
NUC_SUFU	Nuclear SUFU	Represents the nuclear form of SUFU.
NUC_STK36	Nuclear STK36	Represents the nuclear form of STK36
GLI3_A	Activated GLI3 for Transcription	Mediates target gene expression
SKI	Proto-oncogene C-Ski	Functions as a transcriptional co-repressor
NCOR	Nuclear receptor corepressor	Functions as a transcriptional co-repressor
HDAC	Histone deacetylase	Functions as a transcriptional co-repressor
SNO	Ski-like protein or Ski-related oncogene	Functions as a transcriptional co-repressor
SIN3A	Paired amphipathic helix protein Sin3 alpha	Functions as a transcriptional co-repressor
DYRK1	Dual Specificity Tyrosine phosphorylation Regulated Kinase 1A / Dual Specificity Yak1 related Kinase	Known to substantially increase GLI mediated transcription
NUMB	Protein numb homolog	Numb along with ubiquitin ligase such as Itch is able to polyubiquitinate GLI1 and target it for degradation and thus control HH signaling.
ITCH	E3 ubiquitin-protein ligase Itchy homolog	
IV. Output Proteins		
CTNNB_TCF4	Nuclear form of TCF4	Represents the nuclear form of TCF4
CYCLIN_B	G2/mitotic-specific cyclin-B1	Mediates cell cycle regulation
CYCLIN_D	G1/S-specific cyclin-D	Mediates cell cycle regulation
CYCLIN_D2	G1/S-specific cyclin-D2	Mediates cell cycle regulation
CYCLIN_E	G1/S-specific cyclin-E	Mediates cell cycle regulation

FOXM1	Forkhead box protein M1	Implicated in cellular proliferation
PDGFRA	Platelet Derived Growth Factor receptorisoform alpha	Transmembrane receptor
OPN	Osteopontin	Osteopontin is a secreted protein that influences multiple downstream signaling events that allow cancer cells to resist apoptosis, evade host immunity and influence growth of indolent tumors.
CMYC	Myc proto-oncogene protein	Mediates cellular proliferation.
BMI	Polycomb complex protein BMI-1	BMI-1 is a transcriptional repressor belonging to the polycomb gene family and its suppressor functions are involved in maintaining neuronal, haematopoietic and mammary gland stem cells.
SNAI1	Protein snai1 homolog 1	Responsible for the degradation of E-cadherin and initiation of invasion
JAGGED2	Notch ligand Jagged	Stimulates Notch signaling
SFRP	Secreted frizzled-related protein	Wnt antagonist
WNT	Wnt family proteins or ligand	Representative of a WNT ligand
BCL2	Apoptosis regulator Bcl-2	Anti-apoptotic
V. Cellular Responses		
Anti_Apop	Anti apoptosis	These are the cellular responses or phenotypic expressions that have been shown as outcomes of this pathway.
Notch_Signal	Nocth signaling	
Wnt_Signal	Wnt signaling	
Cellcycle_Progression	Cell cycle progression	
Emt	Epithelial to Mesenchymal transition	
Cell_Proliferation	Cellular proliferation	

Appendix Table 5: Master Model of the Hedgehog Signaling Pathway.

Interactions	Documentation
➔DISPATCHED	Inputs to the model. Upstream regulators of these molecules have not been considered. Inputs to the model. Upstream regulators of these molecules have not been considered.
➔HHAT	
➔ CDO	
➔ BOC	
➔ NUC_SUFU	
➔ GAS1	
➔BMP_RUNX3	
➔ULK3	
➔ HFU	
➔ SUFU	
➔ ERK12	
➔ RAS	
➔ TWIST	
➔ DYRK1	
➔ NUMB	
➔ ITCH	
➔ PKA_ALPHA	
➔ BTRCP	
➔ CKI_A	
➔ GSK3	
➔ NUC_STK36	
➔ NOTCH1	
➔ SKI	
➔ SNO	
➔ NCOR	
➔ SIN3 ALPHA	
➔ HDAC	
BMP_RUNX3➔IHH	BMP-RUNX3 signaling induces expression of IHH in surface differentiated epithelial cells of stomach and intestine [337].
CDO and BOC➔SHH	CDO and BOC bind SHH through a high-affinity interaction with a specific fibronectin repeat that is essential for activity. They demonstrate that CDO and BOC are necessary but not sufficient for activation [416]. However, there is no evidence for the exact mechanism and if both are required for the enhancement of signaling.

DISPATCHED and HHAT and not HHIP➔DHH	Dispatched regulates the release and extracellular accumulation of cholesterol-modified hedgehog proteins and is hence required for effective production of the HH signal [608]. HHAT (Hedgehog acyltransferase) is a hedgehog modifier which induces lipid modification to generate mature peptides. HH proteins with lipid modification are then released from producing cells by Dispatched homologues. HHIP can antagonize all types of HH ligands [609].
DISPATCHED and HHAT and not HHIP➔IHH	Dispatched and HHAT system also operates in the same way as during DHH release [608]. HHIP can antagonize all types of HH ligands [609].
DISPATCHED and HHAT and not HHIP and not GAS1➔SHH	Dispatched and HHAT also operate in the same way as during DHH release. HHIP is found to bind directly to SHH and attenuate SHH signaling like PTCH1/2 while its expression was induced by SHH signals [610].
not DHH and not IHH and not SHH➔PTCH1_Free	Negative influence of all the Hedgehog ligands was considered to denote the inactive state of Patched (PTCH1 and PTCH2) receptors. In the absence of Hedgehog ligands Patched receptors are active and suppress the activity of Smoothened [611].
not DHH and not IHH and not SHH➔PTCH2_Free	
DHH and not PTCH1_Free➔SMO	In the absence of a stimulus by Hedgehog, Patched receptor inhibits Smoothened. Upon binding of Hedgehog ligands DHH, SHH or IHH to Patched , Smoothened is activated leading to the transcription of target genes. This is also reported that mutations affecting the transmembrane proteins Patched or Smoothened trigger the ligand independent activity of Hedgehog signaling pathway and are hence associated with human tumors such as basal cell carcinoma and medulloblastoma [611].
IHH and not PTCH1_Free➔SMO	
SHH and not PTCH1_Free➔SMO	
SHH and not PTCH2_Free➔SMO	

SMO➔STK36	SMO binds to STK36 to stabilize GLI proteins [337].
not SMO➔FAS	SMO expression inhibits FAS thereby preventing apoptosis [606].
HFU and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	HFU enhances GLI1 function in a manner that is independent of a functional kinase domain [316]. GSK3 phosphorylates GLI proteins post phosphorylation by PKA and it is known to elicit negative effects [343]. SMO inactivation leads to formation of the cytoplasmic GLI degradation complex, in which GLI family members (GLI1, GLI2 and GLI3) are phosphorylated by casein kinase alpha (CKI_α), glycogen synthase -kinase-3ß (GSK3ß) and protein kinase A (PKA). Phosphorylated GLI is recognized by FBXW1/BTRCP1 and FBXW11/BTRCP2 for ubiquitination, and ubiquitinated GLI is partially degraded to release its intact N-terminal half thereby functioning as transcriptional repressor [337]. The inhibitory interactions have been included with activation interactions using an AND operator. Therefore GLI cannot be activated unless and until all the inhibitors are absent. However, this needs to be checked in *in-vivo* conditions. But in the model it is necessary to introduce these interactions using an "AND" operation to assure signal flow.
ERK12 and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	EGFR signals via ERK potentiate target gene activation via GLI1 [612].
RAS and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	It is reported that oncogenic KRAS/ constitutively active RAS in Pancreatic Cancer cells, increases the transcription of GLI1 levels [345].

TWIST and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	TWIST activates human GLI1 regulatory sequences via two E-boxes in GLI1's first intron. Demonstrated in a murine model and using human GLI sequences. Two critical cis elements in human GLI1 gene: a GC box that binds Sp1 at 195 and two E-boxes that operate at 157 and 482 have also identified. The 157 E-box binds USF1 and USF2, while E-box 482 binds TWIST. Sp1 and USf1/2 are ubiquitiously expressed TFs and can function either as activators or repressors depending upon cellular context. However their roles have not been clearly delineated and hence not incorporated in the model [345].
ULK3 and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	ULK3, a Ser/Threonine kinase present in addition to STK36 is essential for the up-regulation of GLI transcriptional activity. It phosphorylates GLI1 in both N (1-426) and C (754 -1126) terminus but the fragment of gli1 between residues 426 -754 is not phosphorylated by ULK3. Thus ULK3 is a positive activator [343].
STK36 and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	STK36 is a positive regulator of SHH pathway that acts independent on is functional kinase domain. STK36 enhances GLI2 activity but not GLI1 in C3H10T1/2 and HEK293 cells and Gli1 transcriptional activity in NIH3T3C2. Sn480 cells. Hence STK36 expression is cell type specific. As ours is a master model we have nevertheless included this interaction [10].
GLI2 and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI1	GLI1 is a direct target of GLI2. The study was conducted in normal human epidermis and Basal cell carcinoma cells [613].
GLI3_A and not SKI and not SNO and not NCOR and not SIN3A and not HDAC➔GLI1	GLI3 exists in two forms — a full-length transcriptional activator (GLI3A) or an amino-terminal

	fragment that functions as a repressor. This particular activator isoform is GLI3A [320]. On the other hand SKI, SKI related protein SNO, NCOR, SIN3A, HDACs form a transcriptional repressor complex that interfere with GLI1 activation by full length GLI3 [355]. Hence they are included in the AND interaction.
GLI1➔FOXM1	FOXM1/FOXL1 is a direct target of GLI mediated activation [337].
FOXM1➔CELL_PROLIFERATION	FOXM1 is known to mediate cell proliferative functions [614].
CELL_PROLIFERATION➔	Cellular response in our model.
not GLI1➔GLI3_R	Repressor form of GLI3 produced in the absence of GLI1 expression [406].
not GLI3_R ➔ GLI3_ACTIVE	GLI3_ACTIVE form is produced in the absence of GLI3 repressor form GLI3R [406].
GLI1➔PDGFRA	PDGFRA is expressed at high levels in human and murine Basal Cell Carcinoma. It has been found that ectopic expression of GLI increases PDGFRA levels i.e. increases receptor protein levels whereas inhibition of the HH pathway reduces PDGFRA levels [615].
PDGFRA➔	
GLI1➔NUC_GLI1	Cellular location of GLI1 has been found in Cytoplasm as well as in Nucleus. We considered this transportation in our model and named NUC_GLI1 of the nuclear counterpart of GLI1.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔PTCH1	PTCH 1 and HHIP genes transcribed as the transcriptional targets of GLI1 transcription factor in Hedgehog pathway.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔HHIP	

NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔GLI1	It is reported GLI1 is also produced at the end of this pathway and thus create a positive feedback loop [405,616].
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔OPN	Osteopontin (OPN) is a direct transcriptional target of GLI1demonstrated in MDA-MB 435 cell line. OPN is a secreted protein that influences multiple downstream signaling events that allow cancer cells to resist apoptosis, invade through ECM, evade host immunity and influence growth of indolent tumors. OPN is expressed by normal cells. However sustained expression in cancer cells promotes aberrant growth of cells and an invasive phenotype [617].
OPN ➔	
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔CYCLIN_D	Hedgehog signaling regulates the proliferation of distinct cell types via direct activation of genes that are involved in cell cycle progression and mediate G1 to S transition.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔CYCLIN_E	
CYCLIN_D➔CELLCYCLE_PROGRESSION	Cyclins such as CYCLIN D and CYCLIN E are involved in regulation of cell cycle.
CYCLIN_E➔CELLCYCLE_PROGRESSION	
CELLCYCLE_PROGRESSION➔	Cell cycle progression is one of the cellular responses that have been considered in our model.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔CMYC	Expression of GLI1 and C-MYC has also found in various experiments [618].
CMYC➔	
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔BMI	Hedgehog signalling leads to an increased expression of BMI-1 in isolated mammary epithelial stem cells and CSCs. BMI-1 is a transcriptional repressor belonging to the polycomb gene family and its suppressor functions are involved in maintaining neuronal, haematopoietic and mammary

	gland stem cells. It leads to self renewal [612]. Activated STK36 also phosphorylates SUFU to promote the nuclear accumulation of full length GLI [337,612]. It is also reported that activation of hedgehog signaling increases mammosphere- initiating cell number and mammosphere size, whereas inhibition of the pathway results in a reduction of these effects. These effects are mediated by the polycomb gene *bmi* [619].
BMI➔	Output protein considered in our model.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔SNAI1	It is reported that activation of SNAI1 a protein responsible for degradation of cadherin and induction of invasion is directly activated by GLI1 [612]. SNAI1 protein is responsible for epithelial to mesenchymal transition.
SNAI1➔EMT	
EMT➔	Cellular response Epithelial to Mesenchymal Transition.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔JAGGED2	Expression of JAGGED2 has been reported [405].
JAGGED2➔NOTCH_SIGNAL	JAGGED2 is a notch ligand, hence promotes notch signaling.
NOTCH_SIGNAL➔	Cellular response.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔SFRP	Expression of SFRP by GLI1 has been found [405,620].
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔WNT	GLI1 mediates the activation of wnt family proteins and enhances signaling via these pathways, exact wnt ligand is not known and thus not included in our model [406].
WNT and not SFRP➔WNT_SIGNAL	Activation of Wnt signaling depends on the presence of WNT ligand and absence of its antagonist SFRP [621].
WNT_SIGNAL➔	Cellular response.
NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH➔BCL2	In epidermal cells GLI1 can induce the expression of antiapoptotic factor BCL2 [612].
HFU and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI2	HFU enhances GLI2 function in a manner that is independent of a functional kinase domain [316].

STK36 and not PKA_A and not GSK3 and not CKI_A and not BTRCP and not SUFU➔GLI2	STK36 is a positive regulator of GLI2 activity [343]. STK36 enhances GLI2 activity but not GLI1 in C3H10T1/2 and HEK293 cells. GSK3 phosphorylates GLI proteins post phosphorylation by PKA and it is known to elicit negative effects. On the other hand SMO inactivation leads to formation of the cytoplasmic GLI degradation complex, in which GLI family members (GLI1, GLI2 and GLI3) are phosphorylated by casein kinase I (CKI), glycogen synthase kinase-3ß (GSK3ß) and protein kinase A (PKA) [337]. Phosphorylated GLI is recognized by FBXW1/BTRCP1 and FBXW11/BTRCP2 for ubiquitination, and ubiquitinated GLI is partially degraded to release its intact N-terminal half functioning as transcriptional repressor. Thus all the above factors are included in the AND equation as they influence the formation of GLI2
HFU and not NOTCH1➔GLI2	HFU enhances GLI2 function in a manner that is independent of a functional kinase domain [316]. Inactivation of notch1 gene in epidermis induces sustained expression of GLI2 and causes Basal Cell Carcinoma [614].
STK36 and not NOTCH1➔GLI2	We have represented an alternative mode of GLI2 activation by STK36 without the presence of NOTCH1.
GLI2➔NUC_GLI2	GLI2 when transported to the nucleus is represented as NUC_GLI2.
NUC_GLI2 and NUC_STK36 and not NUC_SUFU➔CYCLIN_D2	However, this interaction has been found in a murine model [612,622], we included this interaction as CYCLIN_D2 is one of the important proteins for cell cycle progression. Activated STK36 also phosphorylates SUFU to promote the nuclear accumulation of full length GLI.
CYCLIN_D2➔CELLCYCLE_PROGRESSION	CYCLIN D2 is implicated in cell cycle regulation.

NUC_GLI2 and NUC_STK36 and not NUC_SUFU➔BCL2	Epidermal cells GLI2 can induce the expression of anti apoptotic factor BCL2 [612].
BCL2➔ANTI_APOP	BCL2 is a known anti-apoptotic factor [220].
ANTI_APOP➔	Cellular responses.
not GLI1 ➔ CTNNB_TCF4	Coincident of high-to-low TCF and low-to-high HH-GLI1 transitions in patient Colon Carcinoma have been found. Therefore, we can write that higher level of expression of GLI1 inhibits the activity of TCF complex [623].
CTNNB_TCF4 ➔	
GLI3_A and NUC_STK36 and not NUC_SUFU➔CYCLIN_D2	Similarly active form of GLI3 mediates the activation of CYCLIN_D2.
not PTCH1➔CYCLIN_B	It is reported that PTCH regulates the activity of CYCLIN_B. Interaction with Patched in the cytoplasm blocks cell proliferation by preventing nuclear localization of the activated complex. Ligand induced activation of this complex leads to the nuclear localization of CYCLIN B by disruption of the physical interaction between Patched 1 and CYCLIN B [615].
CYCLIN_B➔CELLCYCLE_PROGRESSION	CYCLIN B is also implicated in cell cycle regulation.

Appendix Table 6: Logical States of the Input and Output Proteins of Hedgehog Signaling in Normal, Glioma, Colon and Pancreatic Cancer Scenarios.

Normal				Glioma				Colon				Pancreatic			
Input		Output		Input		Output		Input		Output		Input		Output	
BMP_RUNX3	0	DHH	0	SHH	1	DHH	1	SHH	1	DHH	0	DHH	1	DHH	1
DISPATCHED	1	IHH	0	DHH	1	IHH	1	IHH	1	SHH	1	SHH	1	SHH	1
HHAT	1	SHH	1	IHH	1	SHH	1	BMP_RUNX3	1	IHH	1	IHH	1	IHH	
CDO	1	PTCH1	1	BMP_RUNX3	1	PTCH1	0	DISPATCHED	1	PTCH1	1	BMP_RUNX3	1	PTCH1	1
BOC	1	PTCH2	0	DISPATCHED	1	PTCH2	0	HHAT	1	PTCH2	0	DISPATCHED	1	PTCH2	0
GAS1	0	SMO	1	HHAT	1	SMO	0	CDO	1	SMO	1	HHAT	1	SMO	1
HFU	0	STK36	1	CDO	1	STK36	0	BOC	1	STK36	1	CDO	1	STK36	1
ULK3	0	GLI1	1	BOC	1	GLI1	0	GAS1	0	GLI1	1	BOC	1	GLI1	1
NOTCH1	0	GLI2	1	GAS1*	0	GLI2	0	HFU	1	GLI2	1	GAS1	0	GLI2	1
SUFU	0	NUC_GLI1	1	GLI1	0	NUC_GLI1	0	ULK3	1	NUC_GLI1	1	HFU	1	NUC_GLI1	1
TWIST	0	NUC_GLI2	1	GLI2	0	NUC_GLI2	0	NOTCH1	0	NUC_GLI2	1	ULK3	1	NUC_GLI2	1
RAS	0	GLI3_A	1	HFU	1	GLI3_A	0	SUFU*	0	GLI3_A	1	NOTCH1	0	GLLI3_A	1
ERK12	0	GLI3_R	0	ULK3	1	GLI3_R	1	TWIST	0	GLI3_R	0	SUFU*	0	GLI3_R	0
PKA_A	0	FAS	0	NOTCH1	0	FAS	1	RAS	1	FAS	0	TWIST	0	FAS	0
BTRCP	0	CYCLIN_B	1	SUFU*	0	CYCLIN_B	0	ERK12	0	CYCLIN_B	1	RAS	1	CYCLIN_B	1
CKI_A	0	CYCLIN_D	1	TWIST	1	CYCLIN_D	0	PKA_A*	0	CYCLIN_D	1	ERK12	1	CYCLIN_D	1
GSK3	0	CYCLIN_D2	1	RAS	1	CYCLIN_D2	0	BTRCP*	0	CYCLIN_D2	1	PKA_A*	0	CYCLIN_D2	1
DYRK1	1	CYCLIN_E	1	ERK12	1	CYCLIN_E	0	CKI_A*	0	CYCLIN_E	1	BTRCP*	0	CYCLIN_E	1
NUMB	0	FOXM1	1	PKA_A*	0	FOXM1	0	GSK3*	0	FOXM1	1	CKI_A*	0	FOXM1	1
ITCH	0	PDGFRA	1	BTRCP*	0	PDGFRA	0	DYRK1	1	PDGFRA	1	GSK3*	0	PDGFRA	1
SKI	0	CTNNB_TCF4	0	CKI_A*	0	CTNNB_TCF4	1	NUMB*	0	CTNNB_TCF4	0	DYRK1	1	CTNNB_TCF4	0
NCOR	0	OPN	1	GSK3*	0	OPN	0	ITCH*	0	OPN	1	NUMB*	0	OPN	1
HDAC	0	CMYC	1	DYRK1	1	CMYC	0	SKI*	0	CMYC	1	ITCH*	0	CMYC	1
SNO	0	BMI	1	NUMB*	0	BMI	0	NCOR*	0	BMI	1	SKI*	0	BMI	1
SIN3A	0	SNAI1	1	ITCH*	0	SNAI1	0	HDAC*	0	SNAI1	1	NCOR*	0	SNAI1	1
NUC_STK36	1	JAGGED2	1	SKI*	0	JAGGED2	0	SNO*	0	JAGGED2	1	HDAC*	0	JAGGED2	1
NUC_SUFU	0	SFRP	1	NCOR*	0	SFRP	0	SIN3A*	0	SFRP	1	SNO*	0	SFRP	1
NA	NA	WNT	1	HDAC*	0	WNT	0	NUC_STK36	1	WNT	1	SIN3A*	0	WNT	1
NA	NA	BCL2	1	SNO*	0	BCL2	0	NUC_SUFU*	0	BCL2	1	NUC_STK36	1	BCL2	1
NA	NA	NA	NA	SIN3A*	0	HHIP	1	NA	NA	HHIP	1	NUC_SUFU*	0	NA	NA
NA	NA	NA	NA	NUC_STK36	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NUC_SUFU*	0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

NA: Not Applicable *** Proteins having Loss of Function in cancer scenario.**

Appendix Table 7: Master Logical Model used for Notch Pathway Simulations.

LOGICAL EQUATIONS	DOCUMENTATION
INPUTS	
JAG1,JAG2, DLL1, DLL3, DLL4, MAGP1, MAGP2,NOV, CNTN1,PRESENILIN1,NICASTRIN, APH1, PEN2, FURIN, NEDD4,ITCH, NUMB,ALPHA_ADAPTIN, O_GLUCOSE, POGLUT_1,XYL, XYLE, NGA,O_FUCOSE, FRINGE, GALACTOSE, GASE,POFUT_1,JIP1,RAS,DVL, JAK2, STAT3,GSK_3BETA, WDR12,P53,FBW7,CDK8, CYCC,DTX1, MAML, EP300, SKIP, HAT,SMAD3, CSL, SMRT, SAP30, HDAC, CIR, SIN3A, YY1, TACE,NICD_ACTIVE	Input proteins of our logical model.
INTERMEDIATE REACTIONS	
JAG1+NOTCH1+TACE=NECD1 JAG1+NOTCH1+TACE=NEXT1 JAG1+NOTCH2+TACE=NECD2 JAG1+NOTCH2+TACE=NEXT2 JAG1+NOTCH3+TACE=NECD3 JAG1+NOTCH3+TACE=NEXT3 JAG1+NOTCH4+TACE=NECD4 JAG1+NOTCH4+TACE=NEXT4	NOTCH receptors (NOTCH1, NOTCH2, NOTCH3, NOTCH4) bind with membrane bound ligand JAG1. Followed by this interaction, a metallo-protease enzyme TACE (TNFalpha-converting enzyme) cleaves the NOTCH receptors and produces NECD (Notch extracellular domain1) and NEXT (Notch Extra cellular Truncated Protein) [624,625].
JAG2+NOTCH1+TACE=NECD1 JAG2+NOTCH1+TACE=NEXT1 JAG2+NOTCH2+TACE=NECD2 JAG2+NOTCH2+TACE=NEXT2 JAG2+NOTCH3+TACE=NECD3 JAG2+NOTCH3+TACE=NEXT3 JAG2+NOTCH4+TACE=NECD4 JAG2+NOTCH4+TACE=NEXT4	NOTCH receptors (NOTCH1, NOTCH2, NOTCH3, NOTCH4) bind with membrane bound ligand JAG2. Followed by this interaction, a metallo-protease enzyme TACE (TNFalpha-converting enzyme) cleaves the NOTCH receptors and produces NECD (Notch extracellular domain 1) and NEXT (Notch Extra cellular Truncated Protein) [624,625].

DLL1+NOTCH1+TACE=NECD1 DLL1+NOTCH1+TACE=NEXT1 DLL1+NOTCH2+TACE=NECD2 DLL1+NOTCH2+TACE=NEXT2 DLL1+NOTCH3+TACE=NECD3 DLL1+NOTCH3+TACE=NEXT3 DLL1+NOTCH4+TACE=NECD4 DLL1+NOTCH4+TACE=NEXT4	NOTCH receptors (NOTCH1, NOTCH2, NOTCH3, NOTCH4) bind with membrane bound ligand DLL1. Followed by this interaction, a metallo-protease enzyme TACE (TNFalpha-converting enzyme) cleaves the NOTCH receptors and produces NECD (Notch extracellular domain 1) and NEXT (Notch Extra cellular Truncated Protein) [624,625].
DLL3+NOTCH1+TACE=NECD1 DLL3+NOTCH1+TACE=NEXT1 DLL3+NOTCH2+TACE=NECD2 DLL3+NOTCH2+TACE=NEXT2 DLL3+NOTCH3+TACE=NECD3 DLL3+NOTCH3+TACE=NEXT3 DLL3+NOTCH4+TACE=NECD4 DLL3+NOTCH4+TACE=NEXT4	NOTCH receptors (NOTCH1, NOTCH2, NOTCH3, NOTCH4) bind with membrane bound ligand DLL3. Followed by this interaction, a metallo-protease enzyme TACE (TNFalpha-converting enzyme) cleaves the NOTCH receptors and produces NECD (Notch extracellular domain 1) and NEXT (Notch Extra cellular Truncated Protein) [624,625].
DLL4+NOTCH1+TACE=NECD1 DLL4+NOTCH1+TACE=NEXT1 DLL4+NOTCH2+TACE=NECD2 DLL4+NOTCH2+TACE=NEXT2 DLL4+NOTCH3+TACE=NECD3 DLL4+NOTCH3+TACE=NEXT3 DLL4+NOTCH4+TACE=NECD4 DLL4+NOTCH4+TACE=NEXT4	NOTCH receptors (NOTCH1, NOTCH2, NOTCH3, NOTCH4) bind with membrane boundligand DLL4. Followed by this interaction, a metallo-protease enzyme TACE (TNFalpha-converting enzyme) cleaves the NOTCH receptors and produces NECD (Notch extracellular domain1) and NEXT (Notch Extra cellular Truncated Protein) [624,625].
NEXT1+GAMMA_SECRETASE=NICD1 NEXT2+GAMMA_SECRETASE=NICD2 NEXT3+GAMMA_SECRETASE=NICD3 NEXT4+GAMMA_SECRETASE=NICD4	Notch extracellular truncated domains (NEXT1, NEXT2, NEXT3 and NEXT4) are cleaved by intracellular proteolytic enzyme called Gamma_Secretase and produces Notch intracellular domains NICD1, NICD2, NICD3 and NICD4 [624,625].
PRESENILIN1+NICASTRIN+APH1+PEN2 =GAMMA_SECRETASE	The component proteins of GAMMA_SECRETASE are PRESENILIN1, NICASTRIN, APH1 and PEN2. A charged aspartate in 19 residues long trans-membrane domain of PRESENILIN1 helps to stabilize the GAMMA_SECRETASE enzyme complex [626].

