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

Sensitivity analysis plays a crucial role in observing the behaviour of output of a variable given variations in the input. In colorectal cancer, β-catenin based transcription complex ($TRCMPLX$) plays a major role in driving the Wnt signaling pathway. In this manuscript, the variation in the effect of the predictive behaviour of $TRCMPLX$ conditional on the evidences of regulated 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 logarithmic psychophysical law crudely, implying deviations in output are proportional to increasing function of deviations in input and showing constancy for higher values of input. This points towards stability in the behaviour of $TRCMPLX$ and is reflected in 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.

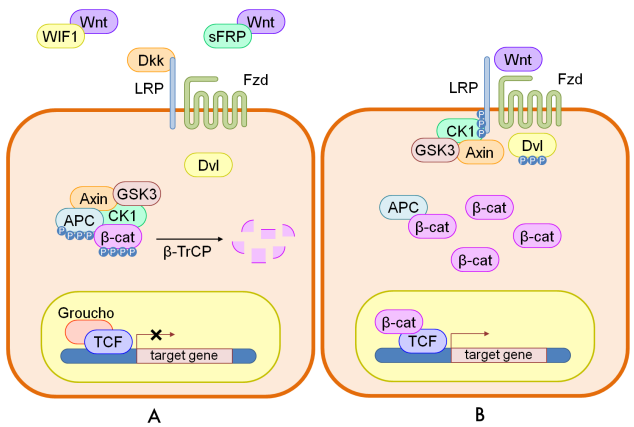

Fig. 1 A cartoon of wnt signaling pathway contributed by Verhaegh *et al.*[2]. Part (A) represents the destruction of β-$catenin$ leading to the inactivation of the wnt target gene. Part (B) represents activation of wnt target gene.

1 Introduction

From Sinha[1], the following two sections help build the background before delving into the problem of sensitivity analysis.

1.1 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[3]). 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[4]). 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

‡ 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] *104-Madhurisha Heights, Risali Sector, Bhilai-490006, India. E-mail: shriprakash.sinha@gmail.com*

active are shown in figures 1(A) and 1(B), respectively.

1.2 Epigenetic factors

One of the widely studied epigenetic factors is methylation (Costello and Plass[5], Das and Singal[6], Issa[7]). Its occurrence 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.*[8]). Similarly, methylation of genes in the Dickkopf ($DKKx$ Niehrs[9], Sato *et al.*[10]), Dapper antagonist of catenin ($DACTx$ Jiang *et al.*[11]) and Wnt inhibitory factor-1 ($WIF1$ Taniguchi *et al.*[12]) 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[13]) can affect gene expression (Peterson *et al.*[14]). 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.*[11]). $DACT1$ 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.*[11]).

Information regarding prior biological knowledge in terms of known influence relations and epigenetic factors have been depicted in the figure represented by figure 2 from Sinha[1].

1.3 Problem statement

In Sinha[1], it has been hypothesized that the activation state of $TRCMPLX$ in the Wnt signaling pathway is not always the same as the state of the test sample (normal/tumorous) 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 Bayesian Network (BN). It was found that the predicted values often increase with an increasing value in the effect of the $TRCMPLX$ on the genes. In a recent development, Goentoro and Kirschner[16] 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's law in sensory physiology.

In accordance with the aforementioned phenomena noted in Sinha[1], it would be important to test the veracity of the observed logarithmic laws and their derivations (like the Weber's law) employed in Goentoro and Kirschner[16]. In the current manuscript, preliminary analysis of results in Sinha[1] shows that the variation in predictive behaviour of $TRCMPLX$ conditional on gene evidences follows power and logarithmic psychophysical law crudely, implying deviations in output are proportional to increasing function of deviations in input and showing constancy for higher values of input. This relates to the work of Adler *et al.*[17] on power and logarithmic law albeit at a coarse level. Also, the result points towards stability in the behaviour of $TRCMPLX$ and is reflected in 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. Note that Weber's law has been found to be a special case of Bernoulli's logarithmic law (Masin *et al.*[18]).

1.4 The logarithmic psychophysical law

Masin *et al.*[18] states the Weber's law as follows -

> Consider a sensation magnitude γ determined by a stimulus magnitude β. Fechner[19] (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[19] (vol 1, p. 65) attributed to the German physiologist Ernst Heinrich Weber the empirical finding Weber[20] that $\Delta\gamma$ remains constant when the relative stimulus increment $\frac{\Delta\beta}{\beta}$ remains constant, and named this finding Weber's law. Fechner[19] (vol 2, p. 10) underlined that Weber's law was empirical.

