Sensitivity analysis of Wnt β-catenin based transcription complex might bolster power-logarithmic psychophysical law and reveal preserved gene gene interactions

Shriprakash Sinha[*a,‡]

Abstract
Let a sensation magnitude γ be determined by a stimulus magnitude β. The Webers laws states that $\Delta\gamma$ remains constant when the relative stimulus increment $\Delta\beta$ remains constant. It has been found that this law is actually a derivation of Bernoullis law were $\Delta\gamma \propto \log \frac{\Delta\beta}{\beta}$. Recently, such psychophysical laws have been observed in the natural behaviour of certain intra/extracellular factors working in the Wnt pathway. This manuscript tests the veracity of the prevalence of such laws, albeit at a coarse level, using sensitivity analysis on biologically inspired computational causal models. Sensitivity analysis plays a crucial role in observing the behaviour of output of a variable given variations in the input. In this work, the variation in the effect of the predictive behaviour of the transcription complex ($TRCMPLX$) conditional on the evidences of gene expressions in normal and tumor samples is observed by varying the initially assigned values of conditional probability tables (cpt) for $TRCMPLX$. Preliminary analysis shows that the variation in predictive behaviour of $TRCMPLX$ conditional on gene evidences follows power and logarithmic psychophysical law crudely. This implies deviations in output are proportional to increasing function of deviations in input and shows constant behaviour for recorded deviations in higher values of the input. This points towards stability in the behaviour of $TRCMPLX$ and is reflected in the preserved gene-gene interactions. These interactions are obtained by thresholding the inferred conditional probabilities of a gene activation given the status of another gene activation. The deviation in the interactions in normal and tumor samples was also observed by varying the initially assigned values of cpt for β-$catenin$ based $TRCMPLX$. Analysis of deviation in interactions show prevalence of power-logarithmic psychophysical law. This prevalence is reported for interaction between elements of pairs ($SFRP3$, MYC), ($SFRP2$, $CD44$) and ($DKK1$, $DACT2$). Dynamic models of Bayesian networks might reveal the phenomena in a better way.

1 Introduction & problem statement

Ever since the accidental discovery of the Wingless in 1973 by Sharma[2], a tremendous amount of research work has been carried out in the related field of Wnt signaling pathway in the past forty years. A majority of the work has focused on issues related to • the discovery of genetic and epigenetic factors affecting the pathway (Thorstensen *et al.*[3] & Baron and Kneissel[4]), • implications of mutations in the pathway and its dominant role on cancer and other diseases (Clevers[5]), • investigation into the pathway's contribution towards embryo development (Sokol[6]), homeostasis (Pinto *et al.*[7], Zhong *et al.*[8]) and apoptosis (Pećina-Šlaus[9]) and • safety and feasibility of drug design for the wnt pathway (Kahn[10], Garber[11], Voronkov and Krauss[12], Blagodatski *et al.*[13] & Curtin and Lorenzi[14]). More recent informative reviews have touched on various issues related to the different types of the Wnt signaling pathway and have stressed not only the activation of the Wnt signaling pathway via the Wnt proteins (Rao and Kühl[15]) but also the on the secretion mechanism that plays a major role in the initiation of the Wnt activity as a prelude (Yu and Virshup[16]).

In a more recent development, there has been the observation and study of psychophysical laws prevailing within the pathway and in this regard Goentoro and Kirschner[17] point to two findings namely, • the robust fold changes of β-$catenin$ and • the transcriptional machinery of the Wnt pathway depends on the fold changes in β-$catenin$ instead of absolute levels of the same and some gene transcription networks must respond to fold changes in signals according to the Weber[18] law in sensory physiology. Note that Weber's law has been found to be a special case of Bernoulli's logarithmic law Masin *et al.*[19]. If a sensation magnitude γ be determined by a stimulus magnitude β, then the Weber's law states that $\Delta\gamma$ remains constant when the relative stimulus increment $\Delta\beta$ remains constant. The law derives from a more general Bernoullis law were $\Delta\gamma \propto \log \frac{\Delta\beta}{\beta}$. In an unrelated work by Sun *et al.*[20], it has been shown that these laws arise at computational level as Bayes optimal under neurobiological constraints at implementational and algorithmic levels. The proposed mathematical framework for understanding the psychophysical scales as Bayes optimal and information theoretically-optimal representation of time sampled continuous valued stimuli is based on established neurobiological assumptions. The authors also show that the psychophysical

‡ This manuscript is an extension of Sinha[1] which was a part of the Netherlands Bioinformatics Centre (NBIC) BioRange-II project BR2.5 (code - IGD71G).

[a] *Independent Researcher, 104-Madhurisha Heights, Risali Sector, Bhilai-490006, India. E-mail: sinha.shriprakash@yandex.com*

laws connect well to quantization frameworks and state that only discrete set of output is distinguishable due to biological constraints. This discretization leads to quantization of stimulus also as the nonlinear scaling of the stimulus that leads to the resultant output is invertible. These mathematical insights might explain the indistinguishable insensitive fold changes in levels of β-$catenin$ shown by Goentoro and Kirschner[17].

Based on the importance of the revealed phenomena, it might be useful to know if these observations could be verified using computational models apart from analysis of results from wet lab experiments. What is needed is a framework that can capture the causal semantics of the signaling pathway where the influence diagrams involving the interacting extra/intracellular factors working in the pathway, represent the biological knowledge/mechanism of the pathway to a certain extent. Once a model representation is available, the desired variation in the activity of an input factor and the observed variation in the output of the activity of factor(s) can be studied. Sensitivity analysis plays a crucial role in observing the behaviour of output of a variable given variations in the input. As will be seen later, probabilistic graphical models or Bayesian networks provide a framework for representing the causal semantics of the pathway under investigation.

To address these issues, the current work uses the Bayesian nework model proposed in Sinha[1] and conducts sensitivity analysis on the model to check the observations regarding the prevalence of the reported psychophysical laws. In Sinha[1], it was shown via hypothesis testing that the active (repressed) state of $TRCMPLX$ in the Wnt signaling pathway for colorectal cancer cases is not always correspond to the tumorous (normal) state of the test sample under consideration. For this, Sinha[1] shows various results on the predicted state of $TRCMPLX$ conditional on the given gene evidences, while varying the assigned probability values of conditional probability tables of $TRCMPLX$ during initialization of the biologically inspired Bayesian Network model. Here, the degree of belief in the activity of $TRCMPLX$ is denoted by the prior probability assigned to the node of $TRCMPLX$ in the network. It was found that the predicted values often increase with an increasing value (in conditional probability tables) of the activity of $TRCMPLX$ on certain genes. What this asks for is that for the recorded deviations due to the changes made in these prior probabilities (i.e the input deviations), is it possible to observe the prevalent logarithmic laws and their deviations (like the Weber's law) as shown by Goentoro and Kirschner[17], using computational causal modeling?

In this manuscript, the preliminary analysis of deviations computed from variation in prior and estimated conditional probability values using Bayesian network model in Sinha[1] show that the variation in predictive behaviour of $TRCMPLX$ conditional on gene evidences (i.e the output deviation) follows power and logarithmic psychophysical law crudely, apropos the variation in assigned priors of $TRCMPLX$ (i.e input deviations). This implies that the deviations in output are proportional to increasing function of deviations in input. This relates to the work of Adler *et al.*[21] on power and logarithmic law albeit at a coarse level. The granularity is obscured due to the use of static data from Jiang *et al.*[22] that is used in Sinha[1] as well as the Bayesian network model that encodes the belief in the factors affecting the pathway in terms of probabilities as well as the inferences made based on the updating of these probabilities conditional on discretized states of gene expression values as evidences. Irrespective of the hurdle posed by the causal models, inferences made based on prior biological knowledge and gene expression evidences coupled with sensitivity analysis sheds light on the prevalent psychophysical laws in the pathway.

Adler *et al.*[21] show in detail that these laws can be studied empirically using models that exhibit the property of fold change detection (FCD). What this means is that the output depends on the relative changes in the input. The biological feedback models employed for these studies consider various parameters like rates of production of a compound, removal removal of a compound, repression of a compound, levels of scaffolds, kinases, etc. that might be responsible for exhibiting these laws. The current work using the static Bayesian network model might not propose feedback loops directly as used by Adler *et al.*[21], yet it could reveal existence of the loops via causal inference even while using static data. The drawback of the current work is its inability to consider cyclic loops. This can be rectified by use of dynamic Bayesian network models that incorporate interaction represented in time series data. Also, the use of Bayesian network models can help in studying the problem from a multiparameter setting as various factors affecting the pathway can be connected in the influence diagrams of the network through the principle of d-connectivity/separability. This connectivity will be explained later in the required theory section.

Note that Goentoro and Kirschner[17] show results for the behaviour of fold change of β-$catenin$ with respect to changes in the single parameter values i.e the Wnt. On similar lines, the current work takes into account the behaviour of $TRCMPLX$ conditional on affects of multiple parameters in the form of evidences of various intra/extracellular gene expression values working in the pathway, based on the changes made in the assinged prior probabilities for $TRCMPLX$. The difference here is that one can analyse changes in nodes of a computational model to explore an inherent law in comparison to use of wet lab experiments. The issue here is that FCD which is recorded with respect to changes in levels of concentration can now be recorded via changes in the strength of belief in the occurrence of an event. For example, suppose it is not known by what degree the $TRCMPLX$ plays a major role in the signaling pathway quantitatively, then it

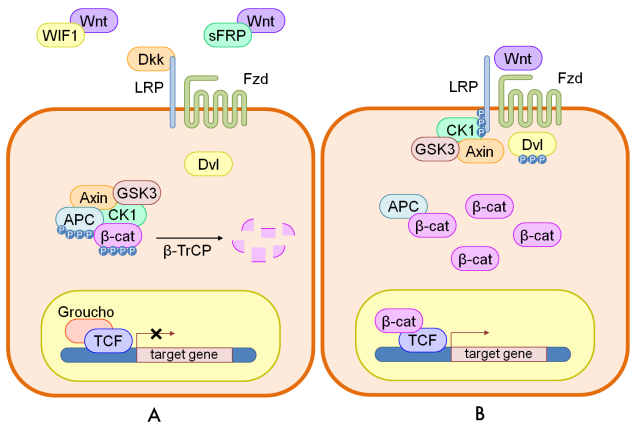


Fig. 1 A cartoon of wnt signaling pathway contributed by Verhaegh *et al.*[23]. Part (A) represents the destruction of β-$catenin$ leading to the inactivation of the wnt target gene. Part (B) represents activation of wnt target gene.

is possible to encode the degree of belief regarding the role of $TRCMPLX$ in the form of prior or conditional probabilities during initialization of the network. By recording the deviations in these probabilities and observing the output deviations, it is possible to study certain psychophysical laws. Finally, this does not mean that probabilities related to actual concentrations cannot be encoded. Thus, Bayesian networks help in captuing the desired biological knowledge via various causal arcs and conditional probabilities and sensitivity analysis aids in the study of the such natural behaviour.

As a second observation, the forgoing result points towards stability in the behaviour of $TRCMPLX$ and this stability is reflected in the preserved gene gene interactions across the changing values of the priors of $TRCMPLX$. The interactions are inferred from conditional probabilities of individual gene activation given the status of another gene activation. Finally, as a third observation, it would be interesting to note if the psychophysical laws are prevalent among the dual gene-gene interactions or not. If the results are affirmative then the following important speculations might hold true • Not just one factor but components of the entire network might be exhibit in such a behavior at some stage or the other. • The psychophysical law might not be restricted to individual intra/extracellular components but also to the interactions among the the intra/extracellular components in the pathway. This might mean that the interactions manifest during the prevalence of power and logarithmic laws. Further wet lab analysis is needed to very these computational claims.

It is important to be aware of the fact that the presented results are derived from a static Bayesian network model. It is speculated that dynamic models might give much better and more realistic results.

2 Revisiting the requisite theory

To understand the logical flow of the current paper, some details of the above related topics from Sinha[1] are revisited here in order and subdivided into descriptions of - (1) general working of canonical Wnt signaling pathway and some of the involved epigenetic factors (2) introduction to Bayesian networks and (3) the intuition behind the Bayesian network model employed. This is followed by Weber's law and its derivation and finally the notations and terminologies to understand the results and discussion section.

2.1 Canonical Wnt signaling pathway

The canonical Wnt signaling pathway is a transduction mechanism that contributes to embryo development and controls homeostatic self renewal in several tissues (Clevers[5]). Somatic mutations in the pathway are known to be associated with cancer in different parts of the human body. Prominent among them is the colorectal cancer case (Gregorieff and Clevers[24]). In a succinct overview, the Wnt signaling pathway works when the Wnt ligand gets attached to the Frizzled(fzd)/LRP coreceptor complex. Fzd may interact with the Dishevelled (Dvl) causing phosphorylation. It is also thought that Wnts cause phosphorylation of the LRP via casein kinase 1 ($CK1$) and kinase $GSK3$. These developments further lead to attraction of Axin which causes inhibition of the formation of the degradation complex. The degradation complex constitutes of $Axin$, the β-$catenin$ transportation complex APC, $CK1$ and $GSK3$. When the pathway is active the dissolution of the degradation complex leads to stabilization in the concentration of β-$catenin$ in the cytoplasm. As β-$catenin$ enters into the nucleus it displaces the $Groucho$ and binds with transcription cell factor TCF thus instigating transcription of Wnt target genes. $Groucho$ acts as lock on TCF and prevents the transcription of target genes which may induce cancer. In cases when the Wnt ligands are not captured by the coreceptor at the cell membrane, $Axin$ helps in formation of the degradation complex. The degradation complex phosphorylates β-$catenin$ which is then recognized by $Fbox/WD$ repeat protein $\beta-TrCP$. $\beta-TrCP$ is a component of ubiquitin ligase complex that helps in ubiquitination of β-$catenin$ thus marking it for degradation via the proteasome. Cartoons depicting the phenomena of Wnt being inactive and active are shown in figures 1(A) and 1(B), respectively.

2.2 Epigenetic factors

One of the widely studied epigenetic factors is methylation (Costello and Plass[25], Das and Singal[26], Issa[27]). Its occur-

rence leads to decrease in the gene expression which affects the working of Wnt signaling pathways. Such characteristic trends of gene silencing like that of secreted frizzled-related proteins ($SFRP$) family in nearly all human colorectal tumor samples have been found at extracellular level (Suzuki *et al.*[28]). Similarly, methylation of genes in the Dickkopf ($DKKx$ Niehrs[29], Sato *et al.*[30]), Dapper antagonist of catenin ($DACTx$ Jiang *et al.*[22]) and Wnt inhibitory factor-1 ($WIF1$ Taniguchi *et al.*[31]) family are known to have significant effect on the Wnt pathway. Also, histone modifications (a class of proteins that help in the formation of chromatin which packs the DNA in a special form Strahl and Allis[32]) can affect gene expression (Peterson *et al.*[33]). In the context of the Wnt signaling pathway it has been found that $DACT$ gene family show a peculiar behavior in colorectal cancer (Jiang *et al.*[22]). $DACT1$ and $DACT2$ showed repression in tumor samples due to increased methylation while $DACT3$ did not show obvious changes to the interventions. It is indicated that $DACT3$ promoter is simultaneously modified by the both repressive and activating (bivalent) histone modifications (Jiang *et al.*[22]).

2.3 Bayesian Networks

In reverse engineering methods for control networks (Gardner and Faith[34]) there exist many methods that help in the construction of the networks from the data sets as well as give the ability to infer causal relations between components of the system. A widely known architecture among these methods is the Bayesian Network (BN). These networks can be used for causal reasoning or diagnostic reasoning or both. It has been shown through reasoning and examples in Roehrig[35] that the probabilistic inference mechanism applied via Bayesian networks are analogous to the structural equation modeling in path analysis problems.

Initial works on BNs in Pearl[36] and Pearl[37] suggest that the networks only need a relatively small amount of marginal probabilities for nodes that have no incoming arcs and a set of conditional probabilities for each node having one or more incoming arcs. The nodes form the driving components of a network and the arcs define the interactive influences that drive a particular process. Under these assumptions of influences the joint probability distribution of the whole network or a part of it can be obtained via a special factorization that uses the concept of direct influence and through dependence rules that define d-connectivity/separability as mentioned in Charniak[38] and Needham *et al.*[39]. This is illustrated through a simple example in Roehrig[35].

The Bayesian networks work by estimating the posterior probability of the model given the data set. This estimation is usually referred to as the Bayesian score of the model conditioned on the data set. Mathematically, let $\mathcal{S}$ represent the model given the data $\mathcal{D}$ and ξ is the background knowledge. Then according to the Bayes Theorem (Bayes and Price[40]):

$$\begin{aligned} \mathcal{P}(\mathcal{S}|\mathcal{D},\xi) &= \frac{\mathcal{P}(\mathcal{S}\cap\mathcal{D}|\xi)}{\mathcal{P}(\mathcal{D}|\xi)} \\ &= \frac{\mathcal{P}(\mathcal{S}|\xi)\times\mathcal{P}(\mathcal{D}|\mathcal{S},\xi)}{\mathcal{P}(\mathcal{D}|\xi)} \\ \text{posterior} &= \frac{prior\times likelihood}{constant} \end{aligned} \tag{1}$$

Thus the Bayesian score is computed by evaluating the **posterior** distribution $\mathcal{P}(\mathcal{S}|\mathcal{D},\xi)$ which is proportional to the **prior** distribution of the model $\mathcal{P}(\mathcal{S}|\xi)$ and the **likelihood** of the data given the model $\mathcal{P}(\mathcal{D}|\mathcal{S},\xi)$. It must be noted that the background knowledge is assumed to be independent of the data. Next, since the evaluation of probabilities require multiplications a simpler way is to take logarithmic scores which boils down to addition. Thus the estimation takes the form:

$$\begin{aligned} \log\mathcal{P}(\mathcal{S}|\mathcal{D},\xi) &= \log\mathcal{P}(\mathcal{S}|\xi)+\log\mathcal{P}(\mathcal{D}|\mathcal{S},\xi)- \\ &\quad \log\mathcal{P}(\mathcal{D}|\xi) \\ &= \log\mathcal{P}(\mathcal{S}|\xi)+\log\mathcal{P}(\mathcal{D}|\mathcal{S},\xi)+ \\ &\quad constant \end{aligned} \tag{2}$$

Finally, the likelihood of the function can be evaluated by averaging over all possible local conditional distributions parameterized by θ_i's that depict the conditioning of parents. This is equated via:

$$\begin{aligned} \mathcal{P}(\mathcal{D}|\mathcal{S},\xi) &= \int_{\theta_1}\ldots\int_{\theta_n}\mathcal{P}(\mathcal{D},\theta_i|\mathcal{S})d\theta_i \\ &= \int_{\theta_1}\ldots\int_{\theta_n}\mathcal{P}(\mathcal{D}|\theta_i\mathcal{S})\mathcal{P}(\theta_i|\mathcal{S})d\theta_i \end{aligned} \tag{3}$$

Work on biological systems that make use of Bayesian networks can also be found in (Friedman *et al.*[41], Hartemink *et al.*[42], Sachs *et al.*[43], Sachs *et al.*[44] and Peer *et al.*[45]). Bayesian networks are good in generating network structures and testing a targeted hypothesis which confine the experimenter to derive causal inferences (Brent and Lok[46]). But a major disadvantage of the Bayesian networks is that they rely heavily on the conditional probability distributions which require good sampling of datasets and are computationally intensive. On the other hand, these networks are quite robust to the existence of the unobserved variables and accommodates noisy datasets. They also have the ability to combine heterogeneous data sets that incorporate different modalities.

