bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 behaviour of certain intra/extracellular factors working in the Wnt pathway and 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. Recently, it has been found that some factors of the Wnt signaling pathway show a natural behaviour that can often be characterized by psychophysical laws. In this work, the variation in the effect of the predictive behaviour of the transcription complex (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 power and logarithmic psychophysical law crudely, implying deviations in output are proportional to increasing function of deviations in input and shows 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. The deviation in the predictive behaviour of gene activation conditional on the evidence of another regulated gene expression in normal and tumor samples was also observed by varying the initially assigned values of conditional probability tables (cpt) for β-catenin based transcription complex. Analysis shows that the deviation in predictive behaviour of gene activation conditional on another gene expression as evidence, follows power-logarithmic psychophysical law crudely. Finally, the computational observations indicate that not only individual factors but the gene-gene interactions in the interaction network might also exhibit psychophysical laws at some stage. Dynamic models of Bayesian networks might reveal the phenomena in a better way.

1 Introduction

1.1 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[2] 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[2]. 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.*[3] on power and logarithmic law albeit at a coarse level. Note that Goentoro and Kirschner[2] shows results for the behaviour of fold change of β-$catenin$ with respect to single parameter values. The current work, takes into account the behaviour of $TRCMPLX$ conditional on affects of multiple parameters in the form of evidences of various intra/extracelluar gene expression values working in the pathway. Bayesian networks help in integrating such multiple

‡ 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*

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

paramater affects via various causal arcs 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 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.*[4]. Finally, as a third observation, it would be interesting to note if these behaviours characterized by 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 (as pointed out by Goentoro and Kirschner[2]) but the entire network might be involved in such a behavior at some stage or the other. • The psychophysical law might not only be restricted to the intra/extracellular components but also to the interactions among the the intra/extracellular components in the pathway.

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.

1.2 The logarithmic psychophysical law

Masin *et al.*[4] states the Weber's law as follows -

> Consider a sensation magnitude γ determined by a stimulus magnitude β. Fechner[5] (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[5] (vol 1, p. 65) attributed to the German physiologist Ernst Heinrich Weber the empirical finding Weber[6] that $\Delta\gamma$ remains constant when the relative stimulus increment $\frac{\Delta\beta}{\beta}$ remains constant, and named this finding Weber's law. Fechner[5] (vol 2, p. 10) underlined that Weber's law was empirical.

It has been found that Bernoulli's principle (Bernoulli[7]) is different from Weber's law (Weber[6]) 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.*[4] shows that Weber's law is a special case of Bernoulli's principle and can be derived as follows - Equation 1 depicts the Bernoulli's principle and increment in sensation represented by $\Delta\gamma$ is proportional to change in stimulus represented by $\Delta\beta$.

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

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

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

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.*[4]. 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[5] 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.

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

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.

Before delving into the details of the experimental setup the following two subsections from Sinha[1] help build the background on Wnt pathway and the computational model used to infer the results in this paper.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

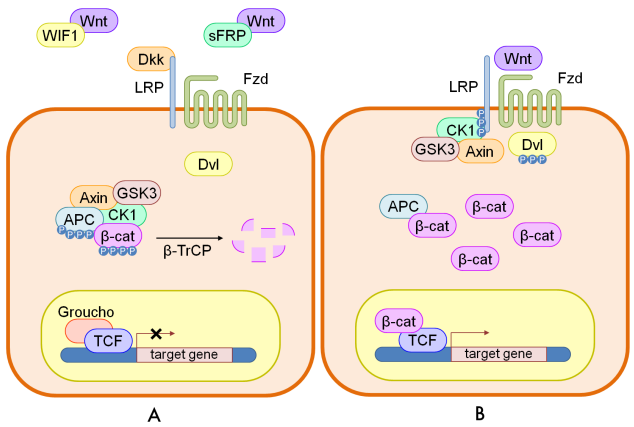

Fig. 1 A cartoon of wnt signaling pathway contributed by Verhaegh *et al.*[8]. Part (A) represents the destruction of β-*catenin* leading to the inactivation of the wnt target gene. Part (B) represents activation of wnt target gene.

1.3 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[9]). 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[10]). 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.

1.4 Epigenetic factors

One of the widely studied epigenetic factors is methylation (Costello and Plass[11], Das and Singal[12], Issa[13]). 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.*[14]). Similarly, methylation of genes in the Dickkopf ($DKKx$ Niehrs[15], Sato *et al.*[16]), Dapper antagonist of catenin ($DACTx$ Jiang *et al.*[17]) and Wnt inhibitory factor-1 ($WIF1$ Taniguchi *et al.*[18]) 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[19]) can affect gene expression (Peterson *et al.*[20]). 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.*[17]). $DACT1$ 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.*[17]).

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].

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.*[17], a training set will contain 46 samples and a test set will contain 2 samples (one

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

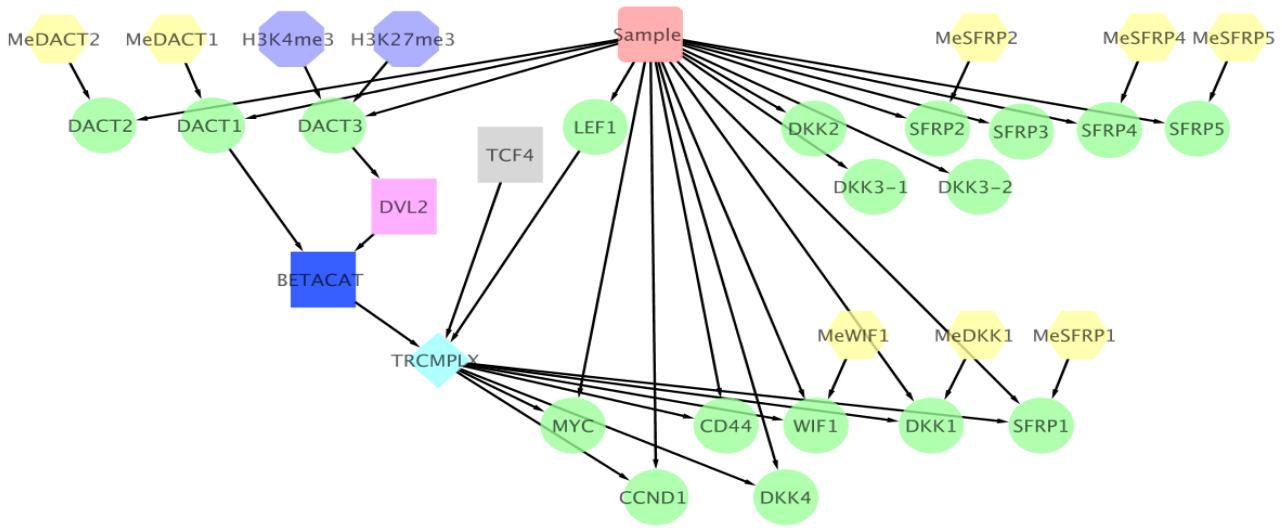

Fig. 2 Influence diagram of $\mathcal{M}_{PBK+EI}$ contains partial prior biological knowledge and epigenetic information in the form of methylation and histone modification. Diagram drawn using Cytoscape Shannon *et al.*[21]. In this model the state of Sample is distinguished from state of $TRCMPLX$ that constitutes the Wnt pathway.

