

RESEARCH ARTICLE

# Connecting signaling and metabolic pathways in EGF receptor-mediated oncogenesis of glioblastoma

Arup K. Bag<sup>1</sup>✉, Sapan Mandloi<sup>2</sup>✉, Saulius Jarmalavicius<sup>3</sup>✉, Susmita Mondal<sup>1</sup>, Krishna Kumar<sup>2</sup>, Chhabinath Mandal<sup>4</sup>, Peter Walden<sup>3\*</sup>, Saikat Chakrabarti<sup>2\*</sup>, Chitra Mandal<sup>1\*</sup>

**1** Cancer Biology and Inflammatory Disorder Division, Indian Institute of Chemical Biology, Kolkata, India, **2** Structural Biology and Bioinformatics Division, Indian Institute of Chemical Biology, Kolkata, India, **3** Department of Dermatology, Venerology and Allergology, Charité– Universitätsmedizin Berlin corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany, **4** National Institute of Pharmaceutical Education and Research, Kolkata, India

✉ These authors contributed equally to this work.

✉ Current address: Department of Biomedical Informatics, University at Buffalo, Buffalo, United States of America

\* [peter.walden@charite.de](mailto:peter.walden@charite.de) (PW); [saikat@iicb.res.in](mailto:saikat@iicb.res.in), [saikat273@gmail.com](mailto:saikat273@gmail.com) (SC); [chitra\\_mandal@yahoo.com](mailto:chitra_mandal@yahoo.com), [cmandal@iicb.res.in](mailto:cmandal@iicb.res.in) (CM)



**OPEN ACCESS**

**Citation:** Bag AK, Mandloi S, Jarmalavicius S, Mondal S, Kumar K, Mandal C, et al. (2019) Connecting signaling and metabolic pathways in EGF receptor-mediated oncogenesis of glioblastoma. *PLoS Comput Biol* 15(8): e1007090. <https://doi.org/10.1371/journal.pcbi.1007090>

**Editor:** Marcel Schilling, German Cancer Research Center (DKFZ), GERMANY

**Received:** January 27, 2018

**Accepted:** May 13, 2019

**Published:** August 6, 2019

**Copyright:** © 2019 Bag et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** AB thanks, Indian Council of Medical Research (ICMR) for the award of Senior Research Fellowships. SM thanks, Department of Science and Technology (DST) for INSPIRE fellowship. CM gratefully acknowledges financial support from J.C. Bose National Fellowship (DST) and Distinguished Biotechnology Research professorship (Department of Biotechnology), Govt. of India. The

## Abstract

As malignant transformation requires synchronization of growth-driving signaling (S) and metabolic (M) pathways, defining cancer-specific S-M interconnected networks (SMINs) could lead to better understanding of oncogenic processes. In a systems-biology approach, we developed a mathematical model for SMINs in mutated EGF receptor (EGFRvIII) compared to wild-type EGF receptor (EGFRwt) expressing glioblastoma multiforme (GBM). Starting with experimentally validated human protein-protein interactome data for S-M pathways, and incorporating proteomic data for EGFRvIII and EGFRwt GBM cells and patient transcriptomic data, we designed a dynamic model for EGFR-driven GBM-specific information flow. Key nodes and paths identified by *in silico* perturbation were validated experimentally when inhibition of signaling pathway proteins altered expression of metabolic proteins as predicted by the model. This demonstrated capacity of the model to identify unknown connections between signaling and metabolic pathways, explain the robustness of oncogenic SMINs, predict drug escape, and assist identification of drug targets and the development of combination therapies.

## Author summary

Complex and highly dynamic interconnected networks allow cancer to take different routes and circumvent chemotherapy. Therefore, understanding these context-specific networks and their dynamics of molecular interactions driven by different oncogenic signaling and metabolic pathways is very much needed to predict drug targets and the effect of therapeutics. We incorporated high-throughput transcriptome and proteome data into

work is supported in part by CSIR-IICB, BIOCERAMIC, CSIR (ESC 0103), (BSC0121) and ICMR (GAP 294), DST (GAP#336 and GAP # 339). CM and SC acknowledge the infrastructural support of CSIR-IICB. SC acknowledges financial support from DBT via SyMeC cluster project (GAP-357). PW and SJ gratefully acknowledge the support by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) for the project (IND 11/050, 01DQ12074), and the German Research Council (Deutsche Forschungsgemeinschaft, DFG) for travel grant in support of the study (WA 1139/22-1). The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

mathematical models to deduce properties of cancer cells through systems biology approach. Here we report the development, testing and validation of an integrated systems biology model of information flow between signaling and metabolic pathways to understand the regulation of the interconnection between them in cancer. Our model efficiently identified unique connections and key nodes important in signaling-metabolic information flow. We predicted some potential novel targets before performing actual drug tests. We have successfully applied this model to identify the interconnections altered in the constitutive signaling of the mutated EGFR by comparing EGF-dependent and wild-type EGFR signaling in glioblastoma multiforme.

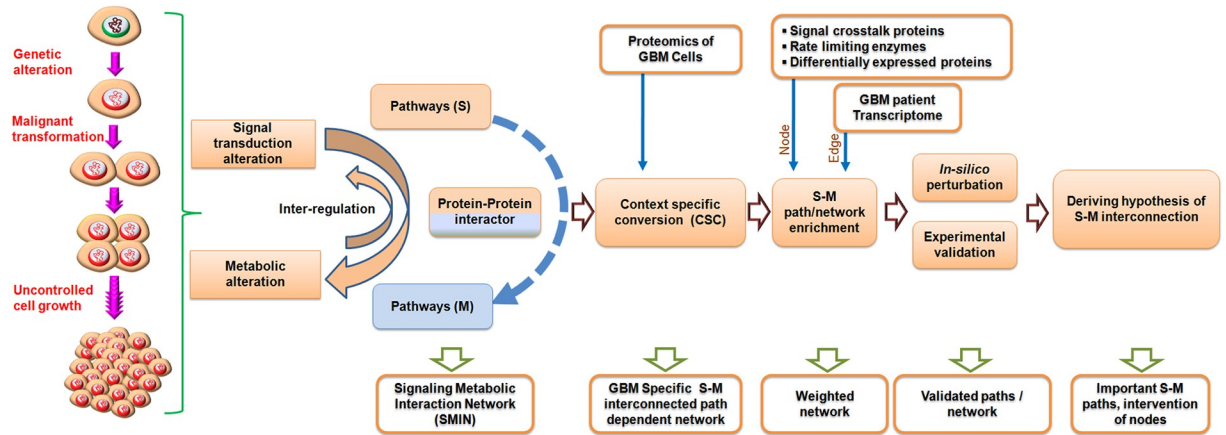
## Introduction

Diseases like cancer involve a large range of components that interact via complex and highly dynamic networks [1–3], and are interconnected with biochemical pathways [4–7]. These multipath interconnections may allow cancer and other diseases to take alternate routes and bypass the effects of therapeutic interventions. Traditional approaches to biological studies which focus on single molecules or pathways may not be able to capture and understand these complex networks of molecular interactions. To predict alternative or escape routes around blockades and to develop effective therapies [8], sophisticated mathematical and computational models are required [9–10].

Transforming traditional drug discovery approaches toward smarter therapeutic strategies, the field of systems biology is emerging [1, 9, 11–20]. Systems approach generally involve large-scale data collections, most often from high-throughput transcriptome or proteome analyses, incorporation of the data into mathematical models to deduce systems properties, model building and finally computational and/or experimental validation of model-derived hypotheses. Systems biology approaches may predict combination therapies for cancers driven by different oncogenic signaling and metabolic pathways.

Signaling and metabolic networks were studied using separate model systems [15, 21–26]. Mathematical models for signaling pathways had been based on logic models [27–30], kinetic models [31–33], decision tree [34], and other differential equation-based models [35]. Computational models of molecular signaling [36–41] have the potential to improve drug discovery and development [32, 42–44]. Analyses of knockdown experiments [45] using mass spectrometry [46] and transcriptomics [47–49] are progressively refined and tuned towards specific physiological situations. While these studies have helped considerably to extend our understanding of tumor biology, they are still restricted to signaling pathways and do not integrate the metabolic pathways, which in some initial studies have been subjected to separate systems biology analysis.

Predicting the effects of multiple targeted drugs [8, 50] with modeling the information flow from new molecular interactions within pathways is challenging [51–55]. Here we report the development, test and validation of an integrated model for signaling and metabolic pathways in cancer using glioblastoma multiforme (GBM) as an example [47, 56–59]. GBM is the most prevalent and most aggressive brain tumor. In the majority of cases, tumor development is dependent on signaling via the epidermal growth factor receptor (EGFR) and requires EGF in lower-grade forms or is EGF-independent in the more aggressive forms. In most cases, the expression of EGFR is up-regulated, often related to the amplification of the EGFR gene. More than fifty per cent of EGFR-amplified GBM cases have in-frame deletions of exons 2–7 that code for the extracellular ligand-binding domain (EGFRvIII mutation) resulting in EGF-



**Fig 1.** A schematic representation of the steps to establish the context-dependent signaling-metabolic interconnection model.

<https://doi.org/10.1371/journal.pcbi.1007090.g001>

independent constitutive signaling and more aggressive tumor growth, higher invasiveness, increased resistance to treatment, and poor prognosis [60–62].

In this study, we choose the cell line U87MG expressing the EGF-dependent EGFRwt and its derivative U87MGvIII expressing the EGF-independent EGFRvIII-mutant, as models of low and high-grade GBM, respectively. Except for the EGFR mutation, the two cell lines have the same genotype but differ in growth behavior suggesting different metabolic requirements, and are experimental examples for the regulation of the interconnection of signaling and metabolic pathways, which is considered among the basic characteristics of cancer [63–64].

We implemented a probabilistic approach based on the Hidden Markov Model (HMM) utilizing the information of experimentally established protein-protein interactions (PPIs) [65–66] to extract novel paths and interconnections between signaling pathway proteins (S) and metabolic pathway proteins (M). To cope with the limitations of PPI identification, for example high error rates of the detection methods [67], incomplete data sets, ignorance of the physiological conditions in the cell or tissue compartments [68], technical problems and study biases [69], we collected information from curated sources (<http://string-db.org>) with high experimental score cut-off, which reduces false positive rates, and used transcriptome data from clinical samples to build more reliable and GBM context-specific PPI networks. To bridge the gap between transcriptome data and cell-biological processes, we incorporated proteome data from the GBM cell lines complemented with transcriptome data to solve the missing data problem caused by the failure of proteomics to capture all proteins in the cells due to sensitivity and reproducibility issues. Including experimental data, our dynamic model can make use of multiple weighted network properties to add biological relevance and can extract novel paths of information propagation in networks. The results of the model were tested in rigorous *in-silico* perturbation experiments and experimentally validated in cell culture systems. Fig 1 depicts the overall strategy implemented with this study.

## Results

### Construction of a signaling to metabolic pathway interconnection network

As a starting model, we constructed an integrated network where signaling (S) and metabolic (M) pathway proteins were connected through protein-protein interactors (PPIs). The signaling pathways were the apoptosis, Akt, EGFR, hedgehog (Hh), JAK-STAT, JNK, MAPK, mTOR, NF-kappa B (NF-κB), Notch, p53, Ras, TGF-β, and Wnt pathways. On the metabolic

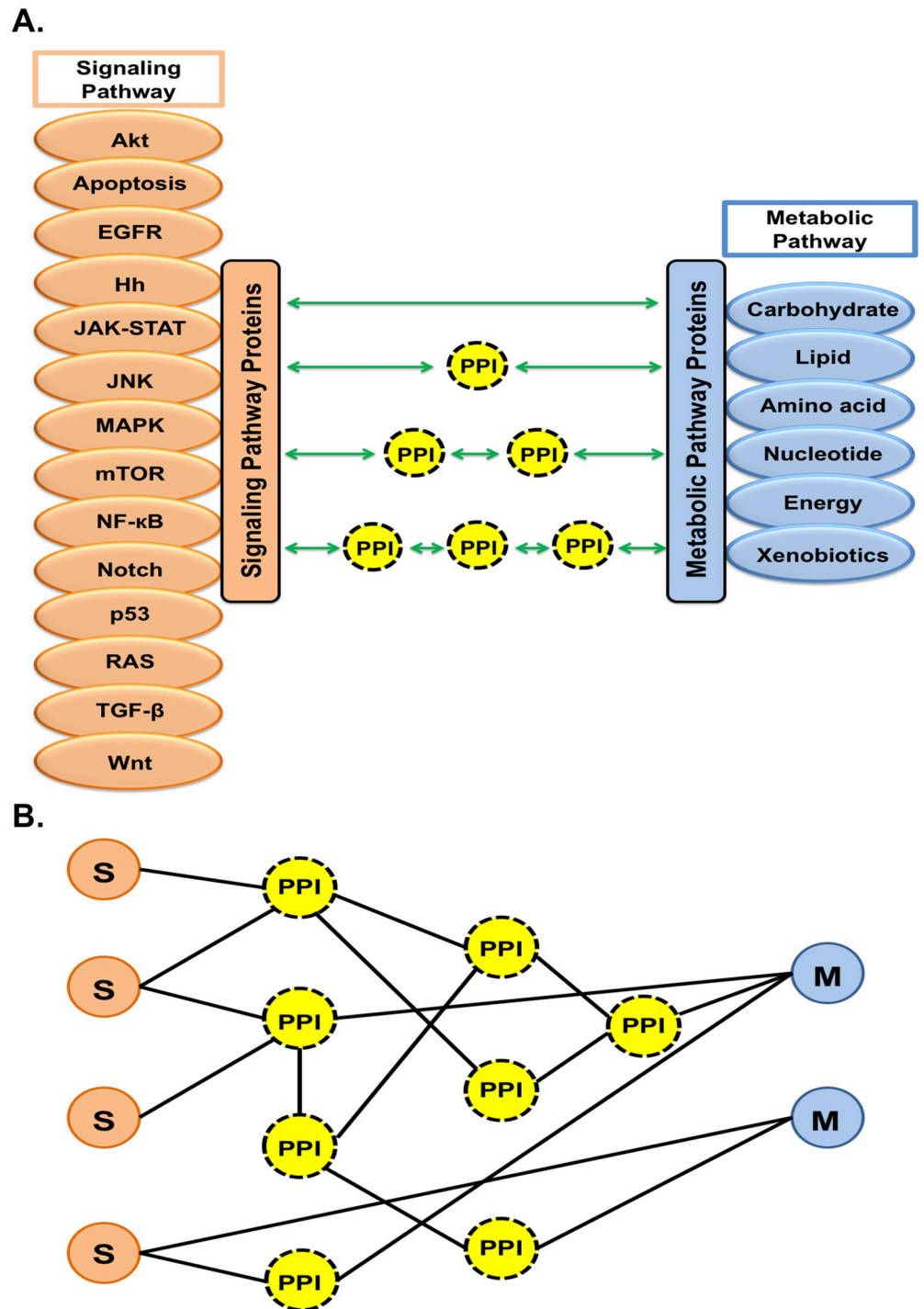
site, there were 81 pathways grouped into the six categories carbohydrate, lipid, amino acid, nucleotide, energy and xenobiotic metabolism (Fig 2).

To build an initial human protein-protein interactome network (HPPIN), the total of all human protein-protein interactions was extracted from the protein-interaction database STRING [70] for experimentally validated interactions including physical and functional associations. Then to structure the signaling-metabolic interaction network (SMIN), cross-connections between the fourteen signal transduction and six groups of metabolic pathway proteins selected from pathway databases (see [Materials and methods](#)) were constructed based on protein-protein interacting proteins (node) and cross-connecting links/paths. All possible connections between any given signaling pathway protein (S) and metabolic pathway protein (M) via protein-protein interconnectors (PPIs) were included (Fig 2A). To derive a simplified but informative network, the interactions were restricted to the second level of protein interactors, i.e. up to interactors of interactors. This led to four different types of cross-connected paths where signaling pathway proteins were either directly connected with metabolic pathway proteins or through one, two or three PPIs, respectively (Fig 2A). These paths connecting all above-mentioned signaling (S) and metabolic (M) pathway proteins were then converted into networks based on the protein-protein interaction status between the involved proteins (Fig 2B).

The resulting signaling-metabolic interaction network (SMIN) is shown in the left panel of Fig 3A with nodes in orange and edges (connection between two nodes) in blue, within the total interactome containing HPPIN network with grey nodes and edges. As a result of afore-said restriction criteria, Fig 3A shows the reduction of the network size (number of nodes/proteins and their interactions/edges) of SMIN (11,059 interactions formed by 2,785 proteins) from HPPIN (16,828 interactions formed by 5,703 proteins). As an example of network links, the detailed connections and interactions between one signaling pathway protein, CSNK2A1, and one metabolic pathway protein, NDUFA13, through different PPIs in the SMIN are shown on the right panel of Fig 3A. These two selected examples are indicated by asterisks. The SMIN was found to contain 158 direct (S-M) linking paths, 4,036 with one interactor (S-P-M), 91,847 with two interactors (S-P-P-M) and 2,110,205 with three interactors (S-P-P-P-M). These paths were formed between 158, 2,967, 22,307, 69,032 S-M pathway protein pairs, respectively (Table 1). Comparisons with respective random networks proved that our selected HPPIN and SMIN are non-random scale-free networks (S1 Fig and S1 Table).