MAGP1+NOTCH1=NEXT1 MAGP1+NOTCH1=NECD1 MAGP2+NOTCH1=NEXT1 MAGP2+NOTCH1=NECD1	MAGP1 and MAGP2 proteins, present on microfibrils can interact with NOTCH1 and form NEXT1 and NECD1 by a furin-like cleavage without the help of TACE metallo- protease enzyme [461].
NOV+NOTCH1=NEXT1 NOV+NOTCH1=NECD1	Nephroblastoma overexpressed protein (NOV) associates with NOTCH1 and induces the subsequent release of Notch extracellular proteins (NEXT1 and NECD1) [462].
CNTN1+NOTCH1=NECD1 CNTN1+NOTCH1=NEXT1 CNTN1+NOTCH2=NEXT2 CNTN1+NOTCH2=NECD2	Trans-extracellular interaction between F3/Contactin (CNTN1) and NOTCH1 or NOTCH2 can trigger the notch signaling pathway [463,464].
FURIN+!NUMB+!ITCH+ALPHA_ADAPTIN +NOTCH1_PRE =NOTCH1 FURIN+!NUMB+!ITCH+ALPHA_ADAPTIN +NOTCH2_PRE=NOTCH2	During maturation procedures, pre-processed NOTCH1 and NOTCH2 molecules (NOTCH1_PRE and NOTCH2_PRE) are cleaved by FURIN like protease and form the processed NOTCH molecules (NOTCH1 and NOTCH2) for further ligand binding and signal transduction [8, 9]. Onco suppressor protein NUMB, with the help of ITCH or ALPHA_ADAPTIN, promotes the degradation of NOTCH1_PRE and NOTCH2_PRE (but not NOTCH3_PRE or NOTCH4_PRE) by recruiting the E3 ubiquitin ligase [627,628].
!NEDD4+NOTCH1_PRE=NOTCH1	Pre-processed NOTCH1 (NOTCH1_PRE) is the direct target of ubiquitin-protein ligase NEED4. Overexpression of NEDD4 in atrophy muscle cell cause down-regulation of NOTCH1 as ubiquitination causes

GSK_3BETA+!DVL+!JIP1 +NICD1 =NUC_NICD1 GSK_3BETA+!DVL+!JIP1 +NICD2 =NUC_NICD2 GSK_3BETA+!DVL+!JIP1 +NICD3 =NUC_NICD3 GSK_3BETA+!DVL+!JIP1 +NICD4 =NUC_NICD4	GSK_3BETA phosphorylates NICD and then phosphorylated NICD goes into the nucleus for further transcription process. For simplicity the phosphorylated NICD are not considered in this model [14]. On the other hand it has also been found that DVL, JIP1 and P53 proteins can also exert inhibitory effect on Notch intracellular domains in cytoplasm (NICD1, NICD2, NICD3 and NICD4) [466,467,469].
RAS+NICD1=NUC_NICD1	Experimental findings have proven the cross talkbetween RAS/MAPK pathways with NOTCH1 intracellular domains. This cross talk results the activation of Notch pathway in various cancer cell line including Glioma, Breast cancer etc. [486].
!P53_P+NUC_NICD1+CSL=NOTCH1_PRE !P53_P+NUC_NICD2+CSL=NOTCH2_PRE !P53_P+NUC_NICD3+CSL=NOTCH3_PRE !P53_P+NUC_NICD4+CSL=NOTCH4_PRE	P53 the tumor suppressor protein has found to bet the suppressor of NOTCH proteins in Glioblastoma cell line. P53 have been considered as the transcription repressor of NUC_NICD and CSL and thus reducing the concentration of NOTCH precursor proteins [487-490].
P53+!NICD1=P53_P	Activated NOTCH1 (or NICD1) interacts with P53 and inhibits its phosphorylation [496].
WDR12+NICD1=NUC_NICD1	WD-repeat protein contains NLS sequence has been found to interact with Notch1 intracellular domain (NICD1). Although the end result of this interaction is still not known, but it is quite intuitive that WDR12 may help to the nuclear translocation of NICD1 from cytoplasm and thereby modulate NOTCH signaling pathway [631].

!FBW7+NICD4=NUC_NICD4	FBW7 expressed in mouse embryo is found to negatively regulate the NOTCH4-HEY1 dependent pathway. The FBW7 degrades intracellular domain of NOTCH4 through its ubiquitin ligase mediated activity [632].
POGLUT_1+O_GLUCOSE+NOTCH1_PRE = NOTCH1 POGLUT_1+O_GLUCOSE+NOTCH2_PRE = NOTCH2 POGLUT_1+O_GLUCOSE+NOTCH3_PRE = NOTCH3 POGLUT_1+O_GLUCOSE+NOTCH4_PRE = NOTCH4	Post-translational modification of NOTCH precursor proteins with O-linked glucose (O_GLUCOSE) molecule by Protein O-glucosyltransferase -1 is a conserved process. This modification is found to be required for NOTCH pathway activation and ligand binding [470].
XYL+O_GLUCOSE+!XYLE+NOTCH1_PRE =NOTCH1 XYL+O_GLUCOSE+!XYLE+NOTCH2_PRE =NOTCH2 XYL+O_GLUCOSE+!XYLE+NOTCH3_PRE =NOTCH3 XYL+O_GLUCOSE+!XYLE+NOTCH4_PRE =NOTCH4	Addition of Xylose (XYL) molecule to the O- GLUCOSE linked NOTCH precursor proteins is mediated by an enzyme Xylosyltransferase (XYLE). Loss or gain of function of XYLE has strongly suggested that Xylose modification is negatively correlated with the notch pathway activation [470].
O_FUCOSE+NGA+FRINGE+POFUT_1 +NOTCH1_PRE=NOTCH1 O_FUCOSE+NGA+FRINGE+POFUT_1 +NOTCH2_PRE=NOTCH2 O_FUCOSE+NGA+FRINGE+POFUT_1 +NOTCH3_PRE=NOTCH3 O_FUCOSE+NGA+FRINGE+POFUT_1 +NOTCH4_PRE=NOTCH4	FRINGE catalyses the addition of N-acetylglucosamine (NGA) to O-fucose in NOTCH precursor proteins. NGA modification plays positive role for ligand receptor binding in Notch signaling pathway [471]. Fucosylation of Notch molecules is mediated by the enzyme POFUT_1 (GDP-fucose protein O-fucosyltransferase 1) [472].
GALACTOSE+GASE+ O_FUCOSE+ NOTCH1_PRE=NOTCH1 GALACTOSE+GASE+ O_FUCOSE +NOTCH2_PRE=NOTCH2 GALACTOSE+GASE+ O_FUCOSE +NOTCH3_PRE=NOTCH3 GALACTOSE+GASE+ O_FUCOSE +NOTCH4_PRE=NOTCH4	GALACTOSE addition to O_FUCOSE linked Notch precursors molecules are mediated by the enzyme GASE (Galactosyltransferase) [473].

NUC_NICD1+YY1=MYC	NUC_NICD1 interacts directly with YY1 transcription factor and regulates the expression of MYC protein [633].
NUC_NICD1+SMAD3+CSL=HES1	NUC_NICD1 and SMAD3 are seen to interact directly and thereafter regulate the expression of HES1 through CSL [483].
HDAC+SAP30+CIR+SIN3A+SMRT=COR EP300+MAML+ HAT +SKIP=COA	On the other hand, the proteins HDAC, SMRT, CIR, SAP30, SIN3A forms a co-repressor complex (COR) of CSL which in turn regulates the expression of Notch target genes [474-477].
NUC_NICD1=NICD_ACTIVE NUC_NICD2=NICD_ACTIVE NUC_NICD3=NICD_ACTIVE NUC_NICD4=NICD_ACTIVE NICD_ACTIVE+CSL+!COR+COA= HES1/HES5/HES7/HEY1/HEY2/HEYL /GATA3/CCND3/CCND1/CD44/KLF5/SOX9 /PTCRA/MKP_1/NFKB/	In order to reduce the complexity of the model, a dummy node NICD_ACTIVE has been considered in place of all NUC_NICD1, 2, 3 and 4. This dummy species is not shown in the main figure. Transcription co-activator complex (COA), consisting of CSL, NICD, Mastermind (MAML), EP300 and histone acetyltransferase (HAT) induces the transcriptional activation of several Notch target genes, such as HES1, HES5, HES7, HEY1, HEY2, HEYL, GATA3, CCND1, CCND3, CD44, KLF5, SOX9, NFKB, [33,478-482,484,634].
NUC_NICD1+COA+CSL+!COR= BCL2 /FLIP/IAP/ P21/P65/P50/C_REL/REL_B	Nuclear NICD1 (NUC_NICD1) has found to activate the anti-apoptosis proteins BCL2, FLIP, IAP as well as other NFkB pathway proteins P65, P50, C_REL, REL_B [484,635]. It also induces the expression of growth arrest factor P21 in primary differentiating keratinocytes cell lines [45].

NUC_NICD1/2+DTX1+CSL+!COR+COA =MAG	F3/contactin trans-extracellular ligand dependent NOTCH pathway promotes oligodendrocyte precursor cell differentiation and upregulates the myelin-related protein MAG. NOTCH1/2 and DTX1 mediated signaling cascade with the help of transcription factor CSL induces the transcription of MAG in OLN-93 cell line [463].
!HES1=MYOD	Ligand-induced Notch signaling in myeloma cell up-regulates HES1 mRNA expression and subsequently reduced expression of MYOD [636].
MAML+!CDK8+!CYCC+!FBW7 =NUC_NICD1/2/3/4	MAML directly interacts with CDK8 and recruits it to hyper-phosphorylate the NICD in nucleus. Followed by the hyper-phosphorylation, NICD undergoes FBW7 dependent ubiquitin degradation [637].
NICD_ACTIVE+CSL+!COR+COA =NRARP	NRARP is the notch target gene which is transcribed by the CSL dependent NOTCH pathway activation [497].
!NRARP+NICD1/2/3/4=NUC_NICD1/2/3/4	NRARP is found to form a ternary complex with NICD in cytoplasm which in turn inhibits the further NICD dependent transcription. This is one of the identified negative feedback loop in NOTCH signaling pathway [497].
NUC_NICD1=CDK2	NUC_NICD1 induces the activation of CDK2 [479].
HES1/5+JAK2+STAT3=STAT3_P STAT3_P=NUC_STAT3	HES1 and HES5 are found to interact with JAK2 and STAT3, and facilitate the complex formation between JAK2/STA3. This complex formation promotes the phosphorylation of STAT3 [445]. Phosphorylated STAT3_P then translocate into the nucleus.

NUC_STAT3=HIF1A	NUC_STAT3 is found to activate HIF1A [491].
HIF1A=NICD1/2/3/4	HIF1A can interact with NICD1/2/3/4 to enhance the NOTCH pathway activity by up regulating the NOTCH pathway target genes [495].
!HES1=PTEN !PTEN=PI3K PI3K=AKT	NOTCH pathway is found to activate the PTEN/AKT pathway by upregulating HES1 production. HES1 is found to inhibit the PTEN dependent suppression of AKT activation [492].

OUTPUT MOLECULES	
NECD1, NECD2, NECD3, NECD4, AKT, CDK2, HEY1, HEY2, MAG, NFKB, MYOD, GATA3, CD44, P21, KLF5, PTCRA, MYC, HES7, HEYL, MKP_1, CCND3, CCND1, FLIP, IAP, BCL2, SOX9, P65, P50, C_REL, REL_B,	Output molecules of the model.

Here '+' sign in the logical equations signifies the 'AND' operation instead of conventional 'OR' logical operator. In CellNetAnalyzer the input equations should contain '+' sign to signify the AND relation among the nodes. Nodes related with OR operations are given by individual logical equations.

Appendix Table 8: Logical Expressions of the Input Molecules used for the Simulation of Notch Pathway under Different Scenarios

MOLECULES	EXPERIMENT (GBE)	SIMULATION (GBS)	GAMMA_SECRETASE INHIBITION	NORMAL NOTCH SCENARIO	TS2	TS1
JAG1	1	1	1	1	1	1
JAG2	0	0	0	1	0	0
DLL1	0	0	0	1	0	0
DLL3	0	0	0	1	0	0
DLL4	N	1	N	1	1	1
MAGP1	1	1	1	0	1	1
MAGP2	N	0	N	0	0	0
NOV	N	1	N	0	1	1
CNTN1	0	0	0	0	0	0
TACE	1	1	1	1	1	1
PRESENILIN1	0*	1	0	1	1	1
NICASTRIN	1	1	1	1	1	1
APH1	1	1	1	1	1	1
PEN2	1	1	1	1	1	1
FURIN	N	1	N	0	1	1
NEDD4	1	1	1	0	1	1
ITCH	1	1	1	0	1	1
NUMB	N	0	N	0	0	0
ALPHA_ADAPTIN	N	0	N	1	0	0
O_GLUCOSE	N	1	N	1	1	1
POGLUT_1	1	1	1	1	1	1
XYL	N	0	N	0	0	0
XYLE	0	0	0	1	0	0
NGA	N	1	N	1	1	1
O_FUCOSE	N	1	N	1	1	1
FRINGE	1	1	1	1	1	1
GALACTOSE	N	1	N	1	1	1
GASE	1	1	1	1	1	1
POFUT_1	1	1	1	1	1	1
JIP1	0	0	0	1	0	0
RAS	0	0	0	1	0	0
DVL	0	0	0	0	0	0
JAK2	N	1	N	0	1	1
STAT3	1	1	1	0	1	1
GSK_3BETA	0	0	0	0	0	0
WDR12	1	1	1	1	1	1
P53	1	1	1	1	1	1
FBW7	0	0	0	0	0	0
CDK8	0	0	0	0	0	0
CYCC	N	0	N	0	0	0
DTX1	0	0	0	0	0	0
MAML	1	1	1	1	0	1
EP300	N	1	N	1	1	1
SKIP	N	1	N	1	1	1
HAT	1	1	1	1	1	1
SMAD3	N	1	N	1	1	1
CSL	N	1	N	1	1	1
SMRT	1	1	1	0	1	1
SAP30	1	1	1	0	1	1
HDAC	N	0	N	0	0	0
CIR	0	0	0	0	0	0
SIN3A	N	0	N	0	0	0
YY1	0	0	0	1	0	0

*** The expression of PRESENILIN1 was found "down regulated" in Glioblastoma cells, but at the time of Glioblastoma Simulation (GBS), logical expression of PRESENILIN1 was considered as '1 to show up regulation of GAMMA SECRETASE. "N" signifies the non significant expression level.**

Appendix Table 9: Notch Signaling Model Constructed for the Analyses of Neural Stem Cell Dynamics

Logical equations of the Notch pathway species
AKT (t+1) = (PI3K(t))
BAX (t+1) = (P53_P(t))
BCL2 (t+1) = (NUC_NICD1(t) and COA(t) and CSL(t) and not COR(t))
CCND1 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
CCND3 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
CD44 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
CDK2 (t+1) = (NICD_ACTIVE(t))
MAML+P53_P (t+1) = (MAML(t) and P53_P(t))
COA (t+1) = (EP300(t) and MAML(t) and HAT (t) and SKIP(t))
COR (t+1) = (HDAC(t) and SAP30(t) and CIR(t) and SIN3A(t) and SMRT(t))
FLIP (t+1) = (NUC_NICD1(t) and COA(t) and CSL(t) and not COR(t))
GAMMA_SECRETASE (t+1) = (PRESENILIN1(t) and NICASTRIN(t) and APH1(t) and PEN2(t)) or (NICASTRIN(t) and APH1(t) and PEN2(t))
HES1_MRNA (t+1) = (NUC_NICD1(t) and SMAD3(t) and CSL(t) and not HES1(t)) or (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t) and not MAML+P53_P(t) and not HES1(t))
HES1 (t+1) = (HES1_MRNA(t))
HES5_MRNA (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t) and not HES5(t))
HES5 (t+1) = (HES5_MRNA(t))
HES7 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
HEY1 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
HEY2 (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
HEYL (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t) and not MAML+P53_P(t))
HIF1A (t+1) = (NUC_STAT3(t))
IAP (t+1) = (NUC_NICD1(t) and COA(t) and CSL(t) and not COR(t))
MYC (t+1) = (NUC_NICD1(t) and YY1(t))
NECD1 (t+1) = (JAG1(t) and NOTCH1(t) and TACE(t)) or (JAG2(t) and NOTCH1(t) and TACE(t)) or (DLL1(t) and NOTCH1(t) and TACE(t)) or (DLL3(t) and NOTCH1(t) and TACE(t)) or (DLL4(t) and NOTCH1(t) and TACE(t)) or (MAGP1(t) and NOTCH1(t)) or (MAGP2(t) and NOTCH1(t)) or (NOV(t) and NOTCH1(t)) or (CNTN1(t) and NOTCH1(t))
NECD2 (t+1) = (JAG1(t) and NOTCH2(t) and TACE(t)) or (JAG2(t) and NOTCH2(t) and TACE(t)) or (DLL1(t) and NOTCH2(t) and TACE(t)) or (DLL3(t) and NOTCH2(t) and TACE(t)) or (DLL4(t) and NOTCH2(t) and TACE(t)) or (CNTN1(t) and NOTCH2(t))
NECD3 (t+1) = (JAG1(t) and NOTCH3(t) and TACE(t)) or (JAG2(t) and NOTCH3(t) and TACE(t)) or (DLL1(t) and NOTCH3(t) and TACE(t)) or (DLL3(t) and NOTCH3(t) and TACE(t)) or (DLL4(t) and NOTCH3(t) and TACE(t))

NECD4 (t+1) = (JAG1(t) and NOTCH4(t) and TACE(t)) or (JAG2(t) and NOTCH4(t) and TACE(t)) or (DLL1(t) and NOTCH4(t) and TACE(t)) or (DLL3(t) and NOTCH4(t) and TACE(t)) or (DLL4(t) and NOTCH4(t) and TACE(t))
NEXT1 (t+1) = (JAG1(t) and NOTCH1(t) and TACE(t)) or (JAG2(t) and NOTCH1(t) and TACE(t)) or (DLL1(t) and NOTCH1(t) and TACE(t)) or (DLL3(t) and NOTCH1(t) and TACE(t)) or (DLL4(t) and NOTCH1(t) and TACE(t)) or (MAGP1(t) and NOTCH1(t)) or (MAGP2(t) and NOTCH1(t)) or (NOV(t) and NOTCH1(t)) or (CNTN1(t) and NOTCH1(t))
NEXT2 (t+1) = (JAG1(t) and NOTCH2(t) and TACE(t)) or (JAG2(t) and NOTCH2(t) and TACE(t)) or (DLL1(t) and NOTCH2(t) and TACE(t)) or (DLL3(t) and NOTCH2(t) and TACE(t)) or (DLL4(t) and NOTCH2(t) and TACE(t)) or (CNTN1(t) and NOTCH2(t))
NEXT3 (t+1) = (JAG1(t) and NOTCH3(t) and TACE(t)) or (JAG2(t) and NOTCH3(t) and TACE(t)) or (DLL1(t) and NOTCH3(t) and TACE(t)) or (DLL3(t) and NOTCH3(t) and TACE(t)) or (DLL4(t) and NOTCH3(t) and TACE(t))
NEXT4 (t+1) = (JAG1(t) and NOTCH4(t) and TACE(t)) or (JAG2(t) and NOTCH4(t) and TACE(t)) or (DLL1(t) and NOTCH4(t) and TACE(t)) or (DLL3(t) and NOTCH4(t) and TACE(t)) or (DLL4(t) and NOTCH4(t) and TACE(t))
NICD_ACTIVE (t+1) = (NUC_NICD1(t)) or (NUC_NICD2(t)) or (NUC_NICD3(t)) or (NUC_NICD4(t))
NICD1 (t+1) = (NEXT1(t) and GAMMA_SECRETASE(t))
NICD2 (t+1) = (NEXT2(t) and GAMMA_SECRETASE(t))
NICD3 (t+1) = (NEXT3(t) and GAMMA_SECRETASE(t))
NICD4 (t+1) = (NEXT4(t) and GAMMA_SECRETASE(t))
NOTCH1 (t+1) = (FURIN(t) and not NUMB(t) and not ITCH(t) and ALPHA_ADAPTIN(t) and NOTCH1_PRE(t)) or (not NEDD4(t) and NOTCH1_PRE(t)) or (POGLUT_1(t) and O_GLUCOSE(t) and NOTCH1_PRE(t)) or (XYL(t) and O_GLUCOSE(t) and not XYLE(t) and NOTCH1_PRE(t)) or (O_FUCOSE(t) and NGA(t) and FRINGE(t) and POFUT_1(t) and NOTCH1_PRE(t)) or (GALACTOSE(t) and GASE(t) and O_FUCOSE(t) and NOTCH1_PRE(t))
NOTCH1_PRE (t+1) = (NUC_NICD1(t) and COA(t) and CSL(t) and not COR(t))
NOTCH2 (t+1) = (FURIN(t) and not NUMB(t) and not ITCH(t) and ALPHA_ADAPTIN(t) and NOTCH2_PRE(t)) or (POGLUT_1(t) and O_GLUCOSE(t) and NOTCH2_PRE(t)) or (XYL(t) and O_GLUCOSE(t) and not XYLE(t) and NOTCH2_PRE(t)) or (O_FUCOSE(t) and NGA(t) and FRINGE(t) and POFUT_1(t) and NOTCH2_PRE(t)) or (GALACTOSE(t) and GASE(t) and O_FUCOSE(t) and NOTCH2_PRE(t))
NOTCH2_PRE (t+1) = (NUC_NICD2(t) and COA(t) and CSL(t) and not COR(t))
NOTCH3 (t+1) = (POGLUT_1(t) and O_GLUCOSE(t) and NOTCH3_PRE(t)) or (XYL(t) and O_GLUCOSE(t) and not XYLE(t) and NOTCH3_PRE(t)) or (O_FUCOSE(t) and NGA(t) and FRINGE(t) and POFUT_1(t) and NOTCH3_PRE(t)) or (GALACTOSE(t) and GASE(t) and O_FUCOSE(t) and NOTCH3_PRE(t))
NOTCH3_PRE (t+1) = (NUC_NICD3(t) and COA(t) and CSL(t) and not COR(t))
NOTCH4 (t+1) = (POGLUT_1(t) and O_GLUCOSE(t) and NOTCH4_PRE(t)) or (XYL(t) and O_GLUCOSE(t) and not XYLE(t) and NOTCH4_PRE(t)) or (O_FUCOSE(t) and NGA(t) and FRINGE(t) and POFUT_1(t) and NOTCH4_PRE(t)) or (GALACTOSE(t) and GASE(t) and O_FUCOSE(t) and NOTCH4_PRE(t))
NOTCH4_PRE (t+1) = (NUC_NICD4(t) and COA(t) and CSL(t) and not COR(t))
NOX (t+1) = (P53_P(t))
NRARP (t+1) = (NICD_ACTIVE(t) and CSL(t) and not COR(t) and COA(t))
NUC_NICD1 (t+1) = (GSK_3BETA(t) and not DVL(t) and not JIP1 (t) and NICD1(t)) or (RAS(t) and NICD1(t)) or (WDR12(t) and NICD1(t)) or (MAML(t) and not CDK8(t) and not CYCC(t) and not FBW7(t) and NICD1(t)) or

(not NRARP(t) and NICD1(t)) or (HIF1A(t) and NICD1(t))
NUC_NICD2 (t+1) = (GSK_3BETA(t) and not DVL(t) and not JIP1 (t) and NICD2(t)) or (MAML(t) and not CDK8(t) and not CYCC(t) and not FBW7(t) and NICD2(t)) or (not NRARP(t) and NICD2(t)) or (HIF1A(t) and NICD2(t))
NUC_NICD3 (t+1) = (GSK_3BETA(t) and not DVL(t) and not JIP1 (t) and NICD3(t)) or (MAML(t) and not CDK8(t) and not CYCC(t) and not FBW7(t) and NICD3(t)) or (not NRARP(t) and NICD3(t)) or (HIF1A(t) and NICD3(t))
NUC_NICD4 (t+1) = (GSK_3BETA(t) and not DVL(t) and not JIP1 (t) and NICD4(t)) or (not FBW7(t) and NICD4(t)) or (MAML(t) and not CDK8(t) and not CYCC(t) and not FBW7(t) and NICD4(t)) or (not NRARP(t) and NICD4(t)) or (HIF1A(t) and NICD4(t))
NUC_STAT3 (t+1) = (STAT3_P(t))
P53_P (t+1) = (P53(t) and not AKT(t) and not NICD1(t))
PI3K (t+1) = (not PTEN(t))
PTEN (t+1) = (not HES1(t))
PUMA (t+1) = (P53_P(t))
STAT3_P (t+1) = (HES1(t) and JAK2(t) and STAT3(t)) or (HES5(t) and JAK2(t) and STAT3(t))
BAD (t+1) = (not AKT(t))
TENASCIN_C (t+1) = (NUC_NICD1(t) and COA(t) and CSL(t) and not COR(t)) or (NUC_NICD2(t) and COA(t) and CSL(t) and not COR(t))
GFAP (t+1) = (NUC_STAT3(t))
NGN1 (t+1) = (not DTX1(t) and EP300(t) and MASH1(t))
BETA_TUBULIN_III (t+1) = (NGN1(t))
NESTIN (t+1) = (not HES1(t) and not HES5(t))
MASH1 (t+1) = (not HES1(t) and not HES5(t))
NEUROD (t+1) = (not HES1(t) and not HES5(t))
**APOPTOSIS (t+1) = (not FLIP(t) and not IAP(t) and not BCL2(t) and BAD(t) and PUMA(t) and NOX(t) and BAX(t))*
**NPC DIFFERENTIATION (t+1) = (BETA_TUBULIN_III(t) and NEUROD(t) and NESTIN(t) and not APOPTOSIS(t) and not ASPC DIFFERENTIATION(t) and not GBM DEVELOPMENT(t))*
**ASPC DIFFERENTIATION (t+1) = (GFAP(t) and not APOPTOSIS(t) and not NPC DIFFERENTIATION(t) and not GBM DEVELOPMENT(t))*
**GBM DEVELOPMENT (t+1) = (CCND3(t) and CCND1(t) and CDK2(t) and MYC(t) and TENASCIN_C(t) and GFAP(t) and not APOPTOSIS(t) and not NPC DIFFERENTIATION(t))*
**NSC RENEWAL (t+1) = (CCND3(t) and CCND1(t) and CDK2(t) and not APOPTOSIS(t) and not NPC DIFFERENTIATION(t) and not GBM DEVELOPMENT(t) and not ASPC_PROLIFERATION(t))*

*** Phenotypes of the model. R.H.S = Logical states of the nodes at t^{th} time point. L.H.S = Logical state of the node at $(t+1)^{th}$ time point.**

Appendix Table 10: Detail Descriptions of the Nodes used in the Logical Model

Input Molecules						
Sr. No.	**Names used in the Model**	**Type of the Molecules**	**Full Name/Gene Name of the Molecules**	**Short Name**	**Uniprot ID**	**Ensemble ID**
1	ALPHA_ADAPTIN	Protein	AP-2 complex subunit beta	AP2B1	P63010	ENSG00000006125
2	APH1	Protein	Gamma-secretase subunit APH-1A	APH1A	Q96BI3	ENSG00000117362
3	CDK8	Protein	Cyclin dependent kinase 8	CDK8	P49336	ENSG00000132964
4	CIR	Protein	Corepressor interacting with RBPJ 1	CIR1	Q86X95	ENSG00000138433
5	CNTN1	Protein	Contactin-1	CNTN1	Q12860	ENSG00000018236
6	CSL	Protein	Recombining binding protein suppressor of hairless	RBPJ	Q06330	ENSG00000168214
7	CYC	Protein	Cyclin-C	CCNC	P24863	ENSG00000112237
8	DLL1	Protein	Delta-like protein 1	DLL1	O00548	ENSG00000198719
9	DLL3	Protein	Delta-like protein 3	DLL3	Q9NYJ7	ENSG00000090932
10	DLL4	Protein	Delta-like protein 4	DLL4	Q9NR61	ENSG00000128917
11	DTX1	Protein	Deltex E3 ubiquitin ligase 1	DTX1	Q86Y01	ENSG00000135144
12	DVL1	Protein	Segment polarity protein dishevelled homolog DVL-1	DVL1	O14640	ENSG00000107404
13	EP300	Protein	E1A binding protein p300	EP300	Q09472	ENSG00000100393
14	FBW7	Protein	F-box/WD repeat-containing protein 7	FBXW7	Q969H0	ENSG00000109670
15	FRINGE	Protein	Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe	LFNG	Q8NES3	ENSG00000106003
16	FURIN	Protein	Furin	FURIN	P09958	ENSG00000140564
17	GALACTOSE	Metabolite	Galactose	NA	NA	NA
18	GASE	Protein	β1,4-galactosyl-transferase	B4GALT1	P15291	ENSG00000086062

19	GSK_3BETA	Protein	Glycogen synthase kinase-3 beta	GSK3B	P49841	ENSG00000082701
20	HAT	Protein	Histone acetyltransferase 1	HAT1	Q09472	ENSG00000128708
21	HDAC	Protein	Histone deacetylase 1	HDAC1	Q13547	ENSG00000116478
22	ITCH	Protein	E3 ubiquitin-protein ligase Itchy homolog	ITCH	Q96J02	ENSG00000078747
23	JAG1	Protein	jagged-1	JAG1	P78504	ENSG00000101384
24	JAG2	Protein	jagged-2	JAG2	Q9Y219	ENSG00000184916
25	JAK2	Protein	Tyrosine-protein kinase JAK2	JAK2	O60674	ENSG00000096968
26	JIP1	Protein	C-Jun-amino-terminal kinase-interacting protein 1	MAPK8IP1	Q9UQF2	ENSG00000121653
27	MAGP1	Protein	Microfibril-associated glycoprotein 1	MFAP2	P55001	ENSG00000117122
28	MAGP2	Protein	Microfibril-associated glycoprotein 2	MFAP5	Q13361	ENSG00000197614
29	MAML	Protein	Mastermind-like protein 1	MAML1	Q92585	ENSG00000161021
30	NEDD4	Protein	E3 ubiquitin-protein ligase NEDD4	NEDD4	P46934	ENSG00000069869
31	NGA	Metabolite	N-acetylglucosamine	NA	NA	NA
32	NICASTRIN	Protein	Nicastrin	NCSTN	Q92542	ENSG00000162736
33	NOV	Protein	Nephroblastoma overexpressed	NOV	P48745	ENSG00000136999
34	NUMB	Protein	Protein numb homolog	NUMB	P49757	ENSG00000133961
35	O_FUCOSE	Metabolite	O-Fucose	NA	NA	NA
36	O_GLUCOSE	Metabolite	O-GLUCOSE	NA	NA	NA
37	P53	Protein	Cellular tumor antigen p53	TP53	P04637	ENSG00000141510
38	PEN2	Protein	Presenilin enhancer gamma-secretase subunit	PSENEN	Q9NZ42	ENSG00000205155
39	POFUT_1	Protein	GDP-fucose protein O-fucosyltransferase 1	POFUT1	Q9H488	ENSG00000101346
40	POGLUT_1	Protein	Protein O-glucosyltransferase 1	POGLUT1	Q8NBL1	ENSG00000163389