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 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. Finally, the **third observation** is to check whether these laws are prevalent among the gene-gene interactions in the network or not.

3 Results and discussion

3.1 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 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$|\forall ge_i$ 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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

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.

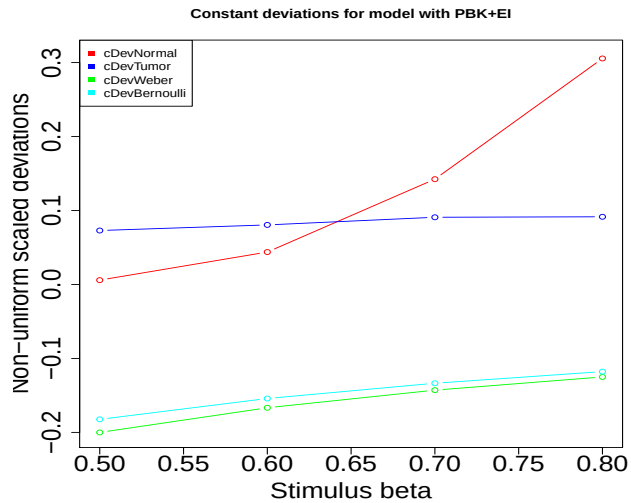

Fig. 3 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse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

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[2], the presented results might be crude in

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

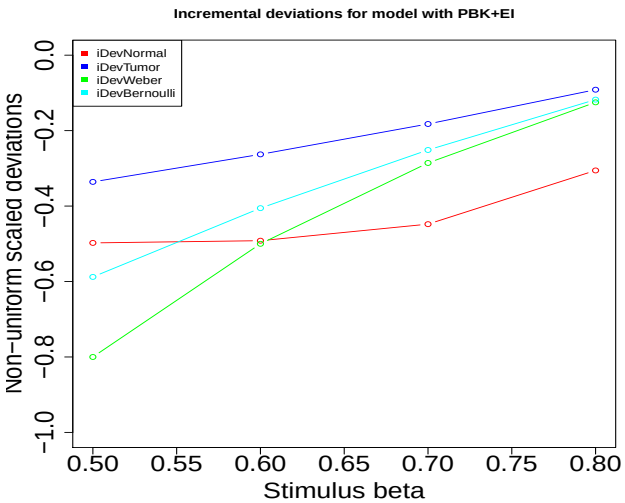

Fig. 4 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually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

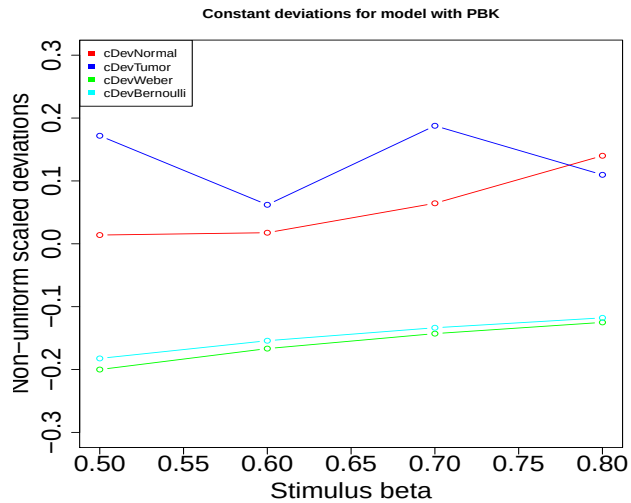

Fig. 5 Constant deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse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

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 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, $PggN$, $PggT$ 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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	ge	aa_N	ar_N	ra_N	rr_N	aa_T	ar_T	ra_T	rr_T	gg_{IN}	gg_{IT}
$SFRP5$ activation status apropos gene evidences in Normal and Tumor samples using $\theta = 0.5$											
1	DKK1	0	360	216	0	360	0	0	216	DKK1 |− <> SFRP5	DKK1 <> − <> SFRP5
2	DKK2	360	216	0	0	0	0	216	360	DKK2 <> − <> SFRP5	DKK2 | − | SFRP5
3	DKK3-1	360	216	0	0	0	0	216	360	DKK3-1 <> − <> SFRP5	DKK3-1 | − | SFRP5
4	DKK3-2	240	336	0	0	0	0	336	240	DKK3-2 |− <> SFRP5	DKK3-2 <> −| SFRP5
5	DKK4	0	480	96	0	460	0	0	116	DKK4 |− <> SFRP5	DKK4 <> − <> SFRP5
6	DACT1	346	230	0	0	0	0	216	360	DACT1 <> − <> SFRP5	DACT1 | − | SFRP5
7	DACT2	312	264	0	0	0	0	264	312	DACT2 <> − <> SFRP5	DACT2 | − | SFRP5
8	DACT3	504	0	0	72	0	507	69	0	DACT3 <> − <> SFRP5	DACT3 |− <> SFRP5
9	SFRP1	552	4	0	20	0	69	46	461	SFRP1 <> − <> SFRP5	SFRP1 | − | SFRP5
10	SFRP2	480	0	0	96	0	480	96	0	SFRP2 <> − <> SFRP5	SFRP2 |− <> SFRP5
11	SFRP3	484	0	0	92	0	480	96	0	SFRP3 <> − <> SFRP5	SFRP3 |− <> SFRP5
12	SFRP4	264	312	0	0	312	264	0	0	SFRP4 |− <> SFRP5	SFRP4 <> − <> SFRP5
13	WIF1	0	408	168	0	398	0	0	178	WIF1 |− <> SFRP5	WIF1 <> − <> SFRP5
14	LEF1	0	480	96	0	484	0	0	92	LEF1 |− <> SFRP5	LEF1 <> − <> SFRP5
15	MYC	0	456	120	0	442	0	0	134	MYC |− <> SFRP5	MYC <> − <> SFRP5
16	CCND1	0	480	96	0	480	0	0	96	CCND1 |− <> SFRP5	CCND1 <> − <> SFRP5
17	CD44	0	376	200	0	384	0	0	192	CD44 |− <> SFRP5	CD44 <> − <> SFRP5
$SFRP5$ activation status apropos gene evidences in Normal and Tumor samples using $\theta = \theta_N$ and $\theta = \theta_T$											
1	DKK1	0	322	216	38	0	0	360	216	DKK1 |− <> SFRP5	DKK1 <> −| SFRP5
2	DKK2	4	0	356	216	0	0	216	360	DKK2 <> −| SFRP5	DKK2 | − | SFRP5
3	DKK3-1	4	0	356	216	0	0	216	360	DKK3-1 <> −| SFRP5	DKK3-1 | − | SFRP5
4	DKK3-2	0	9	240	327	0	0	336	240	DKK3-2 | − | SFRP5	DKK3-2 <> −| SFRP5
5	DKK4	0	434	96	46	327	0	133	116	DKK4 |− <> SFRP5	DKK4 <> − <> SFRP5
6	DACT1	4	230	342	0	0	0	216	360	DACT1 <> −| SFRP5	DACT1 | − | SFRP5
7	DACT2	0	0	312	264	0	0	264	312	DACT2 <> −| SFRP5	DACT2 | − | SFRP5
8	DACT3	504	0	0	72	0	507	69	0	DACT3 <> − <> SFRP5	DACT3 |− <> SFRP5
9	SFRP1	13	0	539	24	0	0	46	530	SFRP1 <> −| SFRP5	SFRP1 | − | SFRP5
10	SFRP2	480	0	0	96	0	363	96	117	SFRP2 <> − <> SFRP5	SFRP2 |− <> SFRP5
11	SFRP3	484	0	0	92	0	350	96	130	SFRP3 <> − <> SFRP5	SFRP3 |− <> SFRP5
12	SFRP4	0	0	264	312	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	450	96	30	394	0	90	92	LEF1 |− <> SFRP5	LEF1 <> − <> SFRP5
15	MYC	0	402	120	54	325	0	117	134	MYC |− <> SFRP5	MYC <> − <> SFRP5
16	CCND1	0	418	96	62	350	0	130	96	CCND1 |− <> SFRP5	CCND1 <> − <> SFRP5
17	CD44	0	65	200	311	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.