In Sinha[1], simple static Bayesian Network models have been developed with an aim to show how • incorporation of

heterogeneous data can be done to increase prediction accuracy of test samples • prior biological knowledge can be embedded to model biological phenomena behind the Wnt pathway in colorectal cancer • to test the hypothesis regarding direct correspondence of active state of β-$catenin$ based transcription complex and the state of the test sample via segregation of nodes in the directed acyclic graphs of the proposed models and • inferences can be made regarding the hidden biological relationships between a particular gene and the β-$catenin$ transcription complex. This work uses Matlab implemented BN toolbox from Murphy *et al.*[47].

2.4 Intuition behind the causal semantics of the biologically inspired Bayesian network

The NB model (Verhaegh *et al.*[23]) assumes that the activation (inactivation) of β-$catenin$ based transcription complex is equivalent to the fact that the sample is cancerous (normal). This assumption needs to be tested and Sinha[1] proposes a newly improvised models based on prior biological knowledge and epigenetic information regarding the signaling pathway with the assumption that sample prediction may not always mean that the β-$catenin$ based transcription complex is activated. These assumptions are incorporated by inserting another node of $Sample$ for which gene expression measurements were available. This is separate from the $TRCMPLX$ node that influences a particular set of known genes in the human colorectal cancer. For those genes whose relation with the $TRCMPLX$ is currently not known or biologically affirmed, indirect paths through the $Sample$ node to the $TRCMPLX$ exist, technical aspect of which is described next.

For those factors whose relations were not yet confirmed but known to be involved in the pathway, the causal arcs were segregated via a latent variable introduced into the Bayesian network. The latent variable in the form of $Sample$ (see figure 2) is extremely valuable as it connects the factors whose relations have not been confirmed till now, to factors whose influences have been confirmed in the pathway. Finally, the introduction of latent variable in a causal model opens the avenue to assume the presence of measurements that haven't been recorded. Intuitively, for cancer samples the hidden measurements might be different from those for normal samples. The connectivity of factors through the variable provides an important route to infer biological relations.

Since all gene expressions have been measured from a sample of subjects the expression of genes is conditional on the state of the $Sample$. Here both tumorous and normal cases are present in equal amounts. The transcription factor $TRCMPLX$ under investigation is known to operate with the help of interaction between β-$catenin$ with $TCF4$ and $LEF1$ (Waterman[49], Kriegl *et al.*[50]). It is also known that the regions in the TSS of MYC (Yochum[51]), $CCND1$ (Schmidt-Ott *et al.*[52]), $CD44$ (Kanwar *et al.*[53]), $SFRP1$ (Caldwell *et al.*[54]), $WIF1$ (Reguart *et al.*[55]), $DKK1$ (González-Sancho *et al.*[56]) and $DKK4$ (Pendas-Franco *et al.*[57], Baehs *et al.*[58]) contain factors that have affinity to β-$catenin$ based $TRCMPLX$. Thus expression of these genes are shown to be influenced by $TRCMPLX$, in figure 2.

Roles of $DKK2$ (Matsui *et al.*[59]) and $DKK3$ (Zitt *et al.*[60], Veeck and Dahl[61]) have been observed in colorectal cancer but their transcriptional relation with β-$catenin$ based $TRCMPLX$ is not known. Similarly, $SFRP2$ is known to be a target of $Pax2$ transcription factor and yet it affects the β-$catenin$ Wnt signaling pathway (Brophy *et al.*[62]). Similarly, $SFRP4$ (Feng Han *et al.*[63], Huang *et al.*[64]) and $SFRP5$ (Suzuki *et al.*[28]) are known to have affect on the Wnt pathway but their role with $TRCMPLX$ is not well studied. $SFRP3$ is known to have a different structure and function with respect to the remaining $SFRPx$ gene family (Hoang *et al.*[65]). Also, the role of $DACT2$ is found to be conflicting in the Wnt pathway (Kivimäe *et al.*[66]). Thus for all these genes whose expression mostly have an extracellular affect on the pathway and information regarding their influence on β-$catenin$ based $TRCMPLX$ node is not available, an indirect connection has been made through the $Sample$ node. This connection will be explained at the end of this section.

Lastly, it is known that concentration of $DVL2$ (a member of Disheveled family) is inversely regulated by the expression of $DACT3$ (Jiang *et al.*[22]). High $DVL2$ concentration and suppression of $DACT1$ leads to increase in stabilization of β-$catenin$ which is necessary for the Wnt pathway to be active (Jiang *et al.*[22]). But in a recent development (Yuan *et al.*[67]) it has been found that expression of $DACT1$ positively regulates β-$catenin$. Both scenarios need to be checked via inspection of the estimated probability values for β-$catenin$ using the test data. Thus there exists direct causal relations between parent nodes $DACT1$ and $DVL2$ and child node, β-$catenin$. Influence of methylation (yellow hexagonal) nodes to their respective gene (green circular) nodes represent the affect of methylation on genes. Influence of histone modifications in $H3K27me3$ and $H3K4me3$ (blue octagonal) nodes to $DACT3$ gene node represents the affect of histone modification on $DACT3$. The β-$catenin$ (blue square) node is influenced by concentration of $DVL2$ (depending on the expression state of $DACT3$) and behavior of $DACT1$.

The aforementioned established prior causal biological knowledge is imposed in the Bayesian network model with the aim to computationally reveal unknown biological relationships. The influence diagram of this model is shown in figure 2 with nodes on methylation and histone modification.

In order to understand indirect connections further it is imperative to know about **d-connectivity/separability**. In a BN model this connection is established via the principle of **d-connectivity** which states that nodes are *connected* in a path

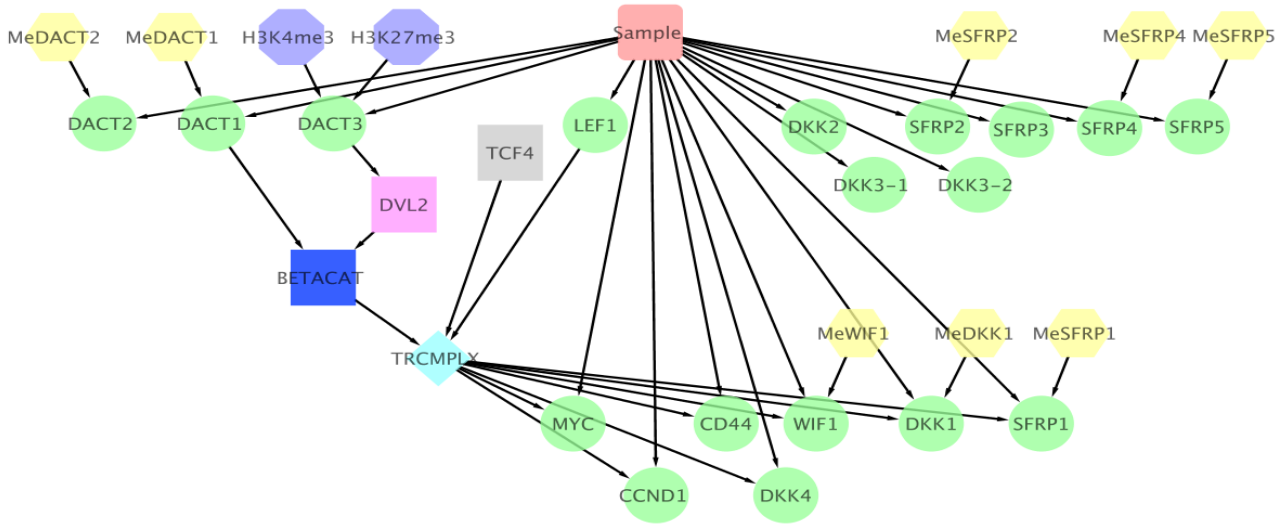


Fig. 2 Influence diagram of $\mathcal{M}_{PBK+EI}$ contains partial prior biological knowledge and epigenetic information in the form of methylation and histone modification. Diagram drawn using Cytoscape Shannon *et al.*[48]. In this model the state of Sample is distinguished from state of $TRCMPLX$ that constitutes the Wnt pathway.

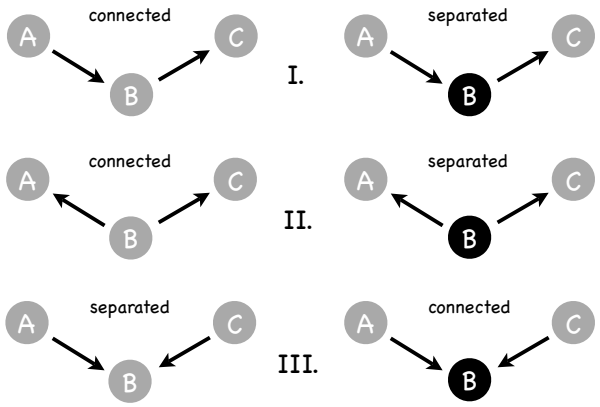


Fig. 3 Cases for d-connectivity and d-separation. Black (Gray) circles mean evidence is available (not available) regarding a particular node.

when there exists no node in the path that has more than one incoming influence edge or there exits nodes in path with more than one incoming influence edge which are observed (i.e evidence regarding such nodes is available) (Charniak[38]). Conversely, via principle of **d-separation** nodes are *separated* in a path when there exists nodes in the path that have more than one incoming influence edge or there exists nodes in the path with at most one incoming influence edge which are observed (i.e evidence regarding such nodes is available). Figure 3 represents three different cases of connectivity and separation between nodes $\mathcal{A}$ and $\mathcal{C}$ when the path between them passes through node $\mathcal{B}$. Connectivity or dependency exists between nodes $\mathcal{A}$ and $\mathcal{C}$ when • evidence is not present regarding node $\mathcal{B}$ in the left graphs of I. and II. in figure 3 or • evidence is present regarding node $\mathcal{B}$ in the right graph of III. in figure 3. Conversely, separation or independence exits between nodes $\mathcal{A}$ and $\mathcal{C}$ when • evidence is present regarding node $\mathcal{B}$ in the right graphs of I. and II. in figure 3 or • evidence is not present regarding node $\mathcal{B}$ in the left graph of III. in figure 3.

It would be interesting to know about the behaviour of $TRCMPLX$ given the evidence of state of $SFRP3$. To reveal such information paths must exist between these nodes. It can be seen that there are multiple paths between $TRCMPLX$ and $SFRP2$ in the BN model in figure 2. These paths are enumerated as follows:

1. $SFRP3, Sample, SFRP1, TRCMPLX$
2. $SFRP3, Sample, DKK1, TRCMPLX$
3. $SFRP3, Sample, WIF1, TRCMPLX$
4. $SFRP3, Sample, CD44, TRCMPLX$
5. $SFRP3, Sample, DKK4, TRCMPLX$
6. $SFRP3, Sample, CCND1, TRCMPLX$
7. $SFRP3, Sample, MYC, TRCMPLX$

8. $SFRP3$, $Sample$, $LEF1$, $TRCMPLX$

9. $SFRP3$, $Sample$, $DACT3$, $DVL2$, β-$catenin$, $TRCMPLX$

10. $SFRP3$, $Sample$, $DACT1$, β-$catenin$, $TRCMPLX$

Knowledge of evidence regarding nodes of $SFRP1$ (path 1), $DKK1$ (path 2), $WIF1$ (path 3), $CD44$ (path 4), $DKK4$ (path 5), $CCND1$ (path 6) and MYC (path 7) makes $Sample$ and $TRCMPLX$ dependent or d-connected. Further, no evidence regarding state of $Sample$ on these paths instigates dependency or connectivity between $SFRP3$ and $TRCMPLX$. On the contrary, evidence regarding $LEF1$, $DACT3$ and $DACT1$ makes $Sample$ (and child nodes influenced by $Sample$) independent or d-separated from $TRCMPLX$ through paths (8) to (10). Due to the dependency in paths (1) to (7) and the given state of $SFRP3$ (i.e evidence regarding it being active or passive), the BN uses these paths during inference to find how $TRCMPLX$ might behave in normal and tumorous test cases. Thus, exploiting the properties of d-connectivity/separability, imposing a biological structure via simple yet important prior causal knowledge and incorporating epigenetic information, BN help in inferring many of the unknown relation of a certain gene expression and a transcription complex.

2.5 The logarithmic psychophysical law

Masin *et al.*[19] states the Weber's law as follows -

> Consider a sensation magnitude γ determined by a stimulus magnitude β. Fechner[68] (vol 2, p. 9) used the symbol $\Delta\gamma$ to denote a just noticeable sensation increment, from γ to $\gamma + \Delta\gamma$, and the symbol $\Delta\beta$ to denote the corresponding stimulus increment, from β to $\beta + \Delta\beta$. Fechner[68] (vol 1, p. 65) attributed to the German physiologist Ernst Heinrich Weber the empirical finding Weber[18] that $\Delta\gamma$ remains constant when the relative stimulus increment $\frac{\Delta\beta}{\beta}$ remains constant, and named this finding Weber's law. Fechner[68] (vol 2, p. 10) underlined that Weber's law was empirical.

It has been found that Bernoulli's principle (Bernoulli[69]) is different from Weber's law (Weber[18]) in that it refers to $\Delta\gamma$ as any possible increment in γ, while the Weber's law refers only to just noticeable increment in γ. Masin *et al.*[19] shows that Weber's law is a special case of Bernoulli's principle and can be derived as follows - Equation 4 depicts the Bernoulli's principle and increment in sensation represented by $\Delta\gamma$ is proportional to change in stimulus represented by $\Delta\beta$.

$$\gamma = b \times \log \frac{\beta}{\alpha} \tag{4}$$

were b is a constant and α is a threshold. To evaluate the increment, the following equation 5 and the ensuing simplification gives -

$$\begin{aligned} \Delta\gamma &= b \times \log \frac{\beta + \Delta\beta}{\alpha} - b \times \log \frac{\beta}{\alpha} \\ &= b \times \log(\frac{\beta + \Delta\beta}{\beta}) \\ &= b \times \log(1 + \frac{\Delta\beta}{\beta}) \end{aligned} \tag{5}$$

Since b is a constant, equation 5 reduces to

$$\Delta\gamma \circ \frac{\Delta\beta}{\beta} \tag{6}$$

were $\circ$ means "is constant when there is constancy of" from Masin *et al.*[19]. The final equation 6 is a formulation of Weber's laws in wordings and thus Bernoulli's principles imply Weber's law as a special case. Using Fechner[68] derivation, it is possible to show the relation between Bernoulli's principles and Weber's law. Starting from the last line of equation 5, the following yields the relation.

$$\begin{aligned} \Delta\gamma &= b \times \log(1 + \frac{\Delta\beta}{\beta}) \\ e^{\Delta\gamma} &= e^{b \times \log(1 + \frac{\Delta\beta}{\beta})} \\ k_p &= e^{\log(1 + \frac{\Delta\beta}{\beta})^b}; \text{ were } k_p = e^{\Delta\gamma} \\ k_p &= (1 + \frac{\Delta\beta}{\beta})^b; \text{ since } e^{\log(x)} = x \\ \sqrt[b]{k_p} &= 1 + \frac{\Delta\beta}{\beta} \\ k_q - 1 &= \frac{\Delta\beta}{\beta}; \text{ were } \sqrt[b]{k_p} = k_q \\ k_r &= \frac{\Delta\beta}{\beta}; \text{ the weber's law s.t. } k_r = \sqrt[b]{e^{\Delta\gamma}} - 1 \end{aligned} \tag{7}$$

Equation 6 holds true given the last line of equation 7. In the current study, observation of deviation recorded in predicted values of state of $TRCMPLX$ conditional on gene evidences show crude logarithmic behaviour which might bolster Weber's law and Bernoulli's principles. But it must be noted that these observations are made on static causal models and observation of the same behaviour in dynamical setting would add more value.

3 Materials and methods

The models purported by Sinha[1] involving the biological knowledge as well as epigenetic information depicted

by $\mathcal{M}_{PBK+EI}$ and biological knowledge excluding epigenetic information $\mathcal{M}_{PBK}$ were used to predict the state of $TRCMPLX$ given the gene evidences. Figure 2 depicts the model $\mathcal{M}_{PBK+EI}$. The predictions were recorded over the varying effect of $TRCMPLX$ on gene regulations via assignment of different values to conditional probability tables (cpt) of $TRCMPLX$ while initializing the aforementioned BN models. This varying effect is represented by the term ETGN in Sinha[1].

As a recapitulation, the design of the experiment is a simple 2-holdout experiment where one sample from the normal and one sample from the tumorous are paired to form a test dataset. Excluding the pair formed in an iteration of 2-hold out experiment the remaining samples are considered for training of a BN model. Thus in a data set of 24 normal and 24 tumorous cases obtained from Jiang *et al.*[22], a training set will contain 46 samples and a test set will contain 2 samples (one of normal and one of tumor). This procedure is repeated for every normal sample which is combined with each of the tumorous sample to form a series of test datasets. In total there will be 576 pairs of test data and 576 instances of training data. Note that for each test sample in a pair, the expression value for a gene is discretized using a threshold computed for that particular gene from the training set. Computation of the threshold has been elucidated in Sinha[1]. This computation is repeated for all genes per test sample. Based on the available evidence from the state of expression of all genes, which constitute the test data, inference regarding the state of both the $TRCMPLX$ and the test sample is made. These inferences reveal information regarding the activation state of the $TRCMPLX$ and the state of the test sample. Finally, for each gene g_i, the conditional probability Pr(g_i = active$|g_k$ evidence) $\forall k$ genes. Note that these probabilities are recorded for both normal and tumor test samples.

Three observations are presented in this manuscript. The **first observation** is regarding the logarithmic deviations in prediction of activation status of $TRCMPLX$ conditional on gene expression evidences. The **second observation** is preservation of some gene gene interactions across different strength of beliefs concerning the affect of $TRCMPLX$. To observe these preservations, first the gene gene interactions have to be constructed from the predicted conditional probabilities of one gene given the evidence of another gene (for all gene evidences taken separately). After the construction, further preprocessing is required before the gene-gene interaction network can be inferred. Finally, the **third observation** is to check whether these laws are prevalent among the gene-gene interactions in the network or not.

Deviation study for model $\mathcal{M}_{PBK+EI}$

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ in Normal	$\Delta\gamma$ in Tumor
0.8	0.1	0.125	0.117783	0.03055427	0.09151151
0.7	0.1	0.1428571	0.1335314	0.01423754	0.09086427
0.6	0.1	0.1666667	0.1541507	0.004384244	0.08052346
0.5	0.1	0.2	0.1823216	0.0005872203	0.07294716
0.8	0.1	0.125	0.117783	0.03055427	0.09151151
0.7	0.2	0.2857143	0.2513144	0.04479181	0.1823758
0.6	0.3	0.5	0.4054651	0.04917605	0.2628992
0.5	0.4	0.8	0.5877867	0.04976327	0.3358464

Table 1 Deviation study for model $\mathcal{M}_{PBK+EI}$. $\Delta\gamma$ - mean value of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) over all runs, γ - Pr($TRCMPLX$ = active|all gene evidences), β - the assigned cpt value of $TRCMPLX$ during initialization of the Bayesian network models and $\Delta\beta$ - the deviation in the assigned values of $TRCMPLX$ during initialization.