It has been found that Bernoulli's principle (Bernoulli[21]) is different from Weber's law (Weber[20]) 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.*[18] shows that Weber's law is a special case of Bernoulli's principle and can be derived as follows - Equation 1 depicts the increment in sensation represented by $\Delta\gamma$ to be proportional to change in stimulus represented by $\Delta\beta$.

$$\gamma = b \times \log \frac{\beta}{\alpha} \tag{1}$$

were b is a constant and α is a threshold. To evaluate the increment, the following equation 2 and the ensuing simplification

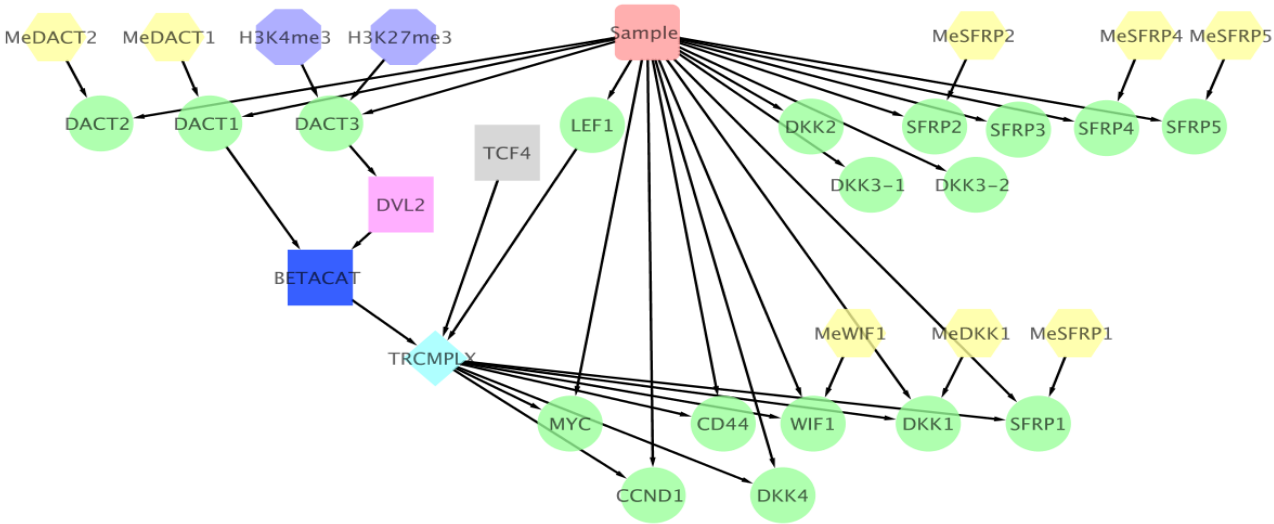

Fig. 2 Influence diagram of $\mathcal{M}_{PBK+EI}$ contains partial prior biological knowledge and epigenetic information in the form of methylation and histone modification. Diagram drawn using Cytoscape[15]. In this model the state of Sample is distinguished from state of $TRCMPLX$ that constitutes the Wnt pathway.

gives -

$$\begin{aligned}\Delta\gamma &= b \times \log\frac{\beta+\Delta\beta}{\alpha} - b \times \log\frac{\beta}{\alpha} \\ &= b \times \log(\frac{\beta+\Delta\beta}{\beta}) \\ &= b \times \log(1+\frac{\Delta\beta}{\beta}) \end{aligned} \tag{2}$$

Since b is a constant, equation 2 reduces to

$$\Delta\gamma \circ \frac{\Delta\beta}{\beta} \tag{3}$$

were $\circ$ means "is constant when there is constancy of" from Masin *et al.*[18]. The final equation 3 is a formulation of Weber's laws in wordings and thus Bernoulli's principles imply Weber's law as a special case. Using Fechner[19] derivation, it is possible to show the relation between Bernoulli's principles and Weber's law. Starting from the last line of equation 2, the following yields the relation.