4. assign a vector of $\widehat{\Pr}_N(g_i|g_k)\ \forall k$ genes to $PggN$ and a vector of $\widehat{\Pr}_T(g_i|g_k)\ \forall k$ 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 5). 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 5).

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

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

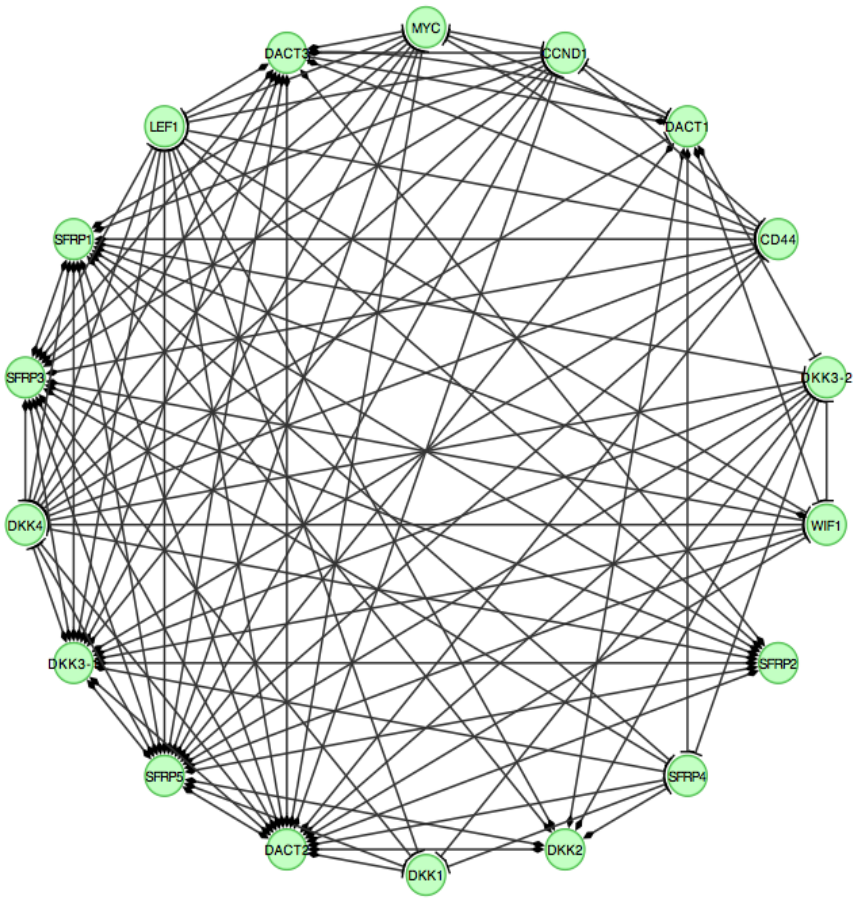

Fig. 7 Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = 0.5$. Note that the effect of initialized cpt for $TRCMPLX$ is 90% in tumorous case and 10% in normal case. Diamond <> means activation and straight bar | means repression.

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 $stats\S g_j <> - | g_i$

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

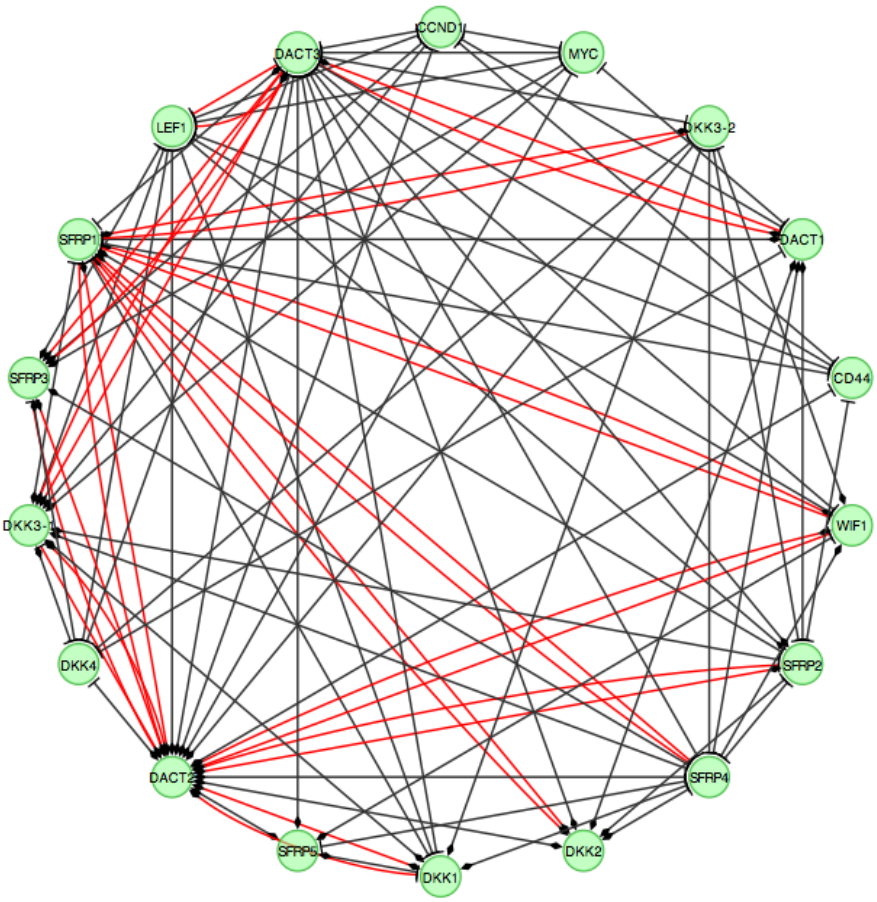

Fig. 8 Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = \theta_N$. Note that the effect of initialized cpt for $TRCMPLX$ is 90% in tumorous case and 10% in normal case. Diamond <> means activation and straight bar | means repression.