4 Results and discussion on observation 1

4.1 Logarithmic-power deviations in predictions of β-$catenin$ transcription complex

Let γ be Pr($TRCMPLX$ = active|all gene evidences), β be the assigned cpt value of $TRCMPLX$ during initialization of the Bayesian network models and $\Delta\beta$ be the deviation in the assigned values of $TRCMPLX$ during initialization. To compute $\Delta\gamma$, the 576 predictions of γ observed at $\beta = 90\%$ is subtracted from the 576 predictions of γ observed at $\beta = 80\%$ and a mean of the deviations recorded. This mean becomes $\Delta\gamma$. The procedure is computed again for different value of β. In this manuscript, the effect of constant and incremental deviations are observed. Tables 1 and 2 represent the deviations for models $\mathcal{M}_{PBK+EI}$ and $\mathcal{M}_{PBK}$, respectively.

Figures 4, 5, 6 and 7 show the deviations represented in tables 1 and 2. Note that the numbers depicted in the tables are scaled in a nonuniform manner for observational purpose in the figures. Unscaled values are represented under the last two columns on the right of tables 1 and 2. Before reading the graphs, note that red indicates deviation of mean of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) in normal test samples, blue indicates deviation of mean of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) in tumor case, green indicates deviations in Weber's law and cyan indicates deviations in Bernoulli's law.

For the case of contant deviations (figure 4) in model $\mathcal{M}_{PBK+EI}$, it was observed that deviations in activation of $TRCMPLX$ conditional on gene evidences for the tumor test samples showed a logarithmic behaviour and were directly proportional to the negative of both the Weber's and Bernoulli's law. This can be seen by the blue curve almost

Deviation study for model $\mathcal{M}_{PBK}$

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ in Normal	$\Delta\gamma$ in Tumor
0.8	0.1	0.125	0.117783	0.1400355	0.1097089
0.7	0.1	0.1428571	0.1335314	0.06442086	0.1877266
0.6	0.1	0.1666667	0.1541507	0.01762791	0.06204044
0.5	0.1	0.2	0.1823216	0.01393517	0.1718198
0.8	0.1	0.125	0.117783	0.1400355	0.1097089
0.7	0.2	0.2857143	0.2513144	0.2044564	0.2974356
0.6	0.3	0.5	0.4054651	0.2220843	0.359476
0.5	0.4	0.8	0.5877867	0.2360195	0.5312958

Table 2 Deviation study for model $\mathcal{M}_{PBK}$. $\Delta\gamma$ - mean value of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) over all runs, γ - Pr($TRCMPLX$ = active|all gene evidences), β - the assigned cpt value of $TRCMPLX$ during initialization of the Bayesian network models and $\Delta\beta$ - the deviation in the assigned values of $TRCMPLX$ during initialization.

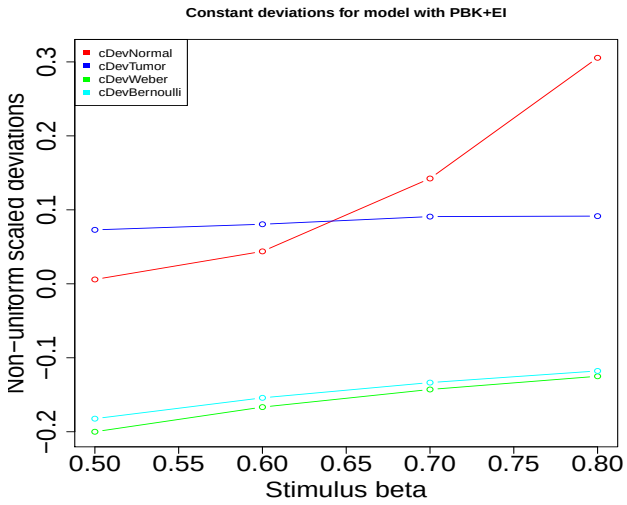


Fig. 4 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - constant deviation in Normal, constant deviation in Tumor, Green - constant deviation in Weber's law, Cyan - constant deviation in Bernoulli's law.

following the green and cyan curves. For the case of deviations in activation of $TRCMPLX$ conditional on gene evidences for the normal test samples showed an exponential behaviour and were proportional to negative of both the Weber's and Bernoulli's law. Similar behaviour was observed for all the coloured curves in case of incremental deviations as shown in figure 5. The exponential behaviour for activation

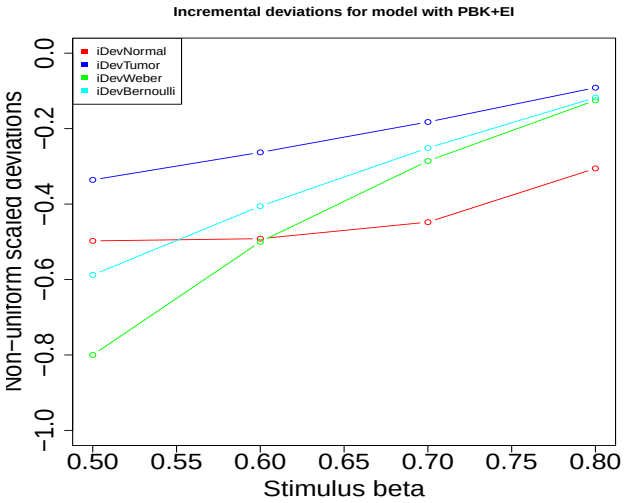


Fig. 5 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - incremental deviation in Normal, incremental deviation in Tumor, Green - incremental deviation in Weber's law, Cyan - incremental deviation in Bernoulli's law.

of $TRCMPLX$ being active conditional on gene evidences correctly supports to the last line of equation 7 which is the derivation of Weber's law from Bernoulli's equation. It actually point to Fechner's derivation of Weber's law from logarithmic formulation.

For model $\mathcal{M}_{PBK}$, the above observations do not yield consistent behaviour. In figure 6, for the case of constant deviations, only the deviations in activation of $TRCMPLX$ conditional on gene evidences for normal test samples exponential in nature and were found to be directly proportional to the negative of both the Weber's and Bernoulli's law. But the deviations in activation of $TRCMPLX$ conditional on gene evidences in tumor test samples show noisy behaviour. But this observation is not the case in incremental deviations for the same model. For the case of incremental deviations as represented in figure 7, the deviations in activation of $TRCMPLX$ conditional on gene evidences is directly proportional to both the Weber's and Bernoulli's law. The figure actually represent the plots with inverted values i.e negative values. A primary reason for this behaviour might be that $\mathcal{M}_{PBK}$ does not capture and constrain the network as much as $\mathcal{M}_{PBK+EI}$ which include epigenetic information. This inclusion of heterogeneous information adds more value to the biologically inspired network and reveals the hidden natural laws occurring in the signaling pathway in both normal and tumor cases.

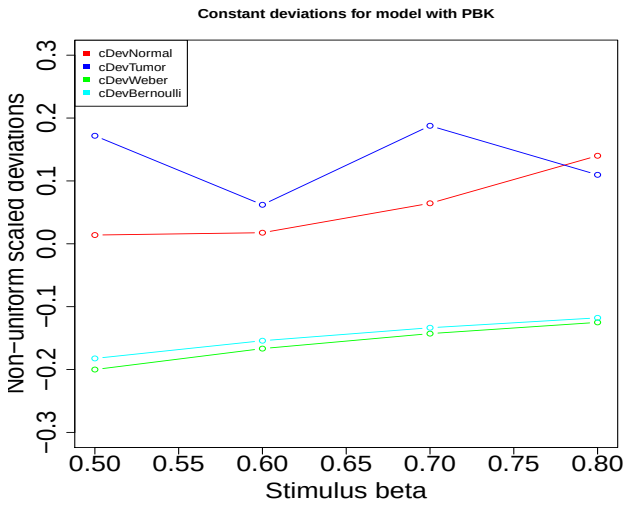


Fig. 6 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active|$\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK}$. Red - constant deviation in Normal, constant deviation in Tumor, Green - constant deviation in Weber's law, Cyan - constant deviation in Bernoulli's law.

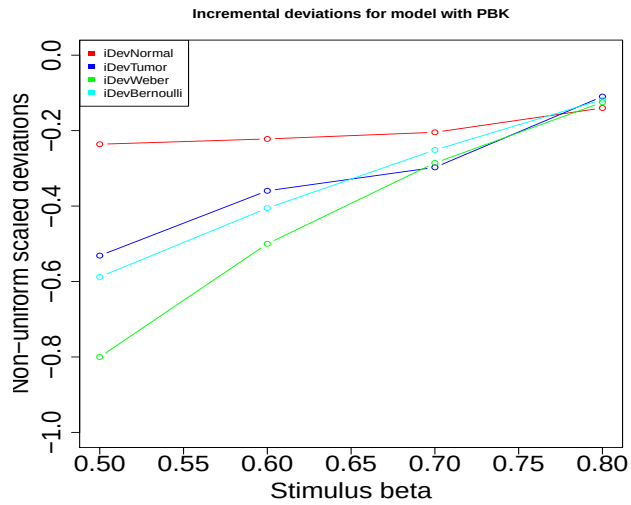


Fig. 7 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active|$\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK}$. Red - incremental deviation in Normal, incremental deviation in Tumor, Green - incremental deviation in Weber's law, Cyan - incremental deviation in Bernoulli's law.

4.2 Intuition behind the curve behaviour

Lastly, the intiutive idea behind the behaviour of the curves generated from constant deviation in table 1 is as follows. It is expected that Pr($TRCMPLX$ = active|all gene evidences) is low (high) in the case of Normal (Tumor) samples. The change ΔPr($TRCMPLX$ = active|all gene evidences) jumps by power of 10 as the β values change from 50% to 90% in Normal cases. It can be observed from the table that there are low deviations in Pr($TRCMPLX$ = active|all gene evidences) when β is low i.e the effect of transcription complex is low and high deviations in Pr($TRCMPLX$ = active|all gene evidences) when β is high i.e the effect of transcription complex is high. But it should be noted that the deviations still tend to be small. This implies that the $TRCMPLX$ is switched off at a constant rate. Thus changes in β leads to exponential curves as in the formulation $\frac{\Delta\beta}{\beta}$, $\Delta\beta \to 0$ and $\beta \to \infty$.

In tumor cases, ΔPr($TRCMPLX$ = active|all gene evidences) behaves near to logarithmic curve as β increases from 50% to 90%. The deviations increase in a slow monotonic way as β increases. Finally, the ratio $\frac{\Delta\beta}{\beta}$ shows monotonically increasing behaviour as $\Delta\beta$ increases proportionally with β. This means that in tumor samples the rate of transcription increases or the effect of rate of transcription complex increases monotonically as β increases. This points to the fold change in β-$catenin$ concentration that might be influencing the transcription rate of the transcription complex. In normal case, the β-$catenin$ concentration remains constant. Due to this, changes in the rate of transcription by the transcription complex might remains constant and near to zero. Change in β values that is the change in initialization of cpt values of transcription complex causes the logarithmic curve in deviations of prediction of transcription complex.

Finally, these observations present a crude yet important picture regarding the downstream transcriptional behaviour of signaling pathway in case of colorectal cancer. Though the current model does not measure the fold changes in the concentration levels of β-$catenin$, it can help in measuring the deviations in activity of the transcription complex conditional on the gene evidences by observing the deviations in the strength of belief assigned as priors in the probability tables of the node representing the transcription complex of the network. Thus sensitivity analysis facilitates in observing such natural phenomena at computational level. In context of the work by Goentoro and Kirschner[17], the presented results are crude in terms of static observations yet they show corresponding behaviour of transcriptional activity in terms of psychophysical laws. Further investigations using dynamic models might reveal more information in comparison to the static models used in Sinha[1]. The observations presented here bolster the existence of behavioural phenomena in terms of logarithmic laws.

$SFRP3$ activation status apropos gene evidences in Normal and Tumor samples using $\theta = 0.5$											
	ge	aa_N	ar_N	ra_N	rr_N	aa_T	ar_T	ra_T	rr_T	gg_{IN}	gg_{IT}
1	DKK1	192	360	24	0	0	216	360	0	DKK1 $\vert-$ <> SFRP3	DKK1 <> $-\vert$ SFRP3
2	DKK2	360	168	0	48	216	217	0	143	DKK2 <> $-$ <> SFRP3	DKK2 $\vert-$ <> SFRP3
3	DKK3-1	360	160	0	56	216	226	0	134	DKK3-1 <> $-$ <> SFRP3	DKK3-1 $\vert-$ <> SFRP3
4	DKK3-2	240	336	0	0	336	240	0	0	DKK3-2 $\vert-$ <> SFRP3	DKK3-2 <> $-$ <> SFRP3
5	DKK4	0	480	96	0	0	116	460	0	DKK4 $\vert-$ <> $SFRP3$	DKK4 <> $-\vert$ SFRP3
6	DACT1	346	230	0	0	216	360	0	0	DACT1 <> $-$ <> SFRP3	DACT1 $\vert-$ <> SFRP3
7	DACT2	312	264	0	0	264	312	0	0	DACT2 <> $-$ <> SFRP3	DACT2 $\vert-$ <> SFRP3
8	DACT3	504	0	0	72	69	0	0	507	DACT3 <> $-$ <> SFRP3	DACT3 $\vert-\vert$ SFRP3
9	SFRP1	552	24	0	0	46	460	0	70	SFRP1 <> $-$ <> SFRP3	SFRP1 $\vert-$ <> SFRP3
10	SFRP2	480	0	0	96	96	480	0	0	SFRP2 <> $-$ <> SFRP3	SFRP2 $\vert-$ <> SFRP3
11	SFRP4	264	312	0	0	312	264	0	0	SFRP4 $\vert-$ <> SFRP3	SFRP4 <> $-$ <> SFRP3
12	SFRP5	460	0	0	116	115	0	0	461	SFRP5 <> $-$ <> SFRP3	SFRP5 $\vert-\vert$ SFRP3
13	WIF1	0	408	168	0	0	178	398	0	WIF1 $\vert-$ <> SFRP3	WIF1 <> $-\vert$ SFRP3
14	LEF1	0	480	96	0	0	92	484	0	LEF1 $\vert-$ <> SFRP3	LEF1 <> $-\vert$ SFRP3
15	MYC	0	456	120	0	0	134	442	0	MYC $\vert-$ <> SFRP3	MYC <> $-\vert$ SFRP3
16	CCND1	0	480	96	0	0	96	480	0	CCND1 $\vert-$ <> SFRP3	CCND1 <> $-\vert$ SFRP3
17	CD44	0	376	200	0	0	192	384	0	CD44 $\vert-$ <> SFRP3	CD44 <> $-\vert$ SFRP3
$SFRP3$ activation status apropos gene evidences in Normal and Tumor samples using $\theta = \theta_N$ and $\theta = \theta_T$											
1	DKK1	0	360	216	0	360	216	0	0	DKK1 $\vert-$ <> SFRP3	DKK1 <> $-$ <> SFRP3
2	DKK2	360	0	0	216	216	360	0	0	DKK2 <> $-$ <> SFRP3	DKK2 $\vert-$ <> SFRP3
3	DKK3-1	360	0	0	216	216	360	0	0	DKK3-1 <> $-$ <> SFRP3	DKK3-1 $\vert-$ <> SFRP3
4	DKK3-2	0	328	240	8	336	240	0	0	DKK3-2 $\vert-$ <> SFRP3	DKK3-2 <> $-$ <> SFRP3
5	DKK4	0	480	96	0	0	116	460	0	DKK4 $\vert-$ <> SFRP3	DKK4 <> $-\vert$ SFRP3
6	DACT1	346	230	0	0	216	360	0	0	DACT1 <> $-$ <> SFRP3	DACT1 $\vert-$ <> SFRP3
7	DACT2	24	0	288	264	264	312	0	0	DACT2 <> $-\vert$ SFRP3	DACT2 $\vert-$ <> SFRP3
8	DACT3	504	0	0	72	69	0	0	507	DACT3 <> $-$ <> SFRP3	DACT3 $\vert-\vert$ SFRP3
9	SFRP1	552	0	0	24	46	530	0	0	SFRP1 <> $-$ <> SFRP3	SFRP1 $\vert-$ <> SFRP3
10	SFRP2	480	0	0	96	96	480	0	0	SFRP2 <> $-$ <> SFRP3	SFRP2 $\vert-$ <> SFRP3
11	SFRP4	0	77	264	235	312	264	0	0	SFRP4 <> $-\vert$ SFRP3	SFRP4 <> $-$ <> SFRP3
12	SFRP5	460	0	0	116	115	411	0	50	SFRP5 <> $-$ <> SFRP3	SFRP5 $\vert-$ <> SFRP3
13	WIF1	0	408	168	0	398	178	0	0	WIF1 $\vert-$ <> SFRP3	WIF1 <> $-$ <> SFRP3
14	LEF1	0	480	96	0	0	92	484	0	LEF1 $\vert-$ <> SFRP3	LEF1 <> $-\vert$ SFRP3
15	MYC	0	456	120	0	0	134	442	0	MYC $\vert-$ <> SFRP3	MYC <> $-\vert$ SFRP3
16	CCND1	0	480	96	0	0	96	480	0	CCND1 $\vert-$ <> SFRP3	CCND1 <> $-\vert$ SFRP3
17	CD44	0	376	200	0	384	192	0	0	CD44 $\vert-$ <> SFRP3	CD44 <> $-$ <> SFRP3

Table 3 $SFRP3$ activation status in test samples conditional on status of individual gene activation (represented by evidence in test data) in Normal and Tumor samples. Measurements are taken over summation of all predicted values across the different runs of the 2-Hold out experiment. Here the notations denote the following: a - active, p - passive, N - Normal, T - Tumor, gg_{IN} - gene-gene interaction with Normal, gg_{IT} - gene-gene interaction with Tumor, <> - active and | - repressed.

5 Preservation of gene gene interactions

The second part of this study was to find interactions between two genes by observing the conditional probability of activation status of one gene given the evidence of another gene. Let g be a gene. To obtain the results, two steps need to be executed in a serial manner. The first step is to construct gene gene interactions based on the available conditional probabilities denoted by Pr(g_i = active/repressed$|g_k$ evidence) $\forall k$ genes. The conditional probabilities are inferred using the junction tree algorithm that employs two-pass message passing scheme. Example code and implementations of the same can be found in Murphy *et al.*[47]. The steps for constructing the gene gene interactions based on these conditional probabilities are documented in the Appendix. The second step is to infer gene gene interaction network based purely on reversible interactions. Note that networks are inferred for gene evidences using normal and tumor test samples separately.

