

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

shriprakash sinha¹

*Faculty of IT and Mathematics
Royal Thimphu College, Ngabiphu, P.O. Box - 1122, Bhutan*

Abstract

Recently, psychophysical laws have been observed to be functional in certain factors working downstream of the Wnt pathway. This work tests the veracity of the prevalence of such laws, albeit at a coarse level, using sensitivity analysis on biologically inspired epigenetically influenced computational causal models. In this work, the variation in the effect of the predictive behaviour of the transcription complex (TRCMPLX) conditional on the evidences of gene expressions in normal/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 follows power-logarithmic psychophysical law, crudely. More recently, wet lab experiments have proved the existence of sensors that behave in a logarithmic fashion thus supporting the earlier proposed postulates based on computational sensitivity analysis of this manuscript regarding the existence of logarithmic behaviour in the signaling pathways. It also signifies the importance of systems biology approach where in silico experiments combined with in vivo/in vitro experiments have the power to explore the deeper mechanisms of a signaling pathway. Additionally, it is hypothesized that these laws are prevalent at gene-gene interaction level also. The interactions were obtained by thresholding the inferred conditional probabilities of a gene activation given the status of another gene activation. The deviation in the interactions in normal/tumor samples was similarly observed by varying the initially assigned values of conditional probability tables (cpt) for TRCMPLX. Analysis of deviation in interactions show prevalence of psychophysical laws and

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

Email address: ssinha@rtc.bt (shriprakash sinha)

¹This work was conducted by Sinha as an independent researcher during the period from 2013 to 2016. ORCID : orcid.org/0000-0001-7027-5788

is reported for interaction between elements of pairs (SFRP3, MYC), (SFRP2, CD44) and (DKK1, DACT2). Based on crude static models, it is assumed that dynamic models of Bayesian networks might reveal the phenomena in a better way.

Keywords: Psychophysical laws; Wnt pathway; Causal models; Inference; Bayesian Network; Sensitivity analysis; Gene interaction network

1. Introduction & problem statement

Ever since the accidental discovery of the Wingless in 1973 by Sharma (1973), a tremendous amount of research work has been carried out in the related field of Wnt signaling pathway in the past forty years. A majority of the work has focused on issues related to • the discovery of genetic and epigenetic factors affecting the pathway Thorstensen et al. (2005) & Baron and Kneissel (2013), • implications of mutations in the pathway and its dominant role on cancer and other diseases Clevers (2006), • investigation into the pathway's contribution towards embryo development Sokol (2011), homeostasis Pinto et al. (2003), Zhong et al. (2014) and apoptosis Pečina-Šlaus (2010) and • safety and feasibility of drug design for the Wnt pathway Kahn (2014), Garber (2009), Voronkov and Krauss (2012), Blagodatski et al. (2014) & Curtin and Lorenzi (2010). More recent informative reviews have touched on various issues related to the different types of the Wnt signaling pathway and have stressed not only the activation of the Wnt signaling pathway via the Wnt proteins Rao and Kühl (2010) but also the on the secretion mechanism that plays a major role in the initiation of the Wnt activity as a prelude Yu and Virshup (2014).

In a more recent development, there has been the observation and study of psychophysical laws prevailing within the pathway and in this regard Goentoro and Kirschner (2009) 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 (1834) law in sensory physiology. Note that Weber's law has been found to be a special case of Bernoulli's logarithmic law Masin et al. (2009). If a sensation magnitude γ be determined by a stimulus magnitude β , then the Weber's law states that $\Delta\gamma$ remains constant when the relative stimulus increment $\Delta\beta$ remains constant. The law derives from a more general Bernoulli's law were $\Delta\gamma \propto \log \frac{\Delta\beta}{\beta}$. In an unrelated work by Sun et al. (2012), it has been shown that these laws arise at computational level as Bayes optimal under neurobiological constraints at implementational and algorithmic levels. The proposed mathematical framework for understanding the psychophysical scales as Bayes optimal and information theoretically-optimal representation of time sampled continuous valued stimuli is based on established neurobiological assumptions. Sun et al. (2012) also show that the psychophysical laws connect well to quantization frameworks and state that only discrete set of output is distinguishable

due to biological constraints. This discretization leads to quantization of stimulus also as the nonlinear scaling of the stimulus that leads to the resultant output is invertible. These mathematical insights might explain the indistinguishable insensitive fold changes in levels of β -catenin shown by Goentoro and Kirschner (2009).