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, $stats$ = $stats$, $PggN$ = $PggN$, $PggT$ = $PggT$)

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 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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

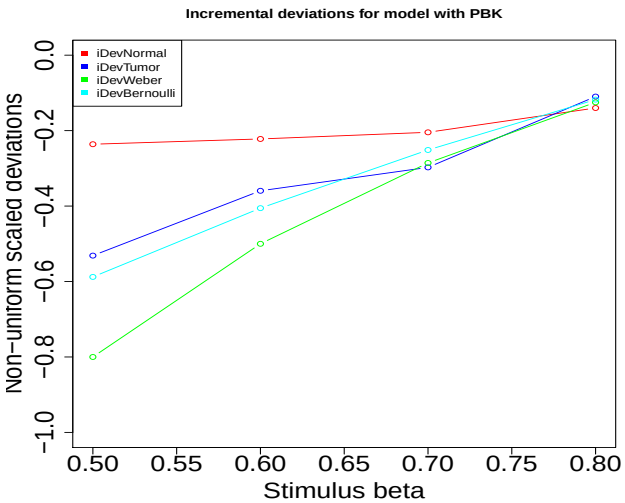

Fig. 6 Incremental deviations in β i.e $ETGN$ and corresponding deviations in Pr($TRCMPLX$ = active$|\forall ge_i$ evidences) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations to visually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

$$\theta_N = \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})}$$
$$\theta_T = \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} \tag{5}$$

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 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. 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 veracity of existence of psychophysical laws in gene-gene interaction network. 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 $\geq \theta$ ($< \theta$). 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.

Case of $\theta = 0.5$ - 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 $DKK2$, $DKK3-1$, $DACT-1/2/3$ 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

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 7 and 8 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.

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).

Case of $\theta = \theta_N^{SFRP5}$ - While employing the weighted mean as the threshold to discretize Pr($SFRP5$ = active$|g_j$ evidence), the $SFRP5$ gene evidences that constitutes the test

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Table 5 Tabulated missing gene gene interactions of figure 7 and 8 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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

data are used. Note that the test evidences for $SFRP5$ are used for two purpose (1) to discretize Pr($SFRP5$ = active$|g_j$ evidence) as discussed above and (2) to compute the probability of activation status of another gene conditional on evidence for $SRFP5$, i.e Pr(g_j = active$|SFRP5$ evidence). Since the test evidence for the 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 is an expression 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 $SFRP5$ test samples then discretizes Pr($SFRP5$ = active$|g_j$ evidence) according to the inherently represented summary. More realistic estimates like kernel density estimates could also be used.

In this preliminary work, the focus is on observing the psycholphysical phenomena that might be prevalent at interaction level also. To this end, while using the weighted mean, it was found that $DKK1$, $SFRP4$ and $WIF1$ showed reversible behaviour with $SFRP5$. This reduction in the reversible reactions is attributed to the use of weighted mean that carries an idiosyncracy of the test 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 might 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. Again, dynamic models might bring more clarity to the picture.

Table 3 shows these interactions using $\theta \in \{0.5, \theta_N, \theta_T\}$.

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. 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($g_j|g_i$)).

Lastly, duplicate interactions are removed from the network for Normal samples. This is repeated for the network based on tumor samples also. 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 complete 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.

Deviation study for $SFRP5$ and MYC for normal case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr($SFPR5\|MYC$)	Pr($MYC\|SFPR5$)
0.8	0.1	0.125	0.117783	0.002803756	0.003196908
0.7	0.1	0.1428571	0.1335314	0.002574599	0.003196908
0.6	0.1	0.1666667	0.1541507	0.002333440	0.003196908
0.5	0.1	0.2	0.1823216	0.002078026	0.003196908
0.8	0.1	0.125	0.117783	0.009789821	0.012787632
0.7	0.2	0.2857143	0.2513144	0.006986065	0.009590724
0.6	0.3	0.5	0.4054651	0.004411466	0.006393816
0.5	0.4	0.8	0.5877867	0.002078026	0.003196908

Table 6 Deviation study for Pr($SFRP5|MYC$) and Pr($MYC|SFRP5$) for normal case

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

In the penultimate section on preservation of gene-gene interaction, it was found that some of the interactions remain preserved as there was change of the effect transcription complex. The transcription complex itself was found to follow a logarithmic-power psychophysical law. It would be interest-

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Deviation study for $SFRP5$ and MYC for tumor case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr($SFPR5\|MYC$)	Pr($MYC\|SFPR5$)
0.8	0.1	0.125	0.117783	0.001049836	0.000000e+00
0.7	0.1	0.1428571	0.1335314	0.001129522	0.000000e+00
0.6	0.1	0.1666667	0.1541507	0.001216178	-5.551115e-17
0.5	0.1	0.2	0.1823216	0.001310739	5.551115e-17
0.8	0.1	0.125	0.117783	0.004706275	0.000000e+00
0.7	0.2	0.2857143	0.2513144	0.003656439	0.000000e+00
0.6	0.3	0.5	0.4054651	0.002526918	0.000000e+00
0.5	0.4	0.8	0.5877867	0.001310739	5.551115e-17

Table 7 Deviation study for Pr($SFRP5|MYC$) and Pr($MYC|SFRP5$) for tumor case

ing to observe if these laws are prevalent among the gene-gene interactions in the network.

3.3.1 Case: $<> -|$ or $|- <>$ with $\theta = 0.5$In Sinha[1], the unknown behaviour of $SFRP5$ in the Wnt pathway has been revealed slightly using computational causal inference. In figure 7, $SFRP5$ 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 $SFRP5$ and MYC, $SFRP5$ showed activation (repression) and MYC 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 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 level. To summarize the results in these tables, graphs were plotted in figures 9 for Pr($SFRP5|MYC$) (constant deviations), 10 for Pr($MYC|SFRP5$) (constant deviations), 11 for Pr($SFRP5|MYC$) (incremental deviations) and 10 for Pr($MYC|SFRP5$) (incremental deviations).