.

$$\begin{aligned}\Delta\gamma &= b \times \log(1+\frac{\Delta\beta}{\beta}) \\ e^{\Delta\gamma} &= e^{b\times\log(1+\frac{\Delta\beta}{\beta})} \\ k_p &= e^{\log(1+\frac{\Delta\beta}{\beta})^b}; \text{ were } k_p = e^{\Delta\gamma} \\ k_p &= (1+\frac{\Delta\beta}{\beta})^b; \text{ since } e^{\log(x)} = x \\ \sqrt[b]{k_p} &= 1+\frac{\Delta\beta}{\beta} \\ k_q \times \beta &= \beta+\Delta\beta; \text{ were } \sqrt[b]{k_p} = k_q \\ k_q \times \beta - \beta &= \Delta\beta \\ (k_q - 1) \times \beta &= \Delta\beta \\ k_q - 1 &= \frac{\Delta\beta}{\beta} \\ k_r &= \frac{\Delta\beta}{\beta}; \text{ the weber's law s.t. } k_r = \sqrt[b]{e^{\Delta\gamma}} - 1 \end{aligned} \tag{4}$$

Equation 3 holds true given the last line of equation 4. 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.

2 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.*[11], 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.

Two 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 and the second observation is preservation of gene gene interactions across deviations. 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.

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

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr in Normal	Pr 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 $\mathcal{M}_{PBK+EI}$. Pr = mean value of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) over all runs.

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

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr in Normal	Pr 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 $\mathcal{M}_{PBK}$. Pr = mean value of Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) over all runs.

3 Results and discussion

3.1 Logarithmic 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 3, 4, 5 and 6 show the deviations represented in tables 1 and 2. Note that the number depicted in the tables are scaled in a nonuniform manner for observational purpose in the figures. Before reading the graphs, note that red indicates deviation of mean of Pr($TRCMPLX$ = active$|\forall ge_i$

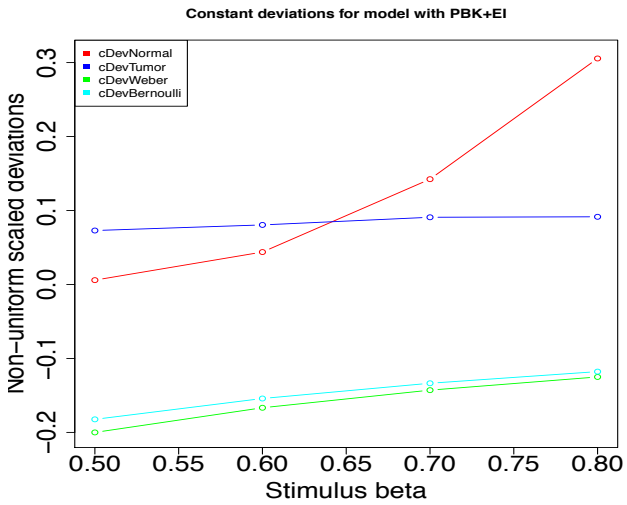

Fig. 3 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse 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.

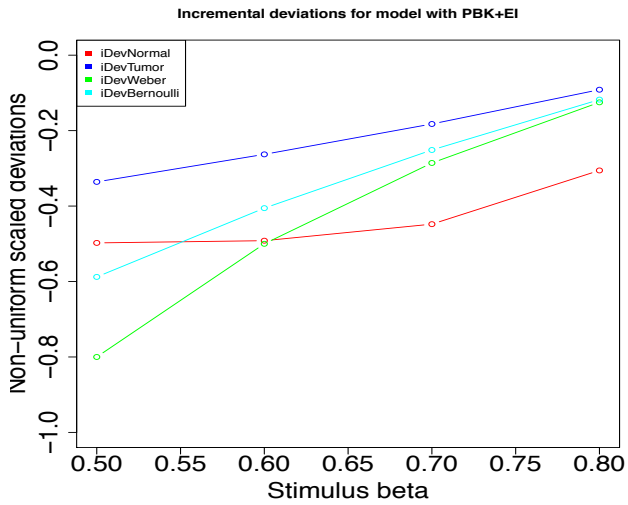

Fig. 4 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually 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.

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 3) 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 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 4. The exponential behaviour for activation of $TRCMPLX$ being active conditional on gene evidences correctly supports to the last line of equation 4 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 5, 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 6, 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.