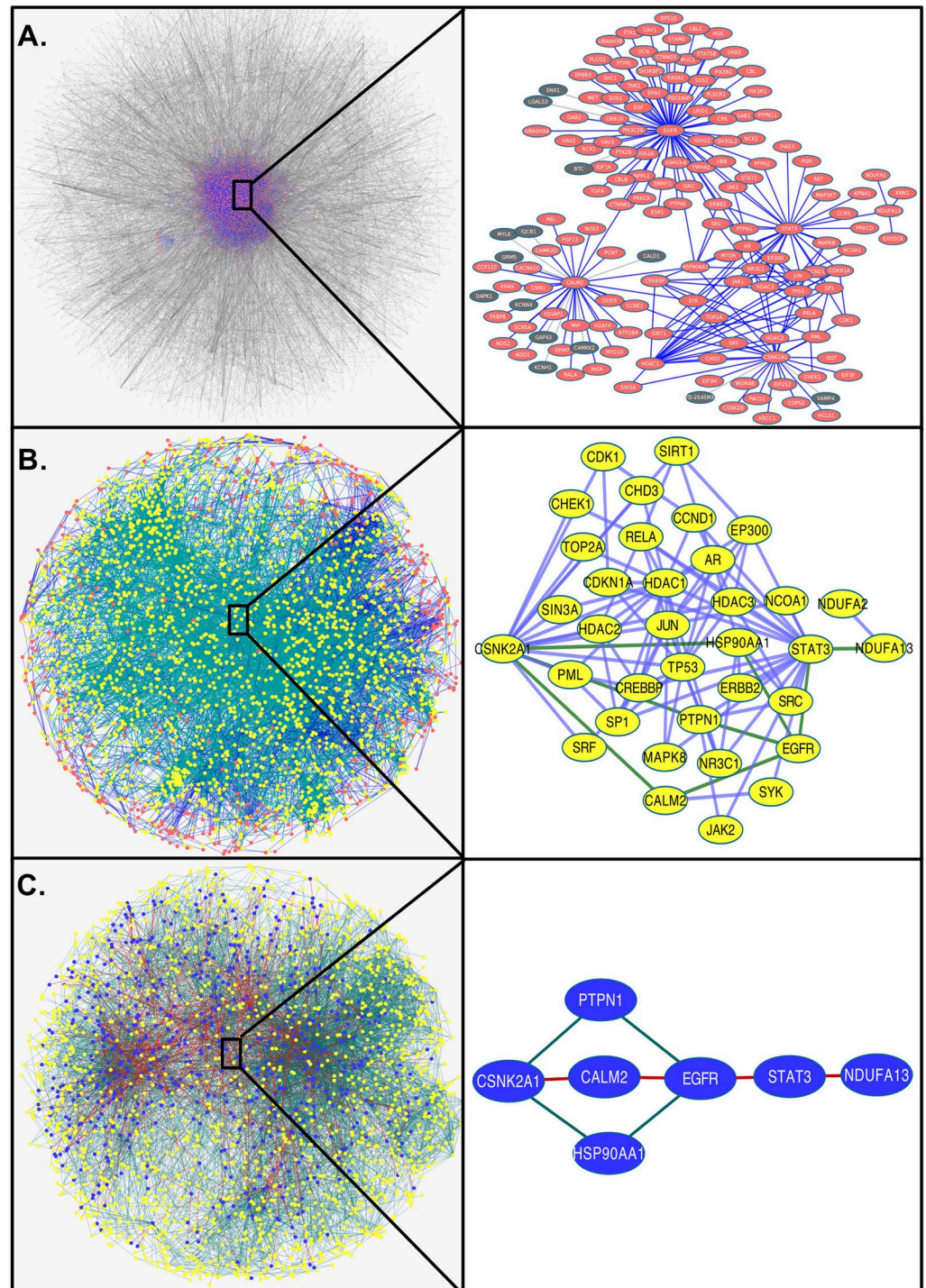
### Expression of signaling and metabolic pathway proteins in GBM cell lines

To render the above condition-independent SMIN GBM-specific, quantitative comparative proteome analysis of the low- and high-grade GBM cell lines U87MG and U87MGvIII were performed and the data incorporated in the model to derive an enriched GBM-specific network. The use of cell lines helped to reduce the noise associated with the usual small size and heterogeneous cellular compositions of clinical samples. Proteomes was chosen as primary expression data to focus the models on the protein level, which is more closely related to the biological processes to be modeled than transcriptome data. Transcriptome data from clinical samples (see below) were subsequently added to reduce the missing data problem of proteomics and link the model to the clinical level. To decrease the complexity of the proteins in the cell extracts and minimize signal suppression by overabundant peptides, the proteins were first separated by SDS-PAGE. Then the gels were cut into one mm slices followed by separate in-gel digestion of the proteins with trypsin. The resulting fragments were extracted from the gel slices as individual samples, separated by reverse-phase nano-HPLC and analyzed on-line by ESI-Q-TOF mass spectrometry (S2 Fig). Quantification was done label-free by calculating



**Fig 2. Establishment of signaling-metabolic pathway protein (S-M) interconnections for construction of SMIN.** (A) Fourteen signaling (left) and six groups of metabolic pathways (right; grouping of 81 individual metabolic pathways) were selected from the KEGG database to develop the S-M interconnected network (SMIN) through PPI. Experimentally validated data from the STRING protein-interaction database were used to construct an initial HPPIN. The levels of interaction were restricted at the interactors of interactors to create the network resulting in four possible linking paths through which selected signaling and metabolic pathway proteins could be connected directly or through one, two and three PPI. (B) Conversion of interconnecting paths into network representation.

<https://doi.org/10.1371/journal.pcbi.1007090.g002>



**Fig 3. Development of a GBM-specific significant network from HPPIN.** (A) A human protein-protein interactome network (HPPIN) was established based on experimentally validated interactome data of STRING database (left panel: the total network where nodes and edges are marked in grey). The SMIN was extracted from the HPPIN as signaling-metabolic interconnected path-dependent network between fourteen signaling and six groups of metabolic pathways as indicated in Fig 2 (highlighted as orange nodes and blue edges within the HPPIN on the left panel). The interconnections in the SMIN between one signaling (CSNK2A1) and one metabolic (NDUFA13) pathway protein through PPIs were highlighted on the right panel. (B) A GBM-specific network (left panel with yellow nodes and green edges) was developed from the SMIN (orange nodes and blue edges within the total background network) considering interconnecting paths with at least one differentially expressed protein identified from comparative proteome analysis of U87MG (EGFRwt)

and U87MGvIII (EGFRvIII) GBM. The interconnecting paths between the same signaling (CSNK2A1) and metabolic (NDUFA13) pathway protein after exclusion of non-GBM specific interconnections are shown on the right panel. (C) The GBM-specific network based on significant nodes ( $z \geq 1$ ) and significant interconnected paths (path score  $\geq 80\%$  of the highest path score of each SM pair) identified from the weighted network were indicated with blue nodes and red edges within the GBM-specific network with yellow node and green edge. The interconnecting paths in the GBM-specific significant network between CSNK2A1 and NDUFA13 are shown as examples on the right panel.

<https://doi.org/10.1371/journal.pcbi.1007090.g003>

the exponentially modified protein abundance index (emPAI) to avoid the drawback of mere signal intensity-based measurements. Only proteins with at least two tryptic fragments were identified by MS/MS with high confidence were considered, which further reduced the noise of protein identification although resulting in lower numbers of hits. The analyses were done with three independent replicates per cell line and the resulting data processed in two different ways. First, each of the three datasets for U87MG was compared to each of the three datasets for U87MGvIII resulting in nine pair-wise comparisons. Second, the average of three datasets of one U87MG was compared with the average of the three datasets for U87MGvIII.

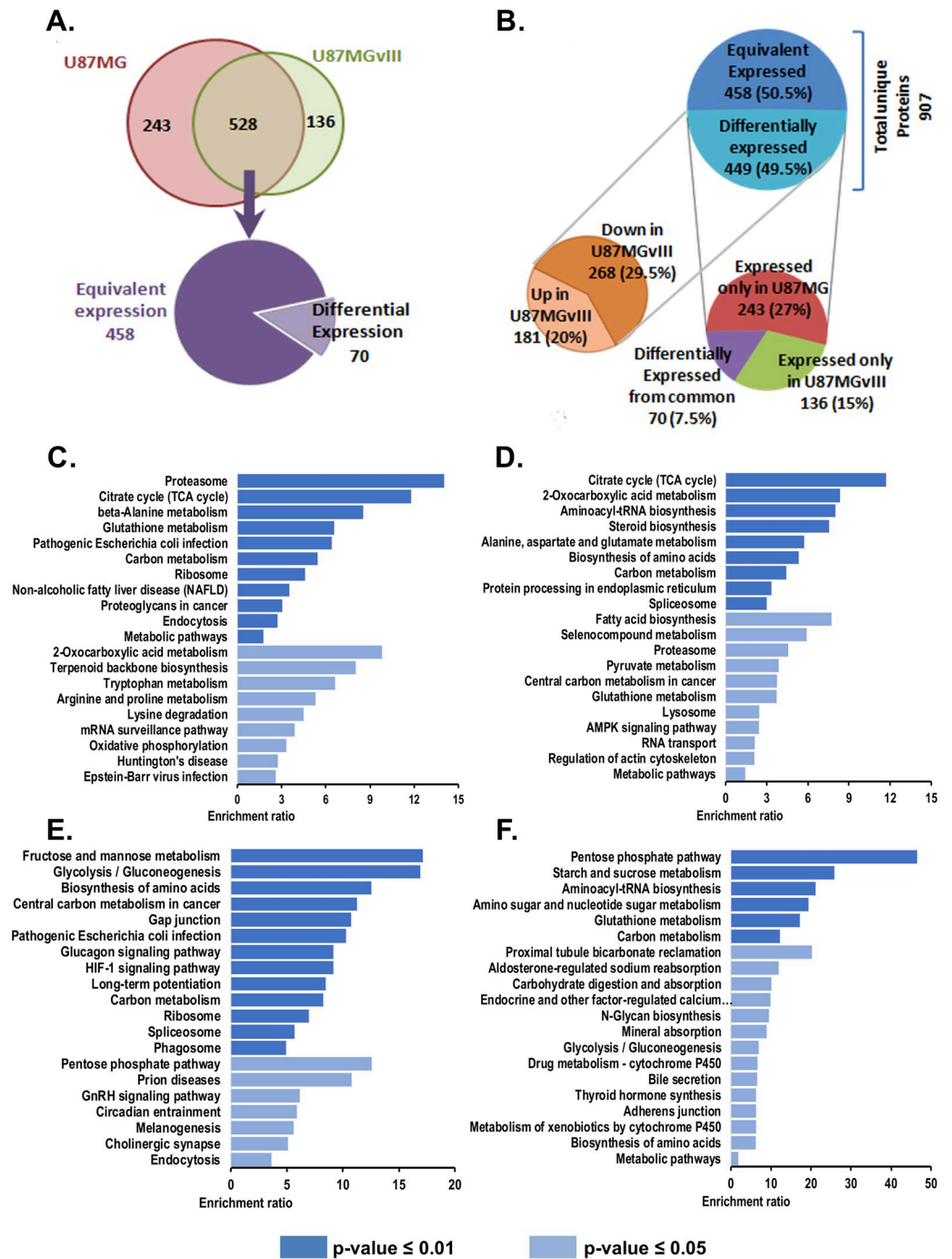
A total of 907 unique proteins were identified, 771 from U87MG and 664 from U87MGvIII. Five hundred twenty-eight proteins were expressed in both cell lines, 243 only in U87MG (EGFRwt) and 136 only in U87MGvIII (EGFRvIII). These distinctively expressed proteins were considered as overexpressed in the respective cell lines in comparison to the other. Compared quantitatively, 458 of the 528 common proteins were expressed at similar levels, 70 proteins were either up (45) or down (25) regulated in U87MGvIII compared to U87MG (Fig 4A). Together, nearly half of the identified proteins (449) were differentially expressed, 268 down-regulated and 181 up-regulated in U87MGvIII versus U87MG (Fig 4B).

Over-representation analysis (ORA) based enrichment for cellular pathways, biological processes and molecular functions were performed using cell line specific and commonly overexpressed proteins. Fig 4C and 4D show the top 20 most enriched pathways for proteins exclusively overexpressed in U87MGvIII (EGFRvIII) and U87MG (EGFRwt), respectively.

**Table 1. Detailed statistics of paths, pairs, and proteins with expression involved in stepwise filtration of the signaling-metabolic interaction network (SMIN) to the significant GBM-specific network with the differential expression states in the EGFR-mutated cell line U87MGvIII compared to U87MG with wild-type EGFR.**

Connection Types		Pairs and Paths							
		Supra Interaction Network (SMIN)		GBM specific Network		GBM specific paths and pairs in weighted network			
Exp. status		Total S-M pairs	Total paths	GBM specific Pairs	GBM specific Paths	Model-1 Pairs selection		Model-2 Path selection	
						$Z \geq 3$ pairs	$Z \geq 1$ pairs	$\geq 80\%$ Paths	$\geq 80\%$ paths
S & M same		51	-	2	-	-	-	-	-
S----M		158	158	22	22	1	1	2	5
S--P--M		2,967	4,036	425	509	2	8	21	84
S--P--P--M		22,307	91,847	5,509	15,042	114	334	228	600
S--P--P--P--M		69,023	2,110,205	33,328	421,876	758	1,961	625	1,564
<b>Total</b>		69,352	2,206,246	33,418	437,449	801	2,055	876	2,253
<b>Total gene</b>		949	2,785	906	2,319	431	494	447	654
UP	UP	17	48	17	48	11	15	17	22
	DN	24	85	24	85	8	12	23	32
	NC	53	162	47	142	11	19	24	40

<https://doi.org/10.1371/journal.pcbi.1007090.t001>



**Fig 4. Comparative proteome analysis between the GBM cell lines U87MG (EGFRwt) and U87MGvIII (EGFRvIII).** Protein identification and quantification were done by nano-HPLC Q-TOF MS/MS and peptide searches using the SwissProt databank for human proteins. Quantitation of the protein expressions was label-free by calculating the emPAI values for each protein. (A) Venn diagram for the statistics of proteins identified in U87MG and U87MGvIII. In three independent biological replicates of the proteome analysis for each cell line 907 proteins were identified, 528 common, 243 only for U87MG and 136 only for U87MGvIII cells. Of the shared proteins, 458 were equivalently and 70 differentially expressed. (B) Distribution of equivalently and differentially expressed proteins. Of the 449 differentially expressed proteins 243 were only present in U87MG and 136 in only U87MGvIII and 70 were present in both but in differential expression (down right). In total 268 were up-regulated and 181 down-regulated in U87MGvIII compared to U87MG. (C-D) Pathway over-representation analysis. Panels C and D plot the enriched pathways for proteins exclusively overexpressed in U87MGvIII (EGFRvIII) and U87MG (EGFRwt), respectively. One thirty four out of 136 and 235 out of



243 proteins overexpressed in U87MGvIII (EGFRvIII) and U87MG (EGFRwt) were mapped onto pathways. (E-F) Panels E and F plot enriched pathways of commonly up-regulated and down-regulated proteins between U87MGvIII and U87MG. Forty five out of 45 and 24 out of 25 proteins commonly overexpressed in U87MGvIII (EGFRvIII) and U87MG (EGFRwt) were mapped onto pathways.

<https://doi.org/10.1371/journal.pcbi.1007090.g004>

Similarly, Fig 4E and 4F provide the top 20 enriched pathways for commonly over and under expressed proteins, respectively.

The most significantly enriched pathways were found to be proteasome (p-value = 0.97763E-07) in U87MGvIII (EGFRvIII) and TCA cycle (p-value = 1.48369E-06) in U87MG (EGFRwt). However, Fructose and mannose metabolism (p-value = 6.7086E-04) and pentose phosphate pathways (p-value = 3.1503E-05) were found to be most significantly enriched for proteins commonly overexpressed and underexpressed, respectively.

Gene ontology (GO) based biological process and molecular function over-representation analysis was also performed using cell line specific overexpressed proteins. Most significantly enriched biological processes and molecular function were found to be Tricarboxylic acid metabolic process (p-value = 2.7304E-04) and Threonine-type peptidase activity (p-value = 0.001053394), respectively in U87MGvIII (EGFRvIII). In U87MG (EGFRwt), TCA metabolic process (p-value = 7.76842E-08) and pre-mRNA binding (p-value = 0.007857451) were found to be the most significantly enriched biological process and molecular functions, respectively (S3 Fig).

### EGFR-mediated GBM-specific network

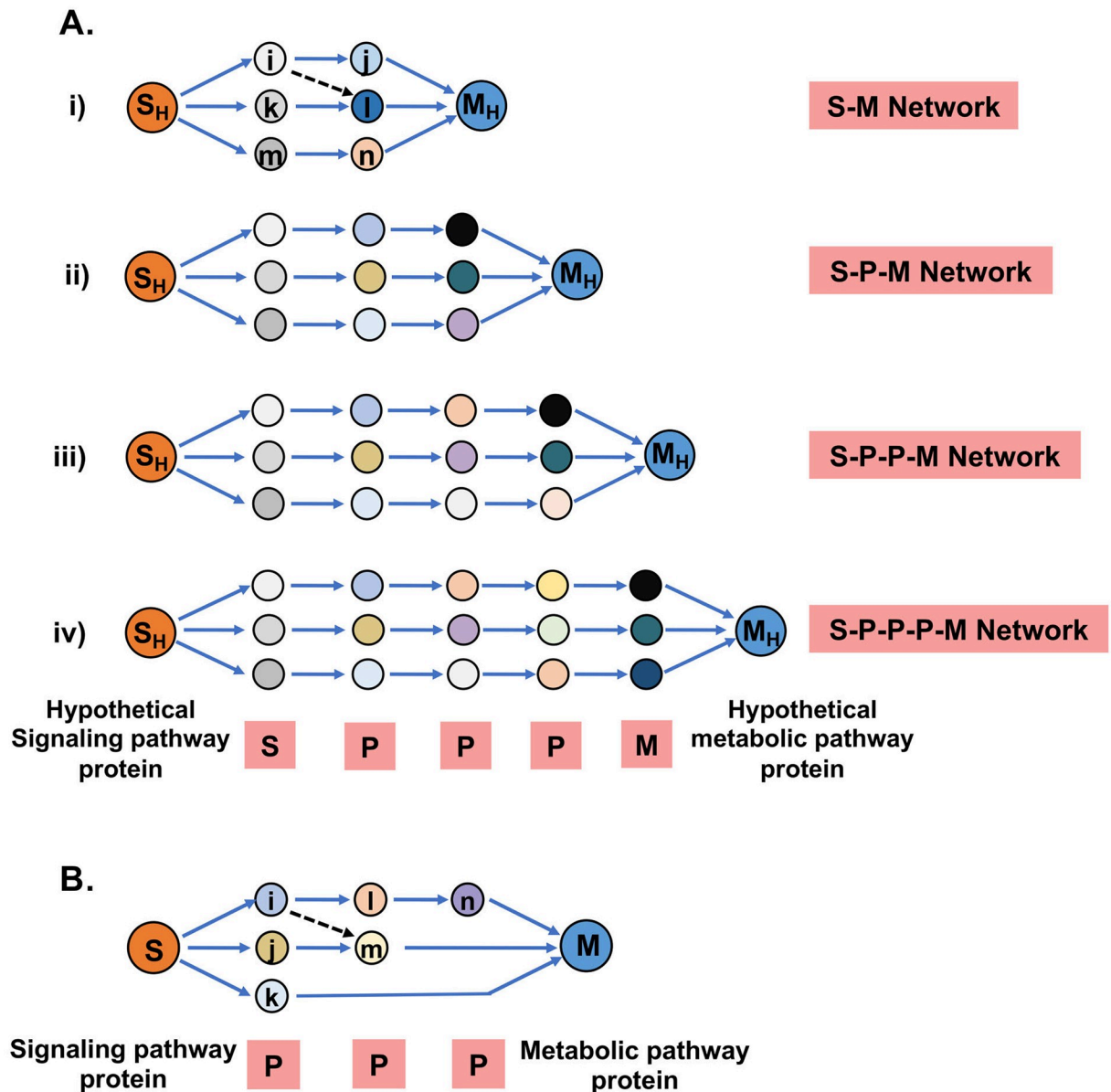
To uncover the signature of the reprogramming of global cellular processes by the EGF-independent constitutively active EGFRvIII in GBM, we mapped the data from the comparative proteomics of U87MGvIII (EGFRvIII) versus U87MG (EGFRwt) onto the above-described SMIN. All S-M interconnecting paths with at least one differentially expressed protein were extracted from the SMIN to generate a GBM-specific network (Fig 3B, left panel: highlighted with yellow nodes and green edges within the orange nodes and blue edges of the SMIN) with the assumption that those paths will have higher probability to be differentially active in mutant GBM condition. As an illustration of proteome data mediated extraction of a GBM-specific network, the interconnections between the aforementioned signaling pathway protein, CSNK2A1 and the metabolic pathway protein, NDUFA13 are highlighted after extraction from SMIN (Fig 3B right panel in comparison to Fig 3A right panel). Table 1 provides the details of the paths/pairs present at different stages of network development.