Considering figure 9, when deviations are constant in both Weber and Bernoulli formulation, the deviations in the prediction of Pr($SFRP5|MYC$) 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($MYC|SFRP5$) as is depicted in figure 10. Note that the interaction for $SFRP5$ given MYC was observed to be reversible in normal and tumor cases. But this is not so with the interaction for MYC given $SFRP5$. It might be expected that the non conformance of logarithmic-power law for Pr($MYC|SFRP5$) may be due to the non preservation

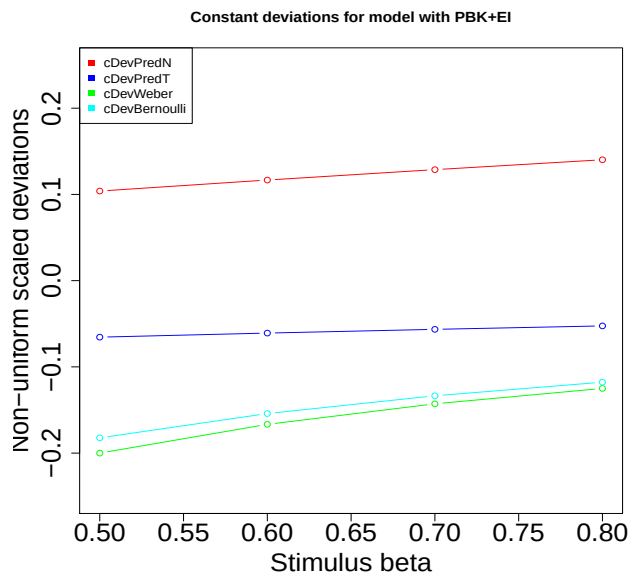

Fig. 9 Constant deviations in β i.e ETGN and corresponding deviations in Pr($SFRP5|MYC$) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - deviation in Pr($SFRP5|MYC$) in Normal case using Weber's law, Blue - deviation in Pr($SFRP5|MYC$) in Tumor using Weber's law, Green - constant deviation in Webers law, Cyan - constant deviation in Bernoullis law.

of the interaction of MYC given $SFRP5$. This is so because Pr($SFRP5|MYC$) depicts a reversible $SFRP5 <> -|MYC$ ($MYC <> -|SFRP5$) in the network on normal (tumor) samples, while Pr($MYC|SFRP5$) does not depict a reversible $MYC|- <> SFRP5$ ($MYC|-|SFRP5$) in the network on normal (tumor) samples.

Similar behaviour was observed in the case of incremental deviations as depicted in figures 11 and 12. Analysis of behaviour of other gene-gene interactions showing $<> -|$ or $|- <>$ can be observed in a similar way and can be found by executing the R code provided in the website.

3.3.2 Case: $|-|$ or $<> - <>$ with $\theta = 0.5$Again, as pointed out in Sinha[1], the unknown behaviour of $DKK3-2$ in the Wnt pathway has been revealed slightly using computational causal inference. In figure 7, $DKK3-2$ 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 $DKK3-2$ and $WIF1$, 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

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

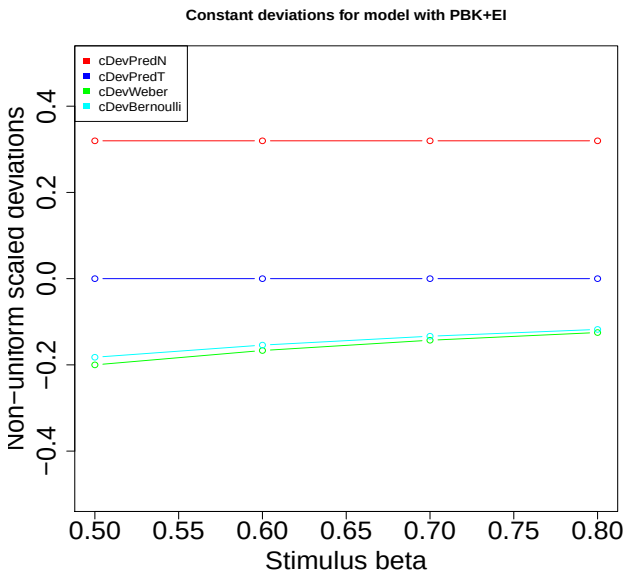

Fig. 10 Same as figure 10 but for Pr($MYC|SFRP5$).

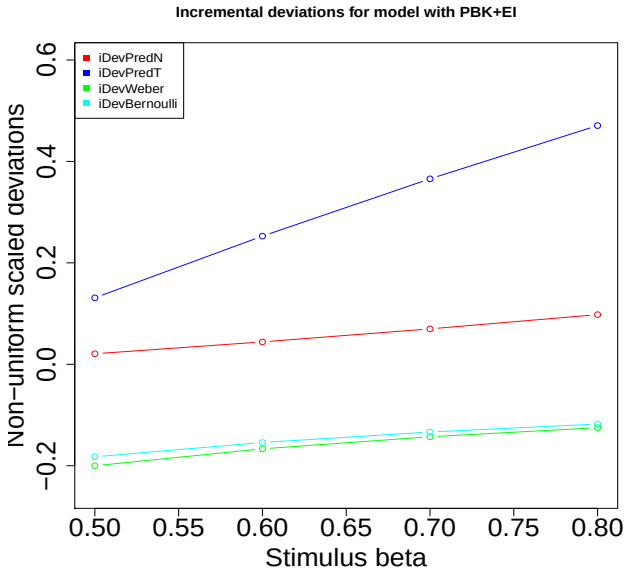

Fig. 11 Same as figure 10 but for Pr($SFRP5|MYC$). Instead of constant deviations, incremental deviations are represented.

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

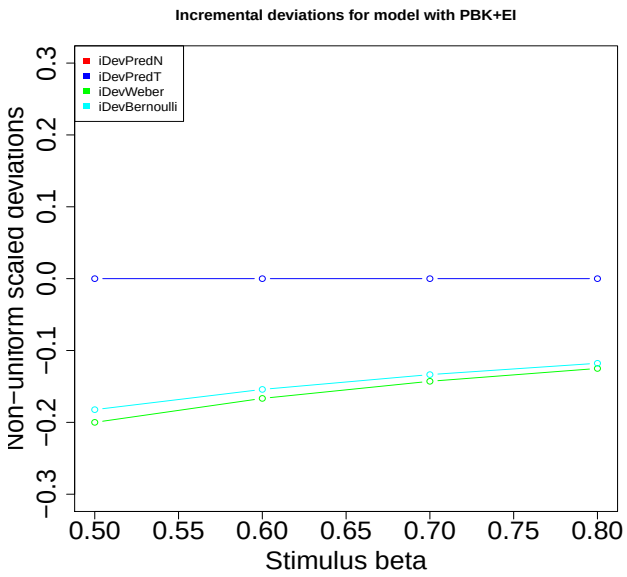

Fig. 12 Same as figure 10 but for Pr($MYC|SFRP5$). Instead of constant deviations, incremental deviations are represented.

incremental level. To summarize the results in these tables, graphs were plotted in figures 13 for Pr($DKK3-2|WIF1$) (constant deviations), 14 for Pr($WIF1|DKK3-2$) (constant deviations), 15 for Pr($DKK3-2|WIF1$) (incremental deviations) and 14 for Pr($WIF1|DKK32$) (incremental deviations).

Considering figure 13, when deviations are constant in both Weber and Bernoulli formulation, the deviations in the prediction of Pr($DKK3-2|WIF1$) 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($WIF1|DKK3-2$) as is depicted in figure 14. It is peculiar to see that the interaction for $DKK3-2$ given $WIF1$ was not observed to be reversible in normal and tumor cases while the interaction for $WIF1$ given $DKK3-2$ was observed to be reversible. This points to a crucial fact that the interactions interpretated from conditional probabilities are not always one 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 15 and 16.