Finally, these observations present a crude yet important picture regarding the downstream transcriptional behaviour of signaling pathway in case of colorectal cancer. Psychophysical laws might not be constrained to a particular domain and as can be seen here, they might play an important role in shedding light on behaviour of the pathway. In context of Goentoro and Kirschner[16], the presented results might be crude in terms of static observations, yet they show corresponding be-

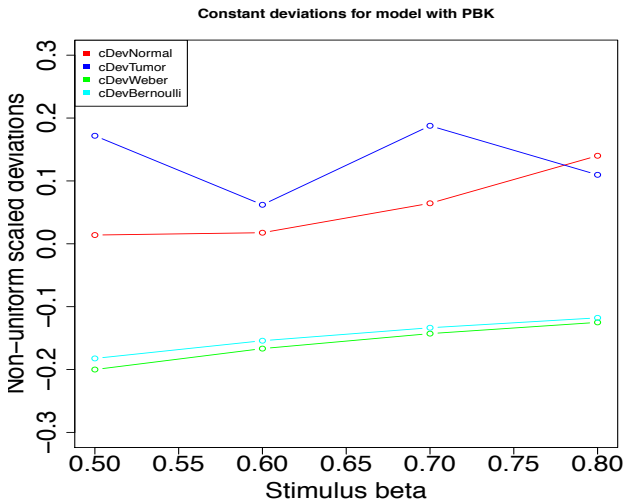

Fig. 5 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse 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.

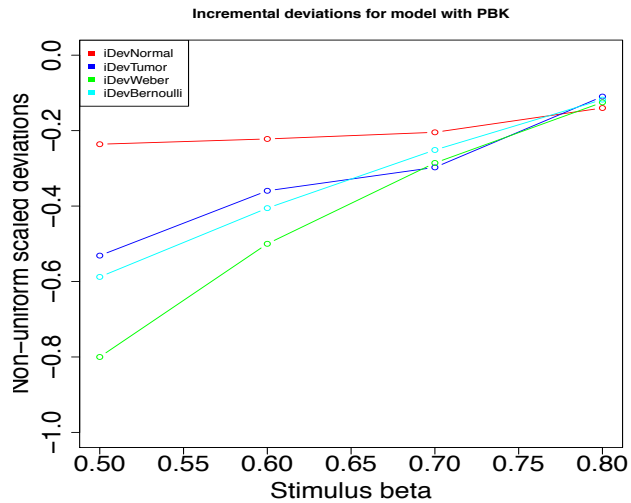

Fig. 6 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually 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.

haviour 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 might bolster the existence of behavioural phenomena in terms of logarithmic laws and its special cases.

3.2 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 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. The following sections elucidate the steps before explaining the implications.

3.2.1 Constructing gene-gene interactions 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 to an empty vector []. Note - a (r) means activation (repression).
3. $\forall k$ 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.

$SFRP5$ activation status apropos to gene evidences in Normal and Tumor samples

	ge	aa_N	ar_N	ra_N	rr_N	aa_T	ar_T	ra_T	rr_T	gg_{IN}	gg_{IT}
1	DKK1	0	360	216	0	0	0	360	216	DKK1 $\|-$ <> SFRP5	DKK1 <> $-\|$ SFRP5
2	DKK2	360	0	0	216	0	0	216	360	DKK2 <> $-$ <> SFRP5	DKK2 $\|-\|$ SFRP5
3	DKK3-1	360	0	0	216	0	0	216	360	DKK3-1 <> $-$ <> SFRP5	DKK3-1 $\|-\|$ SFRP5
4	DKK3-2	133	336	107	0	0	0	336	240	DKK3-2 $\|-$ <> SFRP5	DKK3-2 <> $-\|$ SFRP5
5	DKK4	0	480	96	0	0	0	460	116	DKK4 $\|-$ <> SFRP5	DKK4 <> $-\|$ SFRP5
6	DACT1	346	230	0	0	0	0	216	360	DACT1 <> $-$ <> SFRP5	DACT1 $\|-\|$ SFRP5
7	DACT2	312	218	0	46	0	0	264	312	DACT2 <> $-$ <> SFRP5	DACT2 $\|-\|$ SFRP5
8	DACT3	504	0	0	72	0	0	69	507	DACT3 <> $-$ <> SFRP5	DACT3 $\|-\|$ SFRP5
9	SFRP1	552	0	0	24	0	0	46	530	SFRP1 <> $-$ <> SFRP5	SFRP1 $\|-\|$ SFRP5
10	SFRP2	480	0	0	96	0	0	96	480	SFRP2 <> $-$ <> SFRP5	SFRP2 $\|-\|$ SFRP5
11	SFRP3	484	0	0	92	0	0	96	480	SFRP3 <> $-$ <> SFRP5	SFRP3 $\|-\|$ SFRP5
12	SFRP4	82	312	182	0	312	264	0	0	SFRP4 $\|-$ <> SFRP5	SFRP4 <> $-$ <> SFRP5
13	WIF1	0	408	168	0	0	0	398	178	WIF1 $\|-$ <> SFRP5	WIF1 <> $-\|$ SFRP5
14	LEF1	0	480	96	0	0	0	484	92	LEF1 $\|-$ <> SFRP5	LEF1 <> $-\|$ SFRP5
15	MYC	0	456	120	0	0	0	442	134	MYC $\|-$ <> SFRP5	MYC <> $-\|$ SFRP5
16	CCND1	0	480	96	0	0	0	480	96	CCND1 $\|-$ <> SFRP5	CCND1 <> $-\|$ SFRP5
17	CD44	0	376	200	0	0	0	384	192	CD44 $\|-$ <> SFRP5	CD44 <> $-\|$ SFRP5