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

To address these issues, the current work uses the Bayesian network model proposed in Sinha (2014) and conducts sensitivity analysis on the model to check the observations regarding the prevalence of the reported psychophysical laws. In Sinha (2014), it was shown via hypothesis testing that the active (repressed) state of *TRCMPLX* in the Wnt signaling pathway for colorectal cancer cases is not always correspond to the tumorous (normal) state of the test sample under consideration. For this, Sinha (2014) shows various results on the predicted state of *TRCMPLX* conditional on the given gene evidences, while varying the assigned probability values of conditional probability tables of *TRCMPLX* during initialization of the biologically inspired Bayesian Network model. Here, the degree of belief in the activity of *TRCMPLX* is denoted by the prior probability assigned to the node of *TRCMPLX* in the network. It was found that the predicted values often increase with an increasing value (in conditional probability tables) of the activity of *TRCMPLX* on certain genes. What this asks for is that for the recorded deviations due to the changes made in these prior probabilities (i.e the input deviations), is it possible to observe the prevalent logarithmic laws and their deviations (like the Weber's law) as shown by Goentoro and Kirschner (2009), using computational causal modeling?

In this manuscript, the preliminary analysis of deviations computed from variation in prior and estimated conditional probability values using Bayesian network model in Sinha (2014) show that the variation in predictive behaviour of *TRCMPLX* conditional on gene evidences (i.e the output deviation) follows power and logarithmic psychophysical law crudely, apropos the variation in assigned priors of *TRCMPLX* (i.e input deviations). This implies that the deviations in output are proportional to increasing function of deviations in input. This relates to the work of Adler et al. (2014) on power and logarithmic law albeit at a coarse level. The granularity is obscured due to the use of static data from Jiang et al. (2008) that is used in Sinha (2014) as well as

the Bayesian network model that encodes the belief in the factors affecting the pathway in terms of probabilities as well as the inferences made based on the updating of these probabilities conditional on discretized states of gene expression values as evidences. Irrespective of the hurdle posed by the causal models, inferences made based on prior biological knowledge and gene expression evidences coupled with sensitivity analysis sheds light on the prevalent power-logarithmic psychophysical laws in the pathway. *Note that the foundations of the current work were presented as poster in the International Conference on Systems Biology of Human Disease at the German Cancer Research Center in Heidelberg (Germany) in 2015. Followup of some of the implications were shared with a few labs for verification and it is gladdening to see that in a recent development via wet lab experiments by Olsman and Goentoro (2016), it has been confirmed that there are existence of sensors that behave in a logarithmic fashion. The wet lab work by Olsman and Goentoro (2016) supports the earlier proposed crude postulates based on computational sensitivity analysis of this manuscript regarding the existence of logarithmic behaviour in the signaling pathways. It also signifies the importance of systems biology approach where in silico experiments combined with in vivo/in vitro experiments have the power to explore the deeper mechanisms of a signaling pathway.*

Adler et al. (2014) show in detail that these laws can be studied empirically using models that exhibit the property of fold change detection (FCD). What this means is that the output depends on the relative changes in the input. The biological feedback models employed for these studies consider various parameters like rates of production of a compound, removal removal of a compound, repression of a compound, levels of scaffolds, kinases, etc. that might be responsible for exhibiting these laws. The current work using the static Bayesian network model might not propose feedback loops directly as used by Adler et al. (2014), yet it could reveal existence of the loops via causal inference even while using static data. The drawback of the current work is its inability to consider cyclic loops. This can be rectified by use of dynamic Bayesian network models that incorporate interaction represented in time series data. Also, the use of Bayesian network models can help in studying the problem from a multiparameter setting as various factors affecting the pathway can be connected in the influence diagrams of the network through the principle of d -connectivity/separability. This connectivity will be explained later in the required theory section.