### GBM-specific network weighted for biological significance

To make this network further enriched with potentially disease-relevant paths/pairs, weights specifying disease-related biological properties and expression states of the proteins were assigned to each node (protein/gene) and edge (interaction) of the GBM-specific network. The following three categories of proteins (nodes) were given additional weights. First, proteins cross-talking between different signaling pathways (signaling cross-talk, SC), second, rate-limiting enzymes (RLE) for their roles in regulating metabolic rates and pathways, third, EGFR mutation-specific differentially expressed proteins (dEXP) for their GBM-specific impact. The GBM-specific network included 446 dEXP, 349 SC, and 267 RLE proteins. Of these, 11 SC and 17 RLE proteins were up or down regulated suggesting their involvement in signaling-metabolic cross-connection in EGFR-mutated condition (S4A Fig). For systems-level interpretation and understanding the network property, local signaling entropy ( $S_i$ ) was introduced. Previous studies [71–73] showed that  $S_i$  can be used as a measure of uncertainty in signaling

information flow over a network and to identify important signaling pathways and genes/proteins in cancer. Effect-on-node ( $eff_s$ ) of every protein (node) in the network provided significance of a protein based on SC, RLE and dEXP in its local network. To identify probable paths of information flow from a signaling to metabolic pathways, network entropy ( $S_i$ ) and effect-on-node ( $eff_s$ ) properties were incorporated as node weights into the logic of the Hidden Markov Model (HMM). The edge weight of every two interacting nodes (gene/protein) were defined based on the principle of mass action (assuming that the probability of interaction of two genes in a given sample is proportional to the product of their expression values in the study samples) as probability of interaction ( $p_{ij}$ , where  $i$  and  $j$  are the two nodes) in GBM condition. To assign the expression value of each node present in the GBM specific network, the average expression value of each gene was calculated from the normalized transcriptome data from 239 GBM patients. Incorporating these transcriptome data as edge weights linked the network with biological information from GBM patients. It helped to assign an extra weight other than previously mentioned node weight for all connections made by two nodes based on their expression in clinical GBM patients. Furthermore, it helped to add another level of constraint on over-prediction of information flow for nodes, which got an extra weight based on SC, RLE, and dEXP but are not expressed at higher levels in GBM patients. This helped to incorporate the contribution of those nodes to the disease, which were identified by neither of the three before-mentioned node weights nor by proteomics. Moreover, the much broader coverage of gene expression by genome-wide transcriptomics compared to proteomics helped to overcome some of the missing-data-problem of proteomic datasets.

An HMM-based simple mathematical formalism was used to understand context-specific information propagation from signaling to metabolic pathways in the human biological network. Node weights and edge weights were used to define the two major parameters of the Markov model, emission ( $E_i$ ) and transition ( $T_j$ ) probabilities, respectively. Two model systems were implemented to apply HMM logic, Model 1 for SM pair identification (Fig 5A) and Model 2 for S-M linking path identification (Fig 5B). Model 1 emphasized source (Signaling Pathway Protein, S) and destination (Metabolic Pathway Protein, M) pairs i.e. SM pairs having higher chances of information flow for each type of connections (Figs 2A and 5A). Model 2 was applied to find S-M linking paths between those selected pairs from Model 1. Selection of SM pairs (Model 1) and S-M linking paths (Model 2) was based on  $Path_{score}$  (see Methods for more details) a mathematical function of emission probability ( $E_i$ ) and transition probability ( $T_j$ ). For Model 1 positional emission probability was calculated considering the similar number of proteins because Model 1 was applied after grouping interconnecting links having the similar number of proteins forming the connection (Fig 5Ai–5Aiv). The calculated path scores of linking paths from the individual models were converted to statistical Z scores to identify the paths deviating from the mean. Based on the Z score under the individual models, the signaling-metabolic linking paths were classified as highly significant with Z score  $\geq 3$  (more stringent) or less significant with Z score  $\geq 1$  (less stringent) in EGFR-mutated GBM. The signaling and metabolic pathway proteins from the two ends of linking paths containing significant Z scores of each of the models were defined as the significant SM pairs. Multiple identifications of the same S-M pair from different models i.e. the formation of different types of significant linking paths involving different PPIs or different numbers of PPIs were nullified by considering them as a single. With that we identified 1, 2, 114 and 758 SM pairs meeting the more stringent cut-off (Z score  $\geq 3$ ) and 1, 8, 334 and 1,961 pairs with the less stringent cut-off (Z score  $\geq 1$ ) for the S-M, S-P-M, S-P-P-M and S-P-P-P-M linking types, respectively (Table 1, Model-1). In total 801 Z  $\geq 3$  and 2,055 Z  $\geq 1$  signaling-metabolic cross-connected SM pairs were identified between the 14 signaling and 6 groups of metabolic pathways as potentially important in EGFR-mutated GBM. These SM pairs were categorized according to



**Fig 5. Models for the identification of significant signaling-metabolic (SM) pathway protein pairs and interconnecting paths.** Models were implemented to identify significant SM pairs and paths from the weighted network. Weights of the nodes were defined based on whether the proteins were differentially expressed (dEXP), signal crosstalk (SC) proteins, rate-limiting enzymes (RLE) or none of these. The edge weights were defined as interaction probabilities derived from the gene expression values of two interacting proteins in GBM patients. (A) Model 1 was used to identify significant SM pairs from the GBM-specific weighted network. Four types of interconnection with none, one, two and three PPIs between S-M pairs were considered as separate models (i to iv). Two hypothetical proteins were considered on each side ( $S_H$  and  $M_H$ ) to implement positional probabilities based on the node weights of all proteins including S and M. Using these models and the Hidden Markov Model (HMM) logic, the path scores were calculated from the node and edge weights. The path scores were converted to Z-scores, and scores above the cut-off value of  $Z \geq 1$  or  $Z \geq 3$  were considered as significant and highly significant, respectively, and the respective interconnected SM pairs as significant pairs. (B) Model 2 was used to identify important linking-paths among the total possible paths formed between significant SM pairs from Model 1. Path scores were calculated from the positional probabilities as above. Higher path scores indicate higher significance.

<https://doi.org/10.1371/journal.pcbi.1007090.g005>

the proteomic expression states of the source (S) and destination (M) proteins as UP-DOWN, UP-UP, DOWN-UP, and DOWN-DOWN. Including unidentified proteins in proteomics analysis (NA) of the cell lines, the couplet categories UP-NA, DOWN-NA, NA-UP and NA-UP, and unchanged expression states (NC) in the cell lines with SM categories UP-NC,

DOWN-NC, NC-UP and NC-DOWN, and unidentified and unchanged SM pairs NA-NA and NC-NC were added (S4B Fig).

A number of cross-connections between signaling and metabolic pathways were identified with significant cutoff levels where either one or both pathway proteins (S and/or M) were not identified (NA) and/or unchanged (NC) by mass spectrometry indicating that the integrated network model can identify connections also where intermediate interactors are more important than SC, RLE or dEXP. The identification of pathway cross-connections is thus not dependent on proteomic identification of all constituent members but can be based on signaling crosstalk proteins and their expression status in GBM patients. It is important to include unchanged (NC) proteins in the model building as they might represent nodes in the paths that include other proteins that are differentially expressed. They might also play a role in the crosstalk between signaling paths and pathways or they might become important when known primary paths are blocked, e.g. by therapeutic intervention. The model could thus help to identify potential therapeutic targets for alternative therapies in cases of treatment failures or to design combination therapies that target primary together with potential escape pathways.

Mapping the pathway information of proteins in SM pairs showed which signaling pathways made a higher number of connections with which type of metabolic pathways in EGFR-mutated GBM. Six hundred one significant pairs with  $Z \geq 1$  were cross-connecting the MAP kinase pathway with all six groups of metabolic pathways (S2 Table). Similarly, the Ras, EGFR, AKT and p53 pathways were significantly connected to metabolic pathways through 570, 543, 549 and 179 SM pairs, respectively (S2 Table). As crosstalk between available signaling pathways is common whereas it is less common in between metabolic pathways (S5A Fig), some identified SM pairs and the respective proteins/genes may be shared. Analyzing the shared components in the five most connected signaling pathways revealed that MAPK pathway had the highest number of unique significant SM pairs (256) and genes/proteins (63) involved, followed by 194, 129, 116 and 95 unique significant SM pairs (S5B Fig) and 20, 30, 36, 19 genes/proteins (S5C Fig left) for the EGFR, AKT, p53 and Ras pathways respectively. Twelve cross-connected pathway protein pairs were common to all five pathways and 141 pairs shared by the Ras, EGFR, AKT and MAPK pathways indicating high connectivity between them, and their cross-connection with metabolic pathways suggesting important roles of the respective proteins in EGFR-mutated GBM (S5B Fig for pairs, C for genes left). In turn, the amino acid, carbohydrate, and nucleotide metabolic pathway groups were connected to all fourteen signaling pathways through 327, 289 and 326 cross-connected SM pairs (S2 Table, S5B Fig right) and 296, 260 and 268 genes in significant S-M paths, respectively (S5C Fig right). This indicates that altered cellular signaling related to the EGFR mutation and its constitutive activity affects most strongly these three metabolic pathway groups. As metabolic pathway enzymes interact via their substrates and products, there are few possibilities for interconnection between metabolic pathways except for the end steps, which is confirmed by the analysis of shared proteins. Twenty-three shared pairs were identified between the amino acid and the carbohydrate metabolic pathway, which relates to the low number of amino acids metabolites feeding into the tri-carboxylic acid cycle (S5B Fig right).

Multiple linking paths of different types (S-M, S-P-M, S-P-P-M, and S-P-P-P-M) or of the same type but through different PPIs were possible between SM pairs. Not all of these linking paths could be equally significant in EGFR-mutated GBM. To find the significant linking paths between the above-identified significant SM pairs, all possible paths between a single SM pair were considered under a single model (Model 2, Fig 5B). Since in Model 2 proteins forming interconnections between SM pairs vary, positional emission probabilities were calculated for these proteins. As an example, in Fig 5B the second position contained two proteins and third position one protein. Paths with path scores  $\geq 80\%$  of the highest path score for each SM pair were selected as significant from Model 2. Accordingly, all the significant paths were

identified from high (more stringent  $Z \geq 3$ ) and less (less stringent  $Z \geq 1$ ) significantly specified SM pairs to identify the total of significant linking paths in the network. These analyses showed that 2, 21, 228, 625 and 876 significant paths were present with more stringent cut-off ( $Z \geq 3$ ) and 5, 84, 600, 1,564 and 2,253 paths with the less stringent cut-off ( $Z \geq 1$ ) for the four pathway types (Table 1). By these pathway-based analyses under less stringent condition ( $Z \geq 1$ ), 652 significant linking-paths were identified between 570 cross-connected SM pairs of the Ras pathway with all six groups of metabolic pathways (S2 Table). In addition, 668, 629 and 569 significant linking-paths were identified between 601, 549 and 543 cross-connected S-M pairs between the MAPK, AKT and EGFR pathways, respectively, and all 6 groups of metabolic pathways. Together, in EGFR-mutated GBM, these four signaling pathways were involved in the highest number of SM cross-connections with metabolic pathways: 368, 344 and 298 cross-connecting paths were found between the fourteen signaling pathways and 327, 326 and 289 cross-connected SM pairs involving the amino acid, nucleotide and carbohydrate metabolism, respectively (S2 Table).

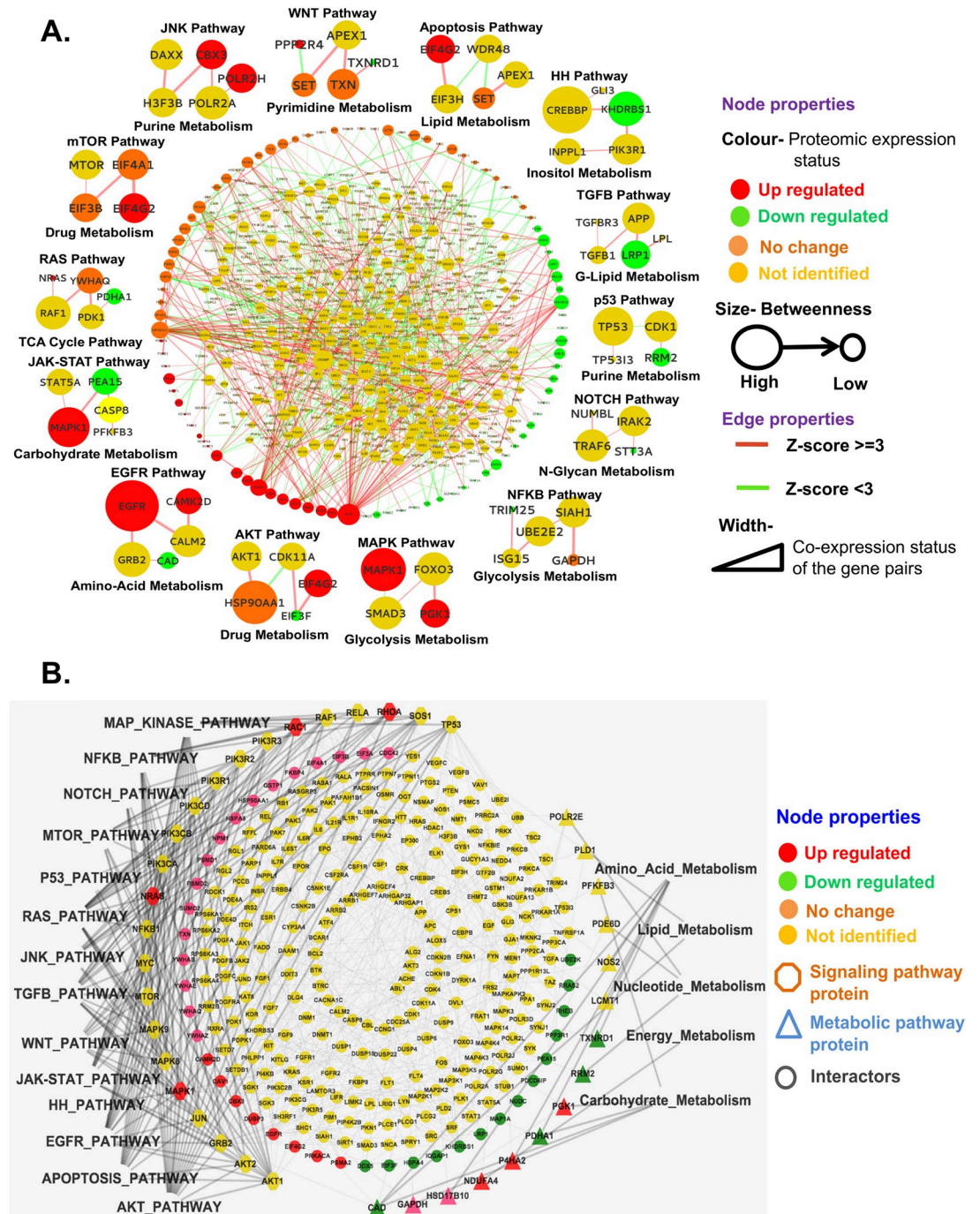
### GBM-specific network based on significant SM pairs and paths

Based on the identified significant (less stringent  $Z \geq 1$ ) SM pairs and the significant linking paths (path score  $\geq 80\%$  of the highest path score of each SM pair), we converted (as in Fig 2B) the significant paths into a network (Fig 3C left panel: highlighted with blue nodes and red edges within yellow nodes and green edges of the GBM network). This filtered network is more specific for EGFR-mutated GBM conditions. The filtration further eliminated non-significant or unimportant SM pairs and PPIs, which is shown, as an example, for the interconnections between signaling pathway protein CSNK2A1 and metabolic pathway protein NDUFA13 (Fig 3C right panel compared to Fig 3B right panel).