Table 3 $SFRP5$ 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.

ii. $\forall$ 576 runs iterated by a counter l

A. if($ge_N[l,k] == 1$ and $x_N[l,k] < 0.5$) increment $tmprr_N$ by 1

B. else if($ge_N[l,k] == 1$ and $x_N[l,k] >= 0.5$) increment $tmpar_N$ by 1

C. else if($ge_N[l,k] == 2$ and $x_N[l,k] < 0.5$) increment $tmpra_N$ by 1

D. else if($ge_N[l,k] == 2$ and $x_N[l,k] >= 0.5$) increment $tmpaa_N$ by 1

E. if($ge_T[l,k] == 1$ and $x_T[l,k] < 0.5$) increment $tmprr_T$ by 1

F. else if($ge_T[l,k] == 1$ and $x_T[l,k] >= 0.5$) increment $tmpar_T$ by 1

G. else if($ge_T[l,k] == 2$ and $x_T[l,k] < 0.5$) increment $tmpra_T$ by 1

H. else if($ge_T[l,k] == 2$ and $x_T[l,k] >= 0.5$) increment $tmpaa_T$ by 1

iii. Comment - store results

iv. 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}

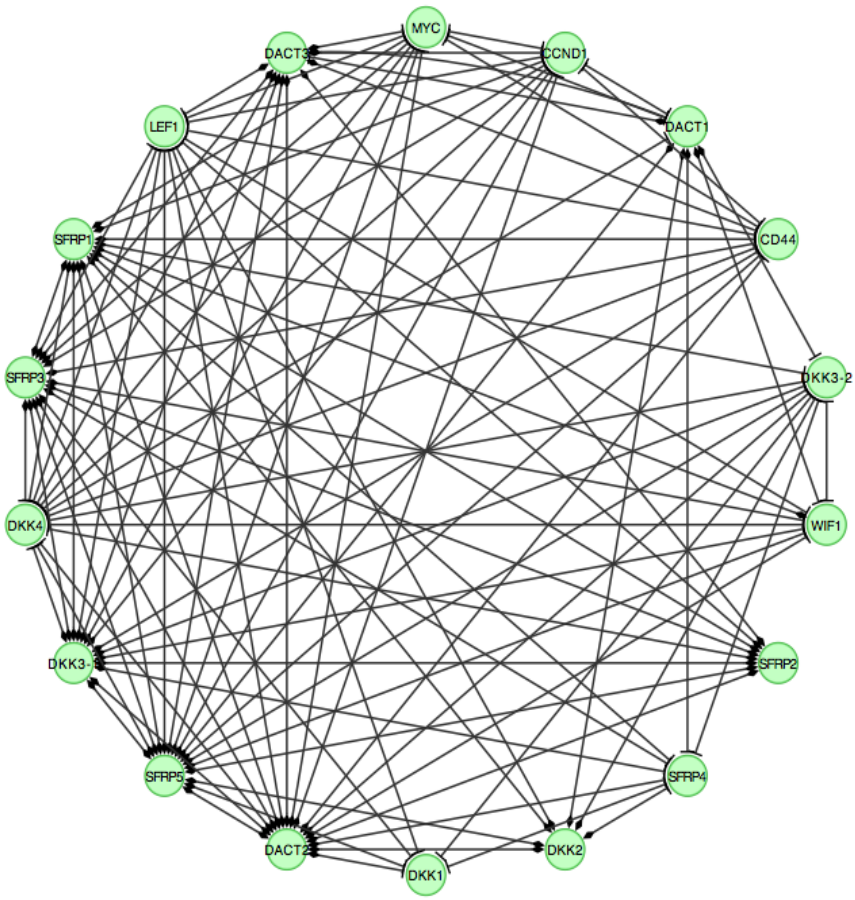

Fig. 7 Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$. Note that the effect of initialized cpt for $TRCMPLX$ is 90% in tumorous case and 10% in normal case. Diamond <> means activation and straight bar | means repression.

(g) assign $stats$§gg_{IT} with gg_{IT}

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

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

Based on the above execution, for each gene a matrix is obtained that shows the statistics of how the status of gene is affected conditional on the individual evidences of the remaining genes. Table 3 represents one such tabulation for gene $SFRP5$. For all runs and all test samples, the following was tabulated in table 3 : aa_N - $SFRP5$ is active (a) when a gene is active (a) in Normal (N) sample, ar_N - $SFRP5$ is active (a) when a gene is repressed (r) in Normal (N) sample, ra_N - $SFRP5$ is repressed (r) when a gene is active (a) in Normal (N) sample, rr_N - $SFRP5$ is repressed (r) when a gene is repressed (r) in Normal (N) sample, aa_T - $SFRP5$ is active (a) when a gene is active (a) in Tumor (T) sample, ar_T - $SFRP5$ is active (a) when a gene is repressed (r) in Tumor (T) sample, pa_T - $SFRP5$ is repressed (r) when a gene is active (a) in Tumor (T) sample, gg_{IN} - interaction of $SFRP5$ given the gene evidence based on majority voting among aa_N, ar_N, ra_N and rr_N and finally, gg_{IT} - interaction of $SFRP5$ 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. Active (repressed) for $SFRP5$ is based on discretization the predicted conditional probability Pr($SFRP5$ = active$|g_j$ evidence) as $\geq$ 0.5 ($<$ 0.5). Active (repressed) 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.