Note that Goentoro and Kirschner (2009) show results for the behaviour of fold change of β -catenin with respect to changes in the single parameter values i.e the Wnt. On similar lines, the current work takes into account the behaviour of *TRCMPLX* conditional on affects of multiple parameters in the form of evidences of various intra/extracellular gene expression values working in the pathway, based on the changes made in the assigned prior probabilities for *TRCMPLX*. The difference here is that one can analyse changes in nodes of a computational model to explore an inherent law in comparison to use of wet lab experiments. The issue here is that FCD which is recorded with respect to changes in levels of concentration can now be recorded via changes in

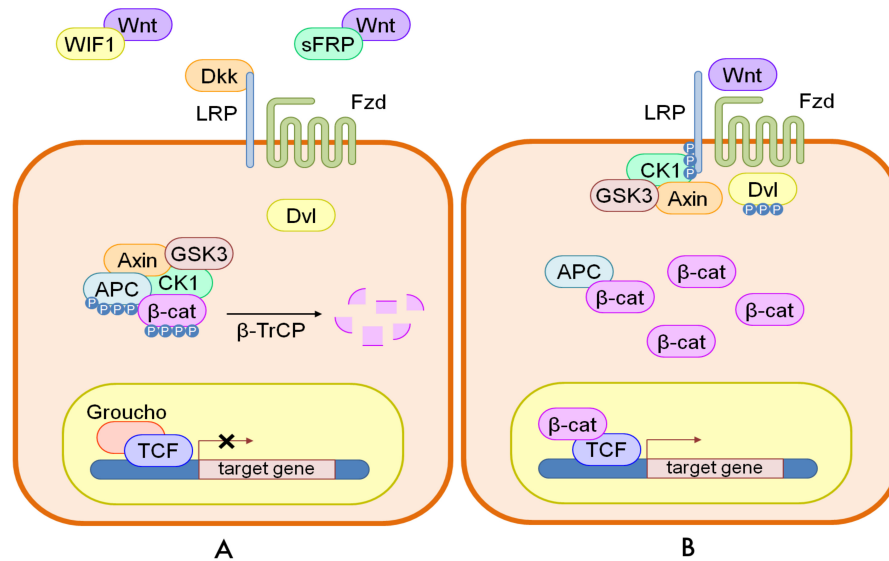


Figure 1: A cartoon of Wnt signaling pathway adapted from Sinha (2014). Part (A) represents the destruction of β -catenin leading to the inactivation of the Wnt target gene. Part (B) represents activation of Wnt target gene.

the strength of belief in the occurrence of an event. For example, suppose it is not known by what degree the *TRCMPLX* plays a major role in the signaling pathway quantitatively, then it is possible to encode the degree of belief regarding the role of *TRCMPLX* in the form of prior or conditional probabilities during initialization of the network. By recording the deviations in these probabilities and observing the output deviations, it is possible to study certain psychophysical laws. Finally, this does not mean that probabilities related to actual concentrations cannot be encoded. Thus, Bayesian networks help in capturing the desired biological knowledge via various causal arcs and conditional probabilities and sensitivity analysis aids in the study of the such natural behaviour.

As a second observation, the forgoing result points towards stability in the behaviour of *TRCMPLX* and this stability is reflected in the preserved gene gene interactions across the changing values of the priors of *TRCMPLX*. The interactions are inferred from conditional probabilities of individual gene activation given the status of another gene activation. Finally, as a third observation, it would be interesting to note if the psychophysical laws are prevalent among the dual gene-gene interactions or not. If the results are affirmative then the following important speculations might hold true • Not just one factor but components of the entire network might be exhibiting such a behavior at some stage or the other. • The psychophysical law might not be restricted to individual intra/extracellular components but also to the interactions among the the

151 intra/extracellular components in the pathway. This might mean that the inter-
152 actions manifest during the prevalence of power and logarithmic laws. Further
153 wet lab analysis is needed to verify these computational claims.

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

157 2. Revisiting the requisite theory

158 To understand the logical flow of the current paper, some details of the above
159 related topics from Sinha (2014) are revisited here in order and subdivided into
160 descriptions of - (1) general working of canonical Wnt signaling pathway and
161 some of the involved epigenetic factors (2) introduction to Bayesian networks
162 and (3) the intuition behind the Bayesian network model employed. This is
163 followed by Weber's law and its derivation and finally the notations and termi-
164 nologies to understand the results and discussion section.