41	PRESENILIN1	Protein	Presenilin 1	PSEN1	P49768	ENSG00000080815
42	RAS	Protein	GTPase HRas	HRAS	P01112	ENSG00000174775
43	SAP30	Protein	Histone deacetylase complex subunit SAP30	SAP30	O75446	ENSG00000164105
44	SIN3A	Protein	Paired amphipathic helix protein Sin3a	SIN3A	Q96ST3	ENSG00000169375
45	SKIP	Protein	SKI Interacting Protein	SNW1	Q13573	ENSG00000100603
46	SMAD3	Protein	Mothers against decapentaplegic homolog 3	SMAD3	P84022	ENSG00000166949
47	SMRT	Protein	Nuclear receptor corepressor 2	NCOR2	Q9Y618	ENSG00000196498
48	STAT3	Protein	Signal transducer and activator of transcription 3	STAT3	P40763	ENSG00000168610
49	TACE	Protein	Disintegrin and metalloproteinase domain-containing protein 17	ADAM17	P78536	ENSG00000151694
50	WDR12	Protein	Ribosome biogenesis protein WDR12	WDR12	Q9GZL7	ENSG00000138442
51	XYL	Metabolite	Xylose	NA	NA	NA
52	XYLE	Protein	α1,3-xylosyltransferase	XXYLT1	Q8NBI6	ENSG00000173950
53	YY1	Protein	YY1 transcription factor	YY1	P25490	ENSG00000100811
Intermediate Molecules						
1	AKT	Protein	AKT serine/threonine kinase 1	AKT1	P31749	ENSG00000142208
2	CD44	Protein	CD44 antigen	CD44	P16070	ENSG00000026508
3	CDK2	Protein	Cyclin-dependent kinase 2	CDK2	P24941	ENSG00000123374
4	NOTCH1	Protein	NOTCH1	NOTCH1	P46531	ENSG00000148400
5	NOTCH2	Protein	NOTCH2	NOTCH2	Q04721	ENSG00000134250
6	NOTCH3	Protein	NOTCH3	NOTCH3	Q9UM47	ENSG00000074181
7	NOTCH4	Protein	NOTCH4	NOTCH4	Q99466	ENSG00000204301
8	PI3K	Protein	Phosphatidylinositol	PIK3R1	P27986	ENSG00000145675

			3-kinase regulatory subunit alpha			
9	PTEN	Protein	Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase	PTEN	P60484	ENSG00000171862
10	MAML+P53_P	Protein Complex	MAML and P53 proteins complex	NA	NA	NA
11	COA	Protein Complex	Transcription Co-activator Complex	NA	NA	NA
12	COR	Protein Complex	Transcription Co-repressor Complex	NA	NA	NA
13	GAMMA_SECRETASE	Protein Complex	Gamm Secreatase enzyme complex	NA	NA	NA
14	HES1_MRNA	mRNA Molecule	mRNA of HES1 Protein	NA	NA	NA
15	HES5_MRNA	mRNA Molecule	mRNA of HES5 Protein	NA	NA	NA
16	NECD1	Truncated Protein	Notch extra-cellular domain 1	NA	NA	NA
17	NECD2	Truncated Protein	Notch extra-cellular domain 2	NA	NA	NA
18	NECD3	Truncated Protein	Notch extra-cellular domain 3	NA	NA	NA
19	NECD4	Truncated Protein	Notch extra-cellular domain 4	NA	NA	NA
20	NEXT1	Truncated Protein	Notch extra-cellular truncated domain 1	NA	NA	NA
21	NEXT2	Truncated Protein	Notch extra-cellular truncated domain 2	NA	NA	NA
22	NEXT3	Truncated Protein	Notch extra-cellular truncated domain 3	NA	NA	NA
23	NEXT4	Truncated Protein	Notch extra-cellular truncated domain 4	NA	NA	NA
24	NICD1	Truncated Protein	Notch intra-cellular domain 1	NA	NA	NA
25	NICD2	Truncated Protein	Notch intra-cellular domain 2	NA	NA	NA
26	NICD3	Truncated Protein	Notch intra-cellular domain 3	NA	NA	NA
27	NICD4	Truncated Protein	Notch intra-cellular domain 4	NA	NA	NA

28	NICD_ACTIVE*	Hypothetical Node*	Transcriptionally active Notch Intra-cellular domain	NA	NA	NA
29	NOTCH1_PRE	Precursor Protein	Precussor of Notch1 receptor protein	NA	NA	NA
30	NOTCH2_PRE	Precursor Protein	Precussor of Notch2 receptor protein	NA	NA	NA
31	NOTCH3_PRE	Precursor Protein	Precussor of Notch3 receptor protein	NA	NA	NA
32	NOTCH4_PRE	Precursor Protein	Precussor of Notch4 receptor protein	NA	NA	NA
33	NUC_NICD1	Nuclear Counterpart	Nuclear counterpart of NICD1	NA	NA	NA
34	NUC_NICD2	Nuclear Counterpart	Nuclear counterpart of NICD2	NA	NA	NA
35	NUC_NICD3	Nuclear Counterpart	Nuclear counterpart of NICD3	NA	NA	NA
36	NUC_NICD4	Nuclear Counterpart	Nuclear counterpart of NICD4	NA	NA	NA
37	NUC_STAT3	Nuclear Counterpart	Nuclear counter of STAT3 protein	NA	NA	NA
38	P53_P	Phosphorylated Protein	Phosphorylated form of P53 protein	NA	NA	NA
39	STAT3_P	Phosphorylated Protein	Phosphorylated form of STAT3 protein	NA	NA	NA
Target Proteins						
1	BAX	Protein	BCL2 associated X	BAX	Q07812	ENSG00000087088
2	BAD	Protein	Bcl2-associated agonist of cell death	BAD	Q92934	ENSG00000002330
3	BCL2	Protein	Apoptosis regulator Bcl-2	BCL2	P10415	ENSG00000171791
4	CCND1	Protein	G1/S-specific cyclin-D1	CCND1	P24385	ENSG00000110092
5	CCND3	Protein	G1/S-specific cyclin-D3	CCND3	P30281	ENSG00000112576

6	FLIP	Protein	CASP8 and FADD-like apoptosis regulator	CFLAR	O15519	ENSG00000003402
7	HES1	Protein	HES1	HES1	Q14469	ENSG00000114315
8	HES5	Protein	HES5	HES5	Q5TA89	ENSG00000197921
9	HES7	Protein	HES7	HES7	Q9BYE0	ENSG00000179111
10	HEY1	Protein	HEY1	HEY1	Q9Y5J3	ENSG00000164683
11	HEY2	Protein	HEY2	HEY2	Q9UBP5	ENSG00000135547
12	HEYL	Protein	Hairy/enhancer-of-split related with YRPW motif-like protein	HEYL	Q9NQ87	ENSG00000163909
13	HIF1A	Protein	Hypoxia-inducible factor 1-alpha	HIF1A	Q16665	ENSG00000100644
14	IAP	Protein	E3 ubiquitin-protein ligase XIAP	XIAP	P98170	ENSG00000101966
15	MYC	Protein	Myc proto-oncogene protein	MYC	P01106	ENSG00000136997
16	NOX	Protein	NADPH oxidase 1	NOX1	Q9Y5S8	ENSG00000007952
17	NRARP	Protein	Notch-regulated ankyrin repeat-containing protein	NRARP	Q7Z6K4	ENSG00000198435
18	PUMA	Protein	Bcl-2-binding component 3	BBC3	Q9BXH1	ENSG00000105327
19	GFAP	Protein	Glial fibrillary acidic protein	GFAP	P14136	ENSG00000131095
20	NGN1	Protein	Neurogenin-1	NEUROG1	Q92886	ENSG00000181965
21	BETA_TUBULIN_III	Protein	Tubulin beta-3 chain	TUBB3	Q13509	ENSG00000258947
22	NESTIN	Protein	Nestin	NES	P48681	ENSG00000132688
23	MASH1	Protein	Achaete-scute homolog 1	ASCL1	P50553	ENSG00000139352
24	NEUROD	Protein	Neurogenic differentiation factor 1	NEUROD1	Q13562	ENSG00000162992
25	TENASCIN_C	Protein	Tenascin	TNC	P24821	ENSG00000041982
Phenotypes						
1	Apoptosis	NA	Apotosis or Natural Cell death	NA	NA	NA

2	NPC Differentiation	NA	Differentiations of Neural progenitor cells	NA	NA	NA
3	NPC Differentiation	NA	Differentiations of Astrocyte progenitor cells	NA	NA	NA
4	GBM Development	NA	Differentiated and developed Glioblastoma cells	NA	NA	NA
5	NSC Renewal	NA	Neural stem cell renewal	NA	NA	NA

*** Hypothetical node is introduced for the simplification of the model simulation**

Appendix Table 11: The Logical Model of ERBB Pathway.

Nodes	S1	S2	S3	S4	S5	S6	S7
EGF	**0**	**3**	**3**	**0**	**3**	**0**	**3**
ERBB1	0	0	0	0	0	0	0
ERBB1_WT	**2**	**2**	**3**	**2**	**2**	**3**	**3**
EGFRvIII	**0**	**0**	**0**	**3**	**3**	**3**	**3**
ERBB2	0	0	0	0	0	0	0
ERBB2_WT	2	2	2	2	2	2	2
ERBB3	0	0	0	0	0	0	0
ERBB3_WT	2	2	2	2	2	2	2
ER_alpha	0	0	0	0	0	0	0
ERBB2_3	0	0	0	0	0	0	0
AKT1	0	0	0	0	0	0	0
MEK1	0	0	0	0	0	0	0
ERBB1_2	0	0	0	0	0	0	0
ERBB1_3	0	0	0	0	0	0	0
IGF1R	0	0	0	0	0	0	0
Cyclin_E1	0	0	0	0	0	0	0
p21	2	2	2	2	2	2	2
p27	2	2	2	2	2	2	2
c_MYC	0	0	0	0	0	0	0
CDK4	0	0	0	0	0	0	0
CDK2	0	0	0	0	0	0	0
CDK6	0	0	0	0	0	0	0
Cyclin_D1	0	0	0	0	0	0	0
pRB	0	0	0	0	0	0	0

Equations
*ERBB1=(EGF and ERBB1_WT) or (ERBB1_WT and EGFRvIII) or (EGF and ERBB1_WT and EGFRvIII)
*ERBB2=(EGF and ERBB2_WT)
*ERBB3=(EGF and ERBB3_WT)
ERBB1_2=(ERBB1 and ERBB2)
ERBB1_3=(ERBB1 and ERBB3)
ERBB2_3=(ERBB2 and ERBB3)
IGF1R=(ER_alpha and not ERBB2_3) or (AKT1 and not ERBB2_3)
ER_alpha=(AKT1) or (MEK1)
c_MYC=(AKT1) or (MEK1) or (ER_alpha)
AKT1=(ERBB1) or (ERBB1_2) or (ERBB1_3) or (ERBB2_3) or (IGF1R)
MEK1=(ERBB1) or (ERBB1_2) or (ERBB1_3) or (ERBB2_3) or (IGF1R)
CDK2=(Cyclin_E1 and not p21 and not p27)
CDK4=(Cyclin_D1 and not p21 and not p27)
CDK6=(Cyclin_D1)
Cyclin_D1=(ER_alpha and c_MYC and AKT1) or (ER_alpha and c_MYC and MEK1)
Cyclin_E1=(c_MYC)
p21=(ER_alpha and not AKT1 and not c_MYC and not CDK4)
p27=(ER_alpha and not CDK4 and not CDK2 and not AKT1 and not c_MYC)
pRB=(CDK4 and CDK6) or (CDK4 and CDK6 and CDK2)

* These equations are modified in our model. Please refer to *Sahin et al.* for the previous model [245].
S1 = No Stimulation; S2 = Regular $EGFR^{WT}$ Expression + EGF; S3 = $EGFR^{WT}$ amplification + EGF; S4 = Regular $EGFR^{WT}$ with EGFRvIII and no stimulation; S5 = Regular $EGFR^{WT}$/EGFRvIII and with stimulation; S6 = Amplified $EGFR^{WT}$/EGFRvIII and no stimulation; S7 = Amplified $EGFR^{WT}$/EGFRvIII and with stimulation.

Appendix Table 12: ExQuLogic-based Logical Equations of the Integrated Pathway Model

Logical Equations
ACTIN_REORG = (RAC1) or (RAC1_CDC42)
ANCHORAGE_INDEPENDENCE = (FGF20 and FGF9)
ANTI_APOP = (BCL2)
APC_CTBP = (APC and CTBP and not DVL_AXIN_APC_GSK3_WTX_INACTIVE)
ASEF = (APC)
AXIN = (CKI_E and not WNT9B_FZD8_LRP6_CKI_G_GSK3 and not WNT1_FZD1_LRP5_CKI_G_GSK3 and not WNT3A_FZD2_LRP6_CKI_G_GSK3 and not WNT7B_FZD10_LRP5_CKI_G_GSK3 and not WNT7B_FZD1_LRP5_CKI_G_GSK3 and not WNT3A_FZD8_LRP6_CKI_G_GSK3 and not WNT2_FZD1_LRP5_CKI_G_GSK3 and not WNT1_FZD8_LRP6_CKI_G_GSK3 and not WNT3A_FZD1_LRP6_CKI_G_GSK3 and not WNT1_FZD1_LRP6_CKI_G_GSK3 and not WNT2_FZD9_LRP6_CKI_G_GSK3 and not WNT5A_FZD4_LRP5_CKI_G_GSK3 and not WNT7A_FZD5_LRP6_CKI_G_GSK3 and not WNT3_FZD1_LRP6_CKI_G_GSK3 and not WNT1_FZD8_RYK_CKI_G_GSK3 and not MACF1 and not HIC) or (DAB2 and not WNT9B_FZD8_LRP6_CKI_G_GSK3 and not WNT1_FZD1_LRP5_CKI_G_GSK3 and not WNT3A_FZD2_LRP6_CKI_G_GSK3 and not WNT7B_FZD10_LRP5_CKI_G_GSK3 and not WNT7B_FZD1_LRP5_CKI_G_GSK3 and not WNT3A_FZD8_LRP6_CKI_G_GSK3 and not WNT2_FZD1_LRP5_CKI_G_GSK3 and not WNT1_FZD8_LRP6_CKI_G_GSK3 and not WNT3A_FZD1_LRP6_CKI_G_GSK3 and not WNT1_FZD1_LRP6_CKI_G_GSK3 and not WNT2_FZD9_LRP6_CKI_G_GSK3 and not WNT5A_FZD4_LRP5_CKI_G_GSK3 and not WNT7A_FZD5_LRP6_CKI_G_GSK3 and not WNT3_FZD1_LRP6_CKI_G_GSK3 and not WNT1_FZD8_RYK_CKI_G_GSK3 and not MACF1 and not HIC)
AXIN_APC_GSK3_WTX_ACTIVE = (AXIN and APC and not GSK3_P and WTX)
AXIN2_P_DUMMY = (AXIN2)
BAX = (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and SMAD2_4 and SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF) or (P53_P)
BCL2 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (NUC_GLI2 and NUC_STK36 and not NUC_SUFU) or (NUC_NICD1 and COA and CSL and not COR)
BIRC5 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
BMI = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
BMP_RUNX3 = (BMP and RUNX3)
BOK = (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and SMAD2_4 and SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF)
BTG2 = (P53_P)
C_REL = (NUC_NICD1 and COA and CSL and not COR)
CA = (IP3)
CALCINEURIN = (CA and not KISS1)
CAMK2 = (CA)
CASR = (CA)

CD44 = (COA and CSL and not COR and NICD_ACTIVE)
CDK2 = (NUC_NICD1) or (CYCLIN_E1 and not P21 and not P27)
CDK4 = (CYCLIN_D1 and not P21 and not P27)
CDK6 = (CYCLIN_D1)
CELL_ADHESION = (CTNNB_MEMBRANE and E_CADHERIN)
CELL_MOTILITY = (ROCK)
CELL_PROLIFERATION = (FOXM1) or (MYC) or (SOX9)
CELL_SURVIVAL = (WISP1 and BIRC5 and COX2)
CELLCYCLE_PROGRESSION = (CYCLIN_D1 and P_RB and ANTI_APOP and not PRO_APOP) or (CYCLIN_E and P_RB and ANTI_APOP and not PRO_APOP) or (CYCLIN_E1 and P_RB and ANTI_APOP and not PRO_APOP) or (CYCLIN_D2 and P_RB and ANTI_APOP and not PRO_APOP) or (CYCLIN_B and P_RB and ANTI_APOP and not PRO_APOP)
CJUN = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
COA = (MAML and EP300 and SKIP and HAT)
COR = (SMRT and SAP30 and HDAC and CIR and SIN3A)
COX2 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
CTNNB = (DVL_AXIN_APC_GSK3_WTX_INACTIVE and PAR1 and not PR61 and not P53 and not DAB2 and not SOX17 and not PKC and not BTRCP) or (SMAD7 and P38 and not P53 and not DAB2 and not SOX17 and not PKC and not BTRCP) or (ERBB1 and not E_CADHERIN and not PKC and not BTRCP) or (PIN1 and not P53 and not DAB2 and not SOX17 and not PKC and not BTRCP) or (CK2A and not P53 and not DAB2 and not SOX17 and not PKC and not BTRCP) or (SIAH2)
CTNNB_MEMBRANE = (WNT4_FZD_LRP and CTNNB)
CYCLIN_B = (CYCLIN_B and not PTCH1)
CYCLIN_D1 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and SMAD2_4 and SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_LEF and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and PIASY and not REPTIN52 and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not DUPLIN and TBL1 and TBLR1 and not NARF) or (ER_ALPHA and MYC and p_AKT) or (ER_ALPHA and MYC and MEK1) or (COA and CSL and not COR and NICD_ACTIVE)
CYCLIN_D2 = (NUC_GLI2 and NUC_STK36 and not NUC_SUFU) or (GLI3_A and NUC_STK36 and not NUC_SUFU) or (NUC_CTNNB_PITX2) or (COA and CSL and not COR and NICD_ACTIVE)
CYCLIN_E = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
CYCLIN_E1 = (MYC)
DAAM1 = (DVL)
DAAM12 = (DAAM1) or (DAAM2)
DAAM2 = (DVL)
DAG = (PLC_G)
DHH = (DISPATCHED and HHAT and not HHIP)
DKK1_KREMEN2 = (DKK1) or (DKK1 and KREMEN2)

DNA_BIND = (not ID1) or (not ID2) or (not ID3)
DVL = (WNT_FZD_LRP and not DAB2 and not DACT3 and not PR72 and not NKD and not IDAX and not AXAM2) or (WNT5A_FZD7_RYK) or (WNT5B_FZD7_RYK) or (WNT3A_FZD1_RYK) or (WNT11_FZD1_RYK) or (WNT5A_FZD3_RYK) or (WNT11_FZD3_RYK)
DVL_AXIN_APC_GSK3_WTX_INACTIVE = (DVL and AXIN_APC_GSK3_WTX_ACTIVE) or (DVL and not APC) or (DVL and not GSK_3BETA) or (DVL and not AXIN) or (DVL and not WTX)
E_CADHERIN = (E_CADHERIN and not IQGAP and not SNAI1)
EMT = (SNAI1)
ER_ALPHA = (p_AKT) or (MEK1)
ERBB1 = (EGF)
ERBB1_2 = (ERBB1 and ERBB2)
ERBB1_3 = (ERBB1 and ERBB3)
ERBB2 = (EGF)
ERBB2_3 = (ERBB2 and ERBB3)
ERBB3 = (EGF)
ERK1 = (SRC)
ERK12 = (ERK1) or (ERK2)
ERK2 = (SRC)
FAS = (FAS and not SMO)
FGF20 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
FGF9 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
FLIP = (NUC_NICD1 and COA and CSL and not COR)
FOLLISTATIN = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
FOXM1 = (GLI1)
FRA1 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
GADD45A = (P53_P)
GAMMA_SECRETASE = (PRESENILIN1 and NICASTRIN and APH1 and PEN2)
GATA3 = (COA and CSL and not COR and NICD_ACTIVE)
GLI1 = (GLI1_INACTIVE and HFU and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (GLI1_INACTIVE and ERK12 and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (GLI1_INACTIVE and RAS and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (GLI1_INACTIVE and TWIST and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (GLI1_INACTIVE and ULK3 and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (GLI1_INACTIVE and STK36 and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or

(GLI1_INACTIVE and GLI2 and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU)
GLI1_INACTIVE = (GLI3_A and not SKI and not SNO and not NCOR and not SIN3A and not HDAC and not HES1) or (GLI3_A and not SKI and not SNO and not NCOR and not SIN3A and not HDAC and not HEY1) or (GLI3_A and not SKI and not SNO and not NCOR and not SIN3A and not HDAC and not HEYL) or (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH and not HES1) or (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH and not HEY1) or (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH and not HEYL)
GLI2 = (HFU and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (STK36 and not PKA_A and not GSK_3BETA and not CKI_A and not BTRCP and not SUFU) or (HFU and not NOTCH1) or (STK36 and not NOTCH1)
GLI3_A = (GLI3_A and not NUC_CTNNB_TCF) or (GLI3_A and not GLI3_R)
GLI3_R = (GLI3_R and not GLI1)
GSK3_P = (p_AKT) or (RAS) or (ERK12) or (DVL_AXIN_APC_GSK3_WTX_INACTIVE)
HATH1 = (HATH1 and not GSK_3BETA)
HES = (HES1) or (HES5) or (HES7) or (HEY1) or (HEY2) or (HEYL)
HES1 = (NUC_NICD1 and SMAD3 and CSL) or (COA and CSL and not COR and NICD_ACTIVE) or (NUC_GLI2 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (NUC_CTNNB_TCF and not NUC_CTNNB_PROP1 and not TLE)
HES5 = (COA and CSL and not COR and NICD_ACTIVE)
HES7 = (COA and CSL and not COR and NICD_ACTIVE)
HEY1 = (COA and CSL and not COR and NICD_ACTIVE)
HEY2 = (COA and CSL and not COR and NICD_ACTIVE)
HEYL = (COA and CSL and not COR and NICD_ACTIVE)
HHIP = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (not HES1) or (not HES5) or (not HES7) or (not HEY1) or (not HEY2) or (not HEYL)
HIF1A = (NUC_STAT3)
IAP = (NUC_NICD1 and COA and CSL and not COR)
ID1 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
ID2 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and BMP and not NARF)
ID3 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
IGF1R = (ER_ALPHA and not ERBB2_3) or (p_AKT and not ERBB2_3)
IHH = (BMP_RUNX3) or (DISPATCHED and HHAT and not HHIP)
INVASION = (NFAT) or (S100A4) or (MMP7)
IP3 = (PLC_G)
JAG1 = (NUC_CTNNB_TCF and PYGO and TRAPP and MLL12 and EP300 and PAF1 and BRG1 and TBL1 and TBLR1 and

PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
JAG2 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (NUC_GLI2 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
JNK = (MKK4 and MKK7)
KISS1 = (KISS1 and not WNT5A_FZD2 and not GRP54)
KLF5 = (COA and CSL and not COR and NICD_ACTIVE)
MAG = (NUC_NICD1 and DTX1 and COA and CSL and not COR) or (NUC_NICD2 and DTX1 and COA and CSL and not COR)
MEK1 = (ERBB1) or (ERBB1_2) or (ERBB1_3) or (ERBB2_3) or (IGF1R)
MEKK1 = (WNT5A_FZD7_RYK and DVL and AXIN and not CCD1 and not GSK_3BETA and not CKI_E and not HIC) or (WNT5B_FZD7_RYK and DVL and AXIN and not CCD1 and not GSK_3BETA and not CKI_E and not HIC) or (WNT3A_FZD1_RYK and DVL and AXIN and not CCD1 and not GSK_3BETA and not CKI_E and not HIC) or (WNT11_FZD1_RYK and DVL and AXIN and not CCD1 and not GSK_3BETA and not CKI_E and not HIC)
MEKK4 = (WNT5A_FZD7_RYK and DVL and AXIN and not CCD1 and not HIC) or (WNT5B_FZD7_RYK and DVL and AXIN and not CCD1 and not HIC) or (WNT3A_FZD1_RYK and DVL and AXIN and not CCD1 and not HIC) or (WNT11_FZD1_RYK and DVL and AXIN and not CCD1 and not HIC)
MKK4 = (MEKK4)
MKK7 = (MEKK1)
MKP_1 = (COA and CSL and not COR and NICD_ACTIVE)
MKRN1 = (p_AKT)
MMP7 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
MSX1 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_LEF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and BMP and not NARF)
MSX2 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and BMP and not NARF) or (NUC_CTNNB_LEF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
MYC = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH) or (NUC_NICD1 and YY1) or (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not SMAD2_4 and not SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_LEF and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and PIASY and not REPTIN52 and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not DUPLIN and TBL1 and TBLR1 and not NARF) or (p_AKT) or (MEK1) or (ER_ALPHA)
MYOD = (MYOD and not HES1)
NECD1 = (JAG1 and NOTCH1 and TACE) or (JAG2 and NOTCH1 and TACE) or (DLL1 and NOTCH1 and TACE) or (DLL3 and

NOTCH1 and TACE) or (DLL4 and NOTCH1 and TACE) or (NOTCH1 and MAGP1) or (NOTCH1 and MAGP2) or (NOTCH1 and NOV) or (NOTCH1 and CNTN1)
NECD2 = (JAG1 and NOTCH2 and TACE) or (JAG2 and NOTCH2 and TACE) or (DLL1 and NOTCH2 and TACE) or (DLL3 and NOTCH2 and TACE) or (DLL4 and NOTCH2 and TACE) or (NOTCH2 and CNTN1)
NECD3 = (JAG1 and NOTCH3 and TACE) or (JAG2 and NOTCH3 and TACE) or (DLL1 and NOTCH3 and TACE) or (DLL3 and NOTCH3 and TACE) or (DLL4 and NOTCH3 and TACE)
NECD4 = (JAG1 and NOTCH4 and TACE) or (JAG2 and NOTCH4 and TACE) or (DLL1 and NOTCH4 and TACE) or (DLL3 and NOTCH4 and TACE) or (DLL4 and NOTCH4 and TACE)
NEURONAL_DIFF = (not REST)
NEXT1 = (JAG1 and NOTCH1 and TACE) or (JAG2 and NOTCH1 and TACE) or (DLL1 and NOTCH1 and TACE) or (DLL3 and NOTCH1 and TACE) or (DLL4 and NOTCH1 and TACE) or (NOTCH1 and MAGP1) or (NOTCH1 and MAGP2) or (NOTCH1 and NOV) or (NOTCH1 and CNTN1)
NEXT2 = (JAG1 and NOTCH2 and TACE) or (JAG2 and NOTCH2 and TACE) or (DLL1 and NOTCH2 and TACE) or (DLL3 and NOTCH2 and TACE) or (DLL4 and NOTCH2 and TACE) or (NOTCH2 and CNTN1)
NEXT3 = (JAG1 and NOTCH3 and TACE) or (JAG2 and NOTCH3 and TACE) or (DLL1 and NOTCH3 and TACE) or (DLL3 and NOTCH3 and TACE) or (DLL4 and NOTCH3 and TACE)
NEXT4 = (JAG1 and NOTCH4 and TACE) or (JAG2 and NOTCH4 and TACE) or (DLL1 and NOTCH4 and TACE) or (DLL3 and NOTCH4 and TACE) or (DLL4 and NOTCH4 and TACE)
NFAT = (CALCINEURIN and not GSK_3BETA)
NFKB = (COA and CSL and not COR and NICD_ACTIVE)
NICD_ACTIVE = (NUC_NICD1 and not HES) or (NUC_NICD2 and not HES) or (NUC_NICD3 and not HES) or (NUC_NICD4 and not HES) or (NUC_NICD1 and WNT3A) or (NUC_NICD2 and WNT3A) or (NUC_NICD3 and WNT3A) or (NUC_NICD4 and WNT3A)
NICD1 = (NEXT1 and GAMMA_SECRETASE) or (HIF1A)
NICD2 = (NEXT2 and GAMMA_SECRETASE) or (HIF1A)
NICD3 = (NEXT3 and GAMMA_SECRETASE) or (HIF1A)
NICD4 = (NEXT4 and GAMMA_SECRETASE) or (HIF1A)
NKD = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
NLK = (TAK1)
NOTCH_SIGNAL = (JAG2)
NOTCH1 = (NOTCH1_PRE and FURIN and not ITCH and not NUMB and ALPHA_ADAPTIN) or (NOTCH1_PRE and not NEDD4) or (NOTCH1_PRE and O_GLUCOSE and POGLUT_1) or (NOTCH1_PRE and O_GLUCOSE and XYL and not XYLE) or (NOTCH1_PRE and NGA and O_FUCOSE and FRINGE and POFUT_1) or (NOTCH1_PRE and O_FUCOSE and GALACTOSE and GASE)
NOTCH1_PRE = (not P53_P and NUC_NICD1 and CSL)
NOTCH2 = (NOTCH2_PRE and FURIN and not ITCH and not NUMB and ALPHA_ADAPTIN) or (NOTCH2_PRE and O_GLUCOSE and POGLUT_1) or (NOTCH2_PRE and O_GLUCOSE and XYL and not XYLE) or (NOTCH2_PRE and NGA and O_FUCOSE and FRINGE and POFUT_1) or (NOTCH2_PRE and O_FUCOSE and GALACTOSE and GASE)
NOTCH2_PRE = (not P53_P and NUC_NICD2 and CSL)
NOTCH3 = (NOTCH3_PRE and O_GLUCOSE and POGLUT_1) or (NOTCH3_PRE and O_GLUCOSE and XYL and not XYLE) or