Finally, once the interaction network is ready, the computational empirical estimates for deviations in gene-gene interaction is recorded and observation on the prevalence of psychophysical laws in these interactions is discussed. An im-

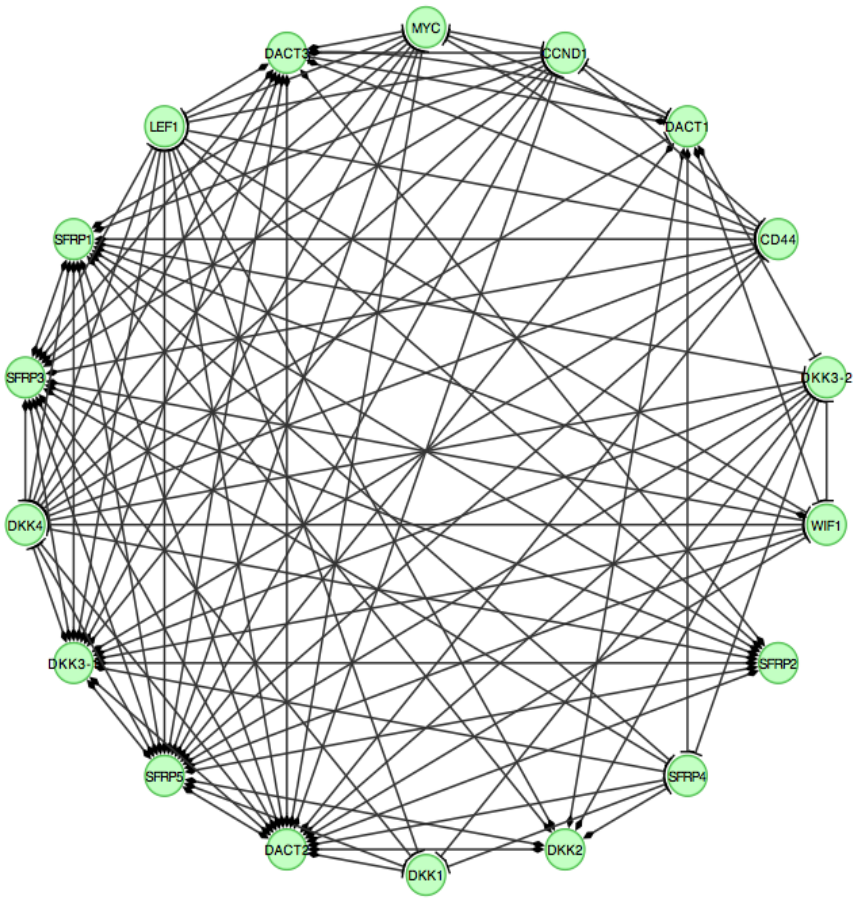


Fig. 8 Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = 0.5$. Note that the effect of initialized cpt for $TRCMPLX$ is 90% in tumorous case and 10% in normal case. Diamond $<>$ means activation and straight bar | means repression.

portant point that needs to be kept in mind is that the inferred interaction network differs based on the choice of the threshold involved (which is a computational issue) but the underlying psychophysical laws remain unchanged (which is a natural phenomena irrespective of the components involved). Thus the while reading the observations on the psychophysical laws, readers must not get confused regarding plots made for interactions from different networks.

5.1 Constructing gene-gene interactions

Gene interactions are constructed by labeling the inferred conditional probability of activation of g_j given the state of g_i, for all $j's \& i's$. Here labels refer to assigning $<>$ for an activated gene and | for repressed gene. Thus the following possible combinations can be inferred - $g_j <> - <> g_i$, $g_j| - |g_i$, $g_j <> -|g_i$ and $g_j|- <> g_i$. Note that all interactions are basically depicting the degree of belief in the state of g_i given or conditional on g_j i.e Pr($g_i|g_j$). The label related to g_i is derived by discretizing Pr($g_i|g_j$) with respect to a weighted mean or the arbitrary value of 0.5. In any interaction, the label associated with g_j is the evidence and the label associated with g_i is the predicted conditional probability. Thus there will always exist two way interactions corresponding to Pr($g_j|g_i$) and Pr($g_i|g_j$) in a Normal case. Similar interactions can be inferred for the Tumor case. Which interactions to select is based on criteria of reversibility and duplication, which is addressed later. To reiterate a final note regarding the interactions - the inferred interactions differ based on the choice of the threshold involved (which is a computational issue) but if prevalent, the underlying psychophysical laws remain unchanged (which is a natural phenomena irrespective of the components involved).

The network obtained by using an arbitrary value like 0.5 for labeling the gene interactions is different from those obtained using a weighted mean. There are advantages of choosing the weighted mean of the training labels for each gene -
• Each gene has an individual threshold that is different from

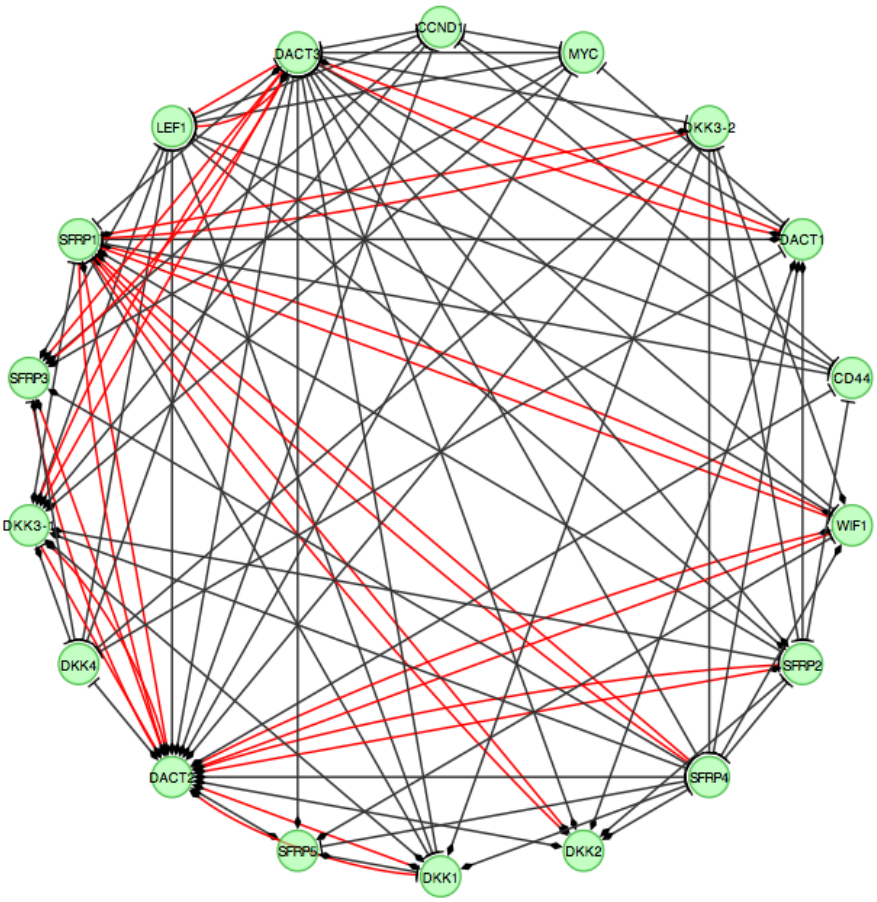


Fig. 9 Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = \theta_N$. Note that the effect of initialized cpt for $TRCMPLX$ is 90% in tumorous case and 10% in normal case. Diamond $<>$ means activation and straight bar | means repression.

the other as the expression values are different and the discretization used to estimate a particluar threshold is based on the median value of the training data for that particular gene under consideration. • The weighted mean assigns appropriate weights to the labels under consideration rather than assigning equal weights which might not represent the actual threshold. • Due to the properties mentioned in the second point, it might be expected that the weighted mean generates a sparse network in comparison to that generated using an arbitrary value of 0.5. • Finally, the weighted mean could reveal interactions between two genes that might be happening at different stages of time. Even though using a static model, capturing such intricate interactions is possible as will be seen later.

There is a formulation for weighted means, but the computation of the weighted mean for training samples belonging to Normal and Tumor is done separately. The separate formulations are given below -

$$\theta_N = \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})}$$
$$\theta_T = \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} \tag{8}$$

were, $n_{1,N}$ is the number of Normal training samples with label 1, $n_{2,N}$ is the number of Normal training samples with label 2, $n_{1,T}$ is the number of Tumor training samples with label 1 and $n_{1,N}$ is the number of Tumor training samples with label 2. Note that the sample labels (i,e evidence of gene expression) were discretized to passive or 1 (active or 2).

Based on the steps described in Appendix, for each gene a matrix is obtained that shows the statistics of how the status of a gene is affected conditional on the individual evidences of the remaining genes. Also, for each of the ith gene the averaged $\widehat{\Pr}_N(g_i|g_k)$ is also stored in vector $PggN$. Same is done for tumor cases. These two vectors are later used to test the

Gene-gene interactions using $\theta = 0.5$
DACT2 <> −| DKK1, SFRP4 | − | DKK1, DACT1 <> − <> DKK2, SFRP1 <> − <> DKK2, LEF1 |− <> DKK2, DKK4 |− <> DKK3-1, DACT3 <> − <> DKK3-1, SFRP2 <> − <> DKK3-1, SFRP3 <> − <> DKK3-1, SFRP5 <> − <> DKK3-1, WIF1 |− <> DKK3-1, LEF1 |− <> DKK3-1, MYC |− <> DKK3-1, CCND1 |− <> DKK3-1, CD44 |− <> DKK3-1, DKK1 | − | DKK3-2, DKK2 <> −| DKK3-2, DKK3-1 <> −| DKK3-2, DACT1 <> −| DKK3-2, DACT2 <> −| DKK3-2, SFRP1 <> −| DKK3-2, SFRP4 | − | DKK3-2, DKK3-2 | − | DKK4, DACT3 <> −| DKK4, SFRP2 <> −| DKK4, SFRP3 <> −| DKK4, SFRP5 <> −| DKK4, WIF1 | − | DKK4, LEF1 | − | DKK4, MYC | − | DKK4, CCND1 | − | DKK4, CD44 | − | DKK4, DKK4 | − | DACT1, DACT3 <> −| DACT1, MYC | − | DACT1, CCND1 | − | DACT1, DKK2 <> − <> DACT2, DKK3-1 <> − <> DACT2, DKK4 |− <> DACT2, DACT3 <> − <> DACT2, SFRP1 <> − <> DACT2, SFRP2 <> − <> DACT2, SFRP3 <> − <> DACT2, SFRP4 |− <> DACT2, SFRP5 <> − <> DACT2, WIF1 |− <> DACT2, LEF1 |− <> DACT2, MYC |− <> DACT2, CCND1 |− <> DACT2, CD44 |− <> DACT2, DACT1 <> −| DACT3, DKK3-1 <> − <> SFRP1, DKK4 |− <> SFRP1, SFRP2 <> − <> SFRP1, SFRP3 <> − <> SFRP1, SFRP4 |− <> SFRP1, SFRP5 <> − <> SFRP1, MYC |− <> SFRP1, CCND1 |− <> SFRP1, CD44 |− <> SFRP1, DACT3 <> − <> SFRP2, SFRP3 <> − <> SFRP2, LEF1 |− <> SFRP2, DKK1 |− <> SFRP3, DACT3 <> − <> SFRP3, SFRP5 <> − <> SFRP3, WIF1 |− <> SFRP3, LEF1 |− <> SFRP3, MYC |− <> SFRP3, CCND1 |− <> SFRP3, CD44 |− <> SFRP3, DKK2 <> −| SFRP4, DKK3-1 <> −| SFRP4, DACT1 <> −| SFRP4, SFRP3 <> −| SFRP4, DKK1 |− <> SFRP5, DKK2 <> − <> SFRP5, DKK3-2 |− <> SFRP5, DACT1 <> − <> SFRP5, DACT3 <> − <> SFRP5, SFRP2 <> − <> SFRP5, WIF1 |− <> SFRP5, LEF1 |− <> SFRP5, MYC |− <> SFRP5, CCND1 |− <> SFRP5, CD44 |− <> SFRP5, DKK3-2 | − | WIF1, DACT1 <> −| WIF1, SFRP1 <> − <> WIF1, DKK1 | − | LEF1, DACT3 <> −| LEF1, WIF1 | − | LEF1, MYC | − | LEF1, CCND1 | − | LEF1, CD44 | − | LEF1, DACT3 <> −| MYC, CCND1 | − | MYC, DACT3 <> −| CCND1, DACT3 <> −| CD44, MYC | − | CD44, CCND1 | − | CD44
Gene interaction using $\theta = \theta_N$
DKK3-1 | − | DKK1, DKK3-2 <> −| DKK1, DACT2 | − | DKK1, SFRP4 <> −| DKK1, DACT1 | − | DKK2, SFRP1 | − | DKK2, DKK4 <> −| DKK3-1, DACT2 |− <> DKK3-1, DACT3 | − | DKK3-1, LEF1 <> −| DKK3-1, MYC <> −| DKK3-1, CCND1 <> −| DKK3-1, SFRP1 | − | DKK3-2, DKK3-2 <> − <> DKK4, DKK4 <> − <> DACT1, DACT3 |− <> DACT1, MYC <> − <> DACT1, CCND1 <> − <> DACT1, DKK1 <> −| DACT2, DKK2 | − | DACT2, DKK3-1 | − | DACT2, DKK3-2 <> −| DACT2, DKK4 <> −| DACT2, SFRP1 | − | DACT2, SFRP2 | − | DACT2, SFRP3 | − | DACT2, SFRP4 <> −| DACT2, SFRP5 | − | DACT2, WIF1 <> −| DACT2, LEF1 <> −| DACT2, MYC <> −| DACT2, CCND1 <> −| DACT2, CD44 <> −| DACT2, DKK1 <> − <> DACT3, DKK2 |− <> DACT3, DKK3-1 |− <> DACT3, DKK3-2 <> − <> DACT3, DKK4 <> − <> DACT3, DACT1 |− <> DACT3, DACT2 |− <> DACT3, SFRP2 |− <> DACT3, SFRP3 |− <> DACT3, SFRP4 <> − <> DACT3, SFRP5 |− <> DACT3, WIF1 <> − <> DACT3, LEF1 <> − <> DACT3, MYC <> − <> DACT3, CCND1 <> − <> DACT3, CD44 <> − <> DACT3, DKK1 <> − <> SFRP1, DKK2 |− <> SFRP1, DKK3-1 |− <> SFRP1, DKK3-2 <> − <> SFRP1, DACT1 |− <> SFRP1, DACT2 |− <> SFRP1, DACT3 |− <> SFRP1, SFRP4 <> − <> SFRP1, WIF1 <> − <> SFRP1, CD44 <> − <> SFRP1, DKK2 |− <> SFRP2, DKK3-1 |− <> SFRP2, DKK3-2 <> − <> SFRP2, DACT1 |− <> SFRP2, DACT2 |− <> SFRP2, SFRP1 |− <> SFRP2, SFRP4 <> − <> SFRP2, LEF1 <> −| SFRP2, CD44 <> − <> SFRP2, DKK4 <> −| SFRP3, DACT2 |− <> SFRP3, DACT3 | − | SFRP3, LEF1 <> −| SFRP3, MYC <> −| SFRP3, CCND1 <> −| SFRP3, DKK2 |− <> SFRP4, DKK3-1 |− <> SFRP4, DKK3-2 <> − <> SFRP4, DACT1 |− <> SFRP4, SFRP1 |− <> SFRP4, SFRP3 |− <> SFRP4, DKK1 <> −| SFRP5, SFRP4 <> − <> SFRP5, DKK3-2 <> −| WIF1, DACT2 | − | WIF1, SFRP1 | − | WIF1, SFRP4 <> −| WIF1, SFRP5 |− <> WIF1, DKK1 <> − <> LEF1, DKK4 <> − <> LEF1, DACT3 |− <> LEF1, WIF1 <> − <> LEF1, CCND1 <> − <> LEF1, CD44 <> − <> LEF1, LEF1 <> − <> MYC, MYC <> − <> CCND1, CCND1 <> − <> CD44

Table 4 Tabulated gene gene interactions of figure 8 and 9 using $\mathcal{M}_{PBK+EI}$ obtained in case of Normal samples. Here, the symbols represent the following - <> activation and \| repression/suppression. Note that for Tumor cases, the interaction roles were found to be reversed, ie. <> −\| in normal became \|− <> in tumor, \|− <> in normal became <> −\| in tumor, <> − <> in normal became \| − \| in tumor and \| − \| in normal became <> − <> in tumor.

veracity of existence of psychophysical laws in gene-gene interaction network. Table 3 represents one such tabulation for gene $SFRP3$. For all runs and all test samples, the following was tabulated in table 3 : aa_N - $SFRP3$ is active (a) when a gene is active (a) in Normal (N) sample, ar_N - $SFRP3$ is active (a) when a gene is repressed (r) in Normal (N) sample, ra_N - $SFRP3$ is repressed (r) when a gene is active (a) in Normal (N) sample, rr_N - $SFRP3$ is repressed (r) when a

Missing gene-gene interactions for different values of ETGN using $\theta = 0.5$		
90N-T1	**80N-T1**	(in **90N-T1**) MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, SFRP2 $<> - <>$ SFRP5, CCND1 $\mid - \mid$ MYC, DACT3 $<> -\mid$ CCND1, MYC $\mid - \mid$ CD44 (in **80N-T1**) SFRP5 $<> - <>$ SFRP2, MYC $\mid - \mid$ CCND1
	70N-T1	(in **90N-T1**) DACT3 $<> -\mid$ DACT1, MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, SFRP2 $<> - <>$ SFRP5, CCND1 $\mid - \mid$ MYC, DACT3 $<> -\mid$ CCND1, DACT3 $<> -\mid$ CD44, MYC $\mid - \mid$ CD44 (in **70N-T1**) SFRP5 $<> - <>$ SFRP2, MYC $\mid - \mid$ CCND1
	60N-T1	(in **90N-T1**) DACT3 $<> -\mid$ DACT1, MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, SFRP2 $<> - <>$ SFRP5, CCND1 $\mid - \mid$ MYC, DACT3 $<> -\mid$ CCND1, DACT3 $<> -\mid$ CD44, MYC $\mid - \mid$ CD44 (in **60N-T1**) SFRP5 $<> - <>$ SFRP2, MYC $\mid - \mid$ CCND1
	50N-T1	(in **90N-T1**) CD44 $\mid- <>$ DKK3-1, SFRP1 $<> -\mid$ DKK3-2, CD44 $\mid - \mid$ DKK4, DACT3 $<> -\mid$ DACT1, MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, DKK3-1 $<> - <>$ SFRP1, DKK4 $\mid- <>$ SFRP1, SFRP2 $<> - <>$ SFRP5, DACT1 $<> -\mid$ WIF1, CCND1 $\mid - \mid$ MYC, DACT3 $<> -\mid$ CCND1, DACT3 $<> -\mid$ CD44, MYC $\mid - \mid$ CD44 (in **50N-T1**) SFRP1 $<> - <>$ DKK3-1, CD44 $\mid - \mid$ DKK3-2, SFRP1 $<> -\mid$ DKK4, DKK3-2 $\mid- <>$ SFRP1, SFRP5 $<> - <>$ SFRP2, MYC $\mid- <>$ SFRP2, CCND1 $\mid- <>$ SFRP2, CD44 $\mid - \mid$ SFRP4, MYC $\mid - \mid$ CCND1
Missing gene-gene interactions for different values of ETGN using $\theta = \theta_N$		
90N-T1	**80N-T1**	(in **90N-T1**) MYC $\mid- <>$ DKK3-1, SFRP1 $<> - <>$ DKK3-2, MYC $\mid - \mid$ DACT1 (in **80N-T1**) MYC $\mid - \mid$ SFRP5
	70N-T1	(in **90N-T1**) DKK4 $\mid- <>$ DKK3-1, MYC $\mid- <>$ DKK3-1, SFRP1 $<> - <>$ DKK3-2, MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, SFRP1 $<> -\mid$ SFRP2, SFRP1 $<> -\mid$ SFRP4, CD44 $\mid - \mid$ LEF1 (in **70N-T1**) DKK4 $\mid - \mid$ SFRP5, MYC $\mid - \mid$ SFRP5, CCND1 $\mid - \mid$ SFRP5, DKK2 $<> - <>$ WIF1, DKK3-1 $<> - <>$ WIF1
	60N-T1	(in **90N-T1**) DKK4 $\mid- <>$ DKK3-1, MYC $\mid- <>$ DKK3-1, CCND1 $\mid- <>$ DKK3-1, SFRP1 $<> - <>$ DKK3-2, DACT3 $<> -\mid$ DACT1, MYC $\mid - \mid$ DACT1, CCND1 $\mid - \mid$ DACT1, DACT3 $<> -\mid$ SFRP1, SFRP1 $<> -\mid$ SFRP2 MYC $\mid- <>$ SFRP3, SFRP1 $<> -\mid$ SFRP4, CD44 $\mid - \mid$ LEF1(in **60N-T1**) MYC $\mid - \mid$ SFRP1, MYC $\mid - \mid$ SFRP2, DKK4 $\mid - \mid$ SFRP5, MYC $\mid - \mid$ SFRP5, CCND1 $\mid - \mid$ SFRP5, DKK2 $<> - <>$ WIF1, DKK3-1 $<> - <>$ WIF1, CD44 $\mid- <>$ WIF1
	50N-T1	(in **90N-T1**) DKK4 $\mid- <>$ DKK3-1, MYC $\mid- <>$ DKK3-1, CCND1 $\mid- <>$ DKK3-1, SFRP1 $<> - <>$ DKK3-2, DACT3 $<> -\mid$ DACT1, MYC $\mid-\mid$ DACT1, CCND1 $\mid-\mid$ DACT1, DACT3 $<> -\mid$ SFRP1, SFRP1 $<> -\mid$ SFRP2, MYC $\mid- <>$ SFRP3, SFRP1 $<> -\mid$ SFRP4, DKK4 $\mid - \mid$ LEF1, CCND1 $\mid - \mid$ LEF1, CD44 $\mid - \mid$ LEF1 (in **50N-T1**) DKK4 $\mid- <>$ DKK1, MYC $\mid- <>$ DKK1, CCND1 $\mid- <>$ DKK1, CD44 $\mid- <>$ DKK1, CD44 $\mid- <>$ DKK3-2, MYC $\mid - \mid$ SFRP1, DKK4 $\mid - \mid$ SFRP2, DACT3 $<> - <>$ SFRP2, MYC $\mid - \mid$ SFRP2, CCND1 $\mid - \mid$ SFRP2, MYC $\mid - \mid$ SFRP5, CCND1 $\mid - \mid$ SFRP5, DKK2 $<> - <>$ WIF1, DKK3-1 $<> - <>$ WIF1, DKK4 $\mid- <>$ WIF1, MYC $\mid- <>$ WIF1, CCND1 $\mid- <>$ WIF1, CD44 $\mid- <>$ WIF1