The GBM-specific network based on significant SM pairs and linking paths was restructured to implement the biological consequences as network properties i.e. color-coded proteomic expression states (up, down, no change and not identified), size of node symbols proportional to the numbers of connections passing through it, colors of the edges as connection formed between more ( $Z \geq 3$ ) or less ( $Z \geq 1$ ) stringently defined SM pairs, width of the edge as the probability of interaction or product of the average expression values of two interacting genes in GBM patients from the transcriptome data (Fig 6A). The resulting network showed the important signaling pathways and their interconnections with metabolic pathways in EGFR-mutated GBM with the significance of every protein (size of the node) and their interactions with interacting partners (width of the edge). Around the network, representative paths between 14 signaling to metabolic pathways are shown as examples (Fig 6A, S6 Fig as more details of RAS pathway). To explore the importance of signal-crosstalk proteins in signaling to metabolic pathway interconnections in EGFR-mutated GBM, the sub-network dependent on the top fifteen crosstalk protein-based interconnecting paths were extracted from the GBM-specific significant network (Fig 6B). These sub-networks showed which signaling pathways are mostly cross-talking and how they are connected with metabolic pathways. This information was used to extract candidate genes/proteins/paths of EGFR-mutated GBM for further analysis.

### *In silico* perturbation test of the network

*In silico* perturbation analysis was performed for identification of paths that significantly change upon removal (e.g. by mutation or down-regulation) of a node (protein). To test the importance of the nodes/proteins in the final weighted network, each of the 654 nodes present in the  $Z \geq 1$  network (Table 1) was removed individually from the human interactome



**Fig 6. GBM-specific significant network.** (A) The SM pairs and interconnecting paths between them that have a significant impact on EGFR-mediated GBM were filtered from the GBM-specific network using the mathematical model of Fig 5 and combined to generate the final signaling-metabolic interconnected network. The network parameters (color code of node and edge, node size, edge width mentioned in the figure) signified the cell-biological consequences of GBM driven by constitutive EGFR signaling. One representative signaling-metabolic connecting path for each of the fourteen signaling pathways is indicated at the side of the network. (B) Sub-network based on the 15 top-ranked signaling cross-talk proteins and their interconnection with all metabolic pathways.

<https://doi.org/10.1371/journal.pcbi.1007090.g006>

(HPPIN) and the node and edge weights were recalculated for the resulting networks and paths by recalculating Model 1 and Model 2 (Fig 2).

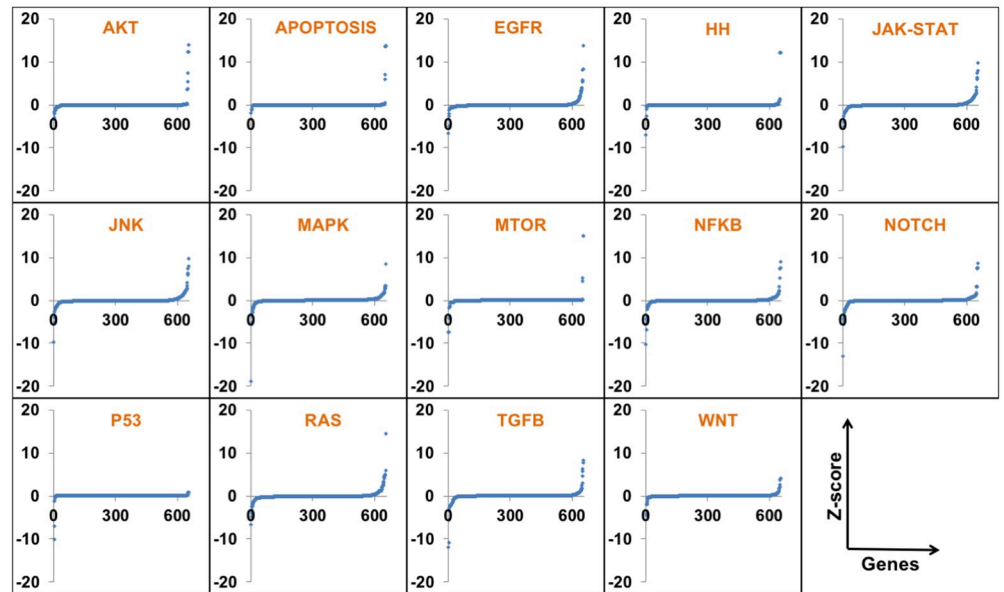
Accordingly, new significant SM pairs ( $Z \geq 3$  or 1) were identified on the basis of Model 1 and significant paths (path score  $\geq 80\%$  of the highest path score) between them on the basis of Model 2. We mapped the pathway details of the SM pairs and calculated the average path scores before and after perturbation for the 14 signaling pathways to all 6 groups of metabolic pathways and *vice-versa*. The difference of values (before *vs.* after perturbation) for the 654 proteins for all pathways were converted to Z-scores and plotted for each perturbed node for each signaling pathway (Fig 7A). The nodes for which the Z-scores deviated from the mean as  $-2 \geq Z \geq 2$  were selected as effective for the respective signaling pathway to all metabolic pathway interconnections in EGFR-mutated GBM (Fig 7B). S3 Table lists the numbers of significant and effective proteins identified for the individual signaling pathways connected to all metabolic pathways and from all signaling pathways to the individual metabolic pathways. As a measure of its effect on signaling-metabolic interconnection, each perturbed node was ranked according to the difference between baseline and perturbed condition. This means that highly ranked proteins have an important role in the connections of the respective signaling pathway to all metabolic pathways or *vice-versa*.

The NOTCH pathway is shown as an example for the level of reduction in the network size when the GBM network is transformed to the significant GBM network to identify significant interconnecting paths and proteins (Fig 8). Fig 8A shows the interconnections of NOTCH pathway proteins with all metabolic pathway proteins present in the signaling-metabolic interaction network (SMIN) and Fig 8B shows only those NOTCH pathway proteins with interconnections to metabolic pathway proteins passing through effective nodes, i.e. nodes identified by the perturbation experiments (Fig 7B). Fig 8C presents the interconnections of the NOTCH pathway to all metabolic pathways in the GBM-specific network filtered on the basis of the weightage parameters for the nodes and edges, and Fig 8D shows the interconnections between the significant proteins only. Fig 8E shows the interconnections of the NOTCH pathway to all metabolic pathways in the GBM-specific significant network and Fig 8F the same only for significant nodes. The comparison of Fig 8B, 8D and 8F on basis of the effective node identified by the perturbation study (red colored) indicate the levels of filtration from the starting SMIN to the final GBM-specific significant network. We found 457,111 paths involving 1941 genes/proteins and 8047 interactions out of a total of 2,206,246 paths in the SMIN connecting the NOTCH signaling pathway to all metabolic pathways, of which 10% were found to have a significant (Z-score cut-off  $-2 \geq Z \geq 2$ ) perturbation impact (PI) (Fig 8B). Comparable reductions of approximate 75% (Fig 8C) and 50% (Fig 8D) for both nodes and interaction were found when going to the GBM-specific condition. After  $Path_{score}$  based filtration (Z-score  $\geq 1$ ), 146 paths involving 125 nodes and 166 interactions were identified in the NOTCH pathway (Fig 8E) of which 51 paths involving 59 nodes and 65 interactions formed by nodes with significant PIs.

## Experimental validation of signaling-metabolic network

The result of perturbation study proved the importance of the respective nodes in the final network for information flow from signaling to metabolic pathways. To validate these findings of interconnections between signaling pathway alterations and metabolic rearrangement, some of SM connections were selected from the final network for experimental validation (Fig 9). The selection was based on the effect of the removal of the respective node in the *in silico* perturbation experiments, the predicted effects on the metabolic pathways and the availability of specific small-molecule inhibitors, excluding transcription factors as their impact is obvious.

**A.**



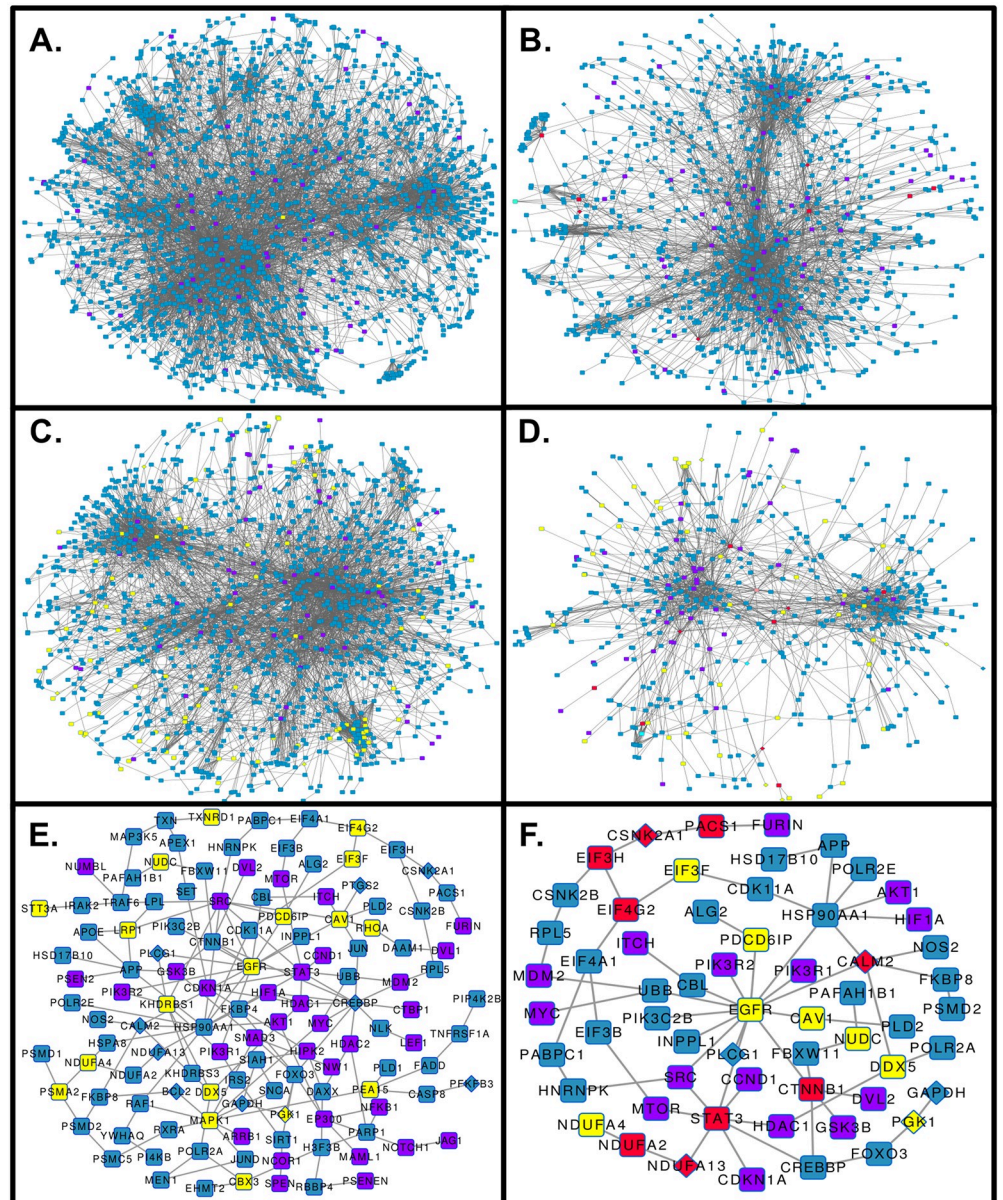
**B.**

AKT	FOXO3	PGK1	GAPDH	HTT	CREBBP	EIF4G2	EIF4E	EIF3H	EIF3F								
APOP	FOXO3	PGK1	GAPDH	HTT	CREBBP												
EGFR	EGFR	CAMK2D	CALM2	PLCG1	SYK	INPPL1	STAT3	NDUFA13	CREBBP	ERBB4							
HH	PRKACA	CREBBP	FBXW11	OPTN													
JAK-STAT	STAT3	NDUFA13	NDUFA2	EGFR	PLCG1	KHDRBS1	CAD	GAPDH	PGK1	STAT1	GRB2						
JNK	IQGAP1	RAC1	CALM2	SYK	CAD	GRB2	EP300	HSP90AA1	RHOA								
MAPK	EGFR	PLCG1	PRKACA	YWHAB	HSP90AA1	APEX1	NTRK2	MAP3K5									
MTOR	EIF4G2	EIF4E	EIF3H	EIF3F	PGK1	GAPDH											
NFKB	GAPDH	SIAH1	UBE2E2	ISG15	EIF4G2	CSNK2A2	CSNK2B	EIF3F	HSP90AA1	RHOA	YWHAB	STUB1	APEX1				
NOTCH	STAT3	NDUFA2	NDUFA13	EIF4G2	CSNK2A1	PACS1	CAMK2D	EIF3H	SYK	CTNNB1	CALM2	CSNK1E					
P53	FOXO3	TP53	CDK1														
RAS	EGFR	BRCA1	PRKACA	PLCG1	INPPL1	CALM2	POLR2A	FOXO3	TP53	CDK1							
TGFB	SMAD4	FOXO3	EGFR	DCN	PLCG1	EIF3H	ERBB4	RHOA	HSP90AA1								
WNT	PGK1	GAPDH	SIAH1	UBB	HSP90AA1	PGK1	GAPDH										
AA-Met	SMAD4	UBE2I	FOXO3														
C-Met	GAPDH	PGK1	FOXO3	HTT	CREBBP	MAP3K5	GAPDH	PGK1									
E-Met	NDUFA13	STAT3	NDUFA2	EGFR													
L-Met	LRP1	BRCA1	PLD1	PEA15	SYK	CALM2											
N-Met	TP53	CDK1															
X-Met	MAPK8	PPP3CA	BCL2	MAP3K5	JUN	SYK	CASP8	NTRK2	FBXW11								

**Fig 7. Importance of the individual nodes in signaling-metabolic pathway interconnection.** (A) Each of the 654 nodes of the final network was tested in perturbation studies by removing it and its associated connections from the network and recalculating all network parameters. The scores for significant SM pairs ( $Z \geq 1$ ) and interconnected paths (path score  $\geq 80\%$  from the highest path score for each SM pair) were redefined for each signaling to all metabolic interconnections. The differences of the sums of all significant path scores before and after perturbation of the individual 654 nodes for each signaling to all metabolic pathway interconnections were converted to Z scores (blue dots) and plotted against the genes/nodes. (B) Summary of the nodes with a significant impact on the entire network. Nodes with significantly altered Z score after perturbation ( $-2 \geq Z \geq 2$ ) were considered as effective nodes with significant impact and are listed in S3 Table for all interconnections of the 14 signaling pathways to the 6 metabolic pathway groups and *vice versa*, the metabolic pathway groups to the signaling pathways. Drugable node/proteins according to the KEGG database are highlighted.

<https://doi.org/10.1371/journal.pcbi.1007090.g007>

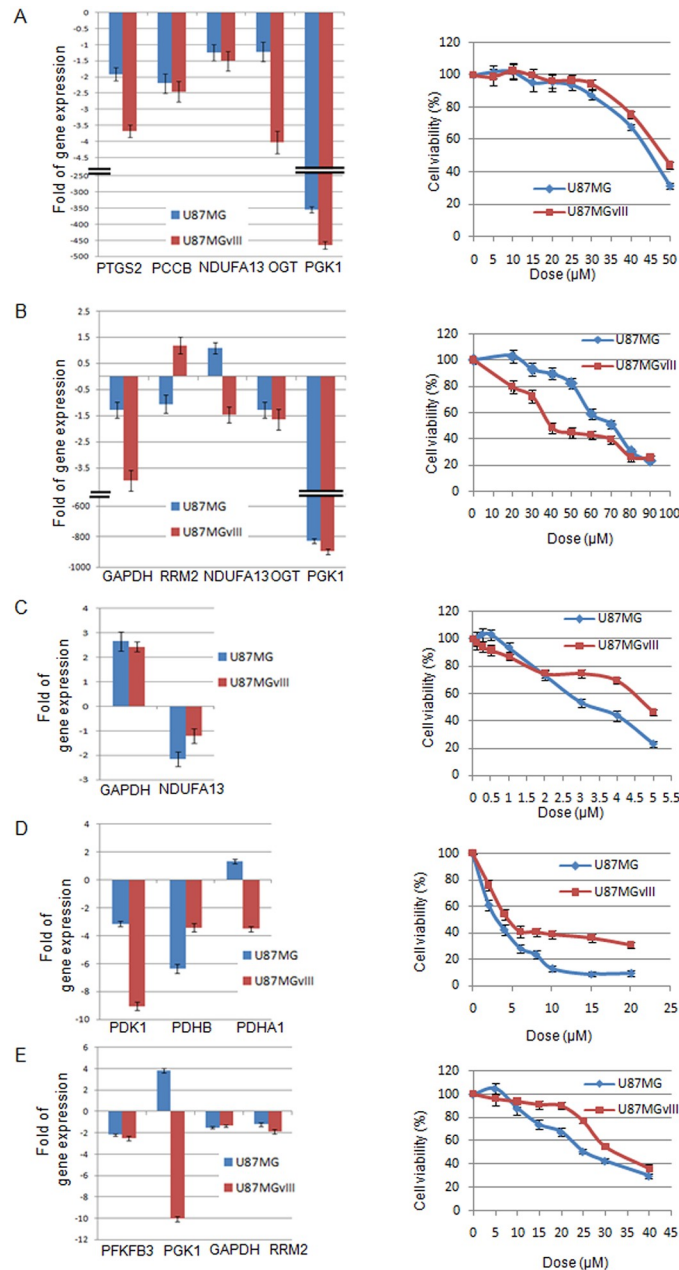




- NOTCH pathway proteins
- Deregulated proteins in GBM
- Protein-protein interactors
- Proteins with significant impact after perturbation
- ◊ Genes/proteins used in experimental validation study

**Fig 8. Enrichment of significant S-M interconnections by successive filtering during the development of the signaling-metabolic pathway network.** The successive reduction of non-significant nodes and paths was followed for the inter-connection of the NOTCH pathway to all metabolic pathways. (A) Starting signaling-metabolic interaction network (SMIN) with the NOTCH pathway proteins in purple; (C) to GBM-specific network, and (E) the final GBM-specific significant ( $Z \geq 1$ ) network. (B) Connections formed through twelve TOP Perturbation Impact (PI) nodes (red color) for NOTCH to all metabolic in SMIN; (D) connections formed through 11 nodes (in red color; out of 12 genes/proteins found with significant perturbation impact [PI] in GBM-specific network and (F) connections formed through 9 nodes (in red color; out of 12) in the GBM-specific significant ( $Z \geq 1$ ) network.