It was found that $DKK1$, $DKK3-2$, $DKK4$ expressed similar repression behaviour as the standard genes $LEF1$, MYC, $CCND1$ and $CD44$ in Normal test samples (and vice versa for Tumor test samples). Also, $DKK2$ and $DKK3-1$ show similar activated behaviour as $DACT-1/2/3$ and $SFRP-1/2/3$ in Normal test samples (and vice versa for Tumor test samples). In comparison to $DKK2$, $DKK3-1$,

Gene-gene interactions
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

Table 4 Tabulated gene gene interactions of figure 7 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.

$DACT-1/2/3$ and $SFRP-1/2/3$, which are activated along with $SRFP5$ in Normal test samples (repressed in Tumor test samples), genes $DKK1$, $DKK3-2$, $DKK4$, $LEF1$, MYC, $CCND1$ and $CD44$ were reversed while $SFRP3$ is activated in Normal test sample (roles reversed in Tumor cases). Genes which showed similar behaviour to $SFRP5$ might be affected by epigenetic factors, i.e these factors might play a role in suppressing the gene expression in Normal test samples. The reverse might be the case for genes that were suppressed in Tumor test samples.

It can also be seen that most of the interactions are reversible except for $SFRP4|- <> SFRP5$ in Normal test sample and $SFRP4 <> - <> SFRP5$ in Tumor test sample. This kind of interaction is deleted the existing set of interactions 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 might corrupt 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 $SFRP5$ in a combined manner and it is not possible to decipher the relation between $SFRP5$ and other genes. This calls for investigation of prediction of $SFRP5$ 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 might not be a problem using Bayesian network models as they are designed to compute conditional probabilities given joint evidences also (except at the cost of high computational time).

3.2.2 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.