(NOTCH3_PRE and NGA and O_FUCOSE and FRINGE and POFUT_1) or (NOTCH3_PRE and O_FUCOSE and GALACTOSE and GASE)
NOTCH3_PRE = (not P53_P and NUC_NICD3 and CSL)
NOTCH4 = (NOTCH4_PRE and O_GLUCOSE and POGLUT_1) or (NOTCH4_PRE and O_GLUCOSE and XYL and not XYLE) or (NOTCH4_PRE and NGA and O_FUCOSE and FRINGE and POFUT_1) or (NOTCH4_PRE and O_FUCOSE and GALACTOSE and GASE)
NOTCH4_PRE = (not P53_P and NUC_NICD4 and CSL)
NRARP = (COA and CSL and not COR and NICD_ACTIVE)
NUC_CTNNB = (CTNNB and IMPORTIN and not ICAT and not CHIBBY and not PEN2) or (CTNNB and IMPORTIN and RAC1 and JNK and not ICAT and not CHIBBY and not PEN2)
NUC_CTNNB_LEF = (NUC_CTNNB and LEF and CKI_E and not TLE and not HDAC) or (NUC_CTNNB and LEF and not HDAC)
NUC_CTNNB_PITX2 = (NUC_CTNNB and PITX2)
NUC_CTNNB_PROP1 = (NUC_CTNNB and PROP1)
NUC_CTNNB_TCF = (NUC_CTNNB_TCF4) or (NUC_CTNNB_TCF3)
NUC_CTNNB_TCF3 = (NUC_CTNNB and TCF3_INACTIVE and CKI_E and not TLE and not HDAC) or (NUC_CTNNB and TCF3_INACTIVE and not TLE and not HDAC)
NUC_CTNNB_TCF4 = (NUC_CTNNB and TCF4_INACTIVE and CKI_E and not TLE and not HDAC) or (NUC_CTNNB and TCF4_INACTIVE and not HDAC)
NUC_GLI1 = (GLI1 and not SNAI1) or (GLI1 and not NUC_CTNNB_TCF)
NUC_GLI2 = (GLI2 and not SNAI1)
NUC_NICD1 = (not JIP1 and not DVL and GSK_3BETA and NICD1) or (RAS and NICD1) or (WDR12 and NICD1) or (not FBW7 and not CDK8 and not CYCLIN_C and MAML) or (NICD1 and not NRARP)
NUC_NICD2 = (not JIP1 and not DVL and GSK_3BETA and NICD2) or (not FBW7 and not CDK8 and not CYCLIN_C and MAML) or (NICD2 and not NRARP)
NUC_NICD3 = (not JIP1 and not DVL and GSK_3BETA and NICD3) or (not FBW7 and not CDK8 and not CYCLIN_C and MAML) or (NICD3 and not NRARP)
NUC_NICD4 = (not JIP1 and not DVL and GSK_3BETA and NICD4) or (NICD4 and not FBW7) or (not FBW7 and not CDK8 and not CYCLIN_C and MAML) or (NICD4 and not NRARP)
NUC_STAT3 = (STAT3_P)
OPN = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
PI3K = (PIK3CA and PIK3R1)
p_AKT = (PI3K and AKT) or (AKT and ERBB1) or (AKT and ERBB1_2) or (AKT and ERBB1_3) or (AKT and ERBB2_3) or (AKT and IGF1R)
GSK_3BETA = (GSK_3BETA and not GSK3_P)
P_RB = (RB and CDK4 and CDK6) or (RB and CDK4 and CDK6 and CDK2)
P14ARF = (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and SMAD2_4 and SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_LEF and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and PIASY and not REPTIN52 and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not DUPLIN and TBL1 and TBLR1 and not NARF) or (TBX3)
P21 = (NUC_NICD1 and COA and CSL and not COR) or (ER_ALPHA and not p_AKT and not MYC and not CDK4) or (P53_P)

P27 = (ER_ALPHA and not CDK4 and not CDK2 and not p_AKT and not MYC)
P50 = (NUC_NICD1 and COA and CSL and not COR)
P53 = (P14ARF and not MKRN1)
P53_P = (not NICD1 and P53)
P65 = (NUC_NICD1 and COA and CSL and not COR)
PDGFRA = (GLI1)
PI3K = (WNT3A_FZD8 and not PTEN) or (WNT3A_FZD2 and not PTEN) or (WNT3A_FZD1 and not PTEN) or (WNT1_FZD1 and not PTEN) or (WNT1_FZD8 and not PTEN)
PIT1 = (NUC_CTNNB_PROP1 and EP300)
PKC = (CA and DAG and not KISS1 and not GRP54)
PLC_G = (WNT5A_FZD2)
PPAR_D = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
PPAR_G = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
PRICKLE12 = (PRICKLE1) or (PRICKLE2)
PRO_APOP = (BOK and BAX)
PTCH1 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
PTCH1_HH = (PTCH1 and DHH) or (PTCH1 and IHH) or (PTCH1 and SHH)
PTCH2_HH = (PTCH2 and DHH) or (PTCH2 and IHH) or (PTCH2 and SHH)
PTCRA = (COA and CSL and not COR and NICD_ACTIVE)
PTEN = (PTEN and not HES1)
RAC1 = (ASEF)
RAC1_CDC42 = (WNT5A_FZD7_RYK and DVL and AXIN and ANKRD6 and NKD and DAB2 and not PAR1 and not CCD1) or (WNT5B_FZD7_RYK and DVL and AXIN and ANKRD6 and NKD and DAB2 and not PAR1 and not CCD1) or (WNT3A_FZD1_RYK and DVL and AXIN and ANKRD6 and NKD and DAB2 and not PAR1 and not CCD1) or (WNT11_FZD1_RYK and DVL and AXIN and ANKRD6 and NKD and DAB2 and not PAR1 and not CCD1) or (APC and IQGAP)
REL_B = (NUC_NICD1 and COA and CSL and not COR)
REST = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
RHOA = (DAAM12 and RHOGEF)
ROCK = (RHOA)
S100A4 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
SFRP = (SFRP1) or (SFRP3)
SFRP1 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
SFRP3 = (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)

SHH = (CDO and BOC) or (DISPATCHED and HHAT and not HHIP and not GAS1)
SIAH2 = (WNT5A_FZD7_RYK and CASR)
SMAD2_4 = (SMAD2) or (SMAD4)
SMAD3_4 = (SMAD3) or (SMAD4)
SMO = (PTCH1_HH and PTCH2_HH)
SNAI1 = (AXIN2_P_DUMMY) or (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
SOX9 = (NUC_CTNNB_TCF4 and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and TNIK and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and SMAD2_4 and SMAD3_4 and not HIPK2 and not KLF4 and not NLK and not AXAM2 and not NARF) or (COA and CSL and not COR and NICD_ACTIVE)
SRC = (WNT3A_FZD8) or (WNT3A_FZD2) or (WNT3A_FZD1) or (WNT1_FZD1) or (WNT1_FZD8)
STAT3_P = (CTNNB and STAT3) or (JAK2 and STAT3 and HES1) or (JAK2 and STAT3 and HES5)
STK36 = (SMO)
TAK1 = (CAMK2)
TBX3 = (NUC_CTNNB_TCF and PYGO and TRAPP and MLL12 and EP300 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
TCF3_INACTIVE = (TCF3)
TCF4 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF)
TCF4_INACTIVE = (TCF4)
TOP2A = (TOP2A and not ub_PTEN)
TPX2 = (NUC_STAT3)
TRANSCRIPTION_REP = (MSX1) or (MSX2)
ub_PTEN = (MKRN1 and PTEN)
VANGL12 = (VANGL1) or (VANGL2)
WISP1 = (NUC_CTNNB_TCF and PYGO and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and TBL1 and TBLR1 and PIASY and SMAD2_4 and SMAD3_4 and not REPTIN52 and not DUPLIN and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not NARF) or (NUC_CTNNB_LEF and PYGO and BCL9 and EP300 and TRAPP and MLL12 and PAF1 and BRG1 and PIASY and not REPTIN52 and not TCF1_R and not APC_CTBP and not HIPK2 and not NLK and not AXAM2 and not DUPLIN and TBL1 and TBLR1 and not NARF)
WNT_FZD_LRP = (WNT3A_FZD8_LRP6) or (WNT9B_FZD8_LRP6) or (WNT1_FZD1_LRP5) or (WNT3A_FZD2_LRP6) or (WNT7B_FZD10_LRP5) or (WNT7B_FZD1_LRP5) or (WNT2_FZD1_LRP5) or (WNT1_FZD8_LRP6) or (WNT3A_FZD1_LRP6) or (WNT1_FZD1_LRP6) or (WNT2_FZD9_LRP6) or (WNT3A_FZD8_LRP6) or (WNT5A_FZD4_LRP5) or (WNT7A_FZD5_LRP6) or (WNT1_FZD8_RYK) or (WNT3_FZD1_LRP6)
WNT1 = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1) or (NUC_GLI1 and NUC_STK36 and DYRK1 and not NUC_SUFU and not NUMB and not ITCH)
WNT1_FZD1 = (WNT1 and FZD1)
WNT1_FZD1_LRP5 = (WNT1 and FZD1 and LRP5 and GAQ and not SHISA and not IGFBP4 and not LRP1 and not DKK1_KREMEN2)
WNT1_FZD1_LRP5_CKI_G_GSK3 = (WNT1_FZD1_LRP5 and CKI_G and GSK_3BETA)

WNT1_FZD1_LRP6 = (WNT1 and FZD1 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2 and not LRP1)
WNT1_FZD1_LRP6_CKI_G_GSK3 = (WNT1_FZD1_LRP6 and CKI_G and GSK_3BETA)
WNT1_FZD8 = (WNT1 and FZD8)
WNT1_FZD8_LRP6 = (WNT1 and FZD8 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT1_FZD8_LRP6_CKI_G_GSK3 = (WNT1_FZD8_LRP6 and CKI_G and GSK_3BETA)
WNT1_FZD8_RYK = (WNT1 and FZD8 and RYK and GAQ and not IGFBP4 and not SHISA)
WNT1_FZD8_RYK_CKI_G_GSK3 = (WNT1_FZD8_RYK and CKI_G and GSK_3BETA)
WNT11 = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT11_FZD1_RYK = (WNT11 and FZD1 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL)
WNT11_FZD3_RYK = (WNT11 and FZD3 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL)
WNT2 = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT2_FZD1_LRP5 = (WNT2 and FZD1 and LRP5 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2 and not LRP1)
WNT2_FZD1_LRP5_CKI_G_GSK3 = (WNT2_FZD1_LRP5 and CKI_G and GSK_3BETA)
WNT2_FZD9_LRP6 = (WNT2 and FZD9 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT2_FZD9_LRP6_CKI_G_GSK3 = (WNT2_FZD9_LRP6 and CKI_G and GSK_3BETA)
WNT3 = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT3_FZD1_LRP6 = (WNT3 and FZD1 and LRP6 and GAQ and not SHISA and not IGFBP4 and not LRP1 and not DKK1_KREMEN2 and NON_CANONICAL)
WNT3_FZD1_LRP6_CKI_G_GSK3 = (WNT3_FZD1_LRP6 and CKI_G and GSK_3BETA)
WNT3A = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT3A_FZD1 = (WNT3A and FZD1)
WNT3A_FZD1_LRP6 = (WNT3A and FZD1 and LRP6 and RYK and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2 and not LRP1)
WNT3A_FZD1_LRP6_CKI_G_GSK3 = (WNT3A_FZD1_LRP6 and CKI_G and GSK_3BETA)
WNT3A_FZD1_RYK = (WNT3A and FZD1 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL)
WNT3A_FZD2 = (WNT3A and FZD2)
WNT3A_FZD2_LRP6 = (WNT3A and FZD2 and LRP6 and RYK and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT3A_FZD2_LRP6_CKI_G_GSK3 = (WNT3A_FZD2_LRP6 and CKI_G and GSK_3BETA)
WNT3A_FZD8 = (WNT3A and FZD8)
WNT3A_FZD8_LRP6 = (WNT3A and FZD8 and LRP6 and RYK and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2) or (WNT3A and FZD8 and LRP6 and PRR and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT3A_FZD8_LRP6_CKI_G_GSK3 = (WNT3A_FZD8_LRP6 and CKI_G and GSK_3BETA)
WNT4 = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT4_FZD_LRP = (WNT4 and LRP5 and FZD3) or (WNT4 and LRP6 and FZD3)
WNT5A = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1) or (CASR and EVI and PORCN and not

SFRP and not CER1 and not SOST and not WIF1)
WNT5A_FZD2 = (WNT5A and FZD2 and GAQ and not SHISA and not IGFBP4 and NON_CANONICAL)
WNT5A_FZD3_RYK = (WNT5A and FZD3 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL) or (WNT5A and FZD3 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and SFRP4 and NON_CANONICAL)
WNT5A_FZD4_LRP5 = (WNT5A and FZD4 and LRP5 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT5A_FZD4_LRP5_CKI_G_GSK3 = (WNT5A_FZD4_LRP5 and CKI_G and GSK_3BETA)
WNT5A_FZD7_RYK = (WNT5A and FZD7 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL) or (WNT5A and FZD7 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and SFRP4 and NON_CANONICAL)
WNT5B = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT5B_FZD7_RYK = (WNT5B and FZD7 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and not IGFBP4 and NON_CANONICAL) or (WNT5B and FZD7 and RYK and ROR2 and PTK7 and VANGL12 and PRICKLE12 and ANKRD6 and NKD and MAGI3 and not SHISA and SFRP4 and NON_CANONICAL)
WNT7A = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT7A_FZD5_LRP6 = (WNT7A and FZD5 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2) or (WNT7A and FZD5 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2 and SFRP4)
WNT7A_FZD5_LRP6_CKI_G_GSK3 = (WNT7A_FZD5_LRP6 and CKI_G and GSK_3BETA)
WNT7B = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT7B_FZD1_LRP5 = (WNT7B and FZD1 and LRP5 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2 and not LRP1)
WNT7B_FZD1_LRP5_CKI_G_GSK3 = (WNT7B_FZD1_LRP5 and CKI_G and GSK_3BETA)
WNT7B_FZD10_LRP5 = (WNT7B and FZD10 and LRP5 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT7B_FZD10_LRP5_CKI_G_GSK3 = (WNT7B_FZD10_LRP5 and CKI_G and GSK_3BETA)
WNT9B = (EVI and PORCN and not SFRP and not CER1 and not SOST and not WIF1)
WNT9B_FZD8_LRP6 = (WNT9B and FZD8 and LRP6 and GAQ and not SHISA and not IGFBP4 and not DKK1_KREMEN2)
WNT9B_FZD8_LRP6_CKI_G_GSK3 = (WNT9B_FZD8_LRP6 and CKI_G and GSK_3BETA)

Appendix Table 13: Descriptions of the Developed Pathway Models

Pathways	Crosstalk with Other Pathways	Modeling Techniques	Statistics (Nodes & Edges)	Study objectives
Hedgehog	RAS, ERK1/2, TWIST	Graph theoretical and Binary Logic-based modeling	57 & 140	Novel drug target identification for GBM treatment
Notch	PI3K/AKT, JAK/STAT, P53, HIF1A	Graph theoretical and Binary Logic-based modeling	115 & 231	1. Novel drug target identification for GBM treatment 2. Study of intra-tumor heterogeneity 3. Study of the Developmental dynamics of neural stem cells, glioma stem cells in GBM ecosystem
Hedgehog, Notch, WNT, EGFR, PI3K/AKT, JAK/STAT, HIF1A, P53	NA	Graph theoretical and ExQuLogic-based modeling	410 & 1695	1. Application of ExQuLogic to measure gene expression pattern 2. Novel drug target identification for GBM treatment

Appendix Table 14: Summary of the Identified Drug-targets

Pathway Models	Drug-targets	Targeted Phenotypes/Cancer Cells	Technique Used
Hedgehog	1. SMO, GLI1, GLI2	Cell Division (Grade-IV GBM Tumor)	Graph theoretical and Binary Logic-based simulation
	2. SMO, HFU, ULK3 and RAS	Cell Division (Colon Cancer)	Graph theoretical and Binary Logic-based simulation
	3. SMO, HFU, ULK3, RAS, ERK1/2	Cell Division (Pancreatic Cancer)	Graph theoretical and Binary Logic-based modeling
Notch	1. NICD, MAML	Cell Division (Grade-IV GBM Tumor)	Graph theoretical and Binary Logic-based simulation
	2. NICD, HIF1A	Cell Division (Grade-IV GBM Tumor)	Graph theoretical and Binary Logic-based simulation
	3. STAT3	Cell Division (Grade-IV GBM Tumor and LGG)	Fast Fourier Transform (FFT) and Binary Logic-based simulation
	4. NGN1, MASH1	Cell Division (Grade-IV GBM Tumor and LGG)	Fast Fourier Transform (FFT) and Binary Logic-based simulation
Hedgehog, Notch, WNT, EGFR, PI3K/AKT, JAK/STAT, HIF1A, P53	1. GLI1, GLI3, NICD, HEY1, NRARP, SNAI1	Cell Division and Metastasis (Grade-IV GBM Tumor)	Strongly Connected Component Analyses and Extended Quaternary Logic-based simulation

Appendix Text 1: *ExQSUS Algorithm*

The description of the ExQSUS algorithm implemented in *Chapter 7* is provided bellow.

Step 1: Start with the logical equation of the node $D_p(t+1) = A_1(t) \wedge A_2(t) \ldots \wedge !I_1(t) \wedge !I_2(t) \ldots$

Step 2: Store the upstream conjunction nodes: $AND[] \leftarrow \{A_1, A_2, A_3, \ldots\ldots, A_m, I_1, I_2, I_3, \ldots\ldots, I_n\}$

Step 3: Compute $\theta_{X=0}$, $\theta_{X=1}$, $\theta_{X=2}$, $\theta_{X=3}$ *from* $AND[]$

Step 4: if $\theta_{X=0} = \theta_{X=1} = \theta_{X=2} = \theta_{X=3} = 0$

$D_p(t+1) \leftarrow AND[A_1, A_2, A_3, \ldots\ldots, A_m, I_1, I_2, I_3, \ldots\ldots, I_n] = 0$

else if $\theta_{X=qbits \in q_1 \notin q_2} > 0 \;\&\; \theta_{X=qbits \in q_2 \notin q_1} = 0$

$D_p(t+1) \leftarrow AND[A_1, A_2, A_3, \ldots\ldots, A_m, I_1, I_2, I_3, \ldots\ldots, I_n] = q_1 \in \{0,1,2,3\}$

else

$D_p(t+1) \leftarrow AND[A_1, A_2, A_3, \ldots\ldots, A_m, I_1, I_2, I_3, \ldots\ldots, I_n] = ExQSUS(\theta_{X=0}, \theta_{X=1}, \theta_{X=2}, \theta_{X=3})$

Step 5: Stop

function $ExQSUS(\theta_{X=0}, \theta_{X=1}, \theta_{X=2}, \theta_{X=3})$:

if $\theta_{X=qbits \in q_0 \notin q_1 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_1 \notin q_0 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_2 \notin q_1 \notin q_3 \notin q_0} > \theta_{X=qbits \in q_3 \notin q_0 \notin q_1 \notin q_2}$

return q_0

else if $\theta_{X=qbits \in q_0 \notin q_1 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_1 \notin q_0 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_2 \notin q_1 \notin q_3 \notin q_0} = \theta_{X=qbits \in q_3 \notin q_0 \notin q_1 \notin q_2} > 0$

return q_0

else if $\theta_{X=qbits \in q_0 \notin q_1 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_1 \notin q_0 \notin q_2 \notin q_3} = \theta_{X=qbits \in q_2 \notin q_1 \notin q_3 \notin q_0} = \theta_{X=qbits \in q_3 \notin q_0 \notin q_1 \notin q_2} > 0$

return q_0

else if $\theta_{X=qbits \in q_0 \notin q_1 \notin q_2 \notin q_3} = \theta_{X=qbits \in q_1 \notin q_0 \notin q_2 \notin q_3} > \theta_{X=qbits \in q_2 \notin q_1 \notin q_3 \notin q_0} > \theta_{X=qbits \in q_3 \notin q_0 \notin q_1 \notin q_2} > 0$

return $(q_0 \wedge q_1)$

else if $\theta_{X=qbits \in q_0 \notin q_1 \notin q_2 \notin q_3} = \theta_{X=qbits \in q_1 \notin q_0 \notin q_2 \notin q_3} = \theta_{X=qbits \in q_2 \notin q_1 \notin q_3 \notin q_0} > \theta_{X=qbits \in q_3 \notin q_0 \notin q_1 \notin q_2} > 0$

return $(q_0 \wedge q_1 \wedge q_2)$

10 REFERENCES

1. Icard P, Lincet H (2012) A global view of the biochemical pathways involved in the regulation of the metabolism of cancer cells. *Biochim Biophys Acta* 1826: 423-433.

2. Schilling CH, Palsson BO (1998) The underlying pathway structure of biochemical reaction networks. *Proc Natl Acad Sci U S A* 95: 4193-4198.

3. Sarkar RR, Maithreye R, Sinha S (2011) Design of regulation and dynamics in simple biochemical pathways. *J Math Biol* 63: 283-307.

4. Tyson JJ, Othmer HG (1978) The dynamics of feedback control circuits in biochemical pathways. *Prog Theor Biol* 5: 62.

5. Michal G, Ed., (1999) *Biochemical pathways*: An Atlas of Biochemistry and Molecular Biology. Heidelberg, Germany, Spektrum Akademischer Verlag.

6. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* 407: 651-654.

7. Breuer EK, Murph MM, Craven RJ (2012) Biochemical pathways in cancer. *Biochem Res Int* 2012: 268504.

8. Chowdhury S, Sarkar RR (2015) Comparison of human cell signaling pathway databases--evolution, drawbacks and challenges. *Database (Oxford)* 2015: bau126.

9. Czarnecka AM, Niedzwiedzka M, Porta C, Szczylik C (2016) Hormone signaling pathways as treatment targets in renal cell cancer (Review). *Int J Oncol* 48: 2221-2235.

10. Davis RJ (2000) Signal transduction by the JNK group of MAP kinases. *Cell* 103: 239-252.

11. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R (2009) Gene regulatory network inference: data integration in dynamic models-a review. *Biosystems* 96: 86-103.

12. Li CW, Chu YH, Chen BS (2007) Construction and clarification of dynamic gene regulatory network of cancer cell cycle via microarray data. *Cancer Inform* 2: 223-241.

13. Zhang B, Xu ZW, Wang KH, Lu TC, Du Y (2013) Complex regulatory network of microRNAs, transcription factors, gene alterations in adrenocortical cancer. *Asian Pac J Cancer Prev* 14: 2265-2268.

14. DeBerardinis RJ, Thompson CB (2012) Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* 148: 1132-1144.

15. Edwards JS, Covert M, Palsson B (2002) Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol* 4: 133-140.

16. Motter AE, Gulbahce N, Almaas E, Barabasi AL (2008) Predicting synthetic rescues in metabolic networks. *Mol Syst Biol* 4: 168.

17. Gormley M, Akella VU, Quong JN, Quong AA (2011) An integrated framework to model cellular phenotype as a component of biochemical networks. *Adv Bioinformatics* 2011: 608295.

18. Jeganathan N, Predescu D, Zhang J, Sha F, Bardita C, et al. (2016) Rac1-mediated cytoskeleton rearrangements induced by intersectin-1s deficiency promotes lung cancer cell proliferation, migration and metastasis. *Mol Cancer* 15: 59.

19. Dandekar T, Schuster S, Snel B, Huynen M, Bork P (1999) Pathway alignment: application to the comparative analysis of glycolytic enzymes. *Biochem J* 343 Pt 1: 115-124.

20. Li Y, de Ridder D, de Groot MJ, Reinders MJ (2008) Metabolic pathway alignment between species using a comprehensive and flexible similarity measure. *BMC Syst Biol* 2: 111.

21. Chang KH, Li R, Papari-Zareei M, Watumull L, Zhao YD, et al. (2011) Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *Proc Natl Acad Sci U S A* 108: 13728-13733.

22. Durmus Tekir S, Yalcin Arga K, Ulgen KO (2009) Drug targets for tumorigenesis: insights from

structural analysis of EGFR signaling network. *J Biomed Inform* 42: 228-236.

23. Hornberg JJ, Bruggeman FJ, Westerhoff HV, Lankelma J (2006) Cancer: a Systems Biology disease. *Biosystems* 83: 81-90.

24. Staines DR (2005) Does dysregulation of key epigenetic and biochemical pathways occur in postulated vasoactive neuropeptide autoimmune disorders? *Med Hypotheses* 65: 1154-1160.

25. Frye RE, Nankova B, Bhattacharyya S, Rose S, Bennuri SC, et al. (2017) Modulation of Immunological Pathways in Autistic and Neurotypical Lymphoblastoid Cell Lines by the Enteric Microbiome Metabolite Propionic Acid. *Front Immunol* 8: 1670.

26. Jimenez-Caliani AJ, Pillich R, Yang W, Diaferia GR, Meda P, et al. (2017) alphaE-Catenin Is a Positive Regulator of Pancreatic Islet Cell Lineage Differentiation. *Cell Rep* 20: 1295-1306.

27. Levy J, Gavin JR, 3rd, Sowers JR (1994) Diabetes mellitus: a disease of abnormal cellular calcium metabolism? *Am J Med* 96: 260-273.

28. Bair SM, Choueiri TK, Moslehi J (2013) Cardiovascular complications associated with novel angiogenesis inhibitors: emerging evidence and evolving perspectives. *Trends Cardiovasc Med* 23: 104-113.

29. Lichtenstein I, Charleston MA, Caetano TS, Gamble JR, Vadas MA (2013) Active subnetwork recovery with a mechanism-dependent scoring function; with application to angiogenesis and organogenesis studies. *BMC Bioinformatics* 14: 59.

30. Warburton D, Schwarz M, Tefft D, Flores-Delgado G, Anderson KD, et al. (2000) The molecular basis of lung morphogenesis. *Mech Dev* 92: 55-81.

31. Yavropoulou MP, Maladaki A, Yovos JG (2015) The role of Notch and Hedgehog signaling pathways in pituitary development and pathogenesis of pituitary adenomas. *Hormones (Athens)* 14: 5-18.

32. Mitsiadis TA, Gayet O, Zhang N, Carroll P (2001) Expression of Deltex1 during mouse embryogenesis: comparison with Notch1, 2 and 3 expression. *Mech Dev* 109: 399-403.

33. Katoh M, Katoh M (2006) Notch ligand, JAG1, is evolutionarily conserved target of canonical WNT signaling pathway in progenitor cells. *Int J Mol Med* 17: 681-685.

34. Novo E, Cannito S, Paternostro C, Bocca C, Miglietta A, et al. (2014) Cellular and molecular mechanisms in liver fibrogenesis. *Arch Biochem Biophys* 548: 20-37.

35. Chaffer CL, Dopheide B, Savagner P, Thompson EW, Williams ED (2007) Aberrant fibroblast growth factor receptor signaling in bladder and other cancers. *Differentiation* 75: 831-842.

36. Kim GH, Ryan JJ, Archer SL (2013) The role of redox signaling in epigenetics and cardiovascular disease. *Antioxid Redox Signal* 18: 1920-1936.

37. Thompson CA, Burcham PC (2008) Genome-wide transcriptional responses to acrolein. *Chem Res Toxicol* 21: 2245-2256.

38. Wu YT, Wu SB, Wei YH (2014) Metabolic reprogramming of human cells in response to oxidative stress: implications in the pathophysiology and therapy of mitochondrial diseases. *Curr Pharm Des* 20: 5510-5526.