<https://doi.org/10.1371/journal.pcbi.1007090.g008>



**Fig 9. Test of the GBM-specific significant network model in cell cultures.** The EGFRwt and EGFRvIII-mutated GBM cell lines U87MG and U87MGvIII were incubated with inhibitors for calmodulin (CALM2) with CGS 9343B (A), casein kinase II subunit alpha (CSNK2A1) with Emodin (B), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 (PLCG1) with U73122 (C), tyrosine-protein kinase ABL1 with Dasatinib(D), and B-cell lymphoma 2 (BCL2) with ABT-199 (E). The bar diagrams on the left show the effects of the inhibition of the expression of metabolic pathway proteins predicted to be interconnected to the targeted signaling pathway proteins. The line graphs on the right show the dose-dependent effects of the signaling protein inhibitors on the viability of the cells.

<https://doi.org/10.1371/journal.pcbi.1007090.g009>

The selected targets were calmodulin (CALM2), casein kinase II subunit alpha (CSNK2A1), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 (PLCG1), the tyrosine-protein kinase ABL1 and B-cell lymphoma 2 (BCL2). The signaling-metabolic interconnecting paths including the selected proteins were extracted from the final GBM-specific significant

network using the most stringent cutoff and the connected metabolic proteins linked to the above-mentioned signaling pathway proteins were identified (S1 File). As the translation of oncogenic signaling to metabolic adjustment is via the expression of metabolic enzymes, the changes of expression of the metabolic pathway proteins upon inhibition of the selected signaling proteins with small molecule inhibitors and cell viability were analyzed. The inhibitors were CGS-9343B for calmodulin (Fig 9A), Emodin for CSNK2A1 (Fig 9B), U73122 for PLCG1 (Fig 9C), Dasatinib for ABL1 (Fig 9D), and ABT199 for BCL2 (Fig 9E). U87MG (EGFRwt) and U87MGvIII (EGFRvIII) cells were incubated with the inhibitors and their effects on the expression of the predicted interconnected metabolic pathway proteins analyzed by quantitative RT-PCR, and the viability of the cells tested in MTT assays. The blockade of the signaling pathway proteins had significant impacts on the expression of the metabolic pathway proteins (Fig 9). In the case of the calmodulin inhibitor, the expression levels were reduced in both cell lines but for prostaglandin-endoperoxide synthase 2 (PTGS2) and *O*-linked  $\beta$ -*N*-acetylglucosamine transferase (OGT) much more pronounced in the EGFR-mutant cell line than in the wild type. The strongest effect was seen for phosphoglycerate kinase 1 (PGK1) with around 400-fold reduction of expression in both cell lines. The effects of the CSNK2A1 inhibitor are more differentiated. While the OGT and PGK1 expression is inhibited uniformly in both cell lines, 800-fold in case of PGK1, the inhibition of glyceraldehyde 3-phosphate dehydrogenase (GAPDH) is much stronger in U87MGvIII than in the wild type, the expression of ribonucleoside-diphosphatereductase 2 (RRM2) is decreased in the wild-type and enhanced in the mutant, and that of NADH dehydrogenase [ubiquinone] 1 alpha sub-complex subunit 13 (NDUFA13) is affected inversely. Inhibition of phospholipase C gamma 1 (PLCG1) results in enhanced expression of GAPDH and reduced expression of NDUFA13. The ABL1 inhibitor reduced the expression of pyruvate dehydrogenase kinase subunit 1 (PDK1) in both cell lines but much more in the mutant cells. In contrast, the expression of pyruvate dehydrogenase (lipoamide) beta (PDHB) is more reduced in the wild type, and that of pyruvate dehydrogenase subunit alpha 1 (PDHA1) is slightly enhanced in the wild-type and reduced in the mutant cells. Finally, the BCL-2 inhibitor reduces the expression of 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3), GAPDH and RRM2 2-fold in both cell lines but enhances the expression of PGK1 in the wild-type 4-fold while reducing it in the mutant cells 10-fold.

All signaling pathway inhibitors have pronounced negative effects on the viability of both GBM tumor cell lines (Fig 9) but U87MGvIII is significantly more sensitive to the inhibition of casein kinase than the wild-type whereas the wild-type is more sensitive to the inhibition of PLCG1, ABL1 and BCL2. Half-maximal reduction of viability is seen for the PLCG1 and the ABL1 inhibitors in the single-digit, for the others in the double-digit micromolar range.

Further, the effect of the signaling molecule inhibitors on cell migration and invasion was tested. The results showed the similar negative effects of the inhibitors with a higher sensitivity of U87MGvIII, which reconfirms the potency of the signaling-metabolic interconnected network model (S7 Fig).

## Discussion

In all cellular condition, signaling, gene expression, and metabolic pathways must be coordinated to maintain cellular integrity and functions [74]. In this study, we propose an integrative systems biology model of information flow between signaling and metabolic pathways that use an interconnected scaffold based on the complete set of human interactome (HPPIN). The model successfully detects probable unique connections of genes/proteins involved in the interconnection of different signaling and metabolic pathways. The identified components of the network (i.e. genes/proteins as nodes and connections through edges) represent known

physical and/or functional associations between proteins/genes. Depending on the oncogenic signaling, the interconnectors involved modulating the metabolic pathways change. We have successfully applied the model to identify the interconnections altered in the constitutive signaling of the mutated EGFR in glioblastoma multiforme (GBM) compared to EGF-dependent and wild-type EGFR (EGFRwt) signaling.

So far, the development of integrated models including all three or any two of signaling pathways, gene regulation and metabolic pathway in any cellular context is still very restricted. Difficulties of integration arise from the mode of information flow in these three processes significantly involving activation/inactivation, inhibition/induction and substrate/product respectively, as well as different extents and time frames or kinetics [74]. Attempts to integrate metabolism and gene regulatory networks were based on regulatory flux balance analysis (rFBA) [75,76], steady-state regulatory FBA (SR-FBA) [77], probabilistic regulation of metabolism (PROM) [78], integrative omics-metabolic analysis (IOMA) [79] and ordinary differential equation (ODE) [80]. There is a straightforward interconnection between signaling and gene regulation through transcription factor (TF) but the little informative interface and the dynamics of cellular localization of TF affect the integrative modeling. A few studies addressed the issues using the integrative logical, influence graph, Boolean and thermodynamics models [81–86]. Attempts to integrate signaling and metabolism is rarer because information in the two process areas flows in different ways via activation/inactivation and substrate/product interactions respectively. The few reports that address this issue are based on ODE, combined dynamic ODE and Boolean models [33, 80, 87]. Here we have taken a novel approach and integrated signaling and metabolic pathways by i) creating a weighted PPI network depending on the path of information flow from signaling pathway molecule to metabolic enzymes, and ii) applying the probabilistic framework of Markov processes to calculate information flow scores. We have used two major functions of the Markov model. First, to estimate the probability of a protein to be present at a particular point in the network to transmit information, transition functions (node weights) were implemented. Second, to estimate the connection strength between two proteins, emission functions (edge weights) were introduced. The two functions utilize different biological parameters to explore the integrated network and to calculate  $S\text{-}M\ Path_{score}$ . In contrast, to other models for integration of signaling and metabolism pathways, the comparative proteomic expression of the signaling molecule, metabolic enzymes and interactors are effectively incorporated to construct a weighted PPI network (one criterion of node weights) and to filter out context-dependent (mutated vs. wild-type EGFR GBM) S-M interconnections. The second attempt toward integration of the two pathway types to make modeling information flow more accurate and render the model context-dependent based on patient-related disease-specific parameter, transcriptomic expression value from GBM patients were implemented for each node and, by using the principle of mass action, edge weight were defined. The implementation of these two layers of information distinguishes the model from others by integrating weighted network properties like network entropy ( $S_i$ ) and effect-on-node ( $eff_s$ ). We believe that these have the potential to capture context-dependent properties of PPI networks in terms of information flow [88]. Genes/proteins having a high impact on the path formation are important in their local network as well as the global context. Unlike considering conventional network centralities (hubs, bottlenecks etc.) of nodes, biological state and expression profiles are used.

Integrating proteome and transcriptome data in the development of the model helps to mitigate problems in translating high-throughput omics into systems biology models. While proteome data more closely related to the biomedical and pathological systems properties of interest, they are less complete and more prone to variation and experimental errors than transcriptome data. The use of proteome data from cell lines helps to some extent to overcome the

experimental problems, however, cell lines are selected cell culture-adapted models and may be quite remote from *in vivo* situations. Transcriptome data from clinical specimens are getting increasingly available and may add a close-to-clinic dimension to the systems biology model, and help to overcome the missing-data problem of proteomics. Furthermore, the two omics data sets were used at different steps of the network establishment and for different biological relevance. The proteome data of the EGFRwt and EGFRvIII-mutated cell lines were used first to filter out the GBM-specific network from signaling-metabolic interaction network (SMIN). On that GBM-specific network, we have incorporated the human GBM patient transcriptomic data as edge weights to make it more relevant to human GBM condition.

The resulting model is probabilistic, based on the assumption that the likelihood of signal flow relates to the expression levels of the nodes/proteins in the different paths/pathways. It is not computing the actual signaling status of the paths and their nodes. Incorporation of phosphoproteome data in this integrated network could provide additional information about the activation/deactivation state of the nodes thus adding elements of the actual information flow. This could help to add some rationale weights for nodes that remained unchanged (NC) or were unidentified (NA) in the proteomic study as well as in GBM patient transcriptome data. Accordingly, it could add to solving the missing-data-problem of proteomics. Although phosphoproteome data could provide relevant information about signaling flow it could be less advantageous in the model of interconnection between the oncogenic signaling and metabolic pathways that are required to sustain the oncogenic processes. It could make the model more complex as expression and phosphorylation of proteins are not correlated directly; phosphorylation states are highly context- and time-dependent and requires to take into account the ratio of phosphorylated and non-phosphorylated signaling proteins at any given position within the cell and the spatial arrangements of active/inactive signaling pathway molecules.

To test the network properties, the dynamic behavior and the robustness of the model, we performed *in silico* perturbation experiments. The *in silico* perturbation experiments were done with the final GBM-specific model with all interactome, proteome and transcriptome data incorporated in this order to render a general interactome map GBM-specific, and calculate node/edge weights and path score. *In silico* perturbation was done by silencing the signaling molecules one by one, and then redeveloping the model and recalculating node/edge weights and path scores to determine the relevance of the respective signaling molecule and the affected signaling path, to identify possible escape paths bypassing the blockade, and to identify targets for experimental validation of the model. Based on the perturbation analysis we postulated that proteins with higher impact as loss-of-function or gain-of-function or *Path-score* difference between signaling-metabolic pathways and *vice-versa* have a higher ranking and chances to transmit information between the studied pair of pathways. The perturbation studies also captured the weighted topological changes in the dynamic network and the distinct sets of interactors that transmit information between proteins. Implementation of path score differences before and after perturbation adds additional accuracy to the model to identify the most important paths of information flow among all probable paths. Applying this model from 437,449 probable GBM-specific S-M paths, we identified 2,253 ( $Z \geq 1$ ) (0.52%) and 876 ( $Z \geq 3$ ) (0.20%) significant paths of information flow in mutated EGFR-dependent compared to wtEGFR EGF-driven GBM. To experimentally test and validate the interconnections predicted by the model, inhibition of signaling pathway proteins (S) present in the SMIN with small molecule inhibitor was performed, which resulted in alteration of metabolic pathway protein expression (M). We found that the performance of the model in predicting the dynamics of the large-scale signaling networks is comparable to state-of-art methods for extracting context-dependent information flow [89]. One node knockout at a time in the *in silico*

perturbation experiments indicated, first, the importance of that protein in the signaling-metabolic interconnected paths in establishing low or high grade GBM condition, and second, alternative signaling routes that can become relevant as escape responses to therapy.

It is important to note that in biological networks including signal transduction, PPI is highly dynamic in nature and may undergo continuous changes [90]. These context-specific network dynamics need to be predicted by systems biology models. Our probabilistic integrated network model captures network dynamics using the topology of PPI networks based on dynamic data under different circumstances. The probabilistic network-based dynamic model uses experimental data and deals with the uncertainty of systems to predict drug targets and understand the effect of therapeutics. In our study, nodes forming paths significant to information flow in a particular biological context is the key to network inference. Nodes present in information flow path with high network weights and connected with higher edge weight can be considered for drug intervention and combinatorial therapy. Our model efficiently identified key nodes which are important in S-M information flow with confirmation of some known and prediction of the potential novel drug targets [89] that can be used alone or in combination to inhibit mutated EGFR-mediated GBM.

This model is now a useful tool and will be used to develop strategies and experiments to study causal relationships. Among the questions to be addressed will be to understand the molecular biology of high-grade GBM versus low-grade GBM, and mechanisms of therapy resistance. However, in its current form, this model is not engineered to measure signal flow in and across signaling pathway(s). It is also not applicable to other regulatory mechanisms such as transcriptional, post-translational and miRNA based modulations of biomolecular interactions. Nonetheless, the model should be robust enough to integrate and utilize large-scale pan-omics data including genomics, transcriptomics, proteomics, metabolomics and post-translational modification information to accurately model the cellular context under a given scenario. These issues are important for assessing systemic changes and we are addressing them in follow-up studies. The strategy for developing the model can also be applied to other cancers and to non-oncological conditions.

Interestingly, GAPDH and PGK1, commonly reported reference [91] and housekeeping genes [92], were identified (Fig 7B) as significant (~40% of 20 networks in the study) by our model and previously reported studies [47] have shown their association with cancer [93–95]. In this study, also significant expression change in two cell lines for GAPDH and PGK1 was found after inhibition of CSNK2A1 and BCL2, respectively.

To our knowledge, this has not been reported before. A study [96] with a mice model has shown increased expression level of *bcl-2* under the control of *pgk-1*, which suggests a possible relation to our findings. Not surprisingly, we have also found FOXO, a key player in cell fate decision [97] as candidate target. In addition, the HSP90 family gene (HSP90AA1) and CREBBP were identified as important agents in forming interconnections between multiple signaling/metabolic pathways. Inhibitors of HSP90 (17-AAG) and CREBBP (ICG-001) have recently shown effects in glioblastoma and other cancer models/cell lines [98–103]. They have entered clinical trials for several cancer types [104, Clinical trial numbers: [NCT01606579](#), [NCT01764477](#), [NCT02413853](#)]. This confirms that our model can identify potential therapeutic targets before performing actual drug tests.

The presented network model can be used to explore novel therapeutic strategies against cancer including combination therapies. This work also represents a step towards finding alternative routes/pathways in cancer or other diseases and thereby predicting the potential path of therapy evasion that can be included in the development of new therapies that aim to prevent therapy resistance in cancer and other diseases.

## Materials and methods

### Cell lines and cell culture

The human grade-IV glioblastoma cell line U87MG expressing wild-type EGFR (U87MG) was purchased from ATCC (USA). The U87MG-derived genetically engineered U87MGvIII cell line with exons 2–7 deleted from the EGFR gene (EGFRvIII) cell was a kind gift from Professors Webster K. Cavenee and Frank B. Furnari, Department of Medicine and Cancer Center, University of California at San Diego, La Jolla, USA. The cells were grown in Iscove's Modified Dulbecco's Medium (IMDM; Gibco, Thermo Fisher Scientific Inc., Schwerte, Germany) supplemented with 10% heat-inactivated fetal calf serum (FCS; Biochrome, Berlin, Germany) and 1% penicillin/streptomycin solution (Gibco, Thermo Fisher Scientific Inc., Schwerte, Germany) in a humidified atmosphere with 8% CO<sub>2</sub> at 37°C.

### MTT assay

GBM cells ( $1 \times 10^4$ ) were plated in 96 well plates and treated with inhibitors of Calmodulin (CGS-9343B, Sigma), Casein kinase II subunit alpha (Emodin, Sigma), 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 (U73122, Sigma), Tyrosine-protein kinase ABL1 (Dasatinib, Santacruz) and B-cell lymphoma 2 (ABT199, Santacruz) for 24hr and subsequently incubated with 3-(4, 5-dimethylthiazol-2-yl)-2, 5-diphenyl tetrazolium bromide (MTT, 100 µg/ml DMSO) in fresh culture medium for 3 hr. Optical density (OD) was taken at 550 nm with an ELISA reader (Thermo) as described elsewhere [105].

### Scratch wound migration assay

U87MG and U87MGvIII were plated separately in 6-well plates with >90% confluence. Scratch-wounds were made with a micropipette tip, washed thrice to remove the floating cells and treated separately with ABL1, PLCG1, BCL2, CALM2 and CSNK IIA inhibitors at their IC<sub>50</sub> dose (5µM, 3µM, 25µM, 45µM and 40µM, respectively) in medium and incubating them for the indicated time periods. Images were taken at 0 hrs and 8 hrs. The wound width was measured for untreated and treated groups from at least five different fields of three separate experiments, and percentage wound healing was calculated from the width at 8 hrs versus the initial width at 0 hrs, all using ImageJ software.

### Invasion assay

U87MG and U87MGvIII cells ( $5 \times 10^4$ ) were seeded separately to the upper part of matrigel-coated invasion chamber in serum-free medium (200 µl) and the lower chamber was filled with medium (600 µl) containing respective inhibitors of ABL1 and PLCG1 at their IC<sub>50</sub> doses. The cells on the lower surface of the insert were stained with the crystal violet after 24 hrs and the numbers of invaded cells were counted by inverted light microscopy from at least three different fields of three different experiments.

### Expression analysis by real-time PCR

U87MG and U87MGvIII cells were treated with the indicated inhibitors at half of the IC<sub>50</sub> dose for 24hr. Total mRNA was isolated by RNeasy Mini Kit (QIAGEN) as per manufacturer's instructions. RNA (1.0 µg) was reverse-transcribed to cDNA with the Reverse Transcriptase kit (Promega). For RT-PCR, the cDNA was amplified with specific primers for the mRNA of the indicated metabolic proteins in a Perkin-Elmer DNA thermal cycler. Real-time PCR analysis was performed by mixing cDNA with 2× SYBR green master mix using Roche Applied Science light cycler 480.0 instruments with the software version 1.5.0. Relative quantification of

each target gene was normalized to two housekeeping genes (18s rRNA and  $\beta$ -actin) and expressed as a fold change compared with untreated control using the comparative cycle threshold (CT) method [6].