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

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Deviation study for $DKK3-2$ and $WIF1$ for normal case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr($DKK3-2\|WIF1$)	Pr($WIF1\|DKK3-2$)
0.8	0.1	0.125	0.117783	7.710071e-05	-0.003833343
0.7	0.1	0.1428571	0.1335314	6.543515e-05	-0.003833343
0.6	0.1	0.1666667	0.1541507	5.389710e-05	-0.003833343
0.5	0.1	0.2	0.1823216	4.247643e-05	-0.003833343
0.8	0.1	0.125	0.117783	2.389094e-04	-0.015333372
0.7	0.2	0.2857143	0.2513144	1.618087e-04	-0.011500029
0.6	0.3	0.5	0.4054651	9.637353e-05	-0.007666686
0.5	0.4	0.8	0.5877867	4.247643e-05	-0.003833343

Table 8 Deviation study for Pr($DKK3-2|WIF1$) and Pr($WIF1|DKK3-2$) for normal case

Deviation study for $DKK3-2$ and $WIF1$ for tumor case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1+\frac{\Delta\beta}{\beta})$	Pr($DKK3-2\|WIF1$)	Pr($WIF1\|DKK3-2$)
0.8	0.1	0.125	0.117783	-0.0008710076	-1.110223e-16
0.7	0.1	0.1428571	0.1335314	-0.0008493470	0.000000e+00
0.6	0.1	0.1666667	0.1541507	-0.0008280790	1.110223e-16
0.5	0.1	0.2	0.1823216	-0.0008071857	-1.110223e-16
0.8	0.1	0.125	0.117783	-0.0033556193	-1.110223e-16
0.7	0.2	0.2857143	0.2513144	-0.0024846117	0.000000e+00
0.6	0.3	0.5	0.4054651	-0.0016352647	0.000000e+00
0.5	0.4	0.8	0.5877867	-0.0008071857	-1.110223e-16

Table 9 Deviation study for Pr($DKK3-2|WIF1$) and Pr($WIF1|DKK3-2$) for tumor case

the threshold. These changes are depicted in the figures 7 and 8 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 8.

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

The dual interactions revealed using weighted means indicate 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

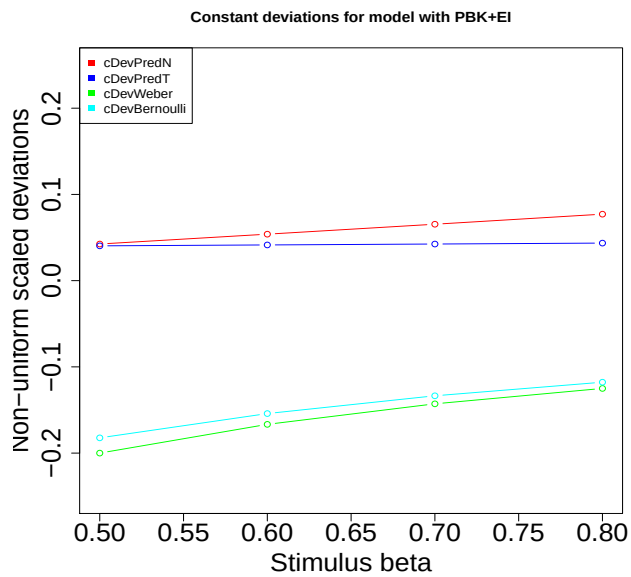

Fig. 13 Constant deviations in β i.e ETGN and corresponding deviations in Pr($DKK32|WIF1$) for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The model used is $\mathcal{M}_{PBK+EI}$. Red - deviation in Pr($DKK32|WIF1$) in Normal case using Weber's law, Blue - deviation in Pr($DKK32|WIF1$) in Tumor using Weber's law, Green - constant deviation in Webers law, Cyan - constant deviation in Bernoullis law.

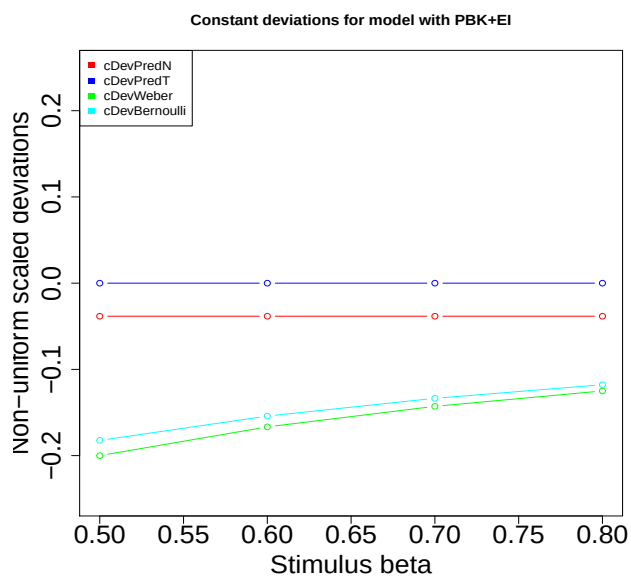

Fig. 14 Same as figure 10 but for Pr($WIF1|DKK32$).

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

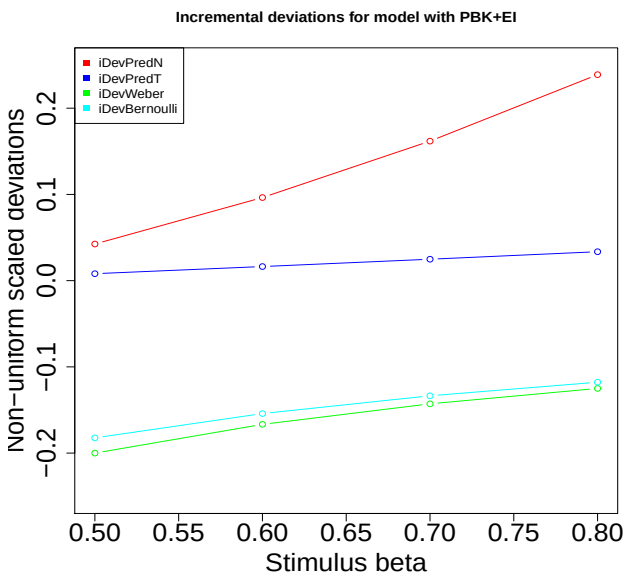

Fig. 15 Same as figure 14 but for Pr($DKK32|WIF1$). Instead of constant deviations, incremental deviations are represented.