39. Garrod AE (2002) The incidence of alkaptonuria: a study in chemical individuality. 1902 [classical article]. *Yale J Biol Med* 75: 221-231.

40. Kovalchuk A, Nersisyan L, Mandal R, Wishart D, Mancini M, et al. (2018) Growth of Malignant Non-CNS Tumors Alters Brain Metabolome. *Front Genet* 9: 41.

41. Elpeleg O, Shaag A, Ben-Shalom E, Schmid T, Bachmann C (2002) N-acetylglutamate synthase deficiency and the treatment of hyperammonemic encephalopathy. *Ann Neurol* 52: 845-849.

42. Chiu CJ, Taylor A (2011) Dietary hyperglycemia, glycemic index and metabolic retinal diseases. *Prog Retin Eye Res* 30: 18-53.

43. Bommer UA (2017) The Translational Controlled Tumour Protein TCTP: Biological Functions

and Regulation. *Results Probl Cell Differ* 64: 69-126.

44. DeStefano GM, Christiano AM (2014) The genetics of human skin disease. *Cold Spring Harb Perspect Med* 4. doi: 10.1101/cshperspect.a015172

45. Rakyan VK, Hildmann T, Novik KL, Lewin J, Tost J, et al. (2004) DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. *PLoS Biol* 2: e405.

46. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562: 203-209.

47. Garrison CB, Lastwika KJ, Zhang Y, Li CI, Lampe PD (2017) Proteomic Analysis, Immune Dysregulation, and Pathway Interconnections with Obesity. *J Proteome Res* 16: 274-287.

48. Wen Y, Alshikho MJ, Herbert MR (2016) Pathway Network Analyses for Autism Reveal Multisystem Involvement, Major Overlaps with Other Diseases and Convergence upon MAPK and Calcium Signaling. *PLoS One* 11: e0153329.

49. Meng Q, Ying Z, Noble E, Zhao Y, Agrawal R, et al. (2016) Systems Nutrigenomics Reveals Brain Gene Networks Linking Metabolic and Brain Disorders. *EBioMedicine* 7: 157-166.

50. Basler G, Nikoloski Z, Larhlimi A, Barabasi AL, Liu YY (2016) Control of fluxes in metabolic networks. *Genome Res* 26: 956-968.

51. Lowe H, Kremling A, Marin-Sanguino A (2016) Time Hierarchies and Model Reduction in Canonical Non-linear Models. *Front Genet* 7: 166.

52. Saez M, Wiuf C, Feliu E (2017) Graphical reduction of reaction networks by linear elimination of species. *J Math Biol* 74: 195-237.

53. Mondal LK, Bhaduri G, Bhattacharya B (2018) Biochemical scenario behind initiation of diabetic retinopathy in type 2 diabetes mellitus. *Indian J Ophthalmol* 66: 535-540.

54. Donatello S, Hudson L, Cottell DC, Blanco A, Aurrekoetxea I, et al. (2011) An imbalance in progenitor cell populations reflects tumour progression in breast cancer primary culture models. *J Exp Clin Cancer Res* 30: 45.

55. Cohen AL, Holmen SL, Colman H (2013) IDH1 and IDH2 mutations in gliomas. *Curr Neurol Neurosci Rep* 13: 345.

56. Salt IP, Palmer TM (2012) Exploiting the anti-inflammatory effects of AMP-activated protein kinase activation. *Expert Opin Investig Drugs* 21: 1155-1167.

57. Hu YS, Xin J, Hu Y, Zhang L, Wang J (2017) Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach. *Alzheimers Res Ther* 9: 29.

58. Yi CQ, Ma CH, Xie ZP, Cao Y, Zhang GQ, et al. (2013) Comparative genome-wide gene expression analysis of rheumatoid arthritis and osteoarthritis. *Genet Mol Res* 12: 3136-3145.

59. Chen YR, Huang HC, Lin CC (2017) Regulatory feedback loops bridge the human gene regulatory network and regulate carcinogenesis. *Brief Bioinform*, bbx166. doi: 10.1093/bib/bbx166.

60. Itzel T, Scholz P, Maass T, Krupp M, Marquardt JU, et al. (2015) Translating bioinformatics in oncology: guilt-by-profiling analysis and identification of KIF18B and CDCA3 as novel driver genes in carcinogenesis. *Bioinformatics* 31: 216-224.

61. Kitano H (2004) Cancer as a robust system: implications for anticancer therapy. *Nat Rev Cancer* 4: 227-235.

62. Che L, Yuan YH, Jia J, Ren J (2012) Activation of sonic hedgehog signaling pathway is an independent potential prognosis predictor in human hepatocellular carcinoma patients. *Chin J Cancer Res* 24: 323-331.

63. Koochekpour S (2010) Androgen receptor signaling and mutations in prostate cancer. *Asian J Androl* 12: 639-657.

64. Bharti R, Dey G, Mandal M (2016) Cancer development, chemoresistance, epithelial to

mesenchymal transition and stem cells: A snapshot of IL-6 mediated involvement. *Cancer Lett* 375: 51-61.

65. Huang WJ, Chen WW, Zhang X (2016) Glioblastoma multiforme: Effect of hypoxia and hypoxia inducible factors on therapeutic approaches. *Oncol Lett* 12: 2283-2288.

66. Grieco L, Calzone L, Bernard-Pierrot I, Radvanyi F, Kahn-Perles B, et al. (2013) Integrative modelling of the influence of MAPK network on cancer cell fate decision. *PLoS Comput Biol* 9: e1003286.

67. Ahmed KA, Chinnaiyan P (2014) Applying metabolomics to understand the aggressive phenotype and identify novel therapeutic targets in glioblastoma. *Metabolites* 4: 740-750.

68. Eaton DL, Gallagher EP (1994) Mechanisms of aflatoxin carcinogenesis. *Annu Rev Pharmacol Toxicol* 34: 135-172.

69. Friedenson B (2011) A common environmental carcinogen unduly affects carriers of cancer mutations: carriers of genetic mutations in a specific protective response are more susceptible to an environmental carcinogen. *Med Hypotheses* 77: 791-797.

70. Mishra P, Tang W, Putluri V, Dorsey TH, Jin F, et al. (2017) ADHFE1 is a breast cancer oncogene and induces metabolic reprogramming. *J Clin Invest* 128(1):323-340.

71. Bu XM, Zhao CH, Dai XW (2008) Aberrant expression of Wnt antagonist SFRP1 in pancreatic cancer. *Chin Med J (Engl)* 121: 952-955.

72. Kaur S, Katsoulidis E, Platanias LC (2008) Akt and mRNA translation by interferons. *Cell Cycle* 7: 2112-2116.

73. Bao B, Azmi AS, Ali S, Ahmad A, Li Y, et al. (2012) The biological kinship of hypoxia with CSC and EMT and their relationship with deregulated expression of miRNAs and tumor aggressiveness. *Biochim Biophys Acta* 1826: 272-296.

74. Singh RR, Kunkalla K, Qu C, Schlette E, Neelapu SS, et al. (2011) ABCG2 is a direct transcriptional target of hedgehog signaling and involved in stroma-induced drug tolerance in diffuse large B-cell lymphoma. *Oncogene* 30: 4874-4886.

75. Chang JT, Carvalho C, Mori S, Bild AH, Gatza ML, et al. (2009) A genomic strategy to elucidate modules of oncogenic pathway signaling networks. *Mol Cell* 34: 104-114.

76. Wu D, Han B, Guo L, Fan Z (2016) Molecular mechanisms associated with breast cancer based on integrated gene expression profiling by bioinformatics analysis. *J Obstet Gynaecol* 36: 615-621.

77. Guo H, Zeng W, Feng L, Yu X, Li P, et al. (2017) Integrated transcriptomic analysis of distance-related field cancerization in rectal cancer patients. *Oncotarget* 8: 61107-61117.

78. Wu Z, Zhao XM, Chen L (2010) A systems biology approach to identify effective cocktail drugs. *BMC Syst Biol* 4 Suppl 2: S7.

79. Chen C, Shen H, Zhang LG, Liu J, Cao XG, et al. (2016) Construction and analysis of protein-protein interaction networks based on proteomics data of prostate cancer. *Int J Mol Med* 37: 1576-1586.

80. Palle K, Mani C, Tripathi K, Athar M (2015) Aberrant GLI1 Activation in DNA Damage Response, Carcinogenesis and Chemoresistance. *Cancers (Basel)* 7: 2330-2351.

81. Pardali K, Moustakas A (2007) Actions of TGF-beta as tumor suppressor and pro-metastatic factor in human cancer. *Biochim Biophys Acta* 1775: 21-62.

82. Chowdhury S, Pradhan RN, Sarkar RR (2013) Structural and logical analysis of a comprehensive hedgehog signaling pathway to identify alternative drug targets for glioma, colon and pancreatic cancer. *PLoS One* 8: e69132.

83. Yang L, Xie G, Fan Q, Xie J (2010) Activation of the hedgehog-signaling pathway in human cancer and the clinical implications. *Oncogene* 29: 469-481.

84. Mischel PS, Cloughesy TF (2003) Targeted molecular therapy of GBM. *Brain Pathol* 13: 52-61.

85. Endersby R, Baker SJ (2008) PTEN signaling in brain: neuropathology and tumorigenesis.

Oncogene 27: 5416-5430.

86. Regl G, Kasper M, Schnidar H, Eichberger T, Neill GW, et al. (2004) Activation of the BCL2 promoter in response to Hedgehog/GLI signal transduction is predominantly mediated by GLI2. *Cancer Res* 64: 7724-7731.

87. Stegh AH, DePinho RA (2011) Beyond effector caspase inhibition: Bcl2L12 neutralizes p53 signaling in glioblastoma. *Cell Cycle* 10: 33-38.

88. Laux H, Tomer R, Mader MT, Smida J, Budczies J, et al. (2004) Tumor-associated E-cadherin mutations do not induce Wnt target gene expression, but affect E-cadherin repressors. *Lab Invest* 84: 1372-1386.

89. Qin CF, Hao K, Tian XD, Xie XH, Yang YM (2012) Combined effects of EGFR and Hedgehog signaling pathway inhibition on the proliferation and apoptosis of pancreatic cancer cells. *Oncol Rep* 28: 519-526.

90. Deng Y, Wang Z, Zhang F, Qiao M, Yan Z, et al. (2016) A Blockade of IGF Signaling Sensitizes Human Ovarian Cancer Cells to the Anthelmintic Niclosamide-Induced Anti-Proliferative and Anticancer Activities. *Cell Physiol Biochem* 39: 871-888.

91. Sferra R, Pompili S, Festuccia C, Marampon F, Gravina GL, et al. (2017) The possible prognostic role of histone deacetylase and transforming growth factor beta/Smad signaling in high grade gliomas treated by radio-chemotherapy: a preliminary immunohistochemical study. *Eur J Histochem* 61: 2732.

92. Schizas D, Mastoraki A, Naar L, Spartalis E, Tsilimigras DI, et al. (2018) Concept of histone deacetylases in cancer: Reflections on esophageal carcinogenesis and treatment. *World J Gastroenterol* 24: 4635-4642.

93. Garmpis N, Damaskos C, Garmpi A, Dimitroulis D, Spartalis E, et al. (2017) Targeting Histone Deacetylases in Malignant Melanoma: A Future Therapeutic Agent or Just Great Expectations? *Anticancer Res* 37: 5355-5362.

94. Bitterman PB, Polunovsky VA (2012) Translational control of cell fate: from integration of environmental signals to breaching anticancer defense. *Cell Cycle* 11: 1097-1107.

95. Huan J, Wang L, Xing L, Qin X, Feng L, et al. (2014) Insights into significant pathways and gene interaction networks underlying breast cancer cell line MCF-7 treated with 17beta-estradiol (E2). *Gene* 533: 346-355.

96. Vargas DA, Sun M, Sadykov K, Kukuruzinska MA, Zaman MH (2016) The Integrated Role of Wnt/beta-Catenin, N-Glycosylation, and E-Cadherin-Mediated Adhesion in Network Dynamics. *PLoS Comput Biol* 12: e1005007.

97. Zhang Y, Fang L, Zang Y, Xu Z (2018) Identification of Core Genes and Key Pathways via Integrated Analysis of Gene Expression and DNA Methylation Profiles in Bladder Cancer. *Med Sci Monit* 24: 3024-3033.

98. Yajun C, Yuan T, Zhong W, Bin X (2018) Investigation of the molecular mechanisms underlying postoperative recurrence in prostate cancer by gene expression profiling. *Exp Ther Med* 15: 761-768.

99. Venkatesh V, Nataraj R, Thangaraj GS, Karthikeyan M, Gnanasekaran A, et al. (2018) Targeting Notch signalling pathway of cancer stem cells. *Stem Cell Investig* 5: 5.

100. Kim YM, Kahn M (2014) The role of the Wnt signaling pathway in cancer stem cells: prospects for drug development. *Res Rep Biochem* 4: 1-12.

101. Aggarwal BB, Sethi G, Baladandayuthapani V, Krishnan S, Shishodia S (2007) Targeting cell signaling pathways for drug discovery: an old lock needs a new key. *J Cell Biochem* 102: 580-592.

102. Besancon R, Valsesia-Wittmann S, Puisieux A, Caron de Fromentel C, Maguer-Satta V (2009) Cancer stem cells: the emerging challenge of drug targeting. *Curr Med Chem* 16: 394-416.

103. Bai XY, Zhang XC, Yang SQ, An SJ, Chen ZH, et al. (2016) Blockade of Hedgehog Signaling Synergistically Increases Sensitivity to Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors in Non-Small-Cell Lung Cancer Cell Lines. *PLoS One* 11: e0149370.

104. Rossi M, Magnoni L, Miracco C, Mori E, Tosi P, et al. (2011) beta-catenin and Gli1 are prognostic markers in glioblastoma. *Cancer Biol Ther* 11: 753-761.

105. Takezaki T, Hide T, Takanaga H, Nakamura H, Kuratsu J, et al. (2011) Essential role of the Hedgehog signaling pathway in human glioma-initiating cells. *Cancer Sci* 102: 1306-1312.

106. Agliano A, Calvo A, Box C (2017) The challenge of targeting cancer stem cells to halt metastasis. *Semin Cancer Biol* 44: 25-42.

107. Katoh M (2017) Canonical and non-canonical WNT signaling in cancer stem cells and their niches: Cellular heterogeneity, omics reprogramming, targeted therapy and tumor plasticity (Review). *Int J Oncol* 51: 1357-1369.

108. Ashizawa T, Miyata H, Iizuka A, Komiyama M, Oshita C, et al. (2013) Effect of the STAT3 inhibitor STX-0119 on the proliferation of cancer stem-like cells derived from recurrent glioblastoma. *Int J Oncol* 43: 219-227.

109. Koury J, Zhong L, Hao J (2017) Targeting Signaling Pathways in Cancer Stem Cells for Cancer Treatment. *Stem Cells Int* 2017: 2925869.

110. Bordonaro M, Lazarova DL (2015) CREB-binding protein, p300, butyrate, and Wnt signaling in colorectal cancer. *World J Gastroenterol* 21: 8238-8248.

111. Klampfer L (2006) Signal transducers and activators of transcription (STATs): Novel targets of chemopreventive and chemotherapeutic drugs. *Curr Cancer Drug Targets* 6: 107-121.

112. Staberg M, Michaelsen SR, Olsen LS, Nedergaard MK, Villingshoj M, et al. (2016) Combined EGFR- and notch inhibition display additive inhibitory effect on glioblastoma cell viability and glioblastoma-induced endothelial cell sprouting in vitro. *Cancer Cell Int* 16: 34.

113. Takahashi-Yanaga F, Kahn M (2010) Targeting Wnt signaling: can we safely eradicate cancer stem cells? *Clin Cancer Res* 16: 3153-3162.

114. Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414: 105-111.

115. Chen D, Wu M, Li Y, Chang I, Yuan Q, et al. (2017) Targeting BMI1(+) Cancer Stem Cells Overcomes Chemoresistance and Inhibits Metastases in Squamous Cell Carcinoma. *Cell Stem Cell* 20: 621-634 e626.

116. Tedesco KL, Lockhart AC, Berlin JD (2004) The epidermal growth factor receptor as a target for gastrointestinal cancer therapy. *Curr Treat Options Oncol* 5: 393-403.

117. Malinowsky K, Wolff C, Gundisch S, Berg D, Becker K (2010) Targeted therapies in cancer - challenges and chances offered by newly developed techniques for protein analysis in clinical tissues. *J Cancer* 2: 26-35.

118. Sawyers C (2004) Targeted cancer therapy. *Nature* 432: 294-297.

119. Rios P, Nunes-Xavier CE, Tabernero L, Kohn M, Pulido R (2014) Dual-specificity phosphatases as molecular targets for inhibition in human disease. *Antioxid Redox Signal* 20: 2251-2273.

120. Krause DS, Van Etten RA (2005) Tyrosine kinases as targets for cancer therapy. *N Engl J Med* 353: 172-187.

121. Hajduk PJ, Huth JR, Tse C (2005) Predicting protein druggability. *Drug Discov Today* 10: 1675-1682.

122. Dai XU, Jing RY, Guo Y, Dong YC, Wang YL, et al. (2015) Predicting the druggability of protein-protein interactions based on sequence and structure features of active pockets. *Curr Pharm Des* 21: 3051-3061.

123. Sanchez RI, Mesia-Vela S, Kauffman FC (2001) Challenges of cancer drug design: a drug

metabolism perspective. *Curr Cancer Drug Targets* 1: 1-32.

124. Bull SC, Doig AJ (2015) Properties of protein drug target classes. *PLoS One* 10: e0117955.

125. Hoelder S, Clarke PA, Workman P (2012) Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Mol Oncol* 6: 155-176.

126. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* 7: 56.

127. Samatar AA, Poulikakos PI (2014) Targeting RAS-ERK signalling in cancer: promises and challenges. *Nat Rev Drug Discov* 13: 928-942.

128. Jones S, Zhang X, Parsons DW, Lin JC-H, Leary RJ, et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* 321: 1801-1806.

129. Parsons DW, Jones S, Zhang X, Lin JC-H, Leary RJ, et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science* 321: 1807-1812.

130. Roessler S, Lin G, Forgues M, Budhu A, Hoover S, et al. (2015) Integrative Genomic and Transcriptomic Characterization of Matched Primary and Metastatic Liver and Colorectal Carcinoma. *International journal of biological sciences* 11: 88.

131. Ng CK, Schultheis AM, Bidard F-C, Weigelt B, Reis-Filho JS (2015) Breast Cancer Genomics from Microarrays to Massively Parallel Sequencing: Paradigms and New Insights. *Journal of the National Cancer Institute* 107: djv015.

132. O'Driscoll L, McMorrow J, Doolan P, McKiernan E, Mehta JP, et al. (2006) Investigation of the molecular profile of basal cell carcinoma using whole genome microarrays. *Molecular cancer* 5: 74.

133. Lin EH, Kao YR, Lin CA, Kuo TY, Yang SP, et al. (2016) Hedgehog pathway maintains cell survival under stress conditions, and drives drug resistance in lung adenocarcinoma. *Oncotarget* 7: 24179-24193.

134. Kuang Y, Rogers A, Yeap BY, Wang L, Makrigiorgos M, et al. (2009) Noninvasive detection of EGFR T790M in gefitinib or erlotinib resistant non-small cell lung cancer. *Clin Cancer Res* 15: 2630-2636.

135. Hu T, Li C (2010) Convergence between Wnt-beta-catenin and EGFR signaling in cancer. *Mol Cancer* 9: 236.

136. Atwood SX, Whitson RJ, Oro AE (2014) Advanced treatment for basal cell carcinomas. *Cold Spring Harb Perspect Med* 4: a013581.

137. Brinkhuizen T, Reinders MG, van Geel M, Hendriksen AJ, Paulussen AD, et al. (2014) Acquired resistance to the Hedgehog pathway inhibitor vismodegib due to smoothened mutations in treatment of locally advanced basal cell carcinoma. *J Am Acad Dermatol* 71: 1005-1008.

138. Fisher R, Pusztai L, Swanton C (2013) Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* 108: 479-485.

139. Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, et al. (2014) Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344: 1396-1401.

140. Gillies RJ, Verduzco D, Gatenby RA (2012) Evolutionary dynamics of carcinogenesis and why targeted therapy does not work. *Nat Rev Cancer* 12: 487-493.

141. Hu YY, Zheng MH, Cheng G, Li L, Liang L, et al. (2011) Notch signaling contributes to the maintenance of both normal neural stem cells and patient-derived glioma stem cells. *BMC Cancer* 11: 82.

142. Sandberg CJ, Altschuler G, Jeong J, Stromme KK, Stangeland B, et al. (2013) Comparison of glioma stem cells to neural stem cells from the adult human brain identifies dysregulated Wnt-signaling and a fingerprint associated with clinical outcome. *Exp Cell Res* 319: 2230-2243.

143. Wheeler DA, Wang L (2013) From human genome to cancer genome: the first decade. *Genome Res* 23: 1054-1062.

144. Ahn AC, Tewari M, Poon CS, Phillips RS (2006) The limits of reductionism in medicine: could systems biology offer an alternative? *PLoS Med* 3: e208.

145. Weinberg RA (1991) Tumor suppressor genes. *Science* 254: 1138-1146.

146. Stanbridge EJ (1990) Identifying tumor suppressor genes in human colorectal cancer. *Science* 247: 12-13.

147. Land H, Parada LF, Weinberg RA (1983) Cellular oncogenes and multistep carcinogenesis. *Science* 222: 771-778.

148. Ullrich A, Schlessinger J (1990) Signal transduction by receptors with tyrosine kinase activity. *Cell* 61: 203-212.

149. Weinstein IB (1988) The origins of human cancer: molecular mechanisms of carcinogenesis and their implications for cancer prevention and treatment--twenty-seventh G.H.A. Clowes memorial award lecture. *Cancer Res* 48: 4135-4143.

150. Ogata H, Goto S, Fujibuchi W, Kanehisa M (1998) Computation with the KEGG pathway database. *Biosystems* 47: 119-128.

151. Bhalla US, Iyengar R (1999) Emergent properties of networks of biological signaling pathways. *Science* 283: 381-387.

152. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proc Natl Acad Sci USA* 99: 7821-7826.

153. Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* 5: 763-769.

154. Teschendorff AE, Banerji CR, Severini S, Kuehn R, Sollich P (2015) Increased signaling entropy in cancer requires the scale-free property of protein interaction networks. *Sci Rep* 5: 9646.

155. Aggarwal A, Guo DL, Hoshida Y, Yuen ST, Chu KM, et al. (2006) Topological and functional discovery in a gene coexpression meta-network of gastric cancer. *Cancer Res* 66: 232-241.

156. Ladha J, Donakonda S, Agrawal S, Thota B, Srividya MR, et al. (2010) Glioblastoma-specific protein interaction network identifies PP1A and CSK21 as connecting molecules between cell cycle-associated genes. *Cancer Res* 70: 6437-6447.

157. Li Y, Chen J, Zhang J, Wang Z, Shao T, et al. (2015) Construction and analysis of lncRNA-lncRNA synergistic networks to reveal clinically relevant lncRNAs in cancer. *Oncotarget* 6: 25003-25016.

158. Hu WL, Zhou XH (2017) Identification of prognostic signature in cancer based on DNA methylation interaction network. *BMC Med Genomics* 10: 63.

159. Santolini M, Barabasi AL (2018) Predicting perturbation patterns from the topology of biological networks. *Proc Natl Acad Sci USA* 115: E6375-E6383.

160. Muller H, Acquati F (2008) Topological properties of co-occurrence networks in published gene expression signatures. *Bioinform Biol Insights* 2: 203-213.

161. Kyoda KM, Morohashi M, Onami S, Kitano H (2000) A gene network inference method from continuous-value gene expression data of wild-type and mutants. *Genome Inform Ser Workshop Genome Inform* 11: 196-204.

162. Tran TD, Kwon YK (2014) Hierarchical closeness efficiently predicts disease genes in a directed signaling network. *Comput Biol Chem* 53PB: 191-197.

163. Liu J, Hua P, Hui L, Zhang LL, Hu Z, et al. (2016) Identification of hub genes and pathways associated with hepatocellular carcinoma based on network strategy. *Exp Ther Med* 12: 2109-2119.

164. Deisboeck TS, Zhang L, Yoon J, Costa J (2009) In silico cancer modeling: is it ready for prime time? *Nat Clin Pract Oncol* 6: 34-42.

165. Ping Y, Zhang H, Deng Y, Wang L, Zhao H, et al. (2014) IndividualizedPath: identifying genetic alterations contributing to the dysfunctional pathways in glioblastoma individuals. *Mol Biosyst*

10: 2031-2042.

166. Eungdamrong NJ, Iyengar R (2004) Modeling cell signaling networks. *Biology of the Cell* 96: 355-362.

167. Saha K, Schaffer DV (2006) Signal dynamics in Sonic hedgehog tissue patterning. *Development* 133: 889-900.

168. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, et al. (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* 3: e163.

169. Hinkelmann F, Murrugarra D, Jarrah AS, Laubenbacher R (2011) A mathematical framework for agent based models of complex biological networks. *Bull Math Biol* 73: 1583-1602.

170. Palsson B (2000) The challenges of in silico biology. *Nat Biotechnol* 18: 1147-1150.

171. Bornholdt S (2008) Boolean network models of cellular regulation: prospects and limitations. *J R Soc Interface* 5 Suppl 1: S85-94.

172. de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 9: 67-103.

173. Qin T, Matmati N, Tsoi LC, Mohanty BK, Gao N, et al. (2014) Finding pathway-modulating genes from a novel Ontology Fingerprint-derived gene network. *Nucleic Acids Res* 42: e138.

174. Simeone P, Trerotola M, Urbanella A, Lattanzio R, Ciavardelli D, et al. (2014) A unique four-hub protein cluster associates to glioblastoma progression. *PLoS One* 9: e103030.

175. Raeymaekers L (2002) Dynamics of Boolean networks controlled by biologically meaningful functions. *J Theor Biol* 218: 331-341.

176. van Riel NA (2006) Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in Bioinformatics* 7: 364-374.

177. Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, et al. (2011) Modeling formalisms in Systems Biology. *AMB Express* 1: 45.

178. Schreck KC, Taylor P, Marchionni L, Gopalakrishnan V, Bar EE, et al. (2010) The Notch target Hes1 directly modulates Gli1 expression and Hedgehog signaling: a potential mechanism of therapeutic resistance. *Clin Cancer Res* 16: 6060-6070.

179. Bi Y, Li H, Yi D, Bai Y, Zhong S, et al. (2018) beta-catenin contributes to cordycepin-induced MGMT inhibition and reduction of temozolomide resistance in glioma cells by increasing intracellular reactive oxygen species. *Cancer Lett* 435: 66-79.

180. Alcedo-Guardia R, Labat E, Blas-Boria D, Vivas-Mejia PE (2016) Diagnosis and New Treatment Modalities for Glioblastoma: Do They Improve Patient Survival? *Curr Mol Med* 16(5), 447-464.

181. Alifieris C, Trafalis DT (2015) Glioblastoma multiforme: Pathogenesis and treatment. *Pharmacol Ther* 152: 63-82.

182. Nagpal J, Jamoona A, Gulati ND, Mohan A, Braun A, et al. (2006) Revisiting the role of p53 in primary and secondary glioblastomas. *Anticancer Res* 26: 4633-4639.

183. Eder K, Kalman B (2014) Molecular heterogeneity of glioblastoma and its clinical relevance. *Pathol Oncol Res* 20: 777-787.

184. Brown DV, Filiz G, Daniel PM, Hollande F, Dworkin S, et al. (2017) Expression of CD133 and CD44 in glioblastoma stem cells correlates with cell proliferation, phenotype stability and intra-tumor heterogeneity. *PLoS One* 12: e0172791.

185. Chowdhury S, Sarkar R (2013) Drug Targets and Biomarker Identification from Computational Study of Human Notch Signaling Pathway. *Clin Exp Pharmacol* 3: 2161-1459.1000137.

186. Funahashi A, Jouraku A, Matsuoka Y, Kitano H (2007) Integration of CellDesigner and SABIO-RK. *In Silico Biol* 7: S81-90.

187. Petri V, Jayaraman P, Tutaj M, Hayman GT, Smith JR, et al. (2014) The pathway ontology - updates and applications. *J Biomed Semantics* 5: 7.

188. McDowall MD, Scott MS, Barton GJ (2009) PIPs: human protein–protein interaction prediction database. *Nucleic acids research* 37: D651-D656.

189. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, et al. (2016) The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Research*: gkw937.

190. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, et al. (2016) The genecards suite: from gene data mining to disease genome sequence analyses. *Current Protocols in Bioinformatics*: 1.30. 31-31.30. 33.

191. Liu X, Yu X, Zack DJ, Zhu H, Qian J (2008) TiGER: a database for tissue-specific gene expression and regulation. *BMC bioinformatics* 9: 271.