Lastly, duplicate interactions are removed from the networks for both Normal and Tumor cases. This removal is done 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 $g_i| - |g_j$). Figure 7 shows one such network after com-

Deviations in gene-gene interactions for different values of ETGN

90N-T1	**80N-T1**	(in **90N-T1**) MYC | − | DACT1, CCND1 | − | DACT1, SFRP2 <> − <> SFRP5, CCND1 | − | MYC, DACT3 <> −| CCND1, MYC | − | CD44 (in **80N-T1**) SFRP5 <> − <> SFRP2, MYC | − | CCND1
	70N-T1	(in **90N-T1**) DACT3 <> −| DACT1, MYC | − | DACT1, CCND1 | − | DACT1, SFRP2 <> − <> SFRP5, CCND1 | − | MYC, DACT3 <> −| CCND1, DACT3 <> −| CD44, MYC | − | CD44 (in **70N-T1**) SFRP5 <> − <> SFRP2, MYC | − | CCND1
	60N-T1	(in **90N-T1**) DACT3 <> −| DACT1, MYC | − | DACT1, CCND1 | − | DACT1, SFRP2 <> − <> SFRP5, CCND1 | − | MYC, DACT3 <> −| CCND1, DACT3 <> −| CD44, MYC | − | CD44 (in **60N-T1**) SFRP5 <> − <> SFRP2, MYC | − | CCND1
	50N-T1	(in **90N-T1**) CD44 |− <> DKK3-1, SFRP1 <> −| DKK3-2, CD44 | − | DKK4, DACT3 <> −| DACT1, MYC | − | DACT1, CCND1 | − | DACT1, DKK3-1 <> − <> SFRP1, DKK4 |− <> SFRP1, SFRP2 <> − <> SFRP5, DACT1 <> −| WIF1, CCND1 | − | MYC, DACT3 <> −| CCND1, DACT3 <> −| CD44, MYC | − | CD44 (in **50N-T1**) SFRP1 <> − <> DKK3-1, CD44 | − | DKK3-2, SFRP1 <> −| DKK4, DKK3-2 |− <> SFRP1, SFRP5 <> − <> SFRP2, MYC |− <> SFRP2, CCND1 |− <> SFRP2, CD44 | − | SFRP4, MYC | − | CCND1

Table 5 Tabulated missing gene gene interactions of figure 7 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 | 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.

plete interaction construction, inference and removal of duplicate interactions in 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 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 figure 7 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.

4 Future Work

In context of the above observations, dynamic models might reveal greater information regarding the psychophysical laws. Work by Goentoro and Kirschner[16] 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 8 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 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'[22] 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[23]). Finally, with respect to the robustness of the gene-gene interaction network, the cur-