Table 5 Tabulated missing gene gene interactions of figure 8 and 9 using $\mathcal{M}_{PBK+EI}$ obtained in case of Normal samples. Interactions found in Normal samples with 80%, 70%, 60% and 50% effect that are not found with 90% and vice versa have been recorded. Here, the symbols represent the following - $<>$ activation and $\mid$ repression/suppression. Note that for Tumor cases, the interaction roles were found to be reversed, ie. $<> -\mid$ in normal became $\mid- <>$ in tumor, $\mid- <>$ in normal became $<> -\mid$ in tumor, $<> - <>$ in normal became $\mid - \mid$ in tumor and $\mid - \mid$ in normal became $<> - <>$ in tumor.

gene is repressed (r) in Normal (N) sample, aa_T - $SFRP3$ is active (a) when a gene is active (a) in Tumor (T) sample, ar_T - $SFRP3$ is active (a) when a gene is repressed (r) in Tumor (T) sample, pa_T - $SFRP3$ is repressed (r) when a gene is active (a) in Tumor (T) sample, gg_{IN} - interaction of $SFRP3$ given the gene evidence based on majority voting among aa_N, ar_N, ra_N and rr_N and finally, gg_{IT} - interaction of $SFRP3$ given the gene evidence based on majority voting among aa_T, ar_T, ra_T and rr_T. The highest score among aa_N, ar_N, ra_N and rr_N (aa_T, ar_T, ra_T and rr_T) confirms the relation between genes using Normal (Tumor) samples. Activation (repression) for $SFRP3$ is based on discretizing the predicted conditional probability Pr($SFRP3$ = active$|g_j$ evidence) as $\geq \theta$ ($< \theta$). Activation (repression) for a particular gene evidence g_j is done using discrete evidence. In table 3, under the columns gg_{IN} and gg_{IT}, $<>$ implies the gene is active and $|$ implies the gene is repressed or passive.

Gene-gene interaction network when $\theta = 0.5$

Considering only reversible interactions, in table 3 it was found that evidence for $DKK1$ and $DKK4$ show similar repression behaviour as the standard genes $WIF1$, $LEF1$, MYC, $CCND1$ and $CD44$ in Normal (Tumor) test samples. Only, $SFRP5$ and $DACT3$ in Normal (Tumor) test samples shows activation (repression). Conditional on the observed activation status of the genes mentioned above, $SFRP3$ shows activated (repressed) state in Normal (Tumor) test samples. $SFRP3$ showed behaviour similar to $SFRP - 1/2/5$. Since it is known that the activation status of the latter is influenced by epigenetic factors, $SFRP3$ might also be influenced by epigenetic factors.

Irreversible interactions present in table 3 are deleted as they do not provide concrete information regarding the functional roles of the genes in normal and tumor cases. This attributes to one of the following facts (1) noise that corrupts prediction values as can be seen in the columns of aa_N (aa_T), ar_N (ar_T), ra_N (ra_T) and rr_N (rr_T) or (2) other multiple genes might be interacting along with $SFRP3$ in a combined manner and it is not possible to decipher the relation between $SFRP3$ and other genes. This calls for investigation of prediction of $SFRP3$ status conditional on joint evidences of two or more genes (a combinatorial problem with a search space order of $2^{17} - 17$, which excludes 17 cases of individual gene evidences which have already been considered here). Incorporating multiple gene evidences is not a problem while using Bayesian network models as they are designed to compute conditional probabilities given joint evidences also (except at the cost of high computational time).

It is evident that an arbitrary value of $\theta = 0.5$ will not generate appropriate networks. This is due to the fact that 0.5 does not encode the biological knowledge of thresholding while using discretization. To over come this, a weighted mean is employed as shown below.

Gene-gene interaction network when $\theta = \theta_N^{SFRP3}$

While employing the weighted mean as the threshold to discretize Pr($SFRP3$ = active$|g_j$ evidence), the $SFRP3$ gene evidences that constitutes the test data are used. See step 5.a.iii in Appendix. Note that the test evidences for $SFRP3$ are used for two purpose (1) to discretize Pr($SFRP3$ = active$|g_j$ evidence) as discussed above and (2) to compute the probability of activation status of another gene conditional on evidence for $SRFP3$, i.e Pr(g_j = active$|SFRP3$ evidence). Why to use test evidence or labels to compute weighted mean? Since the test evidence for a gene (i.e the discretized label) has been derived using the median computed on the corresponding training data for the same gene, it absolutely fine to use the disrectized test labels to further compute the weighted mean. This is because the median of gene expression is a value which is much higher than the probability value of 1 and cannot be used to discretize a predicted conditional probability value. Also, estimating the density estimates from a small population of gene expression values has its own weakness. To converge on a plausible realistic value the discretized test samples can be used to estimate a weighted mean which represents the summary of the distribution of the discretized values. This weighted mean of $SFRP3$ test samples then discretizes Pr($SFRP3$ = active$|g_j$ evidence) according to the inherently represented summary. More realistic estimates like kernel density estimates could also be used.

In comparison to the interactions derived using $\theta = 0.5$ in table 3, it was found that a more restricted list of $DKK4$, $DACT - 2/3$, $LEF1$, MYC and $CCND1$ showed reversible behaviour with $SFRP3$ using the weighted mean. This reduction in the reversible interactions is due to the fact that the weighted mean carries an idiosyncrasy of the test label data distribution and is more restricted in comparison to the use of 0.5 value that was arbitrarily choosen. Finally, using the proposed weighted mean reveals more than one interaction between two genes. These interactions point to important hidden biological phenomena that require further investigation in the form of wet lab experiments and the ensuing in silico analysis. It also points to the fact that a particular gene may be showing different behaviour at different times in the network while interacting with multiple genes. An example of this will be addressed later. Again, dynamic models will bring more clarity to the picture. Table 3 shows these interactions using $\theta \in \{0.5, \theta_N, \theta_T\}$.

5.2 Inferring gene-gene interaction network

Next, after the construction of gene-gene interactions, it is necessary to infer the network. The inference of the estimated gene-gene interactions network is based on explicitly reversible roles in Normal and Tumor test samples. This means that only those interactions are selected which show

the following property - $g_j <> - <> g_i$ in Normal if and only if $g_j| - |g_i$ in Tumor, $g_j <> -|g_i$ in Normal if and only if $g_j|- <> g_i$ in Tumor, $g_j|- <> g_i$ in Normal if and only if $g_j <> -|g_i$ in Tumor and finally, $g_j| - |g_i$ in Normal if and only if $g_j <> - <> g_i$. This restricts the network to only reversible gene-gene interactions in Normal and Tumor cases. Note that an interaction $g_j \mathcal{IR} g_i$ ($g_i \mathcal{IR} g_j$) is depicted by Pr($g_i|g_j$) (Pr($g_j|g_i$)).

Reversibility helps in tracking the behaviour of gene-gene interaction in both normal and tumor case simultaneously and thus give more weightage to confirmatory results. Irreversible reactions here mean that the state of activation of a gene in both normal and tumor sample remains invariant given the evidence of the other gene in the gene-gene interaction. This helps in eliminating the interactions that might not be happening at all from biological perspective. To confirm the computational results wet lab experiments are needed. See table 3 for reversible and irreversible interactions.

Next, duplicate interactions are removed from the network for normal samples. This is repeated for the network based on tumor samples also. This is achieved by removing one of the interactions from the following pairs ($g_j <> - <> g_i$ and $g_i <> - <> g_j$), ($g_j <> -|g_i$ and $g_i|- <> g_j$), ($g_j|- <> g_i$ and $g_i <> -|g_j$) and ($g_j| - |g_i$ and $g_i| - |g_j$). This process is done to remove redundant interactions that are recorded via steps mentioned in construction of gene-gene interaction network. Figure 8 shows one such network after complete network construction, interaction labeling, consideration of reversible interactions and removal of duplicate interactions using Normal test samples with ETGN of 90% in $\mathcal{M}_{PBK+EI}$. For the case of Tumor test samples with ETGN 90% in $\mathcal{M}_{PBK+EI}$, only the reversal of interactions need to be done. Table 4 represents these interactions in figures 8 and 9 in a tabulated form.

Finally, different networks were generated by varying the effect of $TRCMPLX$ (ETGN) and compared for the normal test samples. Table 5 represents the different interactions that were preserved in network from ETGN 90% with respect to networks obtained from ETGN with values of 80%, 70%, 60% and 50%. It was found that most of the genetic interactions depicted in figures 8 and 9 were found to be preserved across the different variations in ETGN as shown in table 5. Out of the total n genes which construct a fully connected graph of $\frac{n\times(n-1)}{2}$, it was observed that lesser number of interconnections were preserved. This preservation indicates towards the robustness of the genetic contributions in the Wnt signaling pathway in both normal and tumor test samples. Note that these observations are made from static models and dynamic models might reveal greater information.

Deviation study for $SFRP3$ and MYC for normal case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ Pr($SFPR3\|MYC$)	$\Delta\gamma$ Pr($MYC\|SFPR3$)
0.8	0.1	0.125	0.117783	0.003014287	0.00324456
0.7	0.1	0.1428571	0.1335314	0.002766111	0.00324456
0.6	0.1	0.1666667	0.1541507	0.002504868	0.00324456
0.5	0.1	0.2	0.1823216	0.002228110	0.00324456
0.8	0.1	0.125	0.117783	0.010513376	0.01297824
0.7	0.2	0.2857143	0.2513144	0.007499089	0.00973368
0.6	0.3	0.5	0.4054651	0.004732978	0.00648912
0.5	0.4	0.8	0.5877867	0.002228110	0.00324456

Table 6 Deviation study for Pr($SFRP3|MYC$) and Pr($MYC|SFRP3$) for normal case

Deviation study for $SFRP3$ and MYC for tumor case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ Pr($SFPR5\|MYC$)	$\Delta\gamma$ Pr($MYC\|SFPR5$)
0.8	0.1	0.125	0.117783	-0.006463410	0.000000e+00
0.7	0.1	0.1428571	0.1335314	-0.006967724	5.551115e-17
0.6	0.1	0.1666667	0.1541507	-0.007515486	-5.551115e-17
0.5	0.1	0.2	0.1823216	-0.008112496	0.000000e+00
0.8	0.1	0.125	0.117783	-0.029059115	0.000000e+00
0.7	0.2	0.2857143	0.2513144	-0.022595705	0.000000e+00
0.6	0.3	0.5	0.4054651	-0.015627982	-5.551115e-17
0.5	0.4	0.8	0.5877867	-0.008112496	0.000000e+00

Table 7 Deviation study for Pr($SFRP3|MYC$) and Pr($MYC|SFRP3$) for tumor case

6 Results and discussion on observations 2 & 3

6.1 Logarithmic-power deviations in prediction of gene-gene interactions

In the previous section, it was found that some of the interactions remain preserved as there was change in the affect of transcription complex. The first observation of this work was that deviations in the activity of the transcription complex followed a logarithmic-power psychophysical law. The manifestation of these laws at transcriptional levels can be attributed to the fold changes in β-$catenin$ levels and the prevalence of Weber's law observed by Goentoro and Kirschner[17]. In this perspective, it would be interesting to observe if these laws are prevalent among the gene-gene interactions in the network or not.

Case: $<> -|$ **or** $|- <>$ **with** $\theta = \theta_N$

In Sinha[1], the unknown behaviour of $SFRP3$ in the Wnt pathway has been revealed slightly using computational causal inference. In figure 8, $SFRP3$ shows preservation in the network and it's interaction with other genetic factors involved in the model proposed in Sinha[1] has been depicted. In one such paired interaction between $SFRP3$ and MYC, $SFRP3$

showed activation (repression) and MYC showed repression (activation) in normal (tumor) samples. As the change in the effect of transcription complex was induced by changing the initially assigned cpt values for $TRCMPLX$ node, the deviations in the prediction of the gene-gene interaction network was observed to follow the logarithmic-power law crudely. What this means is that deviations or fold changes might also be prevalent at the gene-gene interaction level due to the upstream fold changes in β-$catenin$ that induces transcriptional activity. More specifically, the deviation in the joint interaction that is represented by the degree of belief via the conditional probability of status of one gene given the evidence or activation status regarding another gene i.e Pr($g_i|g_j$), is influenced by the fold changes upstream of the pathway and thus exhibit similar psychophysical laws.

Table 6 and 7 show these deviations in the prediction of the interactions for both the normal and the tumor cases. The tables show how deviations are affected when the changes in the effect of the transcription complex are done at constant and incremental rate. To summarize the results in these tables, graphs were plotted in figures 10 for Pr($SFRP3|MYC$) (constant deviations), 11 for Pr($MYC|SFRP3$) (constant deviations), 12 for Pr($SFRP3|MYC$) (incremental deviations) and 11 for Pr($MYC|SFRP3$) (incremental deviations).

Considering figure 10, when deviations are constant in both Weber and Bernoulli formulation, the deviations in the prediction of Pr($SFRP3|MYC$) is observed to be logarithmic in the normal samples (apropos the Weber and Bernoulli deviations represented by green and cyan curves). Deviation in predictions are depicted by the red (blue) curves for normal (tumor) samples. Such a behaviour is not observed for Pr($MYC|SFRP3$) as is depicted in figure 11. Note that the interaction for $SFRP3$ given MYC was observed to be reversible in normal and tumor cases. But this is not so with the interaction for MYC given $SFRP3$. It might be expected that the non conformance of logarithmic-power law for Pr($MYC|SFRP3$) may be due to the non preservation/existence of the interaction of MYC given $SFRP3$. This is so because Pr($SFRP3|MYC$) depicts a reversible $SFRP3 <> -|MYC$ ($MYC <> -|SFRP3$) in the network on normal (tumor) samples, while Pr($MYC|SFRP3$) does not depict a reversible $MYC|- <> SFRP3$ ($MYC|-|SFRP3$) in the network on normal (tumor) samples.

Similar behaviour was observed in the case of incremental deviations as depicted in figures 12 and 13. Analysis of the behaviour of other gene-gene interactions showing $<> \ -|$ or $|- \ <>$ can be observed in a similar way and can be produced by executing the R code in Weber_Fechner_law.r provided in Google drive `https://drive.google.com/folderview?id=0B7Kkv8wlhPU-T05wTTNodWNydjA&usp=sharing`. Note that plots need manual axis and title adjustments. Some

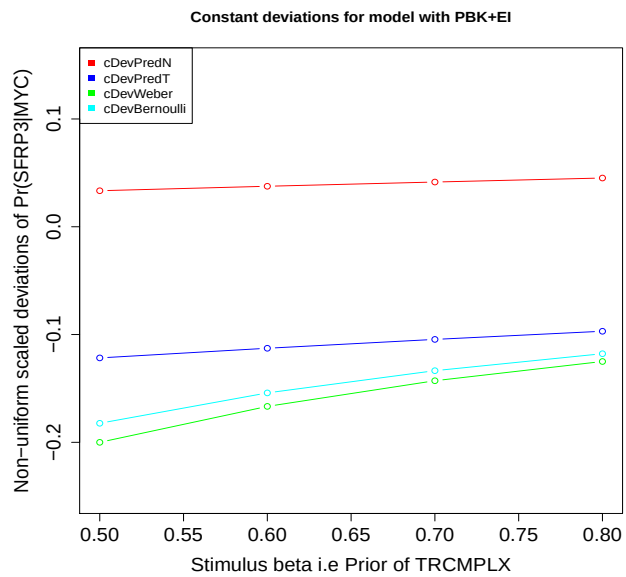


Fig. 10 Constant deviations in β i.e ETGN and corresponding deviations in Pr($SFRP3|MYC$) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - deviation in Pr($SFRP3|MYC$) in Normal case using Weber's law, Blue - deviation in Pr($SFRP3|MYC$) in Tumor using Weber's law, Green - constant deviation in Webers law, Cyan - constant deviation in Bernoullis law.

of the plot results has been compressed in the zip file titled Results-2015.zip.

Case: $|-|$ **or** $<> - <>$ **with** $\theta = \theta_N$

Again, as pointed out in Sinha[1], the unknown behaviour of $SFRP2$ in the Wnt pathway has been captured using computational causal inference. In figure 9, $SFRP2$ shows preservation in the network and it's interaction with other genetic factors involved in the model proposed in Sinha[1] has been depicted. In one such paired interaction between $SFRP2$ and $CD44$, both showed repression (activation) in normal (tumor) samples. As the change in the effect of transcription complex was induced via sensitizing the initially assigned cpt values, the deviations in the prediction of the gene-gene interaction network was observed to follow the logarithmic-power law crudely.