192. Razick S, Magklaras G, Donaldson IM (2008) iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* 9: 405.

193. Stark C, Breitkreutz B-J, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Research* 34: D535-D539.

194. Kerrien S, Aranda B, Breuza L, Bridge A, Broackes-Carter F, et al. (2012) The IntAct molecular interaction database in 2012. *Nucleic acids research* 40: D841-D846.

195. Prasad TK, Goel R, Kandasamy K, Keerthikumar S, Kumar S, et al. (2009) Human protein reference database—2009 update. *Nucleic acids research* 37: D767-D772.

196. Rappaport N, Twik M, Plaschkes I, Nudel R, Stein TI, et al. (2016) MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Research*: gkw1012.

197. Ellson J, Gansner E, Koutsofios L, North SC, Woodhull G. Graphviz—open source graph drawing tools; 2001. Springer. pp. 483-484.

198. Demir E, Cary MP, Paley S, Fukuda K, Lemer C, et al. (2010) The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 28: 935-942.

199. Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, et al. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524-531.

200. Webb RL, Ma'ayan A (2011) Sig2BioPAX: Java tool for converting flat files to BioPAX Level 3 format. *Source Code Biol Med* 6: 5-5.

201. Liu X, Pan L (2015) Identifying Driver Nodes in the Human Signaling Network Using Structural Controllability Analysis. *IEEE/ACM Trans Comput Biol Bioinform* 12: 467-472.

202. Dorogovtsev SN, Mendes JF, Samukhin AN (2001) Giant strongly connected component of directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 64: 025101.

203. Newman ME (2004) Analysis of weighted networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70: 056131.

204. Breitkreutz D, Hlatky L, Rietman E, Tuszynski JA (2012) Molecular signaling network complexity is correlated with cancer patient survivability. *Proc Natl Acad Sci USA* 109: 9209-9212.

205. Carr JM, Trygubenko SA, Wales DJ (2005) Finding pathways between distant local minima. *J Chem Phys* 122: 234903.

206. Batagelj V, Mrvar A. Pajek—analysis and visualization of large networks; 2002. Springer. pp. 477-478.

207. Bastian M, Heymann S, Jacomy M (2009) Gephi: an open source software for exploring and manipulating networks. *ICWSM* 8: 361-362.

208. Csardi G, Nepusz T (2006) The igraph software package for complex network research.

InterJournal, Complex Systems 1695.

209. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 13: 2498-2504.

210. Wang J, Zhong J, Chen G, Li M, Wu FX, et al. (2015) ClusterViz: A Cytoscape APP for Cluster Analysis of Biological Network. *IEEE/ACM Trans Comput Biol Bioinform* 12: 815-822.

211. Nazarieh M, Wiese A, Will T, Hamed M, Helms V (2016) Identification of key player genes in gene regulatory networks. *BMC Syst Biol* 10: 88.

212. Scardoni G, Tosadori G, Pratap S, Spoto F, Laudanna C (2015) Finding the shortest path with PesCa: a tool for network reconstruction. *F1000Res* 4: 484.

213. Verdicchio MP, Kim S (2011) Identifying targets for intervention by analyzing basins of attraction. *Pac Symp Biocomput*: 350-361.

214. Zhou X, Wang X, Dougherty ER (2003) Binarization of microarray data on the basis of a mixture model. *Mol Cancer Ther* 2: 679-684.

215. Mussel C, Schmid F, Blatte TJ, Hopfensitz M, Lausser L, et al. (2016) BiTrinA--multiscale binarization and trinarization with quality analysis. *Bioinformatics* 32: 465-468.

216. Hopfensitz M, Mussel C, Wawra C, Maucher M, Kuhl M, et al. (2012) Multiscale binarization of gene expression data for reconstructing Boolean networks. *IEEE/ACM Trans Comput Biol Bioinform* 9: 487-498.

217. Shmulevich I, Zhang W (2002) Binary analysis and optimization-based normalization of gene expression data. *Bioinformatics* 18: 555-565.

218. Klamt S, Saez-Rodriguez J, Gilles ED (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* 1: 2.

219. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, et al. (2006) Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. *Cancer Cell* 9: 287-300.

220. Anilkumar U, Prehn JH (2014) Anti-apoptotic BCL-2 family proteins in acute neural injury. *Front Cell Neurosci* 8: 281.

221. Liu G, Yuan X, Zeng Z, Tunici P, Ng H, et al. (2006) Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Mol Cancer* 5: 67.

222. Cenciarelli C, Marei HE, Zonfrillo M, Casalbore P, Felsani A, et al. (2017) The interference of Notch1 target Hes1 affects cell growth, differentiation and invasiveness of glioblastoma stem cells through modulation of multiple oncogenic targets. *Oncotarget* 8: 17873-17886.

223. Balenci L, van der Kooy D (2014) Notch signaling induces retinal stem-like properties in perinatal neural retina progenitors and promotes symmetric divisions in adult retinal stem cells. *Stem Cells Dev* 23: 230-244.

224. Gomez-Nicola D, Valle-Argos B, Pallas-Bazarra N, Nieto-Sampedro M (2011) Interleukin-15 regulates proliferation and self-renewal of adult neural stem cells. *Mol Biol Cell* 22: 1960-1970.

225. Alexander JE, Hunt DF, Lee MK, Shabanowitz J, Michel H, et al. (1991) Characterization of posttranslational modifications in neuron-specific class III beta-tubulin by mass spectrometry. *Proc Natl Acad Sci USA* 88: 4685-4689.

226. Aparicio E, Mathieu P, Pereira Luppi M, Almeira Gubiani MF, Adamo AM (2013) The Notch signaling pathway: its role in focal CNS demyelination and apotransferrin-induced remyelination. *J Neurochem* 127: 819-836.

227. Hong S, Song MR (2014) STAT3 but not STAT1 is required for astrocyte differentiation. *PLoS One* 9: e86851.

228. Bizen N, Inoue T, Shimizu T, Tabu K, Kagawa T, et al. (2014) A growth-promoting signaling component cyclin D1 in neural stem cells has antiastrogliogenic function to execute self-renewal. *Stem*

Cells 32: 1602-1615.

229. Faria MH, Khayat AS, Burbano RR, Rabenhorst SH (2008) c -MYC amplification and expression in astrocytic tumors. *Acta Neuropathol* 116: 87-95.

230. Carnemolla B, Castellani P, Ponassi M, Borsi L, Urbini S, et al. (1999) Identification of a glioblastoma-associated tenascin-C isoform by a high affinity recombinant antibody. *Am J Pathol* 154: 1345-1352.

231. Cau E, Blader P (2009) Notch activity in the nervous system: to switch or not switch? *Neural Dev* 4: 36.

232. Wang YZ, Plane JM, Jiang P, Zhou CJ, Deng W (2011) Concise review: Quiescent and active states of endogenous adult neural stem cells: identification and characterization. *Stem Cells* 29: 907-912.

233. Shore J, Johnson R (1980) Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Trans Inform Theory* 26: 26-37.

234. Sottoriva A, Spiteri I, Piccirillo SG, Touloumis A, Collins VP, et al. (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci USA* 110: 4009-4014.

235. Guo W, Yang G, Wu W, He L, Sun M (2014) A parallel attractor-finding algorithm based on Boolean satisfiability for genetic regulatory networks. *PLoS One* 9: e94258.

236. Kauffman S, Peterson C, Samuelsson B, Troein C (2003) Random Boolean network models and the yeast transcriptional network. *Proc Natl Acad Sci USA* 100: 14796-14799.

237. Cedilnik A, Kosmelj K, Blejec A (2004) The distribution of the ratio of jointly normal variables. *Metodoloski zvezki* 1: 99.

238. Baldi P, Benz RW (2008) BLASTing small molecules—statistics and extreme statistics of chemical similarity scores. *Bioinformatics* 24: i357-i365.

239. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 45: 1113-1120.

240. Cancer Genome Atlas Research N (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455: 1061-1068.

241. Dai Z, Sheridan JM, Gearing LJ, Moore DL, Su S, et al. (2014) edgeR: a versatile tool for the analysis of shRNA-seq and CRISPR-Cas9 genetic screens. *F1000Res* 3: 95.

242. Straube J, Huang BE, Cao KL (2017) DynOmics to identify delays and co-expression patterns across time course experiments. *Sci Rep* 7: 40131.

243. Gogna P, Lingalugari M, Chandy J, Heller E, Hasaneen E, et al. (2012) Quaternary logic and applications using multiple quantum well based SWSFETs. *International Journal of VLSI Design & Communication Systems* 3: 27.

244. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, et al. (2007) A Logical Model Provides Insights into T Cell Receptor Signaling. *PLoS Comput Biol* 3: e163.

245. Sahin O, Frohlich H, Lobke C, Korf U, Burmester S, et al. (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst Biol* 3: 1.

246. Kitano H (2002) Systems biology: a brief overview. *Science* 295: 1662-1664.

247. Fiehn O (2001) Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comparative and Functional Genomics* 2: 155-168.

248. Sarkar R (2016) The Big Data Deluge in Biology: Challenges and Solutions. *Journal of Informatics and Data Mining*.

249. Huang DW, Sherman BT, Lempicki RA (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37: 1-13.

250. Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics.

Bioinformatics 23: 2507-2517.

251. Bauer-Mehren A, Furlong LI, Sanz F (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* 5: 290.

252. Ghosh S, Matsuoka Y, Asai Y, Hsin KY, Kitano H (2011) Software for systems biology: from tools to integrated platforms. *Nat Rev Genet* 12: 821-832.

253. Altmäe S, Esteban FJ, Stavreus-Evers A, Simón C, Giudice L, et al. (2014) Guidelines for the design, analysis and interpretation of 'omics' data: focus on human endometrium. *Human Repro Update* 20: 12-28.

254. Collins FS, Morgan M, Patrinos A (2003) The Human Genome Project: lessons from large-scale biology. *Science* 300: 286-290.

255. Collins FS, Green ED, Guttmacher AE, Guyer MS (2003) A vision for the future of genomics research. *Nature* 422: 835-847.

256. Wang K, Li M, Hakonarson H (2010) Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11: 843-854.

257. Hao T, Ma HW, Zhao XM, Goryanin I (2010) Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* 11: 393.

258. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 27: 29-34.

259. Bader GD, Cary MP, Sander C (2006) Pathguide: a pathway resource list. *Nucleic Acids Res* 34: D504-D506.

260. Yi M, Horton JD, Cohen JC, Hobbs HH, Stephens RM (2006) WholePathwayScope: a comprehensive pathway-based analysis tool for high-throughput data. *BMC Bioinformatics* 7: 30.

261. Yang JO, Charny P, Lee B, Kim S, Bhak J, et al. (2007) GS2PATH: a web-based integrated analysis tool for finding functional relationships using gene ontology and biochemical pathway data. *Bioinformation* 2: 194-196.

262. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, et al. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25: 1091-1093.

263. Khatri P, Sirota M, Butte AJ (2012) Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Comput Biol* 8: e1002375.

264. Green M, Karp P (2006) The outcomes of pathway database computations depend on pathway ontology. *Nucleic Acids Res* 34: 3687-3697.

265. Likić VA (2006) Databases of metabolic pathways. *Biochem Mol Biol Edu* 34: 408-412.

266. Karp PD, Riley M. Representations of metabolic knowledge; 1993. pp. 207-215.

267. Karp PD, Paley SM. Representations of metabolic knowledge: pathways; 1994. pp. 203-211.

268. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, et al. (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 40: D742-753.

269. Karp PD, Riley M, Paley SM, Pellegrini-Toole A (2002) The MetaCyc Database. *Nucleic Acids Res* 30: 59-61.

270. Jiang C, Xuan Z, Zhao F, Zhang MQ (2007) TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res* 35: D137-D140.

271. Gerstmann S (2002) Signaling PAthway Database (SPAD)? ms an upcomingonline database on signal transduction. *Signal Transduction* 2: 49-53.

272. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, et al. (2010) NetPath: a public resource of curated signal transduction pathways. *Genome Biol* 11: R3.

273. Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ö, et al. (2010) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res 39(suppl_1): D685-D690*.

274. Croft D (2013) Building models using Reactome pathways as templates. *In Silico Systems Biology*: Springer. pp. 273-283.

275. Croft D (2011) BioPax and SBML output from Reactome: where we are, where we are heading. doi: 10.1038/npre.2011.6376.1

276. Funahashi A, Jouraku A, Kitano H. Converting KEGG pathway database to SBML. In 8th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004).

277. Wrzodek C, Büchel F, Ruff M, Dräger A, Zell A (2013) Precise generation of systems biology models from KEGG pathways. *BMC Syst Biol* 7: 15.

278. Büchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, et al. (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol* 7: 116.

279. Yamamoto S, Sakai N, Nakamura H, Fukagawa H, Fukuda K, et al. (2010) INOH: ontology-based highly structured database of signal transduction pathways. *Database: the journal of biological databases and curation* 2011: bar052-bar052.

280. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, et al. (2009) BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 37: W170-173.

281. McKusick VA (2007) Mendelian Inheritance in Man and its online version, OMIM. *American journal of human genetics* 80: 588.

282. Wilcoxon F, Katti S, Wilcox RA (1970) Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected tables in mathematical statistics* 1: 171-259.

283. Schlamp K, Weinmann A, Krupp M, Maass T, Galle P, et al. (2008) BlotBase: a northern blot database. *Gene* 427: 47-50.

284. Buchkremer S, Hendel J, Krupp M, Weinmann A, Schlamp K, et al. (2010) Library of molecular associations: curating the complex molecular basis of liver diseases. *BMC Genomics* 11: 189.

285. Isserlin R, El-Badrawi RA, Bader GD (2011) The biomolecular interaction network database in PSI-MI 2.5. *Database* 2011: baq037.

286. Kohn KW, Aladjem MI (2006) Circuit diagrams for biological networks. *Molecular Syst Biol* 2: 2006.0002. doi: 10.1038/msb4100044

287. Frank MR, Omiecinski ER, Navathe SB. Adaptive and automated index selection in RDBMS; 1992. Springer. pp. 277-292.

288. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res* 41: D991-D995.

289. Consortium U (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 42: D191-D198.

290. Kanehisa M, Goto S, Kawashima S, Nakaya A (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res* 30: 42 - 46.

291. Pico AR, Kelder T, Van Iersel MP, Hanspers K, Conklin BR, et al. (2008) WikiPathways: pathway editing for the people. *PLoS Biology* 6: e184.

292. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, et al. (2009) PID: the pathway interaction database. *Nucleic Acids Res* 37: D674-D679.

293. Mann RK, Beachy PA (2000) Cholesterol modification of proteins. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids* 1529: 188-202.

294. Cohen P (1988) Review Lecture: Protein Phosphorylation and Hormone Action. *Proceedings of the Royal society of London Series B Biological Sciences* 234: 115-144.

295. Willard FS, Crouch MF (2000) Nuclear and cytoskeletal translocation and localization of heterotrimeric G-proteins. *Immunol Cell Biol* 78: 387-394.

296. Gaston K, Jayaraman P-S (2003) Transcriptional repression in eukaryotes: repressors and repression mechanisms. *Cellular and Molecular Life Sciences CMLS* 60: 721-741.

297. Pribluda VS, Pribluda C, Metzger H (1994) Transphosphorylation as the mechanism by which the high-affinity receptor for IgE is phosphorylated upon aggregation. *Proceedings of the National Academy of Sciences* 91: 11246-11250.

298. Pawson T, Scott JD (1997) Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278: 2075-2080.

299. Crabtree GR (2001) Calcium, calcineurin, and the control of transcription. *Journal of Biological Chemistry* 276: 2313-2316.

300. Komander D (2009) The emerging complexity of protein ubiquitination. *Biochem Society Trans* 37: 937-953.

301. Mullor JL, Sánchez P (2002) Pathways and consequences: Hedgehog signaling in human disease. *Trends Cell Biol* 12: 562-569.

302. Nusse R (2005) Wnt signaling in disease and in development. *Cell Research* 15: 28-32.

303. Polakis P (2000) Wnt signaling and cancer. *Genes Dev* 14: 1837-1851.

304. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5: 101-113.

305. Lee JM, Gianchandani EP, Eddy JA, Papin JA (2008) Dynamic analysis of integrated signaling, metabolic, and regulatory networks. *PLoS Comput Biol* 4: e1000086.

306. Singh A, Nascimento JM, Kowar S, Busch H, Boerries M (2012) Boolean approach to signalling pathway modelling in HGF-induced keratinocyte migration. *Bioinformatics* 28: i495-i501.

307. Fumiã HF, Martins ML (2013) Boolean Network Model for Cancer Pathways: Predicting Carcinogenesis and Targeted Therapy Outcomes. *PLoS One* 8: e69008.

308. Müssel C, Hopfensitz M, Kestler HA (2010) BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 26: 1378-1380.

309. Zhang R, Shah MV, Yang J, Nyland SB, Liu X, et al. (2008) Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci USA* 105: 16308-16313.

310. Hairer E, Wanner G , Nørsett SP (1993) "Runge-Kutta and Extrapolation Methods" in Solving Ordinary Differential Equations I. Berlin, Heidelberg, Springer, pp. 129-353.

311. Ferguson C, Alpern E, Miclau T, Helms JA (1999) Does adult fracture repair recapitulate embryonic skeletal formation? *Mech Dev* 87: 57-66.

312. Infante P, Alfonsi R, Ingallina C, Quaglio D, Ghirga F, et al. (2016) Inhibition of Hedgehog-dependent tumors and cancer stem cells by a newly identified naturally occurring chemotype. *Cell Death Dis* 7: e2376.

313. Fattahi S, Pilehchian Langroudi M, Akhavan-Niaki H (2018) Hedgehog signaling pathway: Epigenetic regulation and role in disease and cancer development. *J Cell Physiol* 233: 5726-5735.

314. Vorechovsky I, Benediktsson KP, Toftgard R (1999) The patched/hedgehog/smoothened signalling pathway in human breast cancer: no evidence for H133Y SHH, PTCH and SMO mutations. *Eur J Cancer* 35: 711-713.

315. Sinha N, Chowdhury S, Sarkar RR (2017) Deciphering structural stability and binding mechanisms of potential antagonists with smoothened protein. *J Biomol Struct Dyn*: 1-21.

316. Osterlund T, Kogerman P (2006) Hedgehog signalling: how to get from Smo to Ci and Gli. *Trends Cell Biol* 16: 176-180.

317. Dillon R, Gadgil C, Othmer HG (2003) Short- and long-range effects of Sonic hedgehog in

limb development. *Proc Natl Acad Sci USA* 100: 10152-10157.

318. Jia X, Min L, Zhu S, Zhang S, Huang X (2018) Loss of sonic hedgehog gene leads to muscle development disorder and megaesophagus in mice. *FASEB J*: fj201701581R.

319. Ogden SK, Ascano M, Jr., Stegman MA, Robbins DJ (2004) Regulation of Hedgehog signaling: a complex story. *Biochem Pharmacol* 67: 805-814.

320. Rohatgi R, Scott MP (2007) Patching the gaps in Hedgehog signalling. *Nat Cell Biol* 9: 1005-1009.

321. Xu X, Ding H, Rao G, Arora S, Saclarides CP, et al. (2012) Activation of the Sonic Hedgehog pathway in thyroid neoplasms and its potential role in tumor cell proliferation. *Endocr Relat Cancer* 19: 167-179.

322. Chen B, Trang V, Lee A, Williams NS, Wilson AN, et al. (2016) Posaconazole, a Second-Generation Triazole Antifungal Drug, Inhibits the Hedgehog Signaling Pathway and Progression of Basal Cell Carcinoma. *Mol Cancer Ther* 15: 866-876.

323. Kiesslich T, Neureiter D (2012) Advances in targeting the Hedgehog signaling pathway in cancer therapy. *Expert Opin Ther Targets* 16: 151-156.

324. Li Y, Maitah MY, Ahmad A, Kong D, Bao B, et al. (2012) Targeting the Hedgehog signaling pathway for cancer therapy. *Expert Opin Ther Targets* 16: 49-66.

325. Wong H, Alicke B, West KA, Pacheco P, La H, et al. (2011) Pharmacokinetic-pharmacodynamic analysis of vismodegib in preclinical models of mutational and ligand-dependent Hedgehog pathway activation. *Clin Cancer Res* 17: 4682-4692.

326. LoRusso PM, Rudin CM, Reddy JC, Tibes R, Weiss GJ, et al. (2011) Phase I trial of hedgehog pathway inhibitor vismodegib (GDC-0449) in patients with refractory, locally advanced or metastatic solid tumors. *Clin Cancer Res* 17: 2502-2511.

327. Li QR, Zhao H, Zhang XS, Lang H, Yu K (2018) Novel-smoothened inhibitors for therapeutic targeting of naive and drug-resistant hedgehog pathway-driven cancers. *Acta Pharmacol Sin*. doi: 10.1038/s41401-018-0019-5

328. Pricl S, Cortelazzi B, Dal Col V, Marson D, Laurini E, et al. (2015) Smoothened (SMO) receptor mutations dictate resistance to vismodegib in basal cell carcinoma. *Mol Oncol* 9: 389-397.

329. Cucchi D, Occhione MA, Gulino A, De Smaele E (2012) Hedgehog signaling pathway and its targets for treatment in basal cell carcinoma. *J Exp Pharmacol* 4: 173-185.

330. Dong X, Wang C, Chen Z, Zhao W (2018) Overcoming the resistance mechanisms of Smoothened inhibitors. *Drug Discov Today* 23: 704-710.

331. Zhang X, Tian Y, Yang Y, Hao J (2017) Development of anticancer agents targeting the Hedgehog signaling. *Cell Mol Life Sci* 74: 2773-2782.

332. Buetti-Dinh A, Jensen R, Friedman R (2018) A computational study of hedgehog signalling involved in basal cell carcinoma reveals the potential and limitation of combination therapy. *BMC Cancer* 18: 569.

333. Samaga R, Klamt S (2013) Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal* 11: 43.

334. Lum L, Zhang C, Oh S, Mann RK, von Kessler DP, et al. (2003) Hedgehog signal transduction via Smoothened association with a cytoplasmic complex scaffolded by the atypical kinesin, Costal-2. *Mol Cell* 12: 1261-1274.

335. Jiang J, Hui CC (2008) Hedgehog signaling in development and cancer. *Dev Cell* 15: 801-812.

336. Shao J, Zhang L, Gao J, Li Z, Chen Z (2006) Aberrant expression of PTCH (patched gene) and Smo (smoothened gene) in human pancreatic cancerous tissues and its association with hyperglycemia. *Pancreas* 33: 38-44.

337. Katoh Y, Katoh M (2006) Hedgehog signaling pathway and gastrointestinal stem cell

signaling network (review). *Int J Mol Med* 18: 1019-1023.

338. Chen W, Tang T, Eastham-Anderson J, Dunlap D, Alicke B, et al. (2011) Canonical hedgehog signaling augments tumor angiogenesis by induction of VEGF-A in stromal perivascular cells. *Proc Natl Acad Sci USA* 108: 9589-9594.

339. Zhang W, Kang JS, Cole F, Yi MJ, Krauss RS (2006) Cdo functions at multiple points in the Sonic Hedgehog pathway, and Cdo-deficient mice accurately model human holoprosencephaly. *Dev Cell* 10: 657-665.

340. Beachy PA, Hymowitz SG, Lazarus RA, Leahy DJ, Siebold C (2010) Interactions between Hedgehog proteins and their binding partners come into view. *Genes Dev* 24: 2001-2012.

341. Trnski D, Sabol M, Gojevic A, Martinic M, Ozretic P, et al. (2015) GSK3beta and Gli3 play a role in activation of Hedgehog-Gli pathway in human colon cancer - Targeting GSK3beta downregulates the signaling pathway and reduces cell proliferation. *Biochim Biophys Acta* 1852: 2574-2584.

342. Osterlund T, Everman DB, Betz RC, Mosca M, Nothen MM, et al. (2004) The FU gene and its possible protein isoforms. *BMC Genomics* 5: 49.

343. Maloverjan A, Piirsoo M, Michelson P, Kogerman P, Osterlund T (2010) Identification of a novel serine/threonine kinase ULK3 as a positive regulator of Hedgehog pathway. *Exp Cell Res* 316: 627-637.

344. Kasper M, Schnidar H, Neill GW, Hanneder M, Klingler S, et al. (2006) Selective modulation of Hedgehog/GLI target gene expression by epidermal growth factor signaling in human keratinocytes. *Mol Cell Biol* 26: 6283-6298.

345. Ji Z, Mei FC, Xie J, Cheng X (2007) Oncogenic KRAS activates hedgehog signaling pathway in pancreatic cancer cells. *J Biol Chem* 282: 14048-14055.

346. Villavicencio EH, Yoon JW, Frank DJ, Fuchtbauer EM, Walterhouse DO, et al. (2002) Cooperative E-box regulation of human GLI1 by TWIST and USF. *Genesis* 32: 247-258.

347. Rao R, Salloum R, Xin M, Lu QR (2016) The G protein Galphas acts as a tumor suppressor in sonic hedgehog signaling-driven tumorigenesis. *Cell Cycle* 15: 1325-1330.

348. Jia J, Zhang L, Zhang Q, Tong C, Wang B, et al. (2005) Phosphorylation by double-time/CKIepsilon and CKIalpha targets cubitus interruptus for Slimb/beta-TRCP-mediated proteolytic processing. *Dev Cell* 9: 819-830.

349. Kise Y, Morinaka A, Teglund S, Miki H (2009) Sufu recruits GSK3beta for efficient processing of Gli3. *Biochem Biophys Res Commun* 387: 569-574.

350. Ruiz i Altaba A, Palma V, Dahmane N (2002) Hedgehog-Gli signalling and the growth of the brain. *Nat Rev Neurosci* 3: 24-33.

351. Mao J, Maye P, Kogerman P, Tejedor FJ, Toftgard R, et al. (2002) Regulation of Gli1 transcriptional activity in the nucleus by Dyrk1. *J Biol Chem* 277: 35156-35161.

352. Di Marcotullio L, Greco A, Mazza D, Canettieri G, Pietrosanti L, et al. (2011) Numb activates the E3 ligase Itch to control Gli1 function through a novel degradation signal. *Oncogene* 30: 65-76.

353. Cheng SY, Bishop JM (2002) Suppressor of Fused represses Gli-mediated transcription by recruiting the SAP18-mSin3 corepressor complex. *Proc Natl Acad Sci USA* 99: 5442-5447.

354. Wang B, Harrison W, Overbeek PA, Zheng H (2011) Transposon mutagenesis with coat color genotyping identifies an essential role for Skor2 in sonic hedgehog signaling and cerebellum development. *Development* 138: 4487-4497.

355. Dai P, Shinagawa T, Nomura T, Harada J, Kaul SC, et al. (2002) Ski is involved in transcriptional regulation by the repressor and full-length forms of Gli3. *Genes Dev* 16: 2843-2848.

356. Chuang PT, McMahon AP (1999) Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* 397: 617-621.

357. Berman DM, Karhadkar SS, Hallahan AR, Pritchard JI, Eberhart CG, et al. (2002) Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science* 297: 1559-1561.

358. Chuang PT, Kawcak T, McMahon AP (2003) Feedback control of mammalian Hedgehog signaling by the Hedgehog-binding protein, Hip1, modulates Fgf signaling during branching morphogenesis of the lung. *Genes Dev* 17: 342-347.

359. Regl G, Neill GW, Eichberger T, Kasper M, Ikram MS, et al. (2002) Human GLI2 and GLI1 are part of a positive feedback mechanism in Basal Cell Carcinoma. *Oncogene* 21: 5529-5539.

360. Costenbader E, Valente TW (2003) The stability of centrality measures when networks are sampled. *Social Networks* 25: 283-307.

361. Barabasi AL (2005) Sociology. Network theory--the emergence of the creative enterprise. *Science* 308: 639-641.

362. Jiang J (2006) Regulation of Hh/Gli signaling by dual ubiquitin pathways. *Cell Cycle* 5: 2457-2463.

363. Ozgur A, Vu T, Erkan G, Radev DR (2008) Identifying gene-disease associations using centrality on a literature mined gene-interaction network. *Bioinformatics* 24: i277-285.

364. Rodrigues FA, Costa Lda F, Barbieri AL (2011) Resilience of protein-protein interaction networks as determined by their large-scale topological features. *Mol Biosyst* 7: 1263-1269.

365. del Sol A, Fujihashi H, Amoros D, Nussinov R (2006) Residues crucial for maintaining short paths in network communication mediate signaling in proteins. *Mol Syst Biol* 2: 2006 0019.

366. Kawahira H, Ma NH, Tzanakakis ES, McMahon AP, Chuang PT, et al. (2003) Combined activities of hedgehog signaling inhibitors regulate pancreas development. *Development* 130: 4871-4879.

367. Heride C, Rigden DJ, Bertsoulaki E, Cucchi D, De Smaele E, et al. (2016) The centrosomal deubiquitylase USP21 regulates Gli1 transcriptional activity and stability. *J Cell Sci* 129: 4001-4013.