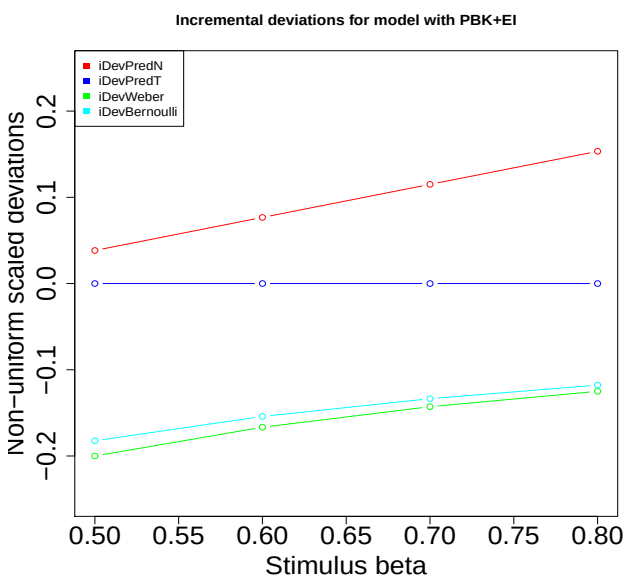

Fig. 16 Same as figure 14 but for Pr($WIF1|DKK32$). Instead of constant deviations, incremental deviations are represented.

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, 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.*[22].

Interaction between DKK1 and DACT2 - Dual interactions $DACT2 <> - <> DKK1$ and $DKK1|- <> DACT2$ ($DACT2|-|DKK1$ and $DKK1 <> -|DACT2$) in normal (tumor) sample were found as depicted in figure 8. Figure 17 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 17. 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 6 -

$$\begin{aligned}
\theta_N^{DKK1} &= \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})} \\
&= \frac{1 \times 264 + 2 \times 312}{3 \times 576} = 0.5138889 \\
\theta_T^{DKK1} &= \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} \\
&= \frac{1 \times 312 + 2 \times 264}{3 \times 576} = 0.4861111
\end{aligned} \tag{6}$$

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

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

following computations in equation 7 -

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

On comparison of these graphs 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 18 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 Pr($DKK1|DACT2$) using weighted mean in Normal case
2. labels of test data ge_N and discretized values Pr($DKK1|DACT2$) using arbitrary value of 0.5 in Normal case
3. labels of test data ge_T and discretized values Pr($DKK1|DACT2$) using weighted mean in Tumor case
4. labels of test data ge_T and discretized values Pr($DKK1|DACT2$) using arbitrary value of 0.5 in Tumor case

In figure 18, 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 weighted mean for Normal case, weighted mean for Tumor case, arbit value of 0.5 for Normal case and arbit value of 0.5 for Tumor case, in order. Respective AUC values for the ROC curves appear on the title of each of the graphs. It can be seen that the ROCs are different for both the interactions and the interaction depicted by Pr($DKK1|DACT2$) are much more steeper while using weighted means in comparison to those using arbitrary value of 0.5. 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.

Finally, 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.

4 Future directions

In context of the above observations, dynamic models might reveal greater information regarding the psychophysical laws. Work by Goentoro and Kirschner[2] 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 19 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 $t+1$.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

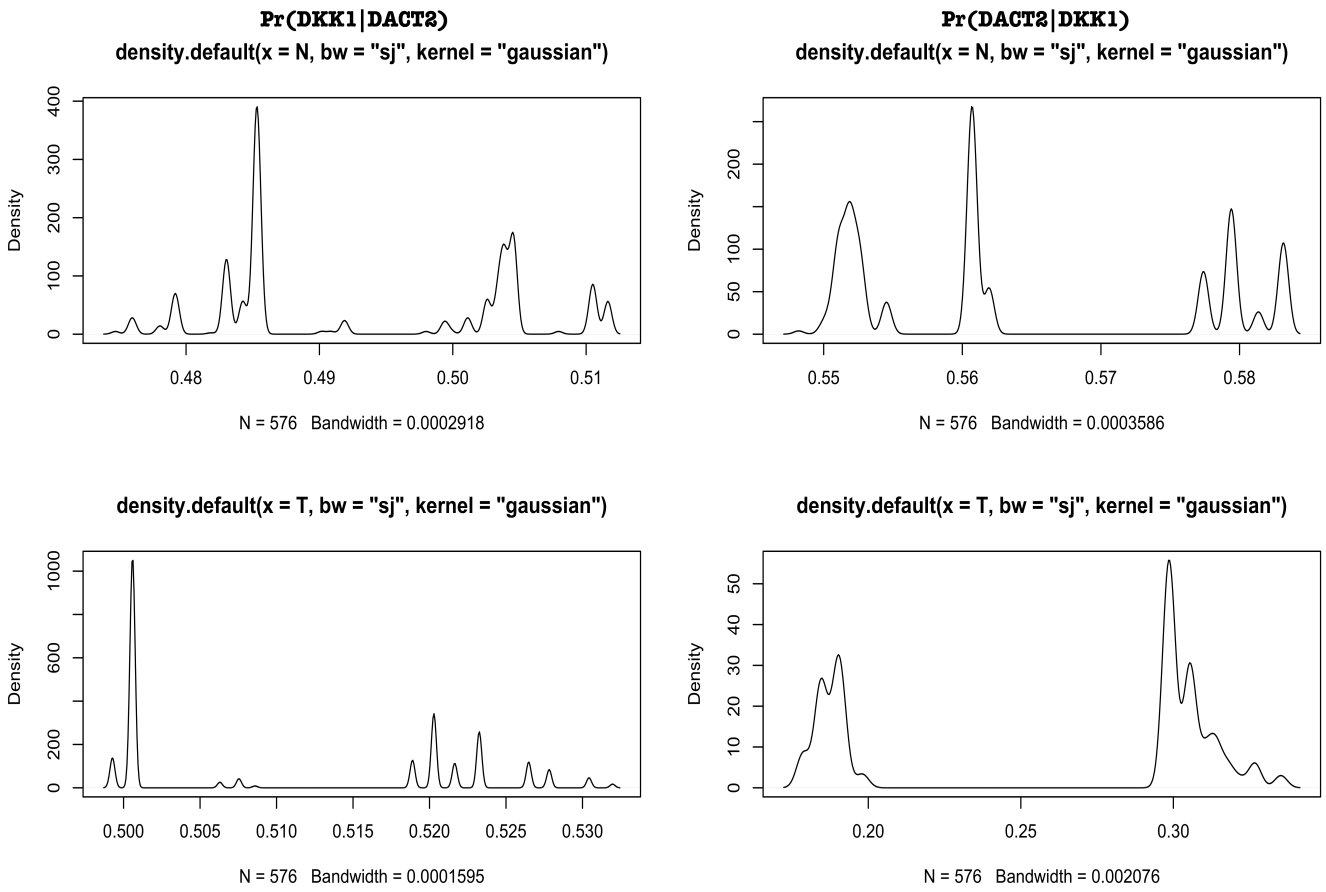

Fig. 17 Kernel density estimates for predicted Pr($DKK1|DACT2$) and Pr($DACT2|DKK1$) in Normal and Tumor cases. Gaussian kernel is used for smoothing the density estimate. The bandwidth of the kernel is selected using the pilot estimation of derivative as proposed by Sheather and Jones[23] and implemented in R programming language.

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'[24] 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[25]). 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.