### Protein preparation

For proteome analysis, total proteins were extracted with SDS-PAGE sample buffer from U87MG and U87MGvIII cells grown to 80% confluency. EGFR expression status of both cell lines was confirmed by western blot analysis. Proteins (100  $\mu$ g) of each cell line were separated by SDS-PAGE (10% acrylamide/0.8% bisacrylamide). After staining with Coomassie blue, the lanes for both cell lysate were sliced into one-mm thick gel pieces. The slices were transferred into 96 well round bottom polypropylene plate and destained with 100  $\mu$ l, 20 mM ammonium carbonate/acetonitrile (60%/40% v/v), dehydrated by adding twice acetonitrile (50  $\mu$ l) and dried under vacuum in a speed vac. The dried gel slides were rehydrated in 20  $\mu$ l trypsin solutions (10 ng/ $\mu$ l) in 30  $\mu$ l  $\text{NH}_4\text{HCO}_3$  (20 mM) for digestion by incubation for 16–18 hr at 37°C. The supernatants were collected in 1.5 ml reaction tubes and the remaining tryptic fragments extracted from the gel slices first with 50  $\mu$ l 50% acetonitrile with 0.1% TFA for 15 min, and then with 5% acetonitrile with 0.1% TFA. The extracts from each gel slice were combined and dried in a speed vac. The peptides were re-dissolved in 12  $\mu$ l 2% acetonitrile with 0.05% TFA.

### LC-MS/MS

For MS/MS, the peptides of the tryptic digests (10  $\mu$ l) were loaded via a pre-column at a flow rate 20  $\mu$ l/min (2% acetonitrile, 0.05%TFA) onto an Acclaim PepMap C18 nano-HPLC column (75 $\mu$ m inner diameter $\times$ 15 cm length; Thermo Fisher Scientific Inc., Schwerte, Germany) using an Ultimate 3000 nano-HPLC system (Dionex, Darmstadt, Germany). The peptides were eluted with a gradient of 5–60% solvent B (0.1% formic acid in 95% acetonitrile/5% water) in solvent A (0.1% formic acid in 1% acetonitrile/water) over 60 min, followed by 60–90% solvent B over 5 min, and 90% solvent B for 5 min at a flow rate of 220 nl/min. The nano HPLC system was directly coupled to a Micro-TOF-Q I mass spectrometer (BrukerDaltonics, Bremen, Germany). Mass spectra were acquired in the m/z range 50–2500 at an acquisition rate of 1.3 per sec. MS/MS spectra were acquired in the data-dependent mode with the fragmentation of the 5 most intensive peaks (absolute threshold 3000) with argon as collision gas, and fragment masses in the range of 400–1400 m/z. The mass spectrometry was run with dynamic exclusion of a time interval established and tested beforehand to avoid or minimize signal suppression by over-abundant peptides. A dynamic exclusion of 1 min was set to avoid repeated fragmentation of the most abundant peptides.

### MS/MS data analysis

The MS and MS/MS spectra were processed with Data Analysis 3.4 and Biotoools 3.1 (Bruker-Daltonics). MS/MS searches for peptide identification were done via Biotoools on a local Mascot (version 2.2) server with a precursor mass tolerance of 50 ppm and a fragment mass tolerance of 0.2 Da. Trypsin was specified as an enzyme, and one allowed missed cleavage and oxidation of methionine as variable modification. Data searches were done in SwissProt database for human proteins. Proteins were identified with at least two peptides with a mascot score of higher than 26. The false discovery rate with these search and filter parameters was below 5% as confirmed with a decoy database. Exponentially modified protein abundance index (emPAI) was employed for protein quantification using the equation

$$emPAI = 10^{\frac{N_{observed}}{N_{observable}}} - 1 \quad (i)$$



where  $N_{\text{observed}}$  is the number of experimentally observed and  $N_{\text{observable}}$  the calculated number of observable peptides for each protein [106]. Comparative proteome analysis was done with proteins identified by at least two peptides based on the emPAI scores for differentially expressed proteins. The proteome analysis was done three times for each cell line as an independent biological replicates.

### Comparative proteomics of U87MG and U87MGvIII

Comparative proteome analysis was done for each proteome dataset (batch) of U87MG with each batch of U87MGvIII using the in-house PROTEOMESTAT-12 software. For this, the emPAI values of all proteins were normalized as

$$emPAI_{norm} = \frac{emPAI_{protein}}{\left[ \sum emPAI_{protein} \right]_{batch}} \times 100 \quad (ii)$$

where  $emPAI_{protein}$  was the emPAI value of a protein in a batch.

Multiple identifications of a protein within a proteome dataset were excluded, only keeping the highest emPAI value for subsequent analysis. Intra-batch analysis was performed to determine the total unique proteins as well as the overlap between the three batches of each cell line. Total unique proteins identified from U87MG and U87MGvIII cells were compared to identify the common and the differentially expressed proteins. Proteins identified only for one of the cell lines were considered up-regulated in that cell line. Commonly expressed proteins were further analyzed in two different ways to identify differentially expressed proteins. In one approach, the  $emPAI_{ave}$  values for each protein were compared between the two cell lines (iii).

$$emPAI_{ave} = \frac{\sum_{batch=1}^3 emPAI_{protein}}{\text{Number of batches with identified protein}} \quad (iii)$$

where  $emPAI_{protein}$  was the emPAI value of a protein in a proteome dataset.

The percentages of up or down-regulation of a protein in U87MGvIII were calculated as %  $Exp_{U87MGvIII}$  using the following the equation

$$\%Exp_{U87MGvIII} = \frac{(emPAI_{ave \ U87MGvIII} - emPAI_{ave \ U87MG})}{emPAI_{ave \ U87MG}} \times 100 \quad (iv)$$

where  $emPAI_{ave \ U87MGvIII}$  and  $emPAI_{ave \ U87MG}$  were the average emPAI value of a protein among the three proteome datasets for U87MGvIII or U87MG. Proteins that were  $\geq 50\%$  up- or down-regulated with p values  $\leq 0.2$  were considered differentially expressed in the U87MGvIII cell.

In a second approach,  $emPAI_{norm}$  of a protein of each batch of one of the cell lines were compared separately with the  $emPAI_{norm}$  for the same protein in each of the three batches of the other cell line. Then these comparative data were normalized by the LOESS method. Up- or down-regulation of a protein in U87MGvIII were calculated using the Eq (iv). From this second approach  $\geq 50\%$  up- or down-regulated proteins with an SE  $\leq 35\%$  were taken as differentially expressed in U87MGvIII. Both the results were combined and duplicates were excluded to identify the differentially expressed proteins.

### Human signaling-metabolic interaction network

A signaling-metabolic cross-connected network was created using 14 signaling and 81 metabolic pathways and protein-protein interactors of the pathway proteins. Signaling pathway databases were created by integrating information from on-line resources including KEGG

[107], Reactome [108] Signallink [109], NetPath [110] and Biocarta (<http://www.biocarta.com>) for extensive coverage of the steps involved in the pathways. The 14 signaling pathways were included based on their association with cancer in general according to KEGG pathway database.

Metabolic pathway datasets were collected from KEGG. Human protein interaction data were taken from the STRING interactome database using only protein-protein interactors [70] established by high-throughput analyses with the high confidence score ( $\geq 0.7$ ) to generate a protein-protein interaction network (PPIN). PPI were restricted up to the second level interactors, i.e. interactors of interactors, which brings about  $S(PPI)_3M$ , meaning  $S-PPI_1-PPI_2-PPI_3-M$  paths.  $PPI_1$  is the first level interactor of signaling molecule and  $PPI_3$  is the first level interactor of metabolic enzyme and  $PPI_2$  the second level interactor of both. Thereby, there are three PPIs between the signaling and the metabolic pathway proteins. The resulting network consisted of 5,703 proteins with 16,828 experimentally confirmed interactions.

### Network centrality

NetworkX (<https://networkx.github.io/>), a network analysis framework in a Python language, was used to calculate the degrees, Estrata Index and betweenness centrality of nodes (genes/proteins).

### GBM gene expression

Gene expression data used for edge weight determination in the weighted-network were collected from five gene expression data sets for GBM patients taken from GEO (Gene Expression Omnibus)[111] GSE4290 (77 samples), GSE53733 (70 samples), GSE50161 (34 samples), GSE36245 (46 samples) and GSE15824 (12 samples) (<http://www.ncbi.nlm.nih.gov/geo/>). All had been acquired with the Affymetrix GPL570 platform [<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL570>]. The raw data of the total of 239 samples were normalized across all arrays within a set using the RMA quintile normalization procedure of the Gene Pattern Expression File Creator module [112]. Averages of normalized raw expression values were used for the study.

### Construction of cross-connected network linking paths

To construct cross-connecting linking paths (CCLPs) between signaling pathway proteins (S) and metabolic pathways proteins (M), common interacting proteins from the HPPIN were used. All possible unique connections between S and M with S-M (direct S-M interaction), and S-P-M, S-P-P-M, and S-P-P-P-M where P is a common interacting protein were established for common interacting proteins up to the second level in HPPIN (Fig 3A). Detailed descriptions of all possible connecting links are provided in the Fig 3B and Table 1. Differentially expressed (i.e. up-regulated or down-regulated) proteins in EGFRvIII versus EGFRwt expressing cells obtained from proteome analyses were mapped onto the cross-connected links [Table 1]. To assess the relevance of signaling to metabolic pathway interconnections for cancer pathogenesis, connections with at least one up-regulated or down-regulated protein (U87MGvIII versus U87MG) were selected to create functional cross-connecting sub-network (CCsN) with, in all, 2,320 protein nodes and 8,914 interactions. This CCsN was used for network topology analysis.

### Calculation of signaling entropy

The local entropy of a protein in CCsN was calculated on the basis of probabilities of an interaction of that protein with its interactors determined by using the principle of mass action.

The calculation of the interaction probabilities is based on the assumption that two proteins known to interact will have a higher probability of interaction when they are highly expressed. Thereby, the interaction probability ( $W_{ij}$ ) of two proteins in a network is proportional to the product of expression values ( $E$  of the corresponding genes ( $E_i$  and  $E_j$ )) [72]:

$$W_{ij} \propto E_i \times E_j \tag{v}$$

with the expression value of CCsN genes calculated as an average expression of a gene in the data set of 239 GBM patients.

Based on  $W_{ij}$ , a stochastic matrix of normalized interaction probabilities ( $P_{ij}$ ) in the network was created for signaling entropy calculation. Probability of interaction between node  $i$  and node  $j$  was calculated as

$$P_{ij} = \frac{W_{ij}}{\sum_{k \in N_i} W_{ik}} \tag{vi}$$

where  $N_i$  is interactors of node  $i$  with  $\sum P_{ij} = 1$ .

The local entropy of node  $i$  was calculated as

$$S_i = -\frac{1}{\log k_i} \sum_{j \in N(i)} p_{ij} \log p_{ij} \tag{vii}$$

where  $k_i$  is the degree of node  $i$  in the CCsN.

### Effect-on-node as CCsN weight

To incorporate the impact of the interactors of a particular protein in the cross-connected network, the node-weight of every node  $i$  was specified based on the categories signaling cross-talk protein (SC), rate limiting enzyme (RLE) and up- or down-regulation in U87MGvIII versus U87MG (dEXP):

$$W_i = \left\{ 1; \text{if} \left( dEXP = \frac{1}{3}, RLE = \frac{1}{3}, SC = \frac{1}{3} \right), 0; \text{else} \right\} \tag{viii}$$

Effect of interactors on a node  $s$  in CCsN was defined as effect-on-node ( $eff_s$ ) depending on the node weight of the interactors up to the second level:

$$eff_s = \sum_j^k \left( \sum_i^n \frac{W_i}{degree_i} + \frac{W_j}{n_j} \right) \tag{ix}$$

where  $k$  is the degree of node  $s$ ,  $n_j$  is the degree of protein  $j$ ,  $w_j$  and  $w_i$  are the node weights of protein  $j$  and  $i$ .

Rate-limiting enzymes (RLE) are the enzymes in metabolic pathways whose kinetics determines the overall kinetics of the entire pathway. It is usually the enzyme with the slowest enzyme kinetics. The identification of the RLE is done based on the Michaelis–Menten equation, and the established Michaelis constants ( $K_M$ ) and  $V_{max}$  of all enzymatic steps in metabolic pathways are listed in the respective biophysics and enzyme kinetic textbooks and databases.

### Calculation of CCLP scores

The CCLP for information flow from signaling pathway protein (S) to metabolic pathway protein (M) was defined to be one of the four types shown in Fig 5A, and path scores were calculated on the basis of node and edge weights of the proteins involved in a path. To select

important S-M pairs, imaginary penultimate signaling and metabolic proteins are considered as starting and ending state, and path score was calculated based on a hidden Markov model (HMM) with a forward algorithm. Emission probability ( $E_j$ ), i.e. the positional probability of a protein at the particular position in that state was calculated as

$$E_j = \frac{S_j + Eff_j}{\sum_i^k (S_i + Eff_i)} \quad (x)$$

where  $k$  is the number of proteins in that state,  $S_i$  and  $S_j$  are normalized local entropy of proteins  $i$  and  $j$ ,  $Eff_i$  is effect-on-node for protein  $i$ . Within the S-M pairs of a path (Fig 5B) information flow is again scored by considering all types of paths formed between single S-M pairs and calculated as

$$Path_{score} = I \prod_{j=1}^n E_j T_j \quad (xi)$$

where  $n$  is number states in a path,  $I$  initial probability (in our case it is equal to one),  $E_j$  is emission probability at state  $j$ ,  $T_j$  is transmission probability at state  $j$  (in our case it is the probability of interaction  $P_{ij}$ ).  $Path_{score}$  is converted into Z-score as

$$Z_{score} = \frac{X - \mu}{\sigma} \quad (xii)$$

where  $X$  is raw  $Path_{score}$ ,  $\mu$  is mean,  $\sigma$  is the standard deviation. A cut-off of  $\geq 1$  is applied to select significant S-M pairs and their cross-connecting links.

To understand the significance of the path scores in comparison to randomly generated paths, we have performed permutations of node weights ( $S_{ij}$ ) and edge weights ( $P_{ij}$ ) and generated 20 random paths with the same number of proteins involved in the identified paths. Averages of 20 path scores are compared with the corresponding original path score. S8 Fig shows significant difference between the original and the randomly placed weighted path scores.

### In silico perturbation

*In silico* perturbation analysis was done by removing nodes/proteins from the network using an in-house programme for nodes/proteins present in the final weighted sub-network. All nodes in this sub-network were removed individually from the SMIN, and the impact was studied by performing the same weighted network analysis with the sub-networks generated after node removal as the starting point. Perturbation score ( $P_s$ ) was calculated in two steps. First, we calculated significant pairs using Model 1 shown in Fig 5A and significant paths using Model 2 shown in Fig 5B using SIN- $N_i$  network where  $N_i$  is the interaction of perturbed node  $N$ . Second, path scores after perturbation ( $Path_{score}'$ ) for significant paths identified from step 1 were calculated. Perturbation score ( $P_s$ ) was defined as

$$P_s = Path_{score} - Path_{score}'$$

To identify nodes with significant impact after perturbation we converted perturbation scores ( $P_s$ ) into Z-score for all signaling to metabolic pathways and vice-versa. In the perturbation analysis, node and edge weights were re-calculated for networks and paths generated after removing a node (protein) from the initial interactome. New weights in Model 1 and Model 2 were used to score the new paths and to compare the scores before and after perturbation thus to identify significant nodes in the network forming connections between signaling and metabolic pathway proteins.

## Functional enrichment by over-representation analysis (ORA)

KEGG pathway based over-representation analysis (ORA) was performed using 136 up-regulated and 243 down-regulated proteins (EGFRvIII vs EGFRwt) and 45 commonly up-regulated and 25 down-regulated proteins using ‘protein coding gene set’ as the reference gene set in WebGestalt [113] web tool.

Additionally, Gene ontology (GO) based biological process and molecular function over-representation analysis was performed for the same genes. Top 20 significant ( $p$ -values  $\leq 0.01$  and  $\leq 0.05$ ) categories were ranked based on the false detection rate (FDR) calculated using Benjamini and Hochberg procedure.

## Supporting information

**S1 Fig. Degree and betweenness distribution of HPPIN and SMIN in comparison with random networks of equivalent sizes.** (A) The degree of distribution (left) and betweenness distribution (right) of the human protein-protein interaction network (HPPIN; red color) and ten random networks (created using Erdős–Rényi model) in 10 different colors. (B) The degree of distribution (left) and betweenness distribution (right) of the signaling-metabolic interaction network (SMIN; red color) and ten random networks (created using Erdős–Rényi model) in 10 different colors. The number of proteins with a given degree ( $k$ ) in the network (in this figure for representation, 20 is set as the upper limit for  $k$ ) approximates a power-law. HPPIN follows power-law degree distribution and is a scale-free network. The betweenness centrality of a node quantifies the communication flowing through the network from node(s)/protein(s) to another using the shortest path.

(TIF)

**S2 Fig. Outline of label-free comparative proteomics by SDS-PAGE plus nano-HPLC ESI-Q-TOF MS/MS.**

(TIF)

**S3 Fig. Gene ontology (GO) based over-representation analysis.** Panels A and B show the enriched GO biological processes, and Panels C and D the enriched molecular functions for proteins exclusively overexpressed in U87MGvIII (EGFRvIII) and U87MG (EGFRwt), respectively.

(TIF)

**S4 Fig. Distribution of weighted nodes and expression states of interconnected SM pairs in the step-wise filtered networks.** (A) The number of differentially expressed proteins (dEXP), signal crosstalk proteins (SC) and rate-limiting enzymes (RLE) present in the signaling-metabolic cross-connected network, and the overlap between each group. (B) The numbers of signaling-metabolic (SM) pairs present in signaling-metabolic interaction network (SMIN), the GBM-specific network and the GBM-specific significant network (threshold cut off  $Z \geq 1$  and  $Z \geq 3$ ), and their expression states in an EGFR-mutated U87MGvIII cell line in comparison to EGFR<sup>wt</sup> U87MG. The expression states are indicated as NC: no change, NA: not identified, UP: up-regulated, DOWN: down-regulated.