Table 8 and 9 show these deviations in the prediction of the interactions for both the normal and the tumor cases. The tables show how deviations are affected when the changes in the effect of the transcription complex are done at constant and incremental level. To summarize the results in these tables, graphs were plotted in figures 14 for Pr($SFRP2|CD44$) (constant deviations), 15 for Pr($CD44|SFRP2$) (constant de-

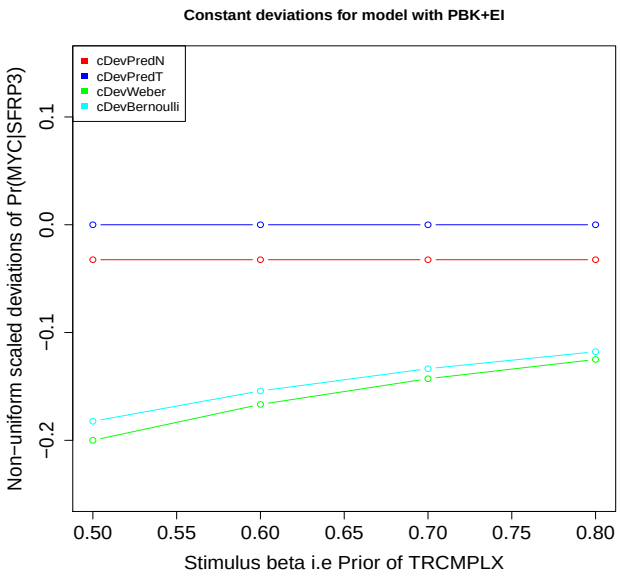


Fig. 11 Same as figure 10 but for Pr($MYC|SFRP3$).

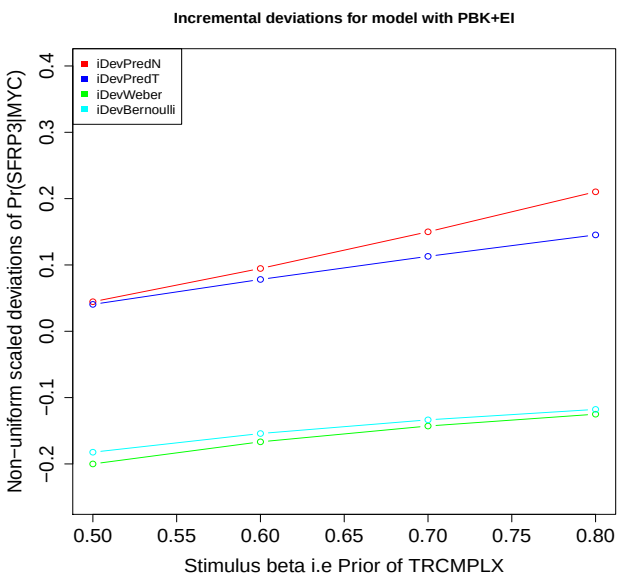


Fig. 12 Same as figure 10 but for Pr($SFRP3|MYC$). Instead of constant deviations, incremental deviations are represented.

viations), 16 for Pr($SFRP2|CD44$) (incremental deviations) and 15 for Pr($CD44|SFRP2$) (incremental deviations).

Considering figure 14, when deviations are constant in both Weber and Bernoulli formulation, the deviations in the prediction of Pr($SFRP2|CD44$) is observed to be logarithmic in the normal samples (apropos the Weber and Bernoulli de-

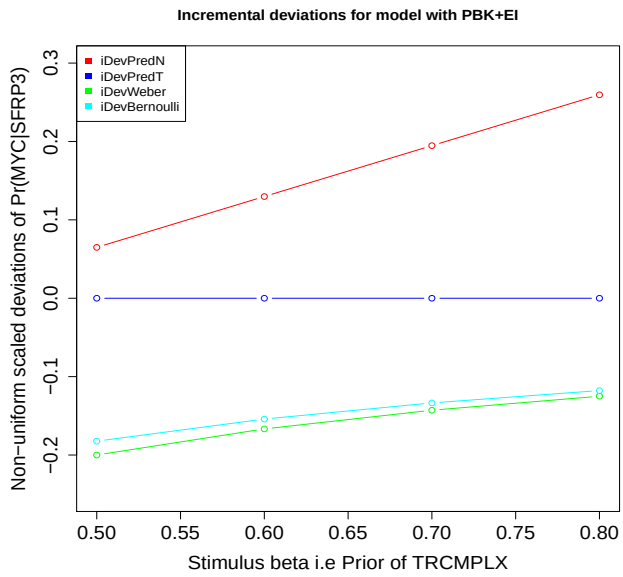


Fig. 13 Same as figure 10 but for Pr($MYC|SFRP3$). Instead of constant deviations, incremental deviations are represented.

viations represented by green and cyan curves). Deviation in predictions are depicted by the red (blue) curves for normal (tumor) samples. Such a behaviour is not observed for Pr($CD44|SFRP2$) as is depicted in figure 15. Even though Pr($CD44|SFRP2$) was computationally estimated through a model, the interaction for $CD44$ given $SFRP2$ was not observed in both normal and tumor cases while the interaction for $SFRP2$ given $CD44$ was observed to be reversible. This points to a crucial fact that the interactions interpretated from conditional probabilities are not always two sided. Thus the interpretation for Pr($g_i|g_j$) is investigated in both directions as $g_i\mathcal{IR}g_j$ and $g_j\mathcal{IR}g_i$ to get a full picture. Not that the results are wrong, but all angles of interpretations need to be investigated to get the picture between any two genes. Similar behaviour was observed in the case of incremental deviations as depicted in figures 16 and 17. Note that graph for incremental deviation in Pr($CD44|SFRP2$) is just a cumulative effect and does not state anything about the logarithmic law.

Finally, note that the predicted conditional probability a gene i given evidence for gene j does not change but the inferred gene-gene interactions do change depending on the choice of the threshold. These changes are depicted in the figures 8 and 9 and table 4. Dual interactions were inferred using the weighted mean as a discretization factor, as is shown next. These are dual interactions are marked in red colour in figure 9.

Case: Dual interactions with $\theta = \theta_N$

The dual interactions revealed using weighted means indi-

Deviation study for $SFRP2$ and $CD44$ for normal case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ Pr($SFRP2\|CD44$)	$\Delta\gamma$ Pr($CD44\|SFRP2$)
0.8	0.1	0.125	0.117783	-0.0007505445	0.0002943409
0.7	0.1	0.1428571	0.1335314	-0.0009398116	0.0002943409
0.6	0.1	0.1666667	0.1541507	-0.0011360011	0.0002943409
0.5	0.1	0.2	0.1823216	-0.0013397022	0.0002943409
0.8	0.1	0.125	0.117783	-0.004166059	0.0011773636
0.7	0.2	0.2857143	0.2513144	-0.003415515	0.0008830227
0.6	0.3	0.5	0.4054651	-0.002475703	0.0005886818
0.5	0.4	0.8	0.5877867	-0.001339702	0.0002943409

Table 8 Deviation study for Pr($SFRP2|CD44$) and Pr($CD44|SFRP2$) for normal case

Deviation study for $SFRP2$ and $CD44$ for tumor case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	$\Delta\gamma$ Pr($SFRP2\|CD44$)	$\Delta\gamma$ Pr($CD44\|SFRP2$)
0.8	0.1	0.125	0.117783	0.02291329	0.01491512
0.7	0.1	0.1428571	0.1335314	0.02132802	0.01491512
0.6	0.1	0.1666667	0.1541507	0.01962443	0.01491512
0.5	0.1	0.2	0.1823216	0.01779600	0.01491512
0.8	0.1	0.125	0.117783	0.08166175	0.05966047
0.7	0.2	0.2857143	0.2513144	0.05874846	0.04474535
0.6	0.3	0.5	0.4054651	0.03742044	0.02983024
0.5	0.4	0.8	0.5877867	0.01779600	0.01491512

Table 9 Deviation study for Pr($SFRP2|CD44$) and Pr($CD44|SFRP2$) for tumor case

cate an important phenomena between any two genes. These interactions reveal that gene activation interplay might not always be constant for normal (tumour) samples. These in silico observations imply that a gene that was found to be actively expressed in normal sample might reverse activity at some stage or the other (an vice versa). Here, one such interaction is discussed in detail. Interpretations of the other dual interactions can be done in the same way. Results for other interactions are available but not presented here.

Also, a point to be observed is that the weighted means show much more crisp discretization during inference of gene-gene interaction in comparison to use of an arbitrary value of 0.5. To determine this distinction between the inferred gene-gene interactions obtained via weighted threshold and the arbitray threshold of 0.5, the receiver operator curves (ROC) along with its corresponding area under the curve (AUC) are plotted. The ROCs are plotted using the discretized predicted values and the discretized labels obtained using the thresholds (computed from the training data) on the test data. The ROC graphs and their respective AUC values indicate how the predictions on the test data behaved under different values assigned to the TRCMPLX while training. Ideally, high values of AUC and steepness in ROC curve indicate good quality results. Finally,

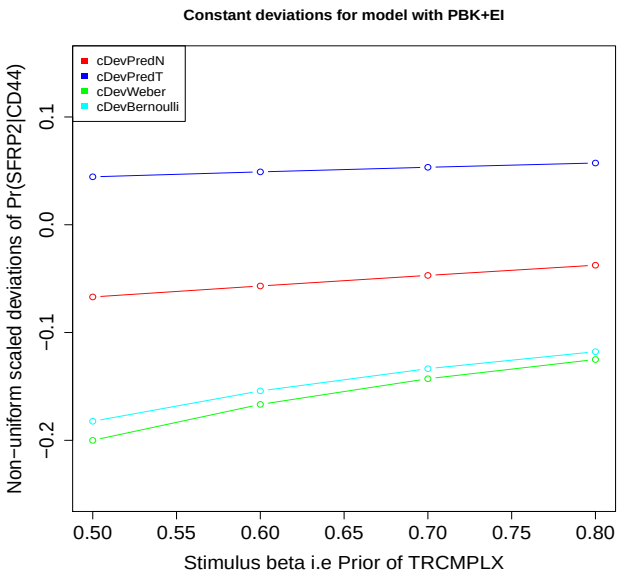


Fig. 14 Constant deviations in β i.e ETGN and corresponding deviations in Pr($SFRP2|CD44$) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - deviation in Pr($SFRP2|CD44$) in Normal case using Weber's law, Blue - deviation in Pr($SFRP2|CD44$) in Tumor using Weber's law, Green - constant deviation in Webers law, Cyan - constant deviation in Bernoullis law.

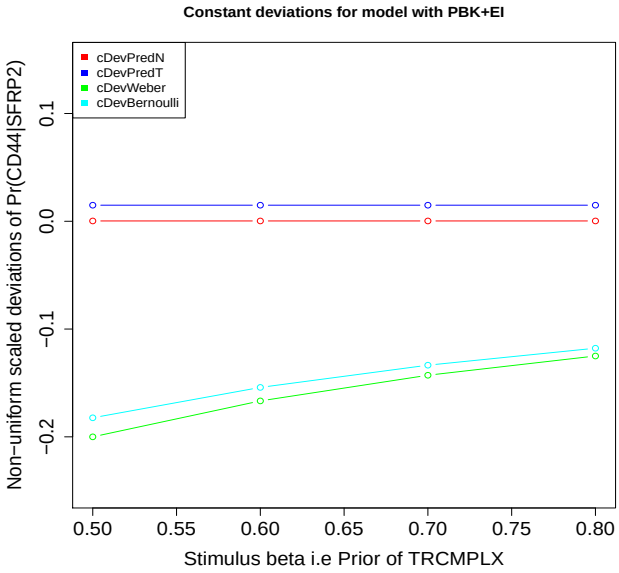


Fig. 15 Same as figure 14 but for Pr($CD44|SFRP2$).

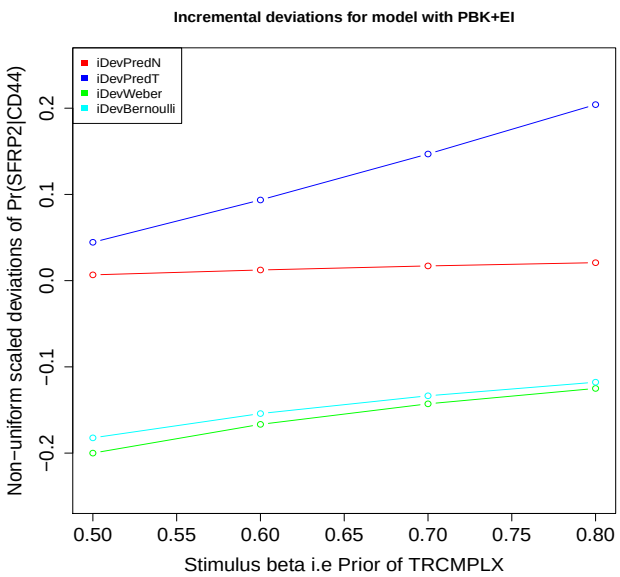


Fig. 16 Same as figure 14 but for Pr($SFRP2|CD44$). Instead of constant deviations, incremental deviations are represented.

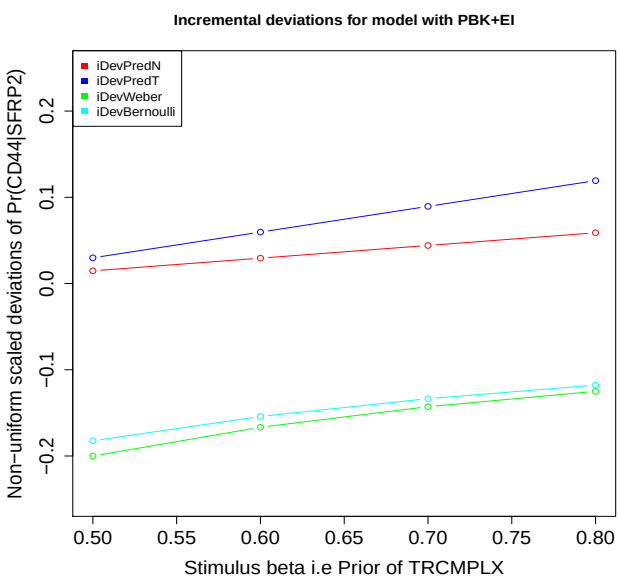


Fig. 17 Same as figure 15 but for Pr($CD44|SFRP2$). Instead of constant deviations, incremental deviations are represented.

two sample Kolmogorov-Smirnov (KS) test was employed to measure the statistical significance between the distribution of predictions. If the cumulative distributions are not similar the KS test returns a small p-value. This small p-value indicates the existing statistical significance between the distributions under consideration.

Finally the ROC plots and AUC values for dual gene-gene interactions are also plotted and KS test is conducted to find the existence of statistical significance if any. These reveal the significance of existence of dual interactions in the signaling pathway which might not have been revealed using the arbitrary threshold value of 0.5. Plots are made using functions from the PRROC package provided by Grau *et al.*[70].

Interaction between DKK1 and DACT2 using $\theta \in \{\theta_N, \theta_T\}$ - Dual interactions $DACT2 <> - <> DKK1$ and $DKK1|- <> DACT2$ ($DACT2| - |DKK1$ and $DKK1 <> -|DACT2$) in normal (tumor) sample were found as depicted in figure 9. Figure 18 shows the kernel density estimate of the predicted conditional probabilities for both normal and tumor test cases. Using the weighted mean of the discretized values of the test samples (discretization done using median estimated from the training data as mentioned before), the predicted Pr($DKK1|DACT2$) and Pr($DACT1|DKK1$) are classified as active or passive. It might be useful to note that instead of using 0.5 as an arbitrary value, the weighted mean captures the distribution of labels in a much more realistic manner and helps infer interactions among the factors in the Wnt pathway.

Note the distributions depicted in figure 18. In the first column of the figure, the median for Pr($DKK1|DACT2$) in normal (tumor) case is 0.4853088 (0.5006437). These medians point to the mid value of the belief in the gene-gene interaction depicted by the range of predicted conditional probabilities. The weighted threshold θ_N^{DKK1} (θ_T^{DKK1}) based on labels for normal (tumor) test case was estimated at 0.5138889 (0.4861111). The estimations come from the following computations in equation 9 -

$$\begin{aligned}
\theta_N^{DKK1} &= \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})} \\
&= \frac{1 \times 264 + 2 \times 312}{3 \times 576} = 0.5138889 \\
\theta_T^{DKK1} &= \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} \\
&= \frac{1 \times 312 + 2 \times 264}{3 \times 576} = 0.4861111
\end{aligned} \tag{9}$$

Similarly, in the second column of the figure, the median for Pr($DACT2|DKK1$) in normal (tumor) case is 0.5606946 (0.2985911). The weighted threshold θ_N^{DACT2} (θ_T^{DACT2}) based on labels for normal (tumor) test case was estimated at 0.4583333 (0.5416667). The estimations come from the

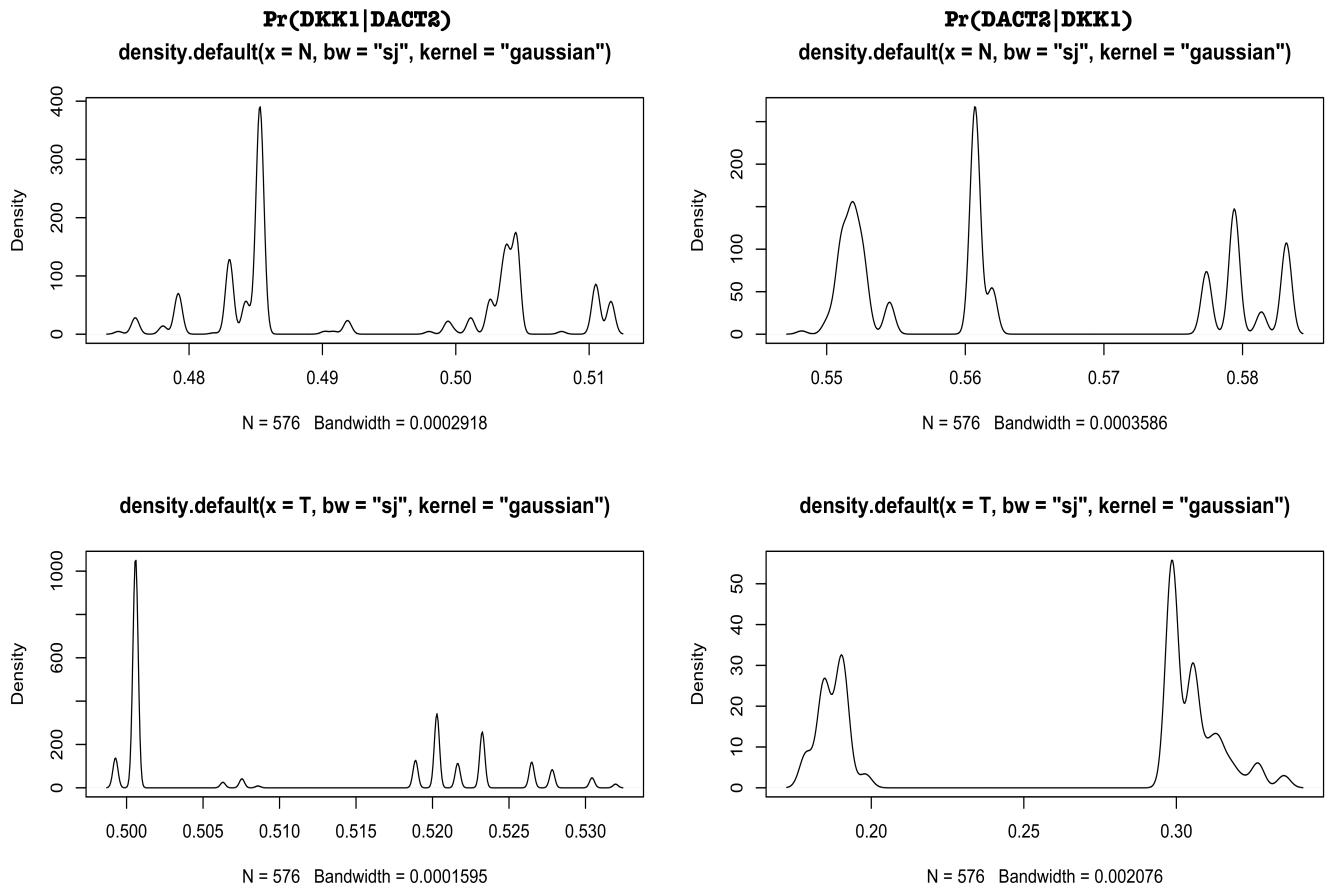


Fig. 18 Kernel density estimates for predicted Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) in Normal and Tumor cases. Gaussian kernel is used for smoothing the density estimate. The bandwidth of the kernel is selected using the pilot estimation of derivative as proposed by Sheather and Jones[71] and implemented in R programming language.