5 Availability

Code with dataset is made available under GNU GPL v3 license at google code project on `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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

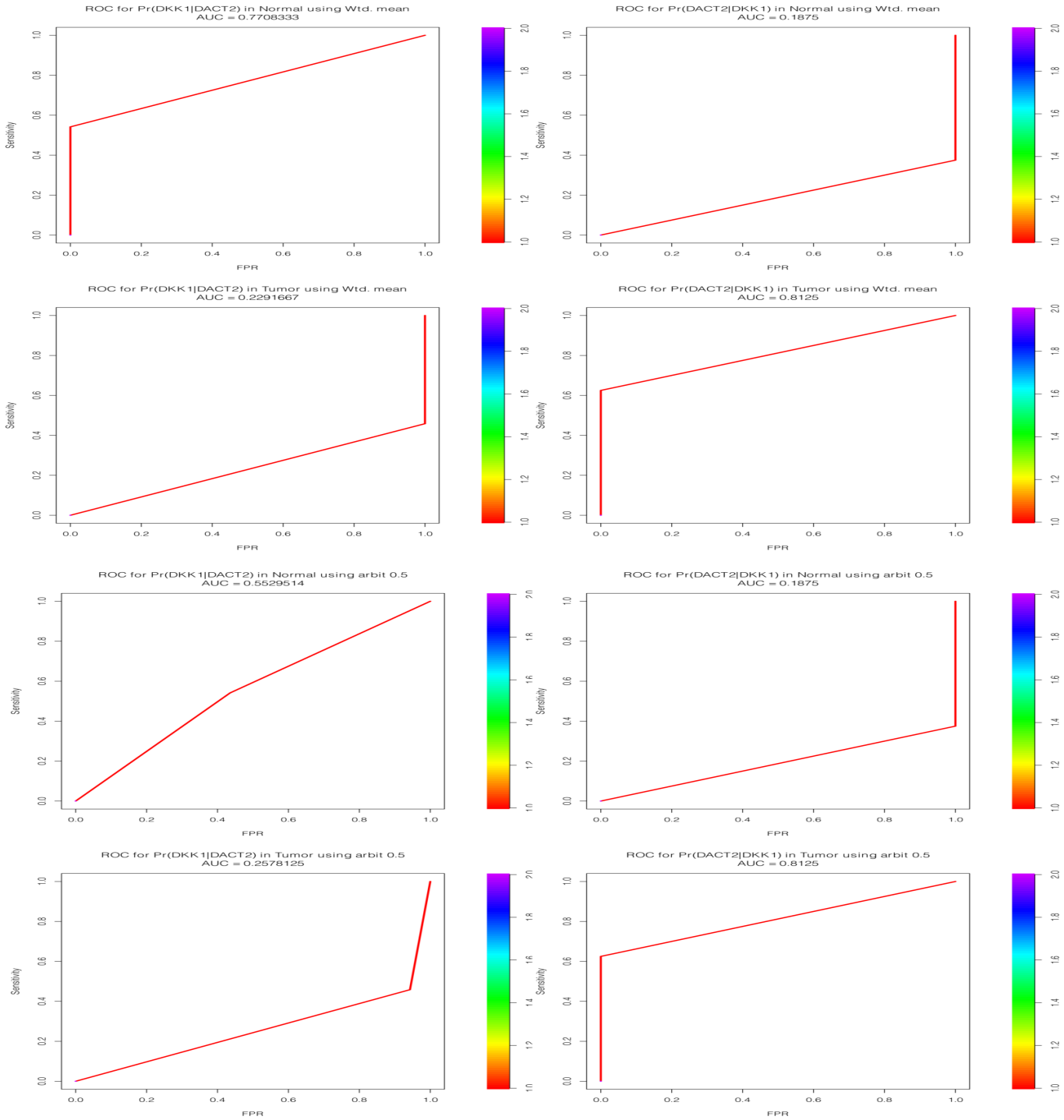

Fig. 18 Column wise ROCs for Pr($DKK1|DACT2$) (1st column) and Pr($DACT2|DKK1$) (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using weighted mean for Normal case, weighted mean for Tumor case, arbit value of 0.5 for Normal case and arbit value of 0.5 for Tumor case. Respective AUC values for the ROC curves appear on the title of each of the graphs.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

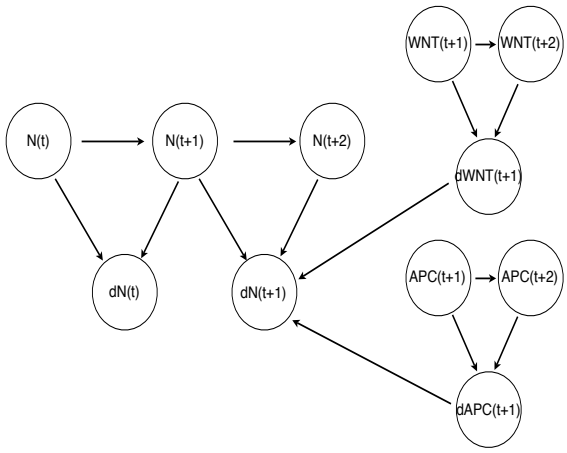

Fig. 19 A schematic diagram of a dynamic bayesian network model that might help study the fold change and the logarithmic psychophysical laws behind the changes.

6 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. 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.

References

1 S. Sinha, *Integr. Biol.*, 2014, **6**, 1034–1048.
2 L. Goentoro and M. W. Kirschner, *Molecular Cell*, 2009, **36**, 872–884.
3 M. Adler, A. Mayo and U. Alon, *PLoS computational biology*, 2014, **10**, e1003781.
4 S. C. Masin, V. Zudini and M. Antonelli, *Journal of History of the Behavioral Sciences*, 2009, **45**, 56–65.
5 G. T. Fechner, *Elemente der Psychophysik (2 vols)*, Breitkopf and Hartel, 1860.
6 E. H. Weber, *De pulsu resorptione, auditu et tactu*, Annotationes anatomicae et physiologicae, 1834.
7 D. Bernoulli, *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 1738, **5**, 175–192.
8 W. Verhaegh, P. Hatzis, H. Clevers and A. van de Stolpe, *Cancer Research, San Antonio Breast Cancer Symposium*, 2011, **71**, 524–525.
9 H. Clevers, *Cell*, 2006, **127**, 469–480.
10 A. Gregorieff and H. Clevers, *Genes & development*, 2005, **19**, 877–890.
11 J. Costello and C. Plass, *Journal of medical genetics*, 2001, **38**, 285–303.
12 P. Das and R. Singal, *Journal of Clinical Oncology*, 2004, **22**, 4632–4642.
13 J. Issa, *Clinical Cancer Research*, 2007, **13**, 1634–1637.
14 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.
15 C. Niehrs, *Oncogene*, 2006, **25**, 7469–7481.
16 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.
17 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.
18 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.
19 B. Strahl and C. Allis, *Nature*, 2000, **403**, 41–45.
20 C. Peterson, M. Laniel *et al.*, *Current Biology*, 2004, **14**, 546–551.
21 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.

bioRxiv preprint doi: https://doi.org/10.1101/015834; this version posted May 16, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

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

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

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