(TIF)

**S5 Fig. Overlapping distribution of SM pairs, S-M path and genes of five selected signaling-metabolic interconnection.** (A) Distribution of pathway-specific genes of five selected signaling pathways (left) and metabolic pathways (right). (B) Distribution of SM pairs of the selected signaling to all metabolic pathways (left) and all signaling to five selected groups of a metabolic pathway (right). (C) Distribution of the genes involved in the significant signaling-

metabolic cross-connected paths of five selected signaling to all metabolic pathways (left) and all signaling to five individual groups of a metabolic pathway (right).

(TIF)

**S6 Fig. RAS pathway proteins with metabolic pathways interconnections.** RAS pathway proteins with their expression states and inter-connections with metabolic pathways via protein-protein interactors in EGFRvIII-mutated GBM mapped onto the GBM-specific significant network.

(TIF)

**S7 Fig. Effect of inhibitors of signaling molecule on U87MG and U87MGvIII cell migration and invasion.** (A) Scratch wound migration assay *in vitro*. Scratch-wounds were made by a micropipette tip in a ~ 90% confluent cell and treated separately with ABL1, PLCG1, BCL2, CALM2 and CSNK IIA inhibitors with their IC50 dose (5  $\mu$ M, 3  $\mu$ M, 25  $\mu$ M, 45  $\mu$ M and 40  $\mu$ M respectively) and incubated with IMDM and 1% FBS and images were taken at 0 hrs and 8 hrs. Comparisons of percent wound healing area of individual treatment group with untreated group are represented as bar diagram for respective cell line. (B) Matrigel coated transwell chamber migration/invasion assay *in vitro*. U87MG and U87MGvIII cells were seeded to the matrigel-coated upper invasion chamber in serum-free medium and lower chamber was filled with medium with ABL1 and PLCG1 inhibitors at their IC50 doses. The cells on the lower surface of the insert were stained with the crystal violet after 24 hrs. Comparisons of number of invaded cells between untreated and treated conditions are presented as bar diagram.

(TIF)

**S8 Fig. Comparison of path scores between the original and randomly assigned weighted paths.** Random node weights ( $S_{ij}$ ) and edge weights ( $P_{ij}$ ) were applied to the path forming proteins to generate 20 random paths for each original path. Scores were compared between the original and the random paths for S-P-M (A), S-P-P-M (B), and S-P-P-P-M (C) connections, respectively. For panels A and B, path scores are in the range of  $10^{-17}$ .

(TIF)

**S1 Table. Network properties of HPPIN and SMIN with 10 random networks of equivalent size.** Clustering coefficient (CC): It is a measure of the degree to which nodes in a graph tend to cluster. Assortivity: It represents to what extent nodes in a network associate with other nodes in the network, being of similar sort or of opposing sort. Estrada Index: It is a measure of the robustness of complex networks by Eigen values and Eigen vectors.

(DOCX)

**S2 Table. Individual signaling or metabolic pathway-wise statistics of a number of paths and pairs formed in different networks.** The number of paths formed between signaling-metabolic pairs for information flow from 14 signaling pathways to all metabolic pathways (magenta) and from all signaling pathways to six groups of metabolic pathways (blue) in signaling-metabolic interaction network (SMIN) and significant GBM-specific network.

(DOCX)

**S3 Table. Identification of important network proteins from *in silico* perturbation studies.** Number of significant ( $-2 \geq Z \geq 2$ ) proteins identified from *in silico* perturbation analysis to impact information flow from 14 analyzed signaling pathways to all metabolic pathways and from all signaling pathways to 6 metabolic pathways. In  $X|Y$ , X is the number of common proteins in the indicated pathway and Y is total number of significant proteins.

(DOCX)

**S1 File. S-MPaths formed through signaling molecules used for inhibition study as shown in Fig 9.**  
(XLSX)

## Author Contributions

**Conceptualization:** Arup K. Bag, Sapan Mandloi, Saulius Jarmalavicius, Peter Walden, Saikat Chakrabarti, Chitra Mandal.

**Data curation:** Arup K. Bag, Sapan Mandloi, Saulius Jarmalavicius, Chhabinath Mandal.

**Formal analysis:** Sapan Mandloi, Krishna Kumar, Saikat Chakrabarti.

**Funding acquisition:** Peter Walden, Chitra Mandal.

**Investigation:** Arup K. Bag, Sapan Mandloi, Susmita Mondal, Peter Walden, Saikat Chakrabarti.

**Methodology:** Arup K. Bag, Sapan Mandloi, Saulius Jarmalavicius, Krishna Kumar.

**Project administration:** Peter Walden, Saikat Chakrabarti, Chitra Mandal.

**Resources:** Peter Walden.

**Software:** Sapan Mandloi, Chhabinath Mandal, Saikat Chakrabarti.

**Supervision:** Saulius Jarmalavicius, Chhabinath Mandal, Peter Walden, Saikat Chakrabarti, Chitra Mandal.

**Validation:** Arup K. Bag, Susmita Mondal.

**Visualization:** Arup K. Bag, Sapan Mandloi, Saulius Jarmalavicius.

**Writing – original draft:** Arup K. Bag, Sapan Mandloi.

**Writing – review & editing:** Arup K. Bag, Sapan Mandloi, Saulius Jarmalavicius, Chhabinath Mandal, Peter Walden, Saikat Chakrabarti, Chitra Mandal.

## References

1. Laubenbacher R, Hower V, Jarrah A, Torti SV, Shulaev V, Mendes P, Torti FM, Akman S. A systems biology view of cancer. *BiochimBiophysActa*. 2009; 1796(2): 129–139. <https://doi.org/10.1016/j.bbcan.2009.06.001> PMID: 19505535
2. Cheng TM, Goehring L, Jeffery L, Lu YE, Hayles J, Novak B, Bates PA. A structural systems biology approach for quantifying the systemic consequences of missense mutations in proteins. *PLoSComputBiol*. 2012; 8(10): e1002738. <https://doi.org/10.1371/journal.pcbi.1002738> PMID: 23093928
3. Gates AJ, Rocha LM. Control of complex networks requires both structure and dynamics. *Sci Rep*. 2016; 6: 24456. <https://doi.org/10.1038/srep24456> PMID: 27087469
4. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. *Nature*. 1999; 402: C47–52. <https://doi.org/10.1038/35011540> PMID: 10591225
5. Zanzoni A, Soler-Lopez M, Aloy P. A network medicine approach to human disease. *FEBS Lett*. 2009; 583(11): 1759–1765. <https://doi.org/10.1016/j.febslet.2009.03.001> PMID: 19269289
6. Bhattacharya K, Maiti S, Mandal C. Phosphatase and tensin homolog protein inhibits mTORC2 formation and signaling in grade IV glioma by stimulating Rictorhyperphosphorylation at Thr1135. *Oncogenesis*. 2016; 5(5): e227.
7. Maiti S, Mondal S, Satyavarapu EM and Mandal C. mTORC2 regulates hedgehog pathway activity by promoting stability to Gli2 protein and its nuclear translocation Cell death and Disease. 2017; 8(7), e2926; <https://doi.org/10.1038/cddis.2017.296> PMID: 28703798
8. Korkut A, Wang W, Demir E, Aksoy BA, Jing X, Molinelli EJ, Babur O, Bemis DL, Onur Sumer S, Solit DB, Pratilas CA, Sander C. Perturbation biology nominates upstream-downstream drug combinations

- in RAF inhibitor resistant melanoma cells. *Elife*. 2015; 4. <https://doi.org/10.7554/eLife.04640> PMID: 26284497
9. Wang E. *Cancer systems biology*, Boca Raton: CRC Press. 2010.
  10. Scharm M, Wolkenhauer O, Waltemath D. An algorithm to detect and communicate the differences in computational models describing biological systems. *Bioinformatics*. 2016; 32 (4): 563–570. <https://doi.org/10.1093/bioinformatics/btv484> PMID: 26490504
  11. Kitano H. *Computational systems biology*. *Nature*. 2002; 420(6912): 206–210. <https://doi.org/10.1038/nature01254> PMID: 12432404
  12. Wang E, Lenferink A, O'Connor-McCourt M. *Cancer systems biology: exploring cancer-associated genes on cellular networks*. *Cell Mol Life Sci*. 2007 64(14):, 1752–1762. <https://doi.org/10.1007/s00018-007-7054-6> PMID: 17415519
  13. Baker SG, Kramer BS. *Systems biology and cancer: promises and perils*. *ProgBiophysMolBiol*. 2011; 106(2): 410–413. <https://doi.org/10.1016/j.pbiomolbio.2011.03.002> PMID: 21419159
  14. Wolkenhauer O. Why model? *Front Physiol*. 2014; 5: 21. <https://doi.org/10.3389/fphys.2014.00021> PMID: 24478728
  15. Le Novere N. Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet*. 2015; 16(3): 146–158. <https://doi.org/10.1038/nrg3885> PMID: 25645874
  16. Henkel R, Hoehndorf R, Kacprowski T, Knupfer C, Liebermeister W, Waltemath D. Notions of similarity for systems biology models. *Brief Bioinform*. 2017; 18(5): 902. <https://doi.org/10.1093/bib/bbw146> PMID: 28096074
  17. Roychowdhury A, Samadder S, Das P, Mandloi S, Addya S, Chakraborty C, Basu PS, Mondal R, Roy A, Chakrabarti S, Roychoudhury S, Panda CK. Integrative genomic and network analysis identified novel genes associated with the development of advanced cervical squamous cell carcinoma. *BiochimBiophysActa*. 2017; 1861: 2899–2911. <https://doi.org/10.1016/j.bbagen.2016.09.014> PMID: 27641506
  18. Alam SK, Yadav VK, Bajaj S, Datta A, Dutta SK, Bhattacharyya M, Bhattacharya S, Debnath S, Roy S, Boardman LA, Smyrk TC, Molina JR, Chakrabarti S, Chowdhury S, Mukhopadhyay D, Roychoudhury S. DNA damage-induced ephrin-B2 reverse signaling promotes chemoresistance and drives EMT in colorectal carcinoma harboring mutant p53. *Cell Death Differ*. 2016; 23(4): 707–722. <https://doi.org/10.1038/cdd.2015.133> PMID: 26494468
  19. Ghosh RD, Ghuwalewala S, Das P, Mandloi S, Alam SK, Chakraborty J, Sarkar S, Chakrabarti S, Panda CK, Roychoudhury S. MicroRNA profiling of cisplatin-resistant oral squamous cell carcinoma cell lines enriched with cancer-stem-cell-like and epithelial-mesenchymal transition-type features. *Sci Rep*. 2016; 6: 23932. <https://doi.org/10.1038/srep23932> PMID: 27045798
  20. Mandloi S, Chakrabarti S. PALM-IST: Pathway Assembly from Literature Mining—an Information Search Tool. *Sci Rep*. 2015; 5: 10021. <https://doi.org/10.1038/srep10021> PMID: 25989388
  21. Puniya BL, Allen L, Hochfelder C, Majumder M, Helikar T. Systems Perturbation Analysis of a Large-Scale Signal Transduction Model Reveals Potentially Influential Candidates for Cancer Therapeutics. *Front Bioeng Biotechnol*. 2016; 4: 10. <https://doi.org/10.3389/fbioe.2016.00010> PMID: 26904540
  22. Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA. Logic-based models for the analysis of cell signaling networks. *Biochemistry*. 2010; 49(15): 3216–3224. <https://doi.org/10.1021/bi902202q> PMID: 20225868
  23. Blair RH, Trichler DL, Gaille DP. Mathematical and statistical modeling in cancer systems biology. *Front Physiol*. 2012; 3: 227. <https://doi.org/10.3389/fphys.2012.00227> PMID: 22754537
  24. Samaga R, Klamt S. Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Commun Signal*. 2013; 11(1): 43. <https://doi.org/10.1186/1478-811X-11-43> PMID: 23803171
  25. Albert R, Thakar J. Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions. *WileyInterdiscip Rev SystBiol Med*. 2014; 6(5): 353–369. PMID: 25269159
  26. Naldi A, Monteiro PT, Mussel C, Consortium for Logical M, Tools, Kestler HA, Thieffry D, Xenarios I, Saez-Rodriguez J, Helikar T, Chaouiya C. Cooperative development of logical modelling standards and tools with CoLoMoTo. *Bioinformatics*. 2015; 31(7): 1154–1159. <https://doi.org/10.1093/bioinformatics/btv013> PMID: 25619997
  27. Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger DA, Sorger PK. Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res*. 2011; 71(16): 5400–5411. <https://doi.org/10.1158/0008-5472.CAN-10-4453> PMID: 21742771



28. MacNamara A, Terfve C, Henriques D, Bernabe BP, Saez-Rodriguez J. State-time spectrum of signal transduction logic models. *Phys Biol*. 2012; 9(4): 045003. <https://doi.org/10.1088/1478-3975/9/4/045003> PMID: 22871648
29. Giacomantonio CE, Goodhill GJ. A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoSComputBiol*. 2010; 6(9). <https://doi.org/10.1371/journal.pcbi.1000936> PMID: 20862356
30. Wang RS, Saadatpour A, Albert R. Boolean modeling in systems biology: an overview of methodology and applications. *Phys Biol*. 2012; 9(5): 055001. <https://doi.org/10.1088/1478-3975/9/5/055001> PMID: 23011283
31. Li C, Donizelli M, Rodríguez N, Dharuri H, Endler L, Chelliah V, Li L, He E, Henry A, Stefan MI, Snoep JL, Hucka M, Le Novère N, Laibe C. BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC SystBiol*. 2010; 4: 92. <https://doi.org/10.1186/1752-0509-4-92> PMID: 20587024
32. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat Biotechnol*. 2002; 20(4): 370–375. <https://doi.org/10.1038/nbt0402-370> PMID: 11923843
33. König M, Bulik S, Holzhütter HG. Quantifying the contribution of the liver to glucose homeostasis: a detailed kinetic model of human hepatic glucose metabolism. *PLoSComputBiol*. 2012; 8(6): e1002577. <https://doi.org/10.1371/journal.pcbi.1002577> PMID: 22761565
34. Hautaniemi S, Kharait S, Iwabu A, Wells A, Lauffenburger DA. Modeling of signal-response cascades using decision tree analysis. *Bioinformatics*. 2005; 21(9): 2027–2035. <https://doi.org/10.1093/bioinformatics/bti278> PMID: 15657095
35. Hughey JJ, Lee TK, Covert MW. Computational modeling of mammalian signaling networks. *WileyInterdiscip Rev SystBiol Med*. 2010; 2(2): 194–209. <https://doi.org/10.1002/wsbm.52> PMID: 20836022
36. Kholodenko B, Yaffe MB, Kolch W. Computational approaches for analyzing information flow in biological networks. *Sci Signal*. 2012; 5(220): re1. <https://doi.org/10.1126/scisignal.2002961> PMID: 22510471
37. Wang DY, Cardelli L, Phillips A, Piterman N, Fisher J. Computational modeling of the EGFR network elucidates control mechanisms regulating signal dynamics. *BMC SystBiol*. 2009; 3: 118. <https://doi.org/10.1186/1752-0509-3-118> PMID: 20028552
38. Chaouiya C. Petri net modelling of biological networks. *Brief Bioinform*. 2007; 8(4): 210–219. <https://doi.org/10.1093/bib/bbm029> PMID: 17626066
39. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol*. 2006; 8(11): 1195–1203. <https://doi.org/10.1038/ncb1497> PMID: 17060902
40. Steffen M, Petti A, Aach J, D'Haeseleer P, Church G. Automated modelling of signal transduction networks. *BMC Bioinformatics*. 2002; 3: 34. PMID: 12413400
41. Christopher R, Dhiman A, Fox J, Gendelman R, Haberichter T, Kagle D, Spizz G, Khalil IG, Hill C. Data-driven computer simulation of human cancer cell. *Ann N Y Acad Sci*. 2004; 1020: 132–153. <https://doi.org/10.1196/annals.1310.014> PMID: 15208190
42. Kumar N, Hendriks BS, Janes KA, de Graaf D, Lauffenburger DA. Applying computational modeling to drug discovery and development. *Drug Discov Today*. 2006; 11(17–18): 806–811. <https://doi.org/10.1016/j.drudis.2006.07.010> PMID: 16935748
43. Cho CR, Labow M, Reinhardt M, van Oostrum J, Peitsch MC. The application of systems biology to drug discovery. *Curr Opin Chem Biol*. 2006; 10(4): 294–302. <https://doi.org/10.1016/j.cbpa.2006.06.025> PMID: 16822703
44. Clermont G, Auffray C, Moreau Y, Rocke DM, Dalevi D, Dubhashi D, Marshall DR, Raasch P, Dehne F, Provero P, Tegner J, Aronow BJ, Langston MA, Benson M. Bridging the gap between systems biology and medicine. *Genome Med*. 2009; 1(9): 88. <https://doi.org/10.1186/gm88> PMID: 19754960
45. Remy E, Rebouissou S, Chaouiya C, Zinoviyev A, Radvanyi F, Calzone L. A Modeling Approach to Explain Mutually Exclusive and Co-Occurring Genetic Alterations in Bladder Tumorigenesis. *Cancer Res*. 2015; 75(19): 4042–4052. <https://doi.org/10.1158/0008-5472.CAN-15-0602> PMID: 26238783
46. Terfve CD, Wilkes EH, Casado P, Cutillas PR, Saez-Rodriguez J. Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data. *Nat Commun*. 2015; 6: 8033. <https://doi.org/10.1038/ncomms9033> PMID: 26354681
47. Huang SS, Clarke DC, Gosline SJ, Labadorf A, Chouinard CR, Gordon W, Lauffenburger DA, Fraenkel E. Linking proteomic and transcriptional data through the interactome and epigenome reveals a map of oncogene-induced signaling. *PLoSComputBiol*. 2013; 9(2): e1002887. <https://doi.org/10.1371/journal.pcbi.1002887> PMID: 23408876