following computations in equation 10 -

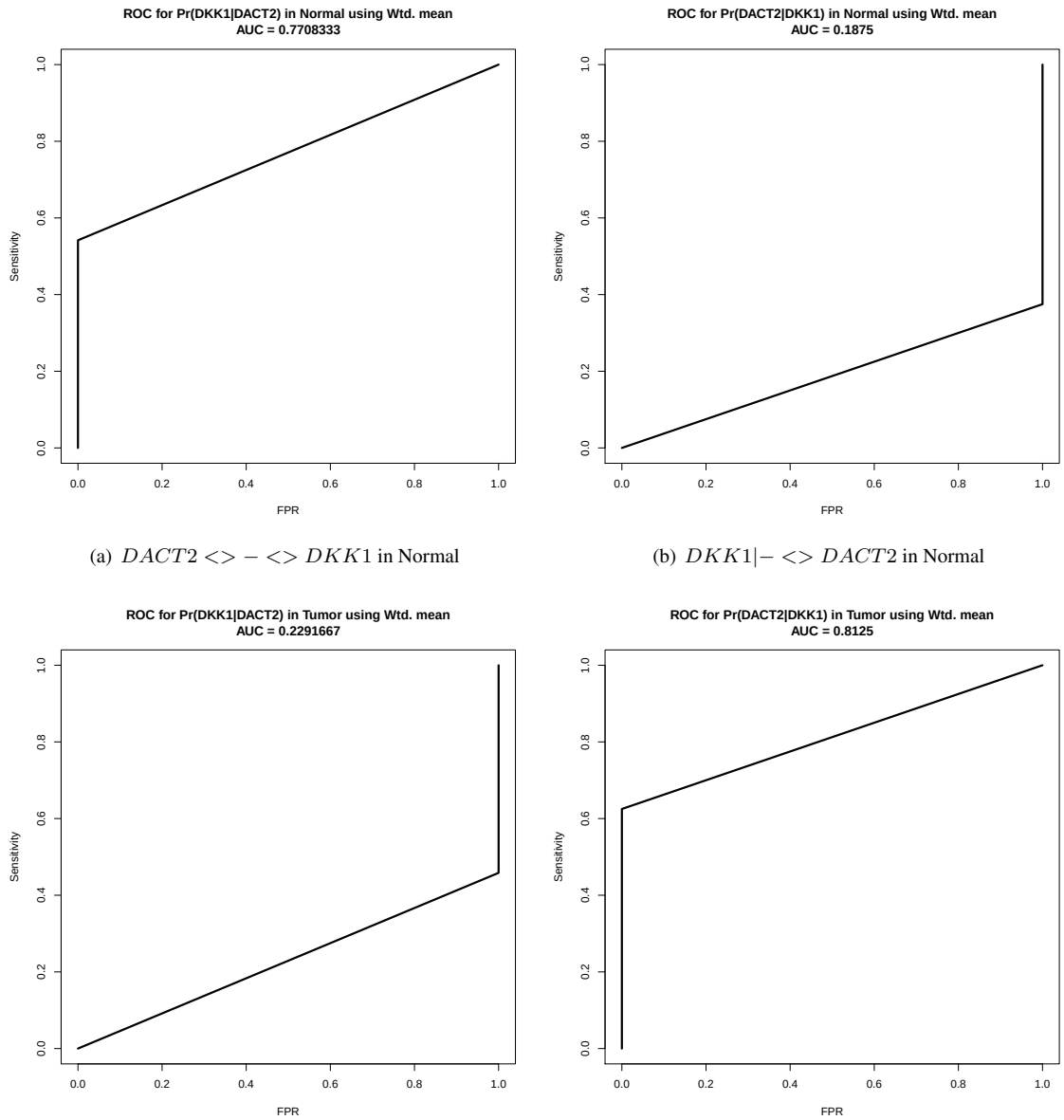
$$\begin{aligned}
\theta_N^{DACT2} &= \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})} \\
&= \frac{1 \times 360 + 2 \times 216}{3 \times 576} = 0.4583333 \\
\theta_T^{DACT2} &= \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} \\
&= \frac{1 \times 216 + 2 \times 360}{3 \times 576} = 0.5416667
\end{aligned}$$

It can be observed that the discretization is more realistic and strict using the weighted threshold rather than using the arbitrary value of 0.5. The multiple peaks point to the different frequencies at which the predicted probabilities were recorded. Note that the probabilities here represent the belief in the activation status and the discretization only caliberates the belief into active and repressed state. To evaluate the results further wet lab tests are needed.

Using these distributions and distributions obtained using artbitrary value, the respective ROC are plotted and corresponding AUC values estimated. Finally, KS test is used to find the existence of statistical significance between the valid permutations of the distributions. These estimates further help derive insights about the interactions at a computational level. Figure 19 shows the ROC plots and the respective AUC values for the dual interactions observed via the in silico experiments. The following are compared -

1. labels of test data ge_N and discretized values of Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) using weighted mean in Normal case
2. labels of test data ge_T and discretized values of Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) using weighted mean in Tumor case

In figure 19 column wise the ROCs for Pr($DKK1|DACT2$)



(a) $DACT2 <> - <> DKK1$ in Normal

(b) $DKK1| - <> DACT2$ in Normal

(c) $DACT2| - |DKK1$ in Tumor

(d) $DKK1 <> -|DACT2$ in Tumor

Fig. 19 Column wise ROCs for Pr($DKK1|DACT2$) (1st column) and Pr($DACT2|DKK1$) (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using weighted mean for Normal case and weighted mean for Tumor case. Respective AUC values for the ROC curves appear on the title of each of the graphs.

(1st column) and Pr($DACT2|DKK1$) (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using weighted mean for Normal case and weighted mean for Tumor case. It can be seen that using the weighted mean, the subfigure 19(a) and 19(d) convey a good guess regarding the type of interaction prevailing in normal and tumor case. Thus $DACT2 <> - <> DKK1$ i.e Pr($DKK1|DACT2$) is highly favoured in Normal case while $DKK1 <> -|DACT2$ i.e Pr($DACT2|DKK1$) is highly favoured in Tumor case. Why this is so is becuase the normal cases show better results in terms of prediction in comparison to the tumor cases. This points to the fact that the interaction $DACT2 <> - <> DKK1$ is strongly supported in the

normal case in comparison to $DACT2| - |DKK1$ which is weakly supported in the tumor case. Even though the algorithm showed that interaction was reversible at computational level, ROC curves and corresponding AUC values indicate weakness in the belief that $DACT2| - |DKK1$ prevails in tumor cases. On the other hand, the interaction depicted by Pr($DACT2|DKK1$) shows higher predictive quality in the tumor case with respect to the normal case. This means that $DKK1 <> -|DACT2$ has more weight in tumor case than its reversible $DKK1|- <> DACT2$ counter part in the normal case. Taken together, the dual interactions do exist but with different strengths of belief as shown conditional probability values. The curves in subfigure 19(b) and 19(c) indicate a bad guess and thus do not support the interactions $DKK1|- <> DACT2$ i.e Pr($DACT2|DKK1$) in Normal case and $DACT2| - |DKK1$ i.e Pr($DKK1|DACT2$) in Tumor case.

Interaction between DKK1 and DACT2 using $\theta = 0.5$

In comparison to use of the weighted θ, the analysis of single interaction using $\theta = 0.5$ is also presented. Figure 8 shows the interaction between $DKK1$ and $DACT2$ as $DACT2 <> -|DKK1$, i.e Pr($DKK1|DACT2$). Using a 0.5 threshold on 18 it is possible to see that discretization of kernel density estimates of Pr($DKK1|DACT2$) induces a degree of belief which is not exactly $0(1)$. This is not the case with Pr($DACT2|DKK1$), were the discretization leads to an exact $0(1)$ which removes the degree of belief. Bayesian networks often represent the degree of belief in terms of some real valued number and exact probabilites of $0(1)$ are considered with suspision.

Figure 20 shows the ROC plots and the respective AUC values for the dual interactions observed via the in silico experiments. The following are compared -

1. labels of test data ge_N and discretized values of Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) using arbitrary value of 0.5 in Normal case

2. labels of test data ge_T and discretized values of Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) using arbitrary value of 0.5 in Tumor case

In figure 20 column wise the ROCs for Pr($DKK1|DACT2$) (1st column) and Pr($DACT2|DKK1$) (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using arbitrary threshold of 0.5 for Normal case and Tumor case. It can be seen that using the a value of 0.5, the subfigure 20(a) conveys a negligibly good guess regarding the type of interaction prevailing in normal. Thus $DACT2 <> -|DKK1$ i.e Pr($DKK1|DACT2$) is highly favoured in Normal case. On the other hand the 20(c) conveys a very bad guess regarding the reversal of interaction in Tumor case for Pr($DKK1|DACT2$). Finally, it was noted that the degree of belief in Pr($DACT2|DKK1$) was not at all recorded via thresholding. Thus even though 20(b) and 20(d) show recorded ROCs but the discretization of 0.5 does not capture the involved interaction. Thus the arbitrary value of 0.5 is not a good factor for inferring interactions.

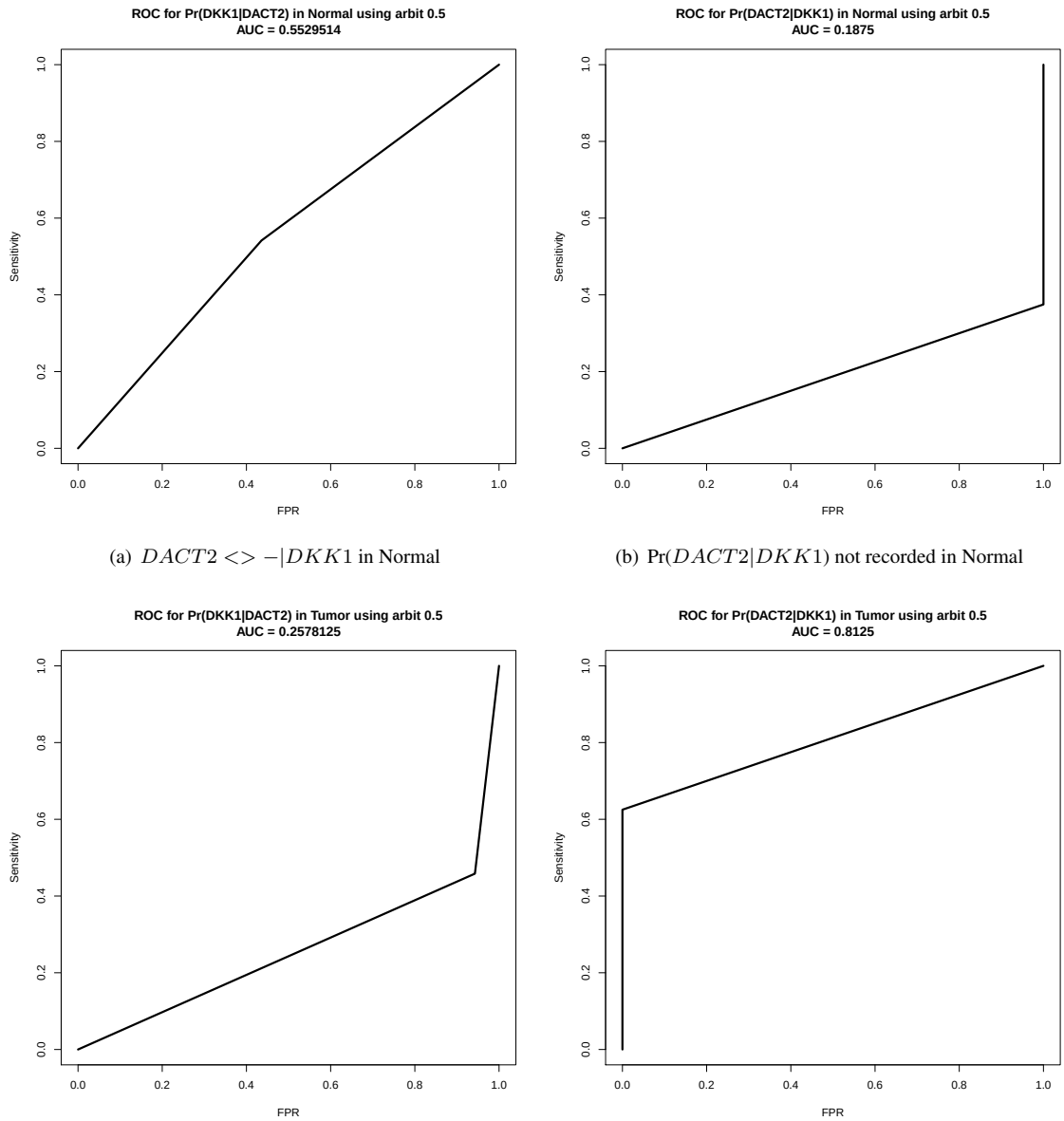
Comparing figures 20 and 19, it is clear that the later gives a better guess in terms of the interpretation of the interaction obtained by discretizing the kernel density estimates of inferred conditional probabilities. To evaluate the statistical significance of the predicted probabilities, the values of the KS test are tabulated and analyzed. Table 10 represents the computed values. The first four rows show the existing significance between the predictions for which the ROC curves have be plotted and described earlier. The next describes the significance between predictions based on thresholds for both normal and tumor cases. Note that some tests show no significance at all as is the case with Pr($DACT2|DKK1$). In general, significance values differ depending on different interactions. Finally, significance values between interactions are also tabulated. It was found that there exists statistical difference between the inferred dual interactions as shown by the low p-values. Similar intepretations can be derived and respective measures can be plotted from the in silico observations.

7 Caveats

This work does not take into account the time series data which contains much more crucial information rather than the static data of gene expression. The inferences have been made regarding a natural phenomena based on the exploration of a computational causal model via sensitivity analysis. The results discussed are based on deviations of inferred conditional probabilities which encode a degree of belief in the occurrence of an event. Even if dynamic bayesian models are used, the observations will be made on degree of beliefs only. Also, the current bayesian network model does not encode the cyclic feedback loops. This has serious implications in the fact that the model might not capture correct interactions. The problem can be overcome to a certain extent by encoding the biological knowledge such that concepts of d-connectivity/separability exploit the inherent prior knowledge and thus help in proper inferences. More specifically, the model captures a snapshot in time but by varying the parameters or the prior/conditional probability tables, it is possible to verify the natural phenomena under investigation.

8 Future directions

In context of the above observations, dynamic models might reveal greater information regarding the psychophysical laws.



(a) $DACT2 <> -|DKK1$ in Normal

(b) Pr($DACT2|DKK1$) not recorded in Normal

(c) $DACT2|- <> DKK1$ in Tumor

(d) Pr($DACT2|DKK1$) not recorded in Tumor

Fig. 20 Column wise ROCs for Pr($DKK1|DACT2$) (1st column) and Pr($DACT2|DKK1$) (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using arbit value of 0.5 for Normal case and arbit value of 0.5 for Tumor case. Respective AUC values for the ROC curves appear on the title of each of the graphs.

Work by Goentoro and Kirschner[17] employs sensitivity analysis methods to reveal such laws by tuning single parameters. There might be a few ways to measure fold change in single an multi parameter settings. Future work might involve deeper study of the phenomena based on multi-parameter setting in a dynamic bayesian network model. If one incorporates nodes in between two time snapshots of β-$catenin$ concentration in a dynamic bayesian network, one might be able to measure the changes at different phases of the signaling pathway. For example, in figure 21 a set of nodes measuring the different concentrations of β-$catenin$ (say N) are depicted. In a dynamic bayesian network, the previous concentration at t is connected to the next concentration at $t+1$. Also, to measure the effect of difference (ΔN), a change in

Kolmogorov-Smirnov test				
Sr. No.	Discretized Val. vs Labels	p-value	Discretized Val. vs Labels	p-value
	Pr($DKK1|DACT2$)		Pr($DACT2|DKK1$)	
1.	wtd. mean (N) vs ge_N	D = 0.5417 p-value $< 2.2e^{-16}$	wtd. mean (N) vs ge_N	D = 0.625 p-value $< 2.2e^{-16}$
2.	wtd. mean (N) vs ge_N	D = 0.1059 p-value = 0.003129	wtd. mean (N) vs ge_N	D = 0.625 p-value $< 2.2e^{-16}$
3.	wtd. mean (N) vs ge_T	D = 0.5417 p-value $< 2.2e^{-16}$	wtd. mean (T) vs ge_T	D = 0.625 p-value $< 2.2e^{-16}$
4.	wtd. mean (N) vs ge_T	D = 0.4844 p-value $< 2.2e^{-16}$	wtd. mean (T) vs ge_T	D = 0.625 p-value $< 2.2e^{-16}$
KS test between predictions using wtd. mean and arbitrary value of 0.5				
Sr. No.	Pr($DKK1|DACT2$)	KS value	Pr($DKK2|DACT1$)	KS value
1.	wtd. mean vs arbit. (N)	D = 0.4358 p-value $< 2.2e^{-16}$	wtd. mean vs arbit. (N)	D = 0 p-value = 1
2.	wtd. mean vs arbit. (T)	D = 0.0573 p-value = 0.3009	wtd. mean vs arbit. (T)	D = 0 p-value = 1
KS test between predictions of interactions I_1 and I_2				
1.	wtd. mean - I_1 (N) vs I_2 (N)	D = 1 p-value $< 2.2e^{-16}$	wtd. mean - I_1 (T) vs I_2 (T)	D = 1 p-value $< 2.2e^{-16}$
2.	arbit. - I_1 (N) vs I_2 (N)	D = 0.5642 p-value $< 2.2e^{-16}$	arbit. - I_1 (T) vs I_2 (T)	D = 0.9427 p-value $< 2.2e^{-16}$

Table 10 Kolmogorov-Smirnov test indicating statistical significance between the distribution of predictions. Statistical significance is evaluated by observing the p-value. Small p-value indicates that significant difference. Significance test is conducted between (1) discretized values of predictions and existing test labels (2) discretized values of predictions based on weighted threshold and discretized values of predictions based on arbit threshold and (3) between predictions representing the dual interactions (obtained using both thresholds). I_1 and I_2 correspond to interactions inferred from Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$), respectively.

concentration can be measured. Computations regarding fold change (ΔN) could then be estimated as posterior probabilities given the two concentrations, which the Bayesian networks can easily handle. In case more parameters need to be involved (say the effect of Wnt and APC together), nodes might be added as shown below. Then the fold change is conditional on $N(t+1)$, $N(t+2)$, ΔWnt and ΔAPC and is estimated as Pr($\Delta N(t+1)|N(t+1), N(t+2), \Delta Wnt, \Delta APC$).