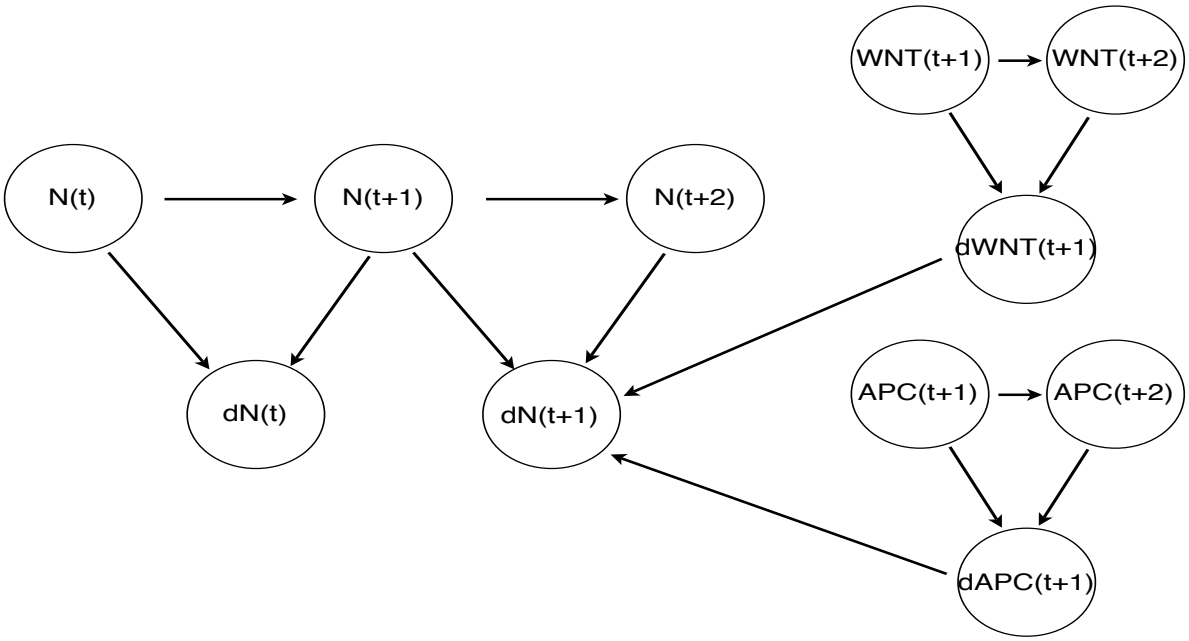

Fig. 8 A schematic diagram of a dynamic bayesian network model that might help study the fold change and the logarithmic psychophysical laws behind the changes.

rent 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.

5 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, implying deviations in output are proportional to increasing function of deviations in 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 might reflect 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.

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. (3) Netherlands Bioinformatics Centre (NBIC) for funding the project. (4) Dr. ir. R. H. J. Fastenau (Dean of Faculty of Electrical Engineering Mathematics and Computer Science), Dr. Prof. Ir. K. Ch. A. M. Luyben (Rector Magnificus) and Drs. D. J. van den Berg (President) at Delft University of Technology, for providing support for conducting this work and giving permission to submit the manuscript. (5) Dr. Wim Verhaegh (faculty at Netherlands Bioinformatics Centre and a principal scientist at Molecular Diagnostic Lab in Philips Research) for providing technical details of Naive Bayes model for replicating experiments in Verhaegh *et al.*[2]. (6) Dr. Marcel J. T. Reinders (scientific director of Bioinformatics Research at Netherlands Bioinformatics Centre and a professor at Delft University of Technology) for refining

the proposed model and suggesting inclusion of methylation data. (7) Dr. Robert P.W. Duin (retired associate professor at Delft University of Technology) for informal discussions on 2-holdout experiments and (8) Dr. Jeroen de Ridder (assistant professor at Delft University of Technology) and (9) Phd candidate Marc Hulsman (Delft University of Technology), for informal discussions on estimation of probability values on methylation data. (10) Lastly, the author is indebted to Mr. Prabhat Sinha and Mrs. Rita Sinha for financially supporting this project while the author was on educational and work leave.

References

1 S. Sinha, *Integr. Biol.*, 2014, **6**, 1034–1048.
2 W. Verhaegh, P. Hatzis, H. Clevers and A. van de Stolpe, *Cancer Research, San Antonio Breast Cancer Symposium*, 2011, **71**, 524–525.
3 H. Clevers, *Cell*, 2006, **127**, 469–480.
4 A. Gregorieff and H. Clevers, *Genes & development*, 2005, **19**, 877–890.
5 J. Costello and C. Plass, *Journal of medical genetics*, 2001, **38**, 285–303.
6 P. Das and R. Singal, *Journal of Clinical Oncology*, 2004, **22**, 4632–4642.
7 J. Issa, *Clinical Cancer Research*, 2007, **13**, 1634–1637.
8 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.
9 C. Niehrs, *Oncogene*, 2006, **25**, 7469–7481.
10 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.
11 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.
12 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.
13 B. Strahl and C. Allis, *Nature*, 2000, **403**, 41–45.
14 C. Peterson, M. Laniel *et al.*, *Current Biology*, 2004, **14**, 546–551.
15 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.
16 L. Goentoro and M. W. Kirschner, *Molecular Cell*, 2009, **36**, 872–884.
17 M. Adler, A. Mayo and U. Alon, *PLoS computational biology*, 2014, **10**, e1003781.
18 S. C. Masin, V. Zudini and M. Antonelli, *Journal of History of the Behavioral Sciences*, 2009, **45**, 56–65.
19 G. T. Fechner, *Elemente der Psychophysik (2 vols)*, Breitkopf and Hartel, 1860.
20 E. H. Weber, *De pulsu resorptione, auditu et tactu*, Annotationes anatomicae et physiologicae, 1834.
21 D. Bernoulli, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 1738, **5**, 175–192.
22 I. M. Sobol', *Matematicheskoe Modelirovanie*, 1990, **2**, 112–118.
23 G. Glen and K. Isaacs, *Environmental Modelling & Software*, 2012, **37**, 157–166.