48. Crespo I, Krishna A, Le Behec A, del Sol A. Predicting missing expression values in gene regulatory networks using a discrete logic modeling optimization guided by network stable states. *Nucleic Acids Res.* 2013; 41(1): e8. <https://doi.org/10.1093/nar/gks785> PMID: 22941654
49. Keller R, Klein M, Thomas M, Drager A, Metzger U, Templin MF, Joos TO, Thasler WE, Zell A, Zanger UM. Coordinating Role of RXRalpha in Downregulating Hepatic Detoxification during Inflammation Revealed by Fuzzy-Logic Modeling. *PLoSComput Biol.* 2016; 12(1): e1004431. <https://doi.org/10.1371/journal.pcbi.1004431> PMID: 26727233
50. Molinelli EJ, Korkut A, Wang W, Miller ML, Gauthier NP, Jing X, Kaushik P, He Q, Mills G, Solit DB, Pratilas CA, Weigt M, Braunstein A, Pagnani A, Zecchina R, Sander C. Perturbation biology: inferring signaling networks in cellular systems. *PLoSComputBiol.* 2013; 9(12): e1003290. <https://doi.org/10.1371/journal.pcbi.1003290> PMID: 24367245
51. Lu S, Zhang F, Chen J, Sze S-H. Finding Pathway Structures in Protein Interaction Networks. *Algorithmica.* 2007; 48: 363–374
52. Bebek G, Yang J. PathFinder: mining signal transduction pathway segments from protein-protein interaction networks. *BMC Bioinformatics.* 2007; 8:335. <https://doi.org/10.1186/1471-2105-8-335> PMID: 17854489
53. Vinayagam A, Stelzl U, Foulle R, Plassmann S, Zenkner M, Timm J, Assmus HE, Andrade-Navarro MA, Wanker EE. A directed protein interaction network for investigating intracellular signal transduction. *Sci Signal.* 2011; 4(189): rs8. <https://doi.org/10.1126/scisignal.2001699> PMID: 21900206
54. Liu W, Li D, Wang J, Xie H, Zhu Y, He F. Proteome-wide prediction of signal flow direction in protein interaction networks based on interacting domains. *Mol Cell Proteomics.* 2009; 8(9): 2063–2070. <https://doi.org/10.1074/mcp.M800354-MCP200> PMID: 19502588
55. Gitter A, Klein-Seetharaman J, Gupta A, Bar-Joseph Z. Discovering pathways by orienting edges in protein interaction networks. *Nucleic Acids Res.* 2011; 39(4): e22. <https://doi.org/10.1093/nar/gkq1207> PMID: 21109539
56. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* 2010; 11(5): R53. <https://doi.org/10.1186/gb-2010-11-5-r53> PMID: 20482850
57. Cerami E, Demir E, Schultz N, Taylor BS, Sander C. Automated network analysis identifies core pathways in glioblastoma. *PLoS One.* 2010; 5(2): e8918. <https://doi.org/10.1371/journal.pone.0008918> PMID: 20169195
58. Friedmann-Morvinski D, Bhargava V, Gupta S, Verma IM, Subramaniam S. Identification of therapeutic targets for glioblastoma by network analysis. *Oncogene.* 2016; 35(5): 608–620. <https://doi.org/10.1038/onc.2015.119> PMID: 25961929
59. Plaisier CL, O'Brien S, Bernard B, Reynolds S, Simon Z, Toledo CM, Ding Y, Reiss DJ, Paddison PJ, Baliga NS. Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis. *Cell Systems.* 2016; 3(2): 172–186. <https://doi.org/10.1016/j.cels.2016.06.006> PMID: 27426982
60. Nishikawa R, Ji XD, Harmon RC, Lazar CS, Gill GN, Cavenee WK, Huang HJ. A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity. *ProcNatlAcadSci U S A.* 1994; 91(16): 7727–7731. <https://doi.org/10.1073/pnas.91.16.7727> PMID: 8052651
61. Gan HK, Kaye AH, Luwor RB. The EGFRvIII variant in glioblastomamultiforme. *JClinNeurosci.* 2009; 16(6): 748–754. <https://doi.org/10.1016/j.jocn.2008.12.005> PMID: 19324552
62. Huang PH, Xu AM, White FM. Oncogenic EGFR signaling networks in glioma. *Sci Signal.* 2009; 2(87): re6. <https://doi.org/10.1126/scisignal.287re6> PMID: 19738203
63. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011; 144(5): 646–674. <https://doi.org/10.1016/j.cell.2011.02.013> PMID: 21376230
64. Ward PS, Thompson CB. Signaling in control of cell growth and metabolism. *ColdSpringHarbPerspect Biol.* 2012; 4(7): a006783. <https://doi.org/10.1101/cshperspect.a006783> PMID: 22687276
65. Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, Neumuller RA, Mohr SE, Perrimon N. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. *Nat Methods.* 2014; 11(1): 94–99. <https://doi.org/10.1038/nmeth.2733> PMID: 24240319
66. Tuncbag N, Braunstein A, Pagnani A, Huang SS, Chayes J, Borgs C, Zecchina R, Fraenkel E. Simultaneous reconstruction of multiple signaling pathways via the prize-collecting steiner forest problem. *J ComputBiol.* 2013; 20(2): 124–136. <https://doi.org/10.1089/cmb.2012.0092> PMID: 23383998
67. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks? *Genome Biol.* 2006; 7(11): 120. <https://doi.org/10.1186/gb-2006-7-11-120> PMID: 17147767

68. Kudla J, Bock R. Lighting the Way to Protein-Protein Interactions: Recommendations on Best Practices for Bimolecular Fluorescence Complementation Analyses. *Plant Cell*. 2016; 28(5): 1002–1008. <https://doi.org/10.1105/tpc.16.00043> PMID: 27099259
69. Gillis J, Ballouz S, Pavlidis P. Bias tradeoffs in the creation and analysis of protein-protein interaction networks. *J Proteomics*. 2014; 100: 44–54. <https://doi.org/10.1016/j.jprot.2014.01.020> PMID: 24480284
70. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C, Jensen LJ. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res*. 2013; 41: D808–815. <https://doi.org/10.1093/nar/gks1094> PMID: 23203871
71. Teschendorff AE, Severini S. Increased entropy of signal transduction in the cancer metastasis phenotype. *BMC SystBiol*. 2010; 4: 104. <https://doi.org/10.1186/1752-0509-4-104> PMID: 20673354
72. Teschendorff AE, Sollich P, Kuehn R. Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional omic data. *Methods*. 2014; 67(3): 282–293. <https://doi.org/10.1016/j.ymeth.2014.03.013> PMID: 24675401
73. West J, Bianconi G, Severini S, Teschendorff AE. Differential network entropy reveals cancer system hallmarks. *Sci Rep*. 2012; 2: 802. <https://doi.org/10.1038/srep00802> PMID: 23150773
74. Goncalves E, Bucher J, Ryll A, Niklas J, Mauch K, Klamt S, Rocha M, Saez-Rodriguez J. Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *MolBiosyst*. 2013; 9(7): 1576–1583. <https://doi.org/10.1039/c3mb25489e> PMID: 23525368
75. Covert MW, Schilling CH, Palsson B. Regulation of gene expression in flux balance models of metabolism. *J TheorBiol*. 2001; 213(1): 73–88. <https://doi.org/10.1006/jtbi.2001.2405> PMID: 11708855
76. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*. 2004; 429(6987), 92–96. <https://doi.org/10.1038/nature02456> PMID: 15129285
77. Shlomi T, Eisenberg Y, Sharan R, Ruppin E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *MolSystBiol*. 2007; 3: 101. <https://doi.org/10.1038/msb4100141> PMID: 17437026
78. Chandrasekaran S, Price ND. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *ProcNatlAcadSci U S A*. 2010; 107(41): 17845–17850. <https://doi.org/10.1073/pnas.1005139107> PMID: 20876091
79. Yizhak K, Benyamini T, Liebermeister W, Ruppin E, Shlomi T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*. 2010; 26(12): i255–260. <https://doi.org/10.1093/bioinformatics/btq183> PMID: 20529914
80. Bettenbrock K, Fischer S, Kremling A, Jahreis K, Sauter T, Gilles ED. A quantitative approach to catabolite repression in *Escherichia coli*. *J BiolChem*. 2006; 281(5): 2578–2584. <https://doi.org/10.1074/jbc.M508090200> PMID: 16263707
81. Nakakuki T, Birtwistle MR, Saeki Y, Yumoto N, Ide K, Nagashima T, Bruschi L, Ogunnaike BA, Okada-Hatakeyama M, Kholodenko BN. Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell*. 2010; 141(5): 884–896. <https://doi.org/10.1016/j.cell.2010.03.054> PMID: 20493519
82. Stelnic-Klotz I, Legewie S, Tchernitsa O, Witzel F, Klinger B, Sers C, Herzel H, Bluthgen N, Schafer R. Reverse engineering a hierarchical regulatory network downstream of oncogenic KRAS. *MolSystBiol*. 2012; 8: 601. <https://doi.org/10.1038/msb.2012.32> PMID: 22864383
83. Zhang F, Liu R, Zheng J. Sig2GRN: a software tool linking signaling pathway with gene regulatory network for dynamic simulation. *BMC SystBiol*. 2016; 10:123. <https://doi.org/10.1186/s12918-016-0365-1> PMID: 28155685
84. Melas IN, Mitsos A, Messinis DE, Weiss TS, Alexopoulos LG. Combined logical and data-driven models for linking signalling pathways to cellular response. *BMC SystBiol*. 2011; 5: 107. <https://doi.org/10.1186/1752-0509-5-107> PMID: 21729292
85. Peng SC, Wong DS, Tung KC, Chen YY, Chao CC, Peng CH, Chuang YJ, Tang CY. Computational modeling with forward and reverse engineering links signaling network and genomic regulatory responses: NF-kappaB signaling-induced gene expression responses in inflammation. *BMC Bioinformatics*. 2010; 11: 308. <https://doi.org/10.1186/1471-2105-11-308> PMID: 20529327
86. Ryll A, Bucher J, Bonin A, Bongard S, Goncalves E, Saez-Rodriguez J, Niklas J, Klamt S. A model integration approach linking signalling and gene-regulatory logic with kinetic metabolic models. *Biosystems*. 2014; 124: 26–38. <https://doi.org/10.1016/j.biosystems.2014.07.002> PMID: 25063553
87. Mosca E, Alfieri R, Maj C, Bevilacqua A, Canti G, Milanese L. Computational modeling of the metabolic States regulated by the kinase akt. *Front Physiol*. 2012; 3: 418. <https://doi.org/10.3389/fphys.2012.00418> PMID: 23181020

88. Anand K, Bianconi G. Entropy measures for networks: toward an information theory of complex topologies. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2009; 80(4 Pt 2): 045102. <https://doi.org/10.1103/PhysRevE.80.045102> PMID: 19905379
89. Huang SS, Fraenkel E. Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks. *Sci Signal.* 2009; 2(81): ra40. <https://doi.org/10.1126/scisignal.2000350> PMID: 19638617
90. Wang J, Peng X, Peng W, Wu FX. Dynamic protein interaction network construction and applications. *Proteomics.* 2014; 14(4–5): 338–352. <https://doi.org/10.1002/pmic.201300257> PMID: 24339054
91. Boda E, Pini A, Hoxha E, Parolisi R, Tempia F. Selection of reference genes for quantitative real-time RT-PCR studies in mouse brain. *J MolNeurosci.* 2009; 37(3): 238–253. <https://doi.org/10.1007/s12031-008-9128-9> PMID: 18607772
92. Lemma S, Avnet S, Salerno M, Chano T, Baldini N. Identification and Validation of Housekeeping Genes for Gene Expression Analysis of Cancer Stem Cells. *PLoS One.* 2016; 11(2): e0149481. <https://doi.org/10.1371/journal.pone.0149481> PMID: 26894994
93. Ahmad SS, Glatzle J, Bajaeifer K, Buhler S, Lehmann T, Konigsrainer I, Vollmer JP, Sipos B, Ahmad SS, Northoff H, Konigsrainer A, Zieker D. Phosphoglycerate kinase 1 as a promoter of metastasis in colon cancer. *Int J Oncol.* 2013; 43(2): 586–590. <https://doi.org/10.3892/ijo.2013.1971> PMID: 23727790
94. Zieker D, Konigsrainer I, Tritschler I, Loffler M, Beckert S, Traub F, Nieselt K, Buhler S, Weller M, Gaedcke J, Taichman RS, Northoff H, Brucher BL, Konigsrainer A. Phosphoglycerate kinase 1 a promoting enzyme for peritoneal dissemination in gastric cancer. *Int J Cancer.* 2010; 126(6): 1513–1520. <https://doi.org/10.1002/ijc.24835> PMID: 19688824
95. Revillion F, Pawlowski V, Hornez L, Peyrat JP. Glyceraldehyde-3-phosphate dehydrogenase gene expression in human breast cancer. *Eur J Cancer.* 2000; 36(8): 1038–1042. [https://doi.org/10.1016/s0959-8049\(00\)00051-4](https://doi.org/10.1016/s0959-8049(00)00051-4) PMID: 10885609
96. Droin NM, Green DR. Role of Bcl-2 family members in immunity and disease. *BiochimBiophysActa.* 2004; 1644(2–3): 179–188. <https://doi.org/10.1016/j.bbamcr.2003.10.011> PMID: 14996502
97. Fu Z, Tindall DJ. FOXOs, cancer and regulation of apoptosis. *Oncogene.* 2008; 27(16): 2312–2319. <https://doi.org/10.1038/onc.2008.24> PMID: 18391973
98. Sarkar S, Dutta D, Samanta S.K, Bhattacharya K, Pal B.C, Li J, Datta K, Mandal CN, Mandal C. Oxidative inhibition of Hsp90 disrupts the super-chaperone complex and attenuates pancreatic adenocarcinoma in vitro and in vivo. *Int. J. Cancer.* 2013; 132(3):695–706. <https://doi.org/10.1002/ijc.27687> PMID: 22729780
99. Liu H, Xiao F, Serebriiskii IG, O'Brien SW, Maglaty MA, Astsaturov I, Litwin S, Martin LP, Proia DA, Golemis EA, Connolly DC. Network analysis identifies an HSP90-central hub susceptible in ovarian cancer. *Clin Cancer Res.* 2013; 19(18): 5053–5067. <https://doi.org/10.1158/1078-0432.CCR-13-1115> PMID: 23900136
100. Arensman MD, Telesca D, Lay AR, Kershaw KM, Wu N, Donahue TR, Dawson DW. The CREB-binding protein inhibitor ICG-001 suppresses pancreatic cancer growth. *Mol Cancer Ther.* 2014; 13(10): 2303–2314. <https://doi.org/10.1158/1535-7163.MCT-13-1005> PMID: 25082960
101. Wiese M, Walther N, Diederichs C, Schill F, Monecke S, Salinas G, Sturm D, Pfister SM, Dressel R, Johnsen SA, Kramm CM. The beta-catenin/CBP-antagonist ICG-001 inhibits pediatric gliomatumorigenicity in a Wnt-independent manner. *Oncotarget.* 2017; 8(16): 27300–27313. <https://doi.org/10.18632/oncotarget.15934> PMID: 28460484
102. Ghadban T, Jessen A, Reeh M, Dibbern JL, Mahner S, Mueller V, Wellner UF, Gungor C, Izbicki JR, Vashist YK. In vitro study comparing the efficacy of the water-soluble HSP90 inhibitors, 17-AEPGA and 17-DMAG, with that of the nonwater-soluble HSP90 inhibitor, 17-AAG, in breast cancer cell lines. *Int J Mol Med.* 2016; 38(4): 1296–1302. <https://doi.org/10.3892/ijmm.2016.2696> PMID: 27498942
103. Sauvageot CM, Weatherbee JL, Kesari S, Winters SE, Barnes J, Dellagatta J, Ramakrishna NR, Stiles CD, Kung AL, Kieran MW, Wen PY. Efficacy of the HSP90 inhibitor 17-AAG in human glioma cell lines and tumorigenic glioma stem cells. *NeuroOncol.* 2009; 11(2): 109–121. <https://doi.org/10.1215/15228517-2008-060> PMID: 18682579
104. Usmani SZ, Bona R, Li Z. 17 AAG for HSP90 inhibition in cancer—from bench to bedside. *CurrMol Med.* 2009; 9(5): 654–664. PMID: 19601813
105. Bhattacharya K, Bag AK, Tripathi R, Samanta SK, Pal BC, Shaha C, Mandal C. Mahanine, a novel mitochondrial complex-III inhibitor induces G0/G1 arrest through redox alteration-mediated DNA damage response and regresses glioblastomamultiforme. *Am J Cancer Res.* 2014; 4(6):629–47. PMID: 25520856
106. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number

- of sequenced peptides per protein. *Mol Cell Proteomics*. 2005; 4(9): 1265–1272. <https://doi.org/10.1074/mcp.M500061-MCP200> PMID: 15958392
107. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016; 44(D1): D457–462. <https://doi.org/10.1093/nar/gkv1070> PMID: 26476454
  108. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S, Matthews L, May B, Milacic M, Rothfels K, Sharmovskiy V, Webber M, Weiser J, Williams M, Wu G, Stein L et al. The Reactome pathway Knowledgebase. *Nucleic Acids Res*. 2016; 44(D1): D481–487. <https://doi.org/10.1093/nar/gkv1351> PMID: 26656494
  109. Fazekas D, Koltai M, Turei D, Modos D, Palfy M, Dul Z, Zsakai L, Szalay-Beko M, Lenti K, Farkas JJ, Vellai T, Csermely P, Korcsmaros T. Signalink 2—a signaling pathway resource with multi-layered regulatory networks. *BMC Syst Biol*. 2013; 7: 7. <https://doi.org/10.1186/1752-0509-7-7> PMID: 23331499
  110. Kandasamy K, Mohan SS, Raju R, Keerthikumar S, Kumar GS, Venugopal AK, Telikicherla D, Navarro JD, Mathivanan S, Pecquet C, Gollapudi SK, Tattikota SG, Mohan S, Padhukasahasram H, Subbannayya Y, Goel R, Jacob HK, Zhong J, Sekhar R, Nanjappa V et al. NetPath: a public resource of curated signal transduction pathways. *Genome Biol*. 2010; 11(1): R3. <https://doi.org/10.1186/gb-2010-11-1-r3> PMID: 20067622
  111. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang NG, Robertson CL, Serova N, Davis S, Soboleva A. NCBI GEO: archive for functional genomics data sets-update. *Nucleic Acids Research*. 2013; 41 (Database issue): D991–D995. <https://doi.org/10.1093/nar/gks1193> PMID: 23193258
  112. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. *Nat Genet*. 2006; 38 (5): 500–501. <https://doi.org/10.1038/ng0506-500> PMID: 16642009
  113. Wang J, Vasaiakar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res*. 2017; 45:W130–7. <https://doi.org/10.1093/nar/gkx356> PMID: 28472511