Regarding sensitivity analysis, in nonlinear problems, it might be useful to use Sobol'[72] indices to estimate the sensitivity of the parameters. These indices are a way to estimate the changes in a multiparameter setting thus helping one to conduct global sensitivity analysis instead of local sensitivity analysis (Glen and Isaacs[73]). Finally, with respect to the robustness of the gene-gene interaction network, the current work employs a very simple algorithm to construct the network and infer preserved interactions across the range of values set for a particular parameter. This helps in eliminating interactions that do not contribute enough biological information in the pathway or are non existant and require further analysis by integration of more data. Work in these lines would require incorporation of bigger datasets.

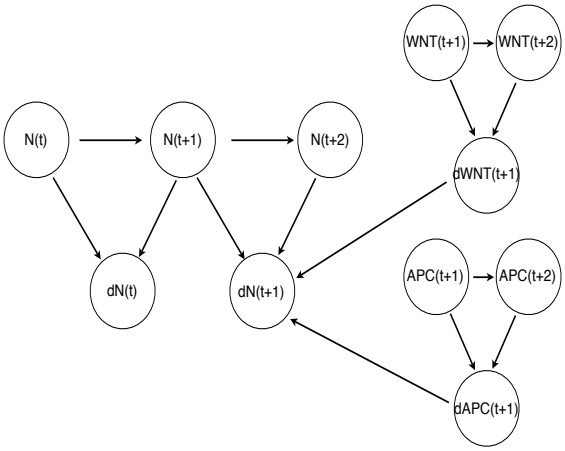


Fig. 21 A schematic diagram of a dynamic bayesian network model that might help study the fold change and the logarithmic psychophysical laws behind the changes.

9 Availability

Code with dataset is made available under GNU GPL v3 license at google code project on `https://sites.google.com/site/shriprakashsinha/shriprakashsinha/projects/static-bn-for-wnt-signaling-pathway` and `https://code.google.com/p/static-bn-for-wnt-signaling-pathway`. Please use the scripts in R as well as the files in zipped directory titled Results-2015.

10 Conclusion

In this preliminary work via sensitivity analysis, the variation in predictive behaviour of β-$catenin$ based transcription complex conditional on gene evidences follows logarithmic psychophysical law crudely. This implies deviations in output are proportional to increasing function of deviations in the input and show constancy for higher values of input. This points towards stability in the behaviour of transcriptional activity downstream of the Wnt pathway. As a further development, this stability reflects the preserved gene-gene interactions of the Wnt pathway inferred from conditional probabilities of individual gene activation given the status of another gene activation derived using biologically inspired Bayesian Network. The prevalence of power-logarithmic psychophysical law is reported for interaction between pairs of ($SFRP3$, MYC), ($SFRP2$, $CD44$) and ($DKK1$, $DACT2$) Finally, based on the sensitivity analysis it was observed that the psychophysical

laws are prevalent among the gene-gene interactions network also.

Conflict of interest

None

Acknowledgement

Thanks to - (1) the Royal Society of Chemistry (RSC) for giving permission to reproduce parts of material in Shriprakash Sinha, Integr. Biol., 2014, DOI: 10.1039/C4IB00124A. (2) all anonymous reviewers who have helped in refining this manuscript.

Appendix

10.1 Steps for construction of gene gene interaction networks

Before starting the construction of interactions from the conditional probabilities, assign a variable gg_I as an empty list (say in R language). Then $\forall i$ genes, execute the following -

1. $\forall$ 576 runs iterated by a counter j
   (a) append x_N with the vector whose elements are Pr(g_i = active$|g_k$ evidence) $\forall k$ genes in the jth run for Normal test sample. This creates a matrix at the end of the runs.
   (b) append x_T with the vector whose elements are Pr(g_i = active$|g_k$ evidence) $\forall k$ genes in the jth run for Tumor test sample. This creates a matrix at the end of the runs.
   (c) append ge_N with the vector whose elements are ge_k evidence $\forall k$ genes in the jth run for Normal test sample. This creates a matrix at the end of the runs.
   (d) append ge_T with the vector whose elements are ge_k evidence $\forall k$ genes in the jth run for Tumor test sample. This creates a matrix at the end of the runs.
2. assign variables ge, aa_N, ar_N, ra_N, rr_N, aa_T, ar_T, ra_T, rr_T, $PggN$, $PggT$ to an empty vector $c()$ (say in R language). Note - a (r) means activation (repression).
3. compute mean across columns of xN and xT to obtain averaged $\widehat{\text{Pr}}_N(g_i|g_k)$ and $\widehat{\text{Pr}}_T(g_i|g_k)$ $\forall k$ gene evidences and $\forall i$ genes. Note $k, i \in 1, ..., n$ if n is the total number of genes.
4. assign a vector of $\widehat{\text{Pr}}_N(g_i|g_k)$ $\forall k$ genes to $PggN$ and a vector of $\widehat{\text{Pr}}_T(g_i|g_k)$ $\forall k$ genes to $PggT$
5. $\forall k$ genes except the ith one
   (a) if($k \neq i$)
      i. assign variables $tmpaa_N$, $tmpar_N$, $tmpra_N$, $tmprr_N$, $tmpaa_T$, $tmpar_T$, $tmpra_T$ and $tmprr_T$ to 0.
      ii. assign threshold values θ to either a fixed value (say 0.5) or a weighted mean.
      iii. if assigning a weighted mean, compute the threshold θ_N as the weighted mean of the labels of the test data i.e evidences for the ith gene, in the case of Normal samples (top formula in equation 8). Similarly, compute the threshold θ_T as the weighted mean of the labels of the test data i.e evidences for the ith gene, in the case of Tumor samples (bottom formula in equation 8).
      iv. $\forall$ 576 runs iterated by a counter l
         A. if($ge_N[l,k] == 1$ and $x_N[l,k] < \theta$) increment $tmprr_N$ by 1
         B. else if($ge_N[l,k] == 1$ and $x_N[l,k] >= \theta$) increment $tmpar_N$ by 1
         C. else if($ge_N[l,k] == 2$ and $x_N[l,k] < \theta$) increment $tmpra_N$ by 1
         D. else if($ge_N[l,k] == 2$ and $x_N[l,k] >= \theta$) increment $tmpaa_N$ by 1
         E. if($ge_T[l,k] == 1$ and $x_T[l,k] < \theta$) increment $tmprr_T$ by 1
         F. else if($ge_T[l,k] == 1$ and $x_T[l,k] >= \theta$) increment $tmpar_T$ by 1
         G. else if($ge_T[l,k] == 2$ and $x_T[l,k] < \theta$) increment $tmpra_T$ by 1
         H. else if($ge_T[l,k] == 2$ and $x_T[l,k] >= \theta$) increment $tmpaa_T$ by 1
      v. Comment - store results
      vi. append ge with g_k, rr_N with $tmprr_N$, ar_N with $tmpar_N$, ra_N with $tmpra_N$, aa_N with $tmpaa_N$, rr_T with $tmprr_T$, ar_T with $tmpar_T$, ra_T with $tmpra_T$ and aa_T with $tmpaa_T$
   (b) store the variables in the previous step to a data frame (say in R language) to a variable $stats$.
   (c) Comment - 1 means aa, 2 means ar, 3 means ra, 4 means rr
   (d) assign variables gg_{IN} and gg_{IT} as empty vector []
   (e) $\forall j$ gene except the ith one under consideration
      i. find the index idx_N in stats that corresponds to 1 or 2 or 3 or 4

ii. if($idx_N == 1$) append gg_{IN} with interaction string $stats\S g_j <> - <> g_i$

iii. else if($idx_N == 2$) append gg_{IN} with interaction string $stats\S ge_j| - <> g_i$

iv. else if($idx_N == 3$) append gg_{IN} with interaction string $stats\S g_j <> -|g_i$

v. else if($idx_N == 4$) append gg_{IN} with interaction string $stats\S g_j| - |g_i$

vi. find the index idx_N in stats that corresponds to 1 or 2 or 3 or 4

vii. if($idx_T == 1$) append gg_{IT} with interaction string $stats\S g_j <> - <> g_i$

viii. else if($idx_T == 2$) append gg_{IT} with interaction string $stats\S g_j| - <> g_i$

ix. else if($idx_T == 3$) append gg_{IT} with interaction string $stats\S g_j <> -|g_i$

x. else if($idx_T == 4$) append gg_{IT} with interaction string $stats\S g_j| - |g_i$

(f) assign $stats\S gg_{IN}$ with gg_{IN}

(g) assign $stats\S gg_{IT}$ with gg_{IT}

(h) Comment - i^{th} gene influenced

(i) $gg_I[[i]] < -$ list(ig = g_i, $stats$ = $stats$, $PggN$ = $PggN$, $PggT$ = $PggT$)

References

1 S. Sinha, *Integr. Biol.*, 2014, **6**, 1034–1048.

2 R. Sharma, *Drosophila information service*, 1973, **50**, 134–134.

3 L. Thorstensen, G. E. Lind, T. Løvig, C. B. Diep, G. I. Meling, T. O. Rognum and R. A. Lothe, *Neoplasia*, 2005, **7**, 99–108.

4 R. Baron and M. Kneissel, *Nature medicine*, 2013, **19**, 179–192.

5 H. Clevers, *Cell*, 2006, **127**, 469–480.

6 S. Sokol, *Wnt Signaling in Embryonic Development*, Elsevier, 2011, vol. 17.

7 D. Pinto, A. Gregorieff, H. Begthel and H. Clevers, *Genes & development*, 2003, **17**, 1709–1713.

8 Z. Zhong, N. J. Ethen and B. O. Williams, *Wiley Interdisciplinary Reviews: Developmental Biology*, 2014, **3**, 489–500.

9 N. Pećina-Šlaus, *Cancer Cell International*, 2010, **10**, 1–5.

10 M. Kahn, *Nature Reviews Drug Discovery*, 2014, **13**, 513–532.

11 K. Garber, *Journal of the National Cancer Institute*, 2009, **101**, 548–550.

12 A. Voronkov and S. Krauss, *Current pharmaceutical design*, 2012, **19**, 634.

13 A. Blagodatski, D. Poteryaev and V. Katanaev, *Mol Cell Ther*, 2014, **2**, 28.

14 J. C. Curtin and M. V. Lorenzi, *Oncotarget*, 2010, **1**, 552.

15 T. P. Rao and M. Kühl, *Circulation research*, 2010, **106**, 1798–1806.

16 J. Yu and D. M. Virshup, *Bioscience reports*, 2014, **34**, 593–607.

17 L. Goentoro and M. W. Kirschner, *Molecular Cell*, 2009, **36**, 872–884.

18 E. H. Weber, *De pulsu resorptione, auditu et tactu*, Annotationes anatomicae et physiologicae, 1834.

19 S. C. Masin, V. Zudini and M. Antonelli, *Journal of History of the Behavioral Sciences*, 2009, **45**, 56–65.

20 J. Z. Sun, G. I. Wang, V. K. Goyal and L. R. Varshney, *Journal of Mathematical Psychology*, 2012, **56**, 495–501.

21 M. Adler, A. Mayo and U. Alon, *PLoS computational biology*, 2014, **10**, e1003781.

22 X. Jiang, J. Tan, J. Li, S. Kivimäe, X. Yang, L. Zhuang, P. Lee, M. Chan, L. Stanton, E. Liu, B. Cheyette and Q. Yu, *Cancer cell*, 2008, **13**, 529–541.

23 W. Verhaegh, P. Hatzis, H. Clevers and A. van de Stolpe, *Cancer Research, San Antonio Breast Cancer Symposium*, 2011, **71**, 524–525.

24 A. Gregorieff and H. Clevers, *Genes & development*, 2005, **19**, 877–890.

25 J. Costello and C. Plass, *Journal of medical genetics*, 2001, **38**, 285–303.

26 P. Das and R. Singal, *Journal of Clinical Oncology*, 2004, **22**, 4632–4642.

27 J. Issa, *Clinical Cancer Research*, 2007, **13**, 1634–1637.

28 H. Suzuki, D. Watkins, K. Jair, K. Schuebel, S. Markowitz, W. Chen, T. Pretlow, B. Yang, Y. Akiyama, M. Van Engeland *et al.*, *Nature genetics*, 2004, **36**, 417–422.

29 C. Niehrs, *Oncogene*, 2006, **25**, 7469–7481.

30 H. Sato, H. Suzuki, M. Toyota, M. Nojima, R. Maruyama, S. Sasaki, H. Takagi, Y. Sogabe, Y. Sasaki, M. Idogawa, T. Sonoda, M. Mori, K. Imai, T. Tokino and Y. Shinomura, *Carcinogenesis*, 2007, **28**, 2459–2466.

31 H. Taniguchi, H. Yamamoto, T. Hirata, N. Miyamoto, M. Oki, K. Nosho, Y. Adachi, T. Endo, K. Imai and Y. Shinomura, *Oncogene*, 2005, **24**, 7946–7952.

32 B. Strahl and C. Allis, *Nature*, 2000, **403**, 41–45.

33 C. Peterson, M. Laniel *et al.*, *Current Biology*, 2004, **14**, 546–551.

34 T. Gardner and J. Faith, *Physics of Life Reviews*, 2005, **2**, 65–88.

35 S. Roehrig, *Decision Support Systems*, 1996, **16**, 55–66.

36 J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*, Morgan Kaufmann, 1988.

37 J. Pearl, *Causality: models, reasoning, and inference*, Cambridge Univ Pr, 2000.

38 E. Charniak, *AI magazine*, 1991, **12**, 50.

39 C. J. Needham, J. R. Bradford, A. J. Bulpitt and D. R. Westhead, *PLoS computational biology*, 2007, **3**, e129.

40 M. Bayes and M. Price, *Philosophical Transactions*, 1763, **53**, 370.

41 N. Friedman, M. Linial, I. Nachman and D. Pe'er, *Journal of computational biology*, 2000, **7**, 601–620.

42 A. Hartemink, D. Gifford, T. Jaakkola and R. Young, Pac. Symp. Biocomput, 2001, pp. 422–433.

43 K. Sachs, D. Gifford, T. Jaakkola, P. Sorger and D. Lauffenburger, *Sci. STKE*, 2002, **148**, 38–42.

44 K. Sachs, O. Perez, D. Pe'er, D. Lauffenburger and G. Nolan, *Science*, 2005, **308**, 523.

45 D. Peer, A. Regev, G. Elidan and N. Friedman, *Bioinformatics*, 2001, **17**, S215.

46 R. Brent and L. Lok, *Science*, 2005, **308**, 504.

47 K. Murphy *et al.*, *Computing science and statistics*, 2001, **33**, 1024–1034.

48 P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, 2003, **13**, 2498–2504.

49 M. Waterman, *Cancer and Metastasis Reviews*, 2004, **23**, 41–52.

50 L. Kriegl, D. Horst, J. Reiche, J. Engel, T. Kirchner, A. Jung *et al.*, *J. Transl. Med*, 2010, **8**, 123.

51 G. Yochum, *PloS one*, 2011, **6**, e18966.

52 K. Schmidt-Ott, T. Masckauchan, X. Chen, B. Hirsh, A. Sarkar, J. Yang, N. Paragas, V. Wallace, D. Dufort, P. Pavlidis *et al.*, *Development*, 2007, **134**, 3177–3190.

53 S. Kanwar, Y. Yu, J. Nautiyal, B. Patel and A. Majumdar, *Molecular cancer*, 2010, **9**, 212.

54 G. Caldwell, C. Jones, P. Taniere, R. Warrack, Y. Soon, G. Matthews and D. Morton, *British journal of cancer*, 2006, **94**, 922–927.

55 N. Reguart, B. He, Z. Xu, L. You, A. Lee, J. Mazieres, I. Mikami, S. Batra, R. Rosell, F. McCormick *et al.*, *Biochemical and biophysical research communications*, 2004, **323**, 229–234.

56 J. González-Sancho, O. Aguilera, J. García, N. Pendás-Franco, C. Peña, S. Cal, A. de Herreros, F. Bonilla and A. Muñoz, *Oncogene*, 2004, **24**, 1098–1103.

57 N. Pendas-Franco, J. Garcia, C. Pena, N. Valle, H. Palmer, M. Heinäniemi, C. Carlberg, B. Jimenez, F. Bonilla, A. Munoz *et al.*, *Oncogene*, 2008, **27**, 4467–4477.

58 S. Baehs, A. Herbst, S. Thieme, C. Perschl, A. Behrens, S. Scheel, A. Jung, T. Brabletz, B. Göke, H. Blum *et al.*, *Cancer letters*, 2009, **276**, 152–159.

59 A. Matsui, T. Yamaguchi, S. Maekawa, C. Miyazaki, S. Takano, T. Uetake, T. Inoue, M. Otaka, H. Otsuka, T. Sato *et al.*, *Cancer science*, 2009, **100**, 1923–1930.

60 M. Zitt, G. Untergasser, A. Amberger, P. Moser, S. Stadlmann, M. Zitt, H. Müller, G. Mühlmann, A. Perathoner, R. Margreiter *et al.*, *Disease Markers*, 2008, **24**, 101–109.

61 J. Veeck and E. Dahl, *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 2012, 18–28.

62 P. Brophy, K. Lang and G. Dressler, *Journal of Biological Chemistry*, 2003, **278**, 52401–52405.

63 Q. Feng Han, W. Zhao, J. Bentel, A. Shearwood, N. Zeps, D. Joseph, B. Iacopetta and A. Dharmarajan, *Cancer letters*, 2006, **231**, 129–137.

64 D. Huang, B. Yu, Y. Deng, W. Sheng, Z. Peng, W. Qin and X. Du, *Journal of cancer research and clinical oncology*, 2010, **136**, 395–401.

65 B. Hoang, M. Moos, S. Vukicevic and F. Luyten, *Journal of Biological Chemistry*, 1996, **271**, 26131–26137.

66 S. Kivimäe, X. Yang and B. Cheyette, *BMC biochemistry*, 2011, **12**, 33.

67 G. Yuan, C. Wang, C. Ma, N. Chen, Q. Tian, T. Zhang and W. Fu, *PloS one*, 2012, **7**, e34004.

68 G. T. Fechner, *Elemente der Psychophysik (2 vols)*, Breitkopf and Hartel, 1860.

69 D. Bernoulli, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 1738, **5**, 175–192.

70 J. Grau, I. Grosse and J. Keilwagen, *Bioinformatics*, 2015, btv153.

71 S. J. Sheather and M. C. Jones, *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991, 683–690.

72 I. M. Sobol', *Matematicheskoe Modelirovanie*, 1990, **2**, 112–118.

73 G. Glen and K. Isaacs, *Environmental Modelling & Software*, 2012, **37**, 157–166.