368. Katoh Y, Katoh M (2005) Hedgehog signaling pathway and gastric cancer. *Cancer Biol Ther* 4: 1050-1054.

369. Di Marcotullio L, Ferretti E, Greco A, De Smaele E, Po A, et al. (2006) Numb is a suppressor of Hedgehog signalling and targets Gli1 for Itch-dependent ubiquitination. *Nat Cell Biol* 8: 1415.

370. Xu Q, Yuan X, Liu G, Black KL, Yu JS (2008) Hedgehog signaling regulates brain tumor-initiating cell proliferation and portends shorter survival for patients with PTEN-coexpressing glioblastomas. *Stem Cells* 26: 3018-3026.

371. Ding H, Roncari L, Shannon P, Wu X, Lau N, et al. (2001) Astrocyte-specific expression of activated p21-ras results in malignant astrocytoma formation in a transgenic mouse model of human gliomas. *Cancer Res* 61: 3826-3836.

372. Hsieh A, Ellsworth R, Hsieh D (2011) Hedgehog/GLI1 regulates IGF dependent malignant behaviors in glioma stem cells. *J Cell Physiol* 226: 1118-1127.

373. Lin Y, Cai Z, Huang S, Yang L, Wang C, et al. (2009) Ptc, Smo, Sufu, and the Hedgehog signaling pathway in amphioxus. *Evol Dev* 11: 710-718.

374. Liu M, Dai B, Kang SH, Ban K, Huang FJ, et al. (2006) FoxM1B is overexpressed in human glioblastomas and critically regulates the tumorigenicity of glioma cells. *Cancer Res* 66: 3593-3602.

375. Joensuu H, Puputti M, Sihto H, Tynninen O, Nupponen NN (2005) Amplification of genes encoding KIT, PDGFRalpha and VEGFR2 receptor tyrosine kinases is frequent in glioblastoma multiforme. *J Pathol* 207: 224-231.

376. Saitoh Y, Kuratsu J, Takeshima H, Yamamoto S, Ushio Y (1995) Expression of osteopontin in human glioma. Its correlation with the malignancy. *Lab Invest* 72: 55-63.

377. Buschges R, Weber RG, Actor B, Lichter P, Collins VP, et al. (1999) Amplification and expression of cyclin D genes (CCND1, CCND2 and CCND3) in human malignant gliomas. *Brain*

Pathol 9: 435-442; discussion 432-433.

378. Zhang X, Zhao M, Huang AY, Fei Z, Zhang W, et al. (2005) The effect of cyclin D expression on cell proliferation in human gliomas. *J Clin Neurosci* 12: 166-168.

379. Chakravarti A, Delaney MA, Noll E, Black PM, Loeffler JS, et al. (2001) Prognostic and pathologic significance of quantitative protein expression profiling in human gliomas. *Clin Cancer Res* 7: 2387-2395.

380. Godlewski J, Nowicki MO, Bronisz A, Williams S, Otsuki A, et al. (2008) Targeting of the Bmi-1 oncogene/stem cell renewal factor by microRNA-128 inhibits glioma proliferation and self-renewal. *Cancer Res* 68: 9125-9130.

381. Han SP, Kim JH, Han ME, Sim HE, Kim KS, et al. (2011) SNAI1 is involved in the proliferation and migration of glioblastoma cells. *Cell Mol Neurobiol* 31: 489-496.

382. Reichrath S, Muller CS, Gleissner B, Pfreundschuh M, Vogt T, et al. (2010) Notch- and vitamin D signaling in 1,25(OH)2D3-resistant glioblastoma multiforme (GBM) cell lines. *J Steroid Biochem Mol Biol* 121: 420-424.

383. Roth W, Wild-Bode C, Platten M, Grimmel C, Melkonyan HS, et al. (2000) Secreted Frizzled-related proteins inhibit motility and promote growth of human malignant glioma cells. *Oncogene* 19: 4210-4220.

384. Lee MJ, Hatton BA, Villavicencio EH, Khanna PC, Friedman SD, et al. (2012) Hedgehog pathway inhibitor saridegib (IPI-926) increases lifespan in a mouse medulloblastoma model. *Proc Natl Acad Sci USA* 109: 7859-7864.

385. Biswas NK, Chandra V, Sarkar-Roy N, Das T, Bhattacharya RN, et al. (2015) Variant allele frequency enrichment analysis in vitro reveals sonic hedgehog pathway to impede sustained temozolomide response in GBM. *Sci Rep* 5: 7915.

386. Qualtrough D, Buda A, Gaffield W, Williams AC, Paraskeva C (2004) Hedgehog signalling in colorectal tumour cells: induction of apoptosis with cyclopamine treatment. *Int J Cancer* 110: 831-837.

387. Douard R, Moutereau S, Pernet P, Chimingqi M, Allory Y, et al. (2006) Sonic Hedgehog-dependent proliferation in a series of patients with colorectal cancer. *Surgery* 139: 665-670.

388. Varnat F, Duquet A, Malerba M, Zbinden M, Mas C, et al. (2009) Human colon cancer epithelial cells harbour active HEDGEHOG-GLI signalling that is essential for tumour growth, recurrence, metastasis and stem cell survival and expansion. *EMBO Mol Med* 1: 338-351.

389. Fan L, Pepicelli CV, Dibble CC, Catbagan W, Zarycki JL, et al. (2004) Hedgehog signaling promotes prostate xenograft tumor growth. *Endocrinology* 145: 3961-3970.

390. Shimoyama A, Wada M, Ikeda F, Hata K, Matsubara T, et al. (2007) Ihh/Gli2 signaling promotes osteoblast differentiation by regulating Runx2 expression and function. *Mol Biol Cell* 18: 2411-2418.

391. Tanaka M, Omura K, Watanabe Y, Oda Y, Nakanishi I (1994) Prognostic factors of colorectal cancer: K-ras mutation, overexpression of the p53 protein, and cell proliferative activity. *J Surg Oncol* 57: 57-64.

392. Kayed H, Kleeff J, Keleg S, Guo J, Ketterer K, et al. (2004) Indian hedgehog signaling pathway: expression and regulation in pancreatic cancer. *Int J Cancer* 110: 668-676.

393. Gysin S, Lee SH, Dean NM, McMahon M (2005) Pharmacologic inhibition of RAF-->MEK-->ERK signaling elicits pancreatic cancer cell cycle arrest through induced expression of p27Kip1. *Cancer Res* 65: 4870-4880.

394. Santoni M, Burattini L, Nabissi M, Morelli MB, Berardi R, et al. (2013) Essential role of Gli proteins in glioblastoma multiforme. *Curr Protein Pept Sci* 14: 133-140.

395. Gratas C, Tohma Y, Van Meir EG, Klein M, Tenan M, et al. (1997) Fas ligand expression in glioblastoma cell lines and primary astrocytic brain tumors. *Brain Pathol* 7: 863-869.

396. Bian YH, Huang SH, Yang L, Ma XL, Xie JW, et al. (2007) Sonic hedgehog-Gli1 pathway in colorectal adenocarcinomas. *World J Gastroenterol* 13: 1659-1665.

397. Pei H, Li L, Fridley BL, Jenkins GD, Kalari KR, et al. (2009) FKBP51 affects cancer cell response to chemotherapy by negatively regulating Akt. *Cancer Cell* 16: 259-266.

398. Kelleher FC, McDermott R (2012) Aberrations and therapeutics involving the developmental pathway Hedgehog in pancreatic cancer. *Vitam Horm* 88: 355-378.

399. Cui D, Chen X, Yin J, Wang W, Lou M, et al. (2012) Aberrant activation of Hedgehog/Gli1 pathway on angiogenesis in gliomas. *Neurol India* 60: 589-596.

400. Ackermann E, Weiel EM, Pfaff T, Drossel B (2012) Boolean versus continuous dynamics in modules with two feedback loops. *Eur Phys J E Soft Matter* 35: 107.

401. Sherriff M, Sarkar R. Computational approaches and modeling of signaling processes in the Immune System; 2008. pp. 187-200.

402. Albert R, Othmer HG (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. *J Theor Biol* 223: 1-18.

403. Subramanian K, Gadgil C (2010) Robustness of the Drosophila segment polarity network to transient perturbations. *IET Syst Biol* 4: 169-176.

404. Du WZ, Feng Y, Wang XF, Piao XY, Cui YQ, et al. (2013) Curcumin suppresses malignant glioma cells growth and induces apoptosis by inhibition of SHH/GLI1 signaling pathway in vitro and vivo. *CNS Neurosci Ther* 19: 926-936.

405. Katoh M (2007) Networking of WNT, FGF, Notch, BMP, and Hedgehog Signaling Pathways during Carcinogenesis. *Stem Cell Reviews* 3: 30-38.

406. Ruiz i Altaba A, Sanchez P, Dahmane N (2002) Gli and hedgehog in cancer: tumours, embryos and stem cells. *Nat Rev Cancer* 2: 361-372.

407. Trent J, Meltzer P, Rosenblum M, Harsh G, Kinzler K, et al. (1986) Evidence for rearrangement, amplification, and expression of c-myc in a human glioblastoma. *Proc Natl Acad Sci USA* 83: 470-473.

408. Arnhold V, Boos J, Lanvers-Kaminsky C (2016) Targeting hedgehog signaling pathway in pediatric tumors: in vitro evaluation of SMO and GLI inhibitors. *Cancer Chemother Pharmacol* 77: 495-505.

409. Atwood SX, Li M, Lee A, Tang JY, Oro AE (2013) GLI activation by atypical protein kinase C iota/lambda regulates the growth of basal cell carcinomas. *Nature* 494: 484-488.

410. Beauchamp E, Bulut G, Abaan O, Chen K, Merchant A, et al. (2009) GLI1 is a direct transcriptional target of EWS-FLI1 oncoprotein. *J Biol Chem* 284: 9074-9082.

411. Hui CC, Angers S (2011) Gli proteins in development and disease. *Annu Rev Cell Dev Biol* 27: 513-537.

412. Feldmann G, Dhara S, Fendrich V, Bedja D, Beaty R, et al. (2007) Blockade of hedgehog signaling inhibits pancreatic cancer invasion and metastases: a new paradigm for combination therapy in solid cancers. *Cancer Res* 67: 2187-2196.

413. Clement V, Sanchez P, de Tribolet N, Radovanovic I, Ruiz i Altaba A (2007) HEDGEHOG-GLI1 signaling regulates human glioma growth, cancer stem cell self-renewal, and tumorigenicity. *Curr Biol* 17: 165-172.

414. Tsuda N, Ishiyama S, Li Y, Ioannides CG, Abbruzzese JL, et al. (2006) Synthetic microRNA designed to target glioma-associated antigen 1 transcription factor inhibits division and induces late apoptosis in pancreatic tumor cells. *Clin Cancer Res* 12: 6557-6564.

415. Agren M, Kogerman P, Kleman MI, Wessling M, Toftgard R (2004) Expression of the PTCH1 tumor suppressor gene is regulated by alternative promoters and a single functional Gli-binding site. *Gene* 330: 101-114.

416. Tenzen T, Allen BL, Cole F, Kang JS, Krauss RS, et al. (2006) The cell surface membrane proteins Cdo and Boc are components and targets of the Hedgehog signaling pathway and feedback network in mice. *Dev Cell* 10: 647-656.

417. Tai CJ, Chang CC, Jiang MC, Yeh CM, Su TC, et al. (2012) Clinical-pathological correlation of K-Ras mutation and ERK phosphorylation in colorectal cancer. *Pol J Pathol* 63: 93-100.

418. Roberts PJ, Der CJ (2007) Targeting the Raf-MEK-ERK mitogen-activated protein kinase cascade for the treatment of cancer. *Oncogene* 26: 3291-3310.

419. Kloog Y, Cox AD (2000) RAS inhibitors: potential for cancer therapeutics. *Mol Med Today* 6: 398-402.

420. Bos JL (1989) ras oncogenes in human cancer: a review. *Cancer Res* 49: 4682-4689.

421. Oh P, Lobry C, Gao J, Tikhonova A, Loizou E, et al. (2013) In vivo mapping of notch pathway activity in normal and stress hematopoiesis. *Cell Stem Cell* 13: 190-204.

422. Gridley T (2007) Notch signaling in vascular development and physiology. *Development* 134: 2709-2718.

423. Roca C, Adams RH (2007) Regulation of vascular morphogenesis by Notch signaling. *Genes Dev* 21: 2511-2524.

424. Rodriguez-Rivera NS, Molina-Hernandez A, Sanchez-Cruz E, Escalante-Alcalde D, Velasco I (2009) Activated Notch1 is a stronger astrocytic stimulus than leukemia inhibitory factor for rat neural stem cells. *Int J Dev Biol* 53: 947-953.

425. Nakayama K, Nagase H, Koh CS, Ohkawara T (2011) gamma-Secretase-regulated mechanisms similar to notch signaling may play a role in signaling events, including APP signaling, which leads to Alzheimer's disease. *Cell Mol Neurobiol* 31: 887-900.

426. Gomez-Lamarca MJ, Falo-Sanjuan J, Stojnic R, Abdul Rehman S, Muresan L, et al. (2018) Activation of the Notch Signaling Pathway In Vivo Elicits Changes in CSL Nuclear Dynamics. *Dev Cell* 44: 611-623 e617.

427. Scholz BA, Harth-Hertle ML, Malterer G, Haas J, Ellwart J, et al. (2013) Abortive lytic reactivation of KSHV in CBF1/CSL deficient human B cell lines. *PLoS Pathog* 9: e1003336.

428. de la Pompa JL, Wakeham A, Correia KM, Samper E, Brown S, et al. (1997) Conservation of the Notch signalling pathway in mammalian neurogenesis. *Development* 124: 1139-1148.

429. Yeh TS, Lin YM, Hsieh RH, Tseng MJ (2003) Association of transcription factor YY1 with the high molecular weight Notch complex suppresses the transactivation activity of Notch. *J Biol Chem* 278: 41963-41969.

430. Suchting S, Freitas C, le Noble F, Benedito R, Breant C, et al. (2007) Negative regulators of vessel patterning. *Novartis Found Symp* 283: 77-80; discussion 80-76, 238-241.

431. Brack AS, Conboy IM, Conboy MJ, Shen J, Rando TA (2008) A temporal switch from notch to Wnt signaling in muscle stem cells is necessary for normal adult myogenesis. *Cell Stem Cell* 2: 50-59.

432. Ferretti E, Tosi E, Po A, Scipioni A, Morisi R, et al. (2008) Notch signaling is involved in expression of thyrocyte differentiation markers and is down-regulated in thyroid tumors. *J Clin Endocrinol Metab* 93: 4080-4087.

433. High FA, Lu MM, Pear WS, Loomes KM, Kaestner KH, et al. (2008) Endothelial expression of the Notch ligand Jagged1 is required for vascular smooth muscle development. *Proc Natl Acad Sci USA* 105: 1955-1959.

434. Marei HE, Ahmed AE (2013) Transcription factors expressed in embryonic and adult olfactory bulb neural stem cells reveal distinct proliferation, differentiation and epigenetic control. *Genomics* 101: 12-19.

435. Li JL, Harris AL (2005) Notch signaling from tumor cells: a new mechanism of angiogenesis. *Cancer Cell* 8: 1-3.

436. Koch U, Radtke F (2007) Notch and cancer: a double-edged sword. *Cell Mol Life Sci* 64: 2746-2762.

437. Shih AH, Holland EC (2006) Notch signaling enhances nestin expression in gliomas. *Neoplasia* 8: 1072-1082.

438. Stylianou S, Clarke RB, Brennan K (2006) Aberrant Activation of Notch Signaling in Human Breast Cancer. *Cancer Res* 66: 1517-1525.

439. Hughes DP (2009) How the NOTCH pathway contributes to the ability of osteosarcoma cells to metastasize. *Cancer Treat Res* 152: 479-496.

440. Villaronga MA, Bevan CL, Belandia B (2008) Notch signaling: a potential therapeutic target in prostate cancer. *Curr Cancer Drug Targets* 8: 566-580.

441. Marsh JC, DeConti RC, Hubbard SP (1971) Treatment of Hodgkin's disease and other cancers with 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU; NSC-409962). *Cancer Chemother Rep* 55: 599-606.

442. Li X, von Boehmer H (2011) Notch Signaling in T-Cell Development and T-ALL. *ISRN Hematol* 2011: 921706.

443. Ge W, Martinowich K, Wu X, He F, Miyamoto A, et al. (2002) Notch signaling promotes astrogliogenesis via direct CSL-mediated glial gene activation. *J Neurosci Res* 69: 848-860.

444. Cui XY, Hu QD, Tekaya M, Shimoda Y, Ang BT, et al. (2004) NB-3/Notch1 pathway via Deltex1 promotes neural progenitor cell differentiation into oligodendrocytes. *J Biol Chem* 279: 25858-25865.

445. Kamakura S, Oishi K, Yoshimatsu T, Nakafuku M, Masuyama N, et al. (2004) Hes binding to STAT3 mediates crosstalk between Notch and JAK-STAT signalling. *Nat Cell Biol* 6: 547-554.

446. Mapstone TB, Galloway PG (1991) Expression of glial fibrillary acidic protein, vimentin, fibronectin, and N-myc oncoprotein in primary human brain tumor cell explants. *Pediatr Neurosurg* 17: 169-174.

447. Patt S, Thiel G, Maas S, Lozanova T, Prosenc N, et al. (1993) Chromosomal changes and correspondingly altered proto-oncogene expression in human gliomas. Value of combined cytogenetic and molecular genetic analysis. *Anticancer Res* 13: 113-118.

448. Pierfelice TJ, Schreck KC, Eberhart CG, Gaiano N (2008) Notch, neural stem cells, and brain tumors. *Cold Spring Harb Symp Quant Biol* 73: 367-375.

449. Xu P, Yu S, Jiang R, Kang C, Wang G, et al. (2009) Differential expression of Notch family members in astrocytomas and medulloblastomas. *Pathol Oncol Res* 15: 703-710.

450. Hovinga KE, Shimizu F, Wang R, Panagiotakos G, Van Der Heijden M, et al. (2010) Inhibition of notch signaling in glioblastoma targets cancer stem cells via an endothelial cell intermediate. *Stem Cells* 28: 1019-1029.

451. Lin J, Zhang XM, Yang JC, Ye YB, Luo SQ (2010) gamma-secretase inhibitor-I enhances radiosensitivity of glioblastoma cell lines by depleting CD133+ tumor cells. *Arch Med Res* 41: 519-529.

452. Fan X (2016) gamma-Secretase inhibitor-resistant glioblastoma stem cells require RBPJ to propagate. *J Clin Invest* 126: 2415-2418.

453. Doody RS, Raman R, Farlow M, Iwatsubo T, Vellas B, et al. (2013) A phase 3 trial of semagacestat for treatment of Alzheimer's disease. *N Engl J Med* 369: 341-350.

454. Yang T, Arslanova D, Xu X, Li YM, Xia W (2010) In vivo manifestation of Notch related phenotypes in zebrafish treated with Alzheimer's amyloid reducing gamma-secretase inhibitors. *J Neurochem* 113: 1200-1209.

455. Hansson CA, Frykman S, Farmery MR, Tjernberg LO, Nilsberth C, et al. (2004) Nicastrin, presenilin, APH-1, and PEN-2 form active gamma-secretase complexes in mitochondria. *J Biol Chem* 279: 51654-51660.

456. Lleo A (2008) Activity of gamma-secretase on substrates other than APP. *Curr Top Med Chem*

8: 9-16.

457. Lee SF, Shah S, Li H, Yu C, Han W, et al. (2002) Mammalian APH-1 interacts with presenilin and nicastrin and is required for intramembrane proteolysis of amyloid-beta precursor protein and Notch. *J Biol Chem* 277: 45013-45019.

458. Rochette MJ, Murphy MP (2002) Gamma-secretase: substrates and inhibitors. *Mol Neurobiol* 26: 81-95.

459. Wong GT, Manfra D, Poulet FM, Zhang Q, Josien H, et al. (2004) Chronic treatment with the gamma-secretase inhibitor LY-411,575 inhibits beta-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation. *J Biol Chem* 279: 12876-12882.

460. Ables JL, Breunig JJ, Eisch AJ, Rakic P (2011) Not(ch) just development: Notch signalling in the adult brain. *Nat Rev Neurosci* 12: 269-283.

461. Miyamoto A, Lau R, Hein PW, Shipley JM, Weinmaster G (2006) Microfibrillar proteins MAGP-1 and MAGP-2 induce Notch1 extracellular domain dissociation and receptor activation. *J Biol Chem* 281: 10089-10097.

462. Sakamoto K, Yamaguchi S, Ando R, Miyawaki A, Kabasawa Y, et al. (2002) The nephroblastoma overexpressed gene (NOV/ccn3) protein associates with Notch1 extracellular domain and inhibits myoblast differentiation via Notch signaling pathway. *J Biol Chem* 277: 29399-29405.

463. Hu QD, Ang BT, Karsak M, Hu WP, Cui XY, et al. (2003) F3/contactin acts as a functional ligand for Notch during oligodendrocyte maturation. *Cell* 115: 163-175.

464. Hu QD, Ma QH, Gennarini G, Xiao ZC (2006) Cross-talk between F3/contactin and Notch at axoglial interface: a role in oligodendrocyte development. *Dev Neurosci* 28: 25-33.

465. Brou C, Logeat F, Gupta N, Bessia C, LeBail O, et al. (2000) A novel proteolytic cleavage involved in Notch signaling: the role of the disintegrin-metalloprotease TACE. *Mol Cell* 5: 207-216.

466. Cotter D, Honavar M, Lovestone S, Raymond L, Kerwin R, et al. (1999) Disturbance of Notch-1 and Wnt signalling proteins in neuroglial balloon cells and abnormal large neurons in focal cortical dysplasia in human cortex. *Acta Neuropathol* 98: 465-472.

467. Kim MY, Ann EJ, Mo JS, Dajas-Bailador F, Seo MS, et al. (2010) JIP1 binding to RBP-Jk mediates cross-talk between the Notch1 and JIP1-JNK signaling pathway. *Cell Death Differ* 17: 1728-1738.

468. Han X, Ju JH, Shin I (2012) Glycogen synthase kinase 3-beta phosphorylates novel S/T-P-S/T domains in Notch1 intracellular domain and induces its nuclear localization. *Biochem Biophys Res Commun* 423: 282-288.

469. Collu GM, Hidalgo-Sastre A, Acar A, Bayston L, Gildea C, et al. (2012) Dishevelled limits Notch signalling through inhibition of CSL. *Development* 139: 4405-4415.

470. Lee TV, Sethi MK, Leonardi J, Rana NA, Buettner FF, et al. (2013) Negative regulation of notch signaling by xylose. *PLoS Genet* 9: e1003547.

471. Bruckner K, Perez L, Clausen H, Cohen S (2000) Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* 406: 411-415.

472. Yamakawa T, Ayukawa T, Matsuno K (2012) Metabolism and transportation pathways of GDP-fucose that are required for the O-fucosylation of Notch. *Adv Exp Med Biol* 727: 37-46.

473. Haltiwanger RS, Stanley P (2002) Modulation of receptor signaling by glycosylation: fringe is an O-fucose-beta1,3-N-acetylglucosaminyltransferase. *Biochim Biophys Acta* 1573: 328-335.

474. Zhou S, Hayward SD (2001) Nuclear localization of CBF1 is regulated by interactions with the SMRT corepressor complex. *Mol Cell Biol* 21: 6222-6232.

475. Hsieh JJ, Zhou S, Chen L, Young DB, Hayward SD (1999) CIR, a corepressor linking the DNA binding factor CBF1 to the histone deacetylase complex. *Proc Natl Acad Sci USA* 96: 23-28.

476. Zhou S, Fujimuro M, Hsieh JJ, Chen L, Miyamoto A, et al. (2000) SKIP, a CBF1-associated

protein, interacts with the ankyrin repeat domain of NotchIC To facilitate NotchIC function. *Mol Cell Biol* 20: 2400-2410.

477. Fang TC, Yashiro-Ohtani Y, Del Bianco C, Knoblock DM, Blacklow SC, et al. (2007) Notch directly regulates Gata3 expression during T helper 2 cell differentiation. *Immunity* 27: 100-110.

478. Joshi I, Minter LM, Telfer J, Demarest RM, Capobianco AJ, et al. (2009) Notch signaling mediates G1/S cell-cycle progression in T cells via cyclin D3 and its dependent kinases. *Blood* 113: 1689-1698.

479. Ronchini C, Capobianco AJ (2001) Induction of cyclin D1 transcription and CDK2 activity by Notch(ic): implication for cell cycle disruption in transformation by Notch(ic). *Mol Cell Biol* 21: 5925-5934.

480. Rodilla V, Villanueva A, Obrador-Hevia A, Robert-Moreno A, Fernandez-Majada V, et al. (2009) Jagged1 is the pathological link between Wnt and Notch pathways in colorectal cancer. *Proc Natl Acad Sci USA* 106: 6315-6320.

481. Liu H, Chi AW, Arnett KL, Chiang MY, Xu L, et al. (2010) Notch dimerization is required for leukemogenesis and T-cell development. *Genes Dev* 24: 2395-2407.

482. Kondoh K, Sunadome K, Nishida E (2007) Notch signaling suppresses p38 MAPK activity via induction of MKP-1 in myogenesis. *J Biol Chem* 282: 3058-3065.

483. Blokzijl A, Dahlqvist C, Reissmann E, Falk A, Moliner A, et al. (2003) Cross-talk between the Notch and TGF-beta signaling pathways mediated by interaction of the Notch intracellular domain with Smad3. *J Cell Biol* 163: 723-728.

484. Sade H, Krishna S, Sarin A (2004) The anti-apoptotic effect of Notch-1 requires p56lck-dependent, Akt/PKB-mediated signaling in T cells. *J Biol Chem* 279: 2937-2944.

485. Terry AJ, Sturrock M, Dale JK, Maroto M, Chaplain MA (2011) A spatio-temporal model of Notch signalling in the zebrafish segmentation clock: conditions for synchronised oscillatory dynamics. *PLoS One* 6: e16980.

486. Mittal S, Subramanyam D, Dey D, Kumar RV, Rangarajan A (2009) Cooperation of Notch and Ras/MAPK signaling pathways in human breast carcinogenesis. *Mol Cancer* 8: 128.

487. Natarajan S, Li Y, Miller EE, Shih DJ, Taylor MD, et al. (2013) Notch1-induced brain tumor models the sonic hedgehog subgroup of human medulloblastoma. *Cancer Res* 73: 5381-5390.

488. Chen L, Zhang R, Li P, Liu Y, Qin K, et al. (2013) P53-induced microRNA-107 inhibits proliferation of glioma cells and down-regulates the expression of CDK6 and Notch-2. *Neurosci Lett* 534: 327-332.

489. Giovannini C, Gramantieri L, Chieco P, Minguzzi M, Lago F, et al. (2009) Selective ablation of Notch3 in HCC enhances doxorubicin's death promoting effect by a p53 dependent mechanism. *J Hepatol* 50: 969-979.

490. Sun Y, Klauzinska M, Lake RJ, Lee JM, Santopietro S, et al. (2011) Trp53 regulates Notch 4 signaling through Mdm2. *J Cell Sci* 124: 1067-1076.

491. Pawlus MR, Wang L, Hu CJ (2014) STAT3 and HIF1alpha cooperatively activate HIF1 target genes in MDA-MB-231 and RCC4 cells. *Oncogene* 33: 1670-1679.

492. Palomero T, Dominguez M, Ferrando AA (2008) The role of the PTEN/AKT Pathway in NOTCH1-induced leukemia. *Cell Cycle* 7: 965-970.

493. Liu X, Zhang Y, Shi M, Wang Y, Zhang F, et al. (2018) Notch1 regulates PTEN expression to exacerbate renal tubulointerstitial fibrosis in diabetic nephropathy by inhibiting autophagy via interactions with Hes1. *Biochem Biophys Res Commun* 497: 1110-1116.

494. Bader GD, Hogue CW (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4: 2.

495. Zheng X, Linke S, Dias JM, Zheng X, Gradin K, et al. (2008) Interaction with factor inhibiting

HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proc Natl Acad Sci USA* 105: 3368-3373.

496. Kim SB, Chae GW, Lee J, Park J, Tak H, et al. (2007) Activated Notch1 interacts with p53 to inhibit its phosphorylation and transactivation. *Cell Death Differ* 14: 982-991.

497. Pirot P, van Grunsven LA, Marine JC, Huylebroeck D, Bellefroid EJ (2004) Direct regulation of the Nrarp gene promoter by the Notch signaling pathway. *Biochem Biophys Res Commun* 322: 526-534.

498. Wang G, Wang JJ, Chen XL, Du SM, Li DS, et al. (2013) The JAK2/STAT3 and mitochondrial pathways are essential for quercetin nanoliposome-induced C6 glioma cell death. *Cell Death Dis* 4: e746.

499. Senft C, Priester M, Polacin M, Schroder K, Seifert V, et al. (2011) Inhibition of the JAK-2/STAT3 signaling pathway impedes the migratory and invasive potential of human glioblastoma cells. *J Neurooncol* 101: 393-403.

500. Saito N, Fu J, Zheng S, Yao J, Wang S, et al. (2014) A high Notch pathway activation predicts response to gamma secretase inhibitors in proneural subtype of glioma tumor-initiating cells. *Stem Cells* 32: 301-312.

501. Kanamori M, Kawaguchi T, Nigro JM, Feuerstein BG, Berger MS, et al. (2007) Contribution of Notch signaling activation to human glioblastoma multiforme. *J Neurosurgery* 106: 417-427.

502. Irshad K, Mohapatra SK, Srivastava C, Garg H, Mishra S, et al. (2015) A combined gene signature of hypoxia and notch pathway in human glioblastoma and its prognostic relevance. *PLoS One* 10: e0118201.

503. Sarkar RR, Chowdhury S (2016) Identification of minimal combinations of oncoproteins in notch pathway to suppress human glioblastoma. US Patent Application Number US14527628.

504. Phillips E, Lang V, Bohlen J, Bethke F, Puccio L, et al. (2016) Targeting atypical protein kinase C iota reduces viability in glioblastoma stem-like cells via a notch signaling mechanism. *Int J Cancer* 139: 1776-1787.

505. Bonavia R, Inda MM, Cavenee WK, Furnari FB (2011) Heterogeneity maintenance in glioblastoma: a social network. *Cancer Res* 71: 4055-4060.

506. Kawamura Y, Takouda J, Yoshimoto K, Nakashima K (2018) New aspects of glioblastoma multiforme revealed by similarities between neural and glioblastoma stem cells. *Cell Biol Toxicol*.

507. Stockhausen MT, Kristoffersen K, Poulsen HS (2010) The functional role of Notch signaling in human gliomas. *Neuro Oncol* 12: 199-211.

508. Zhang X, Chen T, Zhang J, Mao Q, Li S, et al. (2012) Notch1 promotes glioma cell migration and invasion by stimulating β-catenin and NF-κB signaling via AKT activation. *Cancer science* 103: 181-190.

509. Wang L, Rahn JJ, Lun X, Sun B, Kelly JJ, et al. (2008) Gamma-secretase represents a therapeutic target for the treatment of invasive glioma mediated by the p75 neurotrophin receptor. *PLoS Biol* 6: e289.

510. Lee J, Son MJ, Woolard K, Donin NM, Li A, et al. (2008) Epigenetic-mediated dysfunction of the bone morphogenetic protein pathway inhibits differentiation of glioblastoma-initiating cells. *Cancer Cell* 13: 69-80.

511. Sunayama J, Matsuda KI, Sato A, Tachibana K, Suzuki K, et al. (2010) Crosstalk between the PI3K/mTOR and MEK/ERK pathways involved in the maintenance of self-renewal and tumorigenicity of glioblastoma stem-like cells. *Stem Cells* 28: 1930-1939.

512. Hede SM, Nazarenko I, Nister M, Lindstrom MS (2011) Novel Perspectives on p53 Function in Neural Stem Cells and Brain Tumors. *J Oncol* 2011: 852970.

513. Prasetyanti PR, Medema JP (2017) Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol Cancer* 16: 41.

514. Prestegarden L, Svendsen A, Wang J, Sleire L, Skaftnesmo KO, et al. (2010) Glioma cell populations grouped by different cell type markers drive brain tumor growth. *Cancer Res* 70: 4274-4279.

515. Brennan C, Momota H, Hambardzumyan D, Ozawa T, Tandon A, et al. (2009) Glioblastoma subclasses can be defined by activity among signal transduction pathways and associated genomic alterations. *PLoS One* 4: e7752.

516. Esdar C, Milasta S, Maelicke A, Herget T (2001) Differentiation-associated apoptosis of neural stem cells is effected by Bcl-2 overexpression: impact on cell lineage determination. *Euro J Cell Biol* 80: 539-553.

517. Pfeuty B (2015) A computational model for the coordination of neural progenitor self-renewal and differentiation through Hes1 dynamics. *Development* 142: 477-485.

518. Kageyama R, Ohtsuka T, Shimojo H, Imayoshi I (2008) Dynamic Notch signaling in neural progenitor cells and a revised view of lateral inhibition. *Nat Neurosci* 11: 1247-1251.

519. Barton A, Fendrik AJ (2013) Sustained vs. oscillating expressions of Ngn2, Dll1 and Hes1: a model of neural differentiation of embryonic telencephalon. *J Theor Biol* 328: 1-8.

520. Claus EB, Walsh KM, Wiencke JK, Molinaro AM, Wiemels JL, et al. (2015) Survival and low-grade glioma: the emergence of genetic information. *Neurosurg Focus* 38: E6.

521. Armesilla-Diaz A, Bragado P, Del Valle I, Cuevas E, Lazaro I, et al. (2009) p53 regulates the self-renewal and differentiation of neural precursors. *Neuroscience* 158: 1378-1389.

522. Fu S, Jiang W, Gao X, Zeng A, Cholger D, et al. (2016) Aberrant Adult Neurogenesis in the Subventricular Zone-Rostral Migratory Stream-Olfactory Bulb System Following Subchronic Manganese Exposure. *Toxicol Sci* 150: 347-368.

523. O'Brien ER, Howarth C, Sibson NR (2013) The role of astrocytes in CNS tumors: pre-clinical models and novel imaging approaches. *Front Cell Neurosci* 7: 40.

524. Willan AR, O'Brien BJ (1996) Confidence intervals for cost-effectiveness ratios: an application of Fieller's theorem. *Health Econ* 5: 297-305.

525. Birner P, Toumangelova-Uzeir K, Natchev S, Guentchev M (2010) STAT3 tyrosine phosphorylation influences survival in glioblastoma. *J Neurooncol* 100: 339-343.

526. Kim JE, Patel M, Ruzevick J, Jackson CM, Lim M (2014) STAT3 Activation in Glioblastoma: Biochemical and Therapeutic Implications. *Cancers (Basel)* 6: 376-395.

527. Chiba T, Mack L, Delis N, Brill B, Groner B (2012) Stat3 inhibition in neural lineage cells. *Horm Mol Biol Clin Investig* 10: 255-263.

528. Park NI, Guilhamon P, Desai K, McAdam RF, Langille E, et al. (2017) ASCL1 Reorganizes Chromatin to Direct Neuronal Fate and Suppress Tumorigenicity of Glioblastoma Stem Cells. *Cell Stem Cell* 21: 411.

529. Ball DW (2004) Achaete-scute homolog-1 and Notch in lung neuroendocrine development and cancer. *Cancer Lett* 204: 159-169.

530. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, et al. (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17: 98-110.

531. Patel VN, Gokulrangan G, Chowdhury SA, Chen Y, Sloan AE, et al. (2013) Network signatures of survival in glioblastoma multiforme. *PLoS Comput Biol* 9: e1003237.

532. Bansod S, Kageyama R, Ohtsuka T (2017) Hes5 regulates the transition timing of neurogenesis and gliogenesis in mammalian neocortical development. *Development* 144: 3156-3167.

533. Basak O, Giachino C, Fiorini E, Macdonald HR, Taylor V (2012) Neurogenic subventricular zone stem/progenitor cells are Notch1-dependent in their active but not quiescent state. *J Neurosci* 32: 5654-5666.

534. de Almeida Sassi F, Lunardi Brunetto A, Schwartsmann G, Roesler R, Abujamra AL (2012) Glioma revisited: from neurogenesis and cancer stem cells to the epigenetic regulation of the niche. *J Oncol* 2012: 537861.

535. de Oliveira ME, Neto LM (2016) Directional entropy based model for diffusivity-driven tumor growth. *Math Biosci Eng* 13: 333-341.

536. Wei Z, Fu W, Liu Q, Jing H, Jin C, et al. (2018) Construction of Boolean logic gates based on dual-vector circuits of multiple gene regulatory elements. *Mol Genet Genomics*.

537. Yachie-Kinoshita A, Onishi K, Ostblom J, Langley MA, Posfai E, et al. (2018) Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. *Mol Syst Biol* 14: e7952.

538. Dobrin A, Saxena P, Fussenegger M (2016) Synthetic biology: applying biological circuits beyond novel therapies. *Integr Biol (Camb)* 8: 409-430.

539. Khan FM, Schmitz U, Nikolov S, Engelmann D, Putzer BM, et al. (2014) Hybrid modeling of the crosstalk between signaling and transcriptional networks using ordinary differential equations and multi-valued logic. *Biochim Biophys Acta* 1844: 289-298.

540. Ganguli P, Chowdhury S, Bhowmick R, Sarkar R (2015) Temporal Protein Expression Pattern in Intracellular Signaling Cascade during T-cell Activation: A Computational Study. *Journal of Biosciences* 40: 769-789.

541. Furukawa K, Hohmann S (2015) A fungicide-responsive kinase as a tool for synthetic cell fate regulation. *Nucleic Acids Res* 43: 7162-7170.

542. Ganguli P, Chowdhury S, Chowdhury S, Sarkar RR (2015) Identification of Th1/Th2 regulatory switch to promote healing response during leishmaniasis: a computational approach. *EURASIP J Bioinform Syst Biol* 2015: 13.

543. Zhang F, Liu R, Zheng J (2016) Sig2GRN: a software tool linking signaling pathway with gene regulatory network for dynamic simulation. *BMC Syst Biol* 10: 123.

544. Whelan KE, King RD (2008) Using a logical model to predict the growth of yeast. *BMC Bioinformatics* 9: 97.

545. Traynard P, Faure A, Fages F, Thieffry D (2016) Logical model specification aided by model-checking techniques: application to the mammalian cell cycle regulation. *Bioinformatics* 32: i772-i780.

546. Baldazzi V, Bertin N, de Jong H, Genard M (2012) Towards multiscale plant models: integrating cellular networks. *Trends Plant Sci* 17: 728-736.

547. Faure A, Naldi A, Chaouiya C, Thieffry D (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 22: e124-131.

548. Levy N, Naldi A, Hernandez C, Stoll G, Thieffry D, et al. (2018) Prediction of Mutations to Control Pathways Enabling Tumor Cell Invasion with the CoLoMoTo Interactive Notebook (Tutorial). *Front Physiol* 9: 787.

549. Gupta S, Silveira DA, Mombach JCM (2018) Modeling the role of microRNA-449a in the regulation of the G2/M cell cycle checkpoint in prostate LNCaP cells under ionizing radiation. *PLoS One* 13: e0200768.

550. Kaufman M, Andris F, Leo O (1999) A logical analysis of T cell activation and anergy. *Proc Natl Acad Sci USA* 96: 3894-3899.

551. Dnyane PA, Puntambekar SS, Gadgil CJ (2018) Method for identification of sensitive nodes in Boolean models of biological networks. *IET Syst Biol* 12: 1-6.

552. Bloomingdale P, Nguyen VA, Niu J, Mager DE (2018) Boolean network modeling in systems pharmacology. *J Pharmacokinet Pharmacodyn* 45: 159-180.

553. Schwab JD, Siegle L, Kuhlwein SD, Kuhl M, Kestler HA (2017) Stability of Signaling Pathways during Aging-A Boolean Network Approach. *Biology (Basel)* 6.

554. Franke R, Theis FJ, Klamt S (2010) From binary to multivalued to continuous models: the lac operon as a case study. *J Integr Bioinform* 7.

555. Sur I, Taipale J (2016) The role of enhancers in cancer. *Nat Rev Cancer* 16: 483-493.

556. An Z, Aksoy O, Zheng T, Fan QW, Weiss WA (2018) Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies. *Oncogene* 37: 1561-1575.

557. Ghasimi S, Wibom C, Dahlin AM, Brannstrom T, Golovleva I, et al. (2016) Genetic risk variants in the CDKN2A/B, RTEL1 and EGFR genes are associated with somatic biomarkers in glioma. *J Neurooncol* 127: 483-492.

558. Makarov EM, Shtam TA, Kovalev RA, Pantina RA, Varfolomeeva EY, et al. (2017) The rare nonsense mutation in p53 triggers alternative splicing to produce a protein capable of inducing apoptosis. *PLoS One* 12: e0185126.

559. Koso H, Yi H, Sheridan P, Miyano S, Ino Y, et al. (2016) Identification of RNA-Binding Protein LARP4B as a Tumor Suppressor in Glioma. *Cancer Res* 76: 2254-2264.

560. Chang JJ, Woods M, Lindsay RJ, Doyle EH, Griesbeck M, et al. (2013) Higher expression of several interferon-stimulated genes in HIV-1-infected females after adjusting for the level of viral replication. *J Infect Dis* 208: 830-838.

561. Bahrami S, Drablos F (2016) Gene regulation in the immediate-early response process. *Adv Biol Regul* 62: 37-49.

562. Muire PJ, Hanson LA, Wills R, Petrie-Hanson L (2017) Differential gene expression following TLR stimulation in rag1-/- mutant zebrafish tissues and morphological descriptions of lymphocyte-like cell populations. *PLoS One* 12: e0184077.

563. Liu D, Chang X, Liu Z, Chen L, Wang R (2011) Bistability and oscillations in gene regulation mediated by small noncoding RNAs. *PLoS One* 6: e17029.

564. Mitrophanov AY, Groisman EA (2008) Positive feedback in cellular control systems. *Bioessays* 30: 542-555.

565. Gordley RM, Williams RE, Bashor CJ, Toettcher JE, Yan S, et al. (2016) Engineering dynamical control of cell fate switching using synthetic phospho-regulons. *Proc Natl Acad Sci USA* 113: 13528-13533.

566. Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, et al. (2009) ON/OFF and beyond--a boolean model of apoptosis. *PLoS Comput Biol* 5: e1000595.

567. Sedghamiz H, Morris M, Craddock TJA, Whitley D, Broderick G (2018) High-fidelity discrete modeling of the HPA axis: a study of regulatory plasticity in biology. *BMC Syst Biol* 12: 76.

568. Wollbold J, Jaster R, Muller S, Rateitschak K, Wolkenhauer O (2014) Anti-inflammatory effects of reactive oxygen species - a multi-valued logical model validated by formal concept analysis. *BMC Syst Biol* 8: 101.

569. Guebel DV, Schmitz U, Wolkenhauer O, Vera J (2012) Analysis of cell adhesion during early stages of colon cancer based on an extended multi-valued logic approach. *Mol Biosyst* 8: 1230-1242.

570. Naldi A, Berenguier D, Faure A, Lopez F, Thieffry D, et al. (2009) Logical modelling of regulatory networks with GINsim 2.3. *Biosystems* 97: 134-139.

571. Didier G, Remy E, Chaouiya C (2011) Mapping multivalued onto Boolean dynamics. *J Theor Biol* 270: 177-184.

572. Morris MK, Clarke DC, Osimiri LC, Lauffenburger DA (2016) Systematic Analysis of Quantitative Logic Model Ensembles Predicts Drug Combination Effects on Cell Signaling Networks. *CPT Pharmacometrics Syst Pharmacol* 5: 544-553.

573. Liu H, Zhang F, Mishra SK, Zhou S, Zheng J (2016) Knowledge-guided fuzzy logic modeling to infer cellular signaling networks from proteomic data. *Sci Rep* 6: 35652.

574. Bernardo-Faura M, Massen S, Falk CS, Brady NR, Eils R (2014) Data-derived modeling

characterizes plasticity of MAPK signaling in melanoma. *PLoS Comput Biol* 10: e1003795.

575. Nauck D, Kruse R (1999) Obtaining interpretable fuzzy classification rules from medical data. *Artif Intell Med* 16: 149-169.

576. Atalay KD, Can GF, Eraslan E (2018) Evaluation of effect of different membership functions on risk assessment. *Int J Occup Saf Ergon* 24: 373-385.

577. Su B, Jacinto E, Hibi M, Kallunki T, Karin M, et al. (1994) JNK is involved in signal integration during costimulation of T lymphocytes. *Cell* 77: 727-736.

578. Harlin H, Podack E, Boothby M, Alegre ML (2002) TCR-independent CD30 signaling selectively induces IL-13 production via a TNF receptor-associated factor/p38 mitogen-activated protein kinase-dependent mechanism. *J Immunol* 169: 2451-2459.

579. Gravestein LA, Amsen D, Boes M, Calvo CR, Kruisbeek AM, et al. (1998) The TNF receptor family member CD27 signals to Jun N-terminal kinase via Traf-2. *Eur J Immunol* 28: 2208-2216.

580. Yano S, Kondo K, Yamaguchi M, Richmond G, Hutchison M, et al. (2003) Distribution and function of EGFR in human tissue and the effect of EGFR tyrosine kinase inhibition. *Anticancer Res* 23: 3639-3650.

581. Sachdeva UM, O'Brien JM (2012) Understanding pRb: toward the necessary development of targeted treatments for retinoblastoma. *J Clin Invest* 122: 425-434.

582. Comisso E, Scarola M, Rosso M, Piazza S, Marzinotto S, et al. (2017) OCT4 controls mitotic stability and inactivates the RB tumor suppressor pathway to enhance ovarian cancer aggressiveness. *Oncogene* 36: 4253-4266.

583. Peng J, Jordan VC (2010) Expression of estrogen receptor alpha with a Tet-off adenoviral system induces G0/G1 cell cycle arrest in SKBr3 breast cancer cells. *Int J Oncol* 36: 451-458.

584. Batra SK, Castelino-Prabhu S, Wikstrand CJ, Zhu X, Humphrey PA, et al. (1995) Epidermal growth factor ligand-independent, unregulated, cell-transforming potential of a naturally occurring human mutant EGFRvIII gene. *Cell Growth Differ* 6: 1251-1259.

585. Hetmanski JH, Zindy E, Schwartz JM, Caswell PT (2016) A MAPK-Driven Feedback Loop Suppresses Rac Activity to Promote RhoA-Driven Cancer Cell Invasion. *PLoS Comput Biol* 12: e1004909.

586. Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS computational biology* 5: e1000438.

587. Vardi L, Ruppin E, Sharan R (2012) A linearized constraint-based approach for modeling signaling networks. *J Comput Biol* 19: 232-240.

588. Ramis G, Thomas-Moya E, Fernandez de Mattos S, Rodriguez J, Villalonga P (2012) EGFR inhibition in glioma cells modulates Rho signaling to inhibit cell motility and invasion and cooperates with temozolomide to reduce cell growth. *PLoS One* 7: e38770.

589. Xue J, Zhou A, Tan C, Wu Y, Lee HT, et al. (2015) Forkhead Box M1 Is Essential for Nuclear Localization of Glioma-associated Oncogene Homolog 1 in Glioblastoma Multiforme Cells by Promoting Importin-7 Expression. *J Biol Chem* 290: 18662-18670.

590. Carrasco-Garcia E, Martinez-Lacaci I, Mayor-Lopez L, Tristante E, Carballo-Santana M, et al. (2018) PDGFR and IGF-1R Inhibitors Induce a G2/M Arrest and Subsequent Cell Death in Human Glioblastoma Cell Lines. *Cells* 7.

591. Guichet PO, Guelfi S, Teigell M, Hoppe L, Bakalara N, et al. (2015) Notch1 stimulation induces a vascularization switch with pericyte-like cell differentiation of glioblastoma stem cells. *Stem Cells* 33: 21-34.

592. Zurawel RH, Allen C, Chiappa S, Cato W, Biegel J, et al. (2000) Analysis of PTCH/SMO/SHH pathway genes in medulloblastoma. *Genes Chromosomes Cancer* 27: 44-51.

593. Taipale J, Beachy PA (2001) The Hedgehog and Wnt signalling pathways in cancer. *nature* 411: 349-354.

594. Kim JH, Park S, Chung H, Oh S (2015) Wnt5a attenuates the pathogenic effects of the Wnt/beta-catenin pathway in human retinal pigment epithelial cells via down-regulating beta-catenin and Snail. *BMB Rep* 48: 525-530.

595. Sheng X, Sun X, Sun K, Sui H, Qin J, et al. (2016) Inhibitory effect of bufalin combined with Hedgehog signaling pathway inhibitors on proliferation and invasion and metastasis of liver cancer cells. *Int J Oncol* 49: 1513-1524.

596. Espinoza I, Pochampally R, Xing F, Watabe K, Miele L (2013) Notch signaling: targeting cancer stem cells and epithelial-to-mesenchymal transition. *Onco Targets Ther* 6: 1249-1259.

597. Espinoza I, Miele L (2013) Deadly crosstalk: Notch signaling at the intersection of EMT and cancer stem cells. *Cancer Lett* 341: 41-45.

598. Balakrishnan S, Bhat FA, Raja Singh P, Mukherjee S, Elumalai P, et al. (2016) Gold nanoparticle-conjugated quercetin inhibits epithelial-mesenchymal transition, angiogenesis and invasiveness via EGFR/VEGFR-2-mediated pathway in breast cancer. *Cell Prolif* 49: 678-697.

599. Tran PV, Lachke SA, Stottmann RW (2013) Toward a systems-level understanding of the Hedgehog signaling pathway: defining the complex, robust, and fragile. *Wiley Interdiscip Rev Syst Biol Med* 5: 83-100.

600. De Francesco EM, Maggiolini M, Musti AM (2018) Crosstalk between Notch, HIF-1alpha and GPER in Breast Cancer EMT. *Int J Mol Sci* 19.

601. Kandyba E, Leung Y, Chen YB, Widelitz R, Chuong CM, et al. (2013) Competitive balance of intrabulge BMP/Wnt signaling reveals a robust gene network ruling stem cell homeostasis and cyclic activation. *Proc Natl Acad Sci USA* 110: 1351-1356.

602. Kim TH, Monsefi N, Song JH, von Kriegsheim A, Vandamme D, et al. (2015) Network-based identification of feedback modules that control RhoA activity and cell migration. *J Mol Cell Biol* 7: 242-252.

603. Warrell J, Mhlanga M (2017) Stability and structural properties of gene regulation networks with coregulation rules. *J Theor Biol* 420: 304-317.

604. Manesso E, Kueh HY, Freedman G, Rothenberg EV, Peterson C (2016) Irreversibility of T-Cell Specification: Insights from Computational Modelling of a Minimal Network Architecture. *PLoS One* 11: e0161260.

605. Zanudo JG, Albert R (2013) An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos* 23: 025111.

606. Athar M, Li C, Tang X, Chi S, Zhang X, et al. (2004) Inhibition of smoothened signaling prevents ultraviolet B-induced basal cell carcinomas through regulation of Fas expression and apoptosis. *Cancer Res* 64: 7545-7552.

607. Rosland GV, Engelsen AS (2015) Novel points of attack for targeted cancer therapy. *Basic Clin Pharmacol Toxicol* 116: 9-18.

608. Evangelista M, Tian H, de Sauvage FJ (2006) The hedgehog signaling pathway in cancer. *Clin Cancer Res* 12: 5924-5928.

609. Olsen CL, Hsu PP, Glienke J, Rubanyi GM, Brooks AR (2004) Hedgehog-interacting protein is highly expressed in endothelial cells but down-regulated during angiogenesis and in several human tumors. *BMC Cancer* 4: 43.

610. Tojo M, Kiyosawa H, Iwatsuki K, Kaneko F (2002) Expression of a sonic hedgehog signal transducer, hedgehog-interacting protein, by human basal cell carcinoma. *Br J Dermatol* 146: 69-73.

611. Taipale J, Cooper MK, Maiti T, Beachy PA (2002) Patched acts catalytically to suppress the activity of Smoothened. *Nature* 418: 892-897.

612. Kasper M, Jaks V, Fiaschi M, Toftgard R (2009) Hedgehog signalling in breast cancer. *Carcinogenesis* 30: 903-911.

613. Ikram MS, Neill GW, Regl G, Eichberger T, Frischauf AM, et al. (2004) GLI2 is expressed in normal human epidermis and BCC and induces GLI1 expression by binding to its promoter. *J Invest Dermatol* 122: 1503-1509.

614. Gong A, Huang S (2012) FoxM1 and Wnt/beta-catenin signaling in glioma stem cells. *Cancer Res* 72: 5658-5662.

615. Pasca di Magliano M, Hebrok M (2003) Hedgehog signalling in cancer formation and maintenance. *Nat Rev Cancer* 3: 903-911.

616. Sicklick JK, Li YX, Jayaraman A, Kannangai R, Qi Y, et al. (2006) Dysregulation of the Hedgehog pathway in human hepatocarcinogenesis. *Carcinogenesis* 27: 748-757.

617. Das S, Harris LG, Metge BJ, Liu S, Riker AI, et al. (2009) The hedgehog pathway transcription factor GLI1 promotes malignant behavior of cancer cells by up-regulating osteopontin. *J Biol Chem* 284: 22888-22897.

618. Louro ID, Bailey EC, Li X, South LS, McKie-Bell PR, et al. (2002) Comparative gene expression profile analysis of GLI and c-MYC in an epithelial model of malignant transformation. *Cancer Res* 62: 5867-5873.

619. Liu S, Dontu G, Mantle ID, Patel S, Ahn NS, et al. (2006) Hedgehog signaling and Bmi-1 regulate self-renewal of normal and malignant human mammary stem cells. *Cancer Res* 66: 6063-6071.

620. He J, Sheng T, Stelter AA, Li C, Zhang X, et al. (2006) Suppressing Wnt signaling by the hedgehog pathway through sFRP-1. *J Biol Chem* 281: 35598-35602.

621. Elzi DJ, Song M, Hakala K, Weintraub ST, Shiio Y (2012) Wnt antagonist SFRP1 functions as a secreted mediator of senescence. *Mol Cell Biol* 32: 4388-4399.

622. Buttitta L, Mo R, Hui CC, Fan CM (2003) Interplays of Gli2 and Gli3 and their requirement in mediating Shh-dependent sclerotome induction. *Development* 130: 6233-6243.

623. Varnat F, Siegl-Cachedenier I, Malerba M, Gervaz P, Ruiz i Altaba A (2010) Loss of WNT-TCF addiction and enhancement of HH-GLI1 signalling define the metastatic transition of human colon carcinomas. *EMBO Mol Med* 2: 440-457.

624. Al-Hussaini H, Subramanyam D, Reedijk M, Sridhar SS (2011) Notch signaling pathway as a therapeutic target in breast cancer. *Mol Cancer Ther* 10: 9-15.

625. Tagami S, Okochi M, Yanagida K, Ikuta A, Fukumori A, et al. (2008) Regulation of Notch signaling by dynamic changes in the precision of S3 cleavage of Notch-1. *Mol Cell Biol* 28: 165-176.

626. Fassler M, Li X, Kaether C (2011) Polar transmembrane-based amino acids in presenilin 1 are involved in endoplasmic reticulum localization, Pen2 protein binding, and gamma-secretase complex stabilization. *J Biol Chem* 286: 38390-38396.

627. Kadesch T (2004) Notch signaling: the demise of elegant simplicity. *Curr Opin Genet Dev* 14: 506-512.

628. Beres BJ, George R, Lougher EJ, Barton M, Verrelli BC, et al. (2011) Numb regulates Notch1, but not Notch3, during myogenesis. *Mech Dev* 128: 247-257.

629. Nagpal P, Plant PJ, Correa J, Bain A, Takeda M, et al. (2012) The ubiquitin ligase Nedd4-1 participates in denervation-induced skeletal muscle atrophy in mice. *PLoS One* 7: e46427.

630. Koncarevic A, Jackman RW, Kandarian SC (2007) The ubiquitin-protein ligase Nedd4 targets Notch1 in skeletal muscle and distinguishes the subset of atrophies caused by reduced muscle tension. *FASEB J* 21: 427-437.

631. Nal B, Mohr E, Silva MI, Tagett R, Navarro C, et al. (2002) Wdr12, a mouse gene encoding a novel WD-Repeat Protein with a notchless-like amino-terminal domain. *Genomics* 79: 77-86.

632. Tsunematsu R, Nakayama K, Oike Y, Nishiyama M, Ishida N, et al. (2004) Mouse Fbw7/Sel-

10/Cdc4 is required for notch degradation during vascular development. *J Biol Chem* 279: 9417-9423.

633. Liao WR, Hsieh RH, Hsu KW, Wu MZ, Tseng MJ, et al. (2007) The CBF1-independent Notch1 signal pathway activates human c-myc expression partially via transcription factor YY1. *Carcinogenesis* 28: 1867-1876.

634. Ang HL, Tergaonkar V (2007) Notch and NFkappaB signaling pathways: Do they collaborate in normal vertebrate brain development and function? *Bioessays* 29: 1039-1047.

635. Cheng P, Zlobin A, Volgina V, Gottipati S, Osborne B, et al. (2001) Notch-1 regulates NF-kappaB activity in hemopoietic progenitor cells. *J Immunol* 167: 4458-4467.

636. Kuroda K, Tani S, Tamura K, Minoguchi S, Kurooka H, et al. (1999) Delta-induced Notch Signaling Mediated by RBP-J Inhibits MyoD Expression and Myogenesis. *Journal of Biological Chemistry* 274: 7238-7244.

637. Fryer CJ, White JB, Jones KA (2004) Mastermind recruits CycC:CDK8 to phosphorylate the Notch ICD and coordinate activation with turnover. *Mol Cell* 16: 509-520.