165 2.1. Canonical Wnt signaling pathway

166 The canonical Wnt signaling pathway is a transduction mechanism that con-
167 tributes to embryo development and controls homeostatic self renewal in several
168 tissues Clevers (2006). Somatic mutations in the pathway are known to be as-
169 sociated with cancer in different parts of the human body. Prominent among
170 them is the colorectal cancer case Gregorieff and Clevers (2005). In a succinct
171 overview, the Wnt signaling pathway works when the Wnt ligand gets attached
172 to the Frizzled(*fzd*)/LRP coreceptor complex. *Fzd* may interact with the Dis-
173 shevelled (*Dvl*) causing phosphorylation. It is also thought that Wnts cause
174 phosphorylation of the LRP via casein kinase 1 (*CK1*) and kinase *GSK3*.
175 These developments further lead to attraction of Axin which causes inhibition
176 of the formation of the degradation complex. The degradation complex consti-
177 tutes of *Axin*, the β -*catenin* transportation complex *APC*, *CK1* and *GSK3*.
178 When the pathway is active the dissolution of the degradation complex leads to
179 stabilization in the concentration of β -*catenin* in the cytoplasm. As β -*catenin*
180 enters into the nucleus it displaces the *Groucho* and binds with transcription
181 cell factor *TCF* thus instigating transcription of Wnt target genes. *Groucho*
182 acts as lock on *TCF* and prevents the transcription of target genes which may
183 induce cancer. In cases when the Wnt ligands are not captured by the corecep-
184 tor at the cell membrane, *Axin* helps in formation of the degradation complex.
185 The degradation complex phosphorylates β -*catenin* which is then recognized
186 by *Fbx/WD* repeat protein β -*TrCP*. β -*TrCP* is a component of ubiq-
187 uitin ligase complex that helps in ubiquitination of β -*catenin* thus marking it
188 for degradation via the proteasome. Cartoons depicting the phenomena of Wnt
189 being inactive and active are shown in figures 1(A) and 1(B), respectively.

190 2.2. Epigenetic factors

191 One of the widely studied epigenetic factors is methylation Costello and Plass
192 (2001), Das and Singal (2004), Issa (2007). Its occurrence leads to decrease in
193 the gene expression which affects the working of Wnt signaling pathways. Such
194 characteristic trends of gene silencing like that of secreted frizzled-related pro-
195 teins (*SFRP*) family in nearly all human colorectal tumor samples have been
196 found at extracellular level Suzuki et al. (2004). Similarly, methylation of genes
197 in the Dickkopf (*DKKx* Niehrs (2006), Sato et al. (2007), Dapper antagonist
198 of catenin (*DACTx* Jiang et al. (2008) and Wnt inhibitory factor-1 (*WIF1*
199 Taniguchi et al. (2005) family are known to have significant effect on the Wnt
200 pathway. Also, histone modifications (a class of proteins that help in the for-
201 mation of chromatin which packs the DNA in a special form Strahl and Allis
202 (2000) can affect gene expression Peterson et al. (2004). In the context of the
203 Wnt signaling pathway it has been found that *DACT* gene family show a pe-
204 culiar behavior in colorectal cancer Jiang et al. (2008). *DACT1* and *DACT2*
205 showed repression in tumor samples due to increased methylation while *DACT3*
206 did not show obvious changes to the interventions. It is indicated that *DACT3*
207 promoter is simultaneously modified by the both repressive and activating (bi-
208 valent) histone modifications Jiang et al. (2008).

209 2.3. Bayesian Networks

210 In reverse engineering methods for control networks Gardner and Faith
211 (2005) there exist many methods that help in the construction of the networks
212 from the data sets as well as give the ability to infer causal relations between
213 components of the system. A widely known architecture among these methods
214 is the Bayesian Network (BN). These networks can be used for causal reasoning
215 or diagnostic reasoning or both. It has been shown through reasoning and ex-
216 amples in Roehrig (1996) that the probabilistic inference mechanism applied via
217 Bayesian networks are analogous to the structural equation modeling in path
218 analysis problems.

219 Initial works on BNs in Pearl (1988) and Pearl (2000) suggest that the net-
220 works only need a relatively small amount of marginal probabilities for nodes
221 that have no incoming arcs and a set of conditional probabilities for each node
222 having one or more incoming arcs. The nodes form the driving components of
223 a network and the arcs define the interactive influences that drive a particular
224 process. Under these assumptions of influences the joint probability distribution
225 of the whole network or a part of it can be obtained via a special factorization
226 that uses the concept of direct influence and through dependence rules that de-
227 fine d-connectivity/separability as mentioned in Charniak (1991) and Needham
228 et al. (2007). This is illustrated through a simple example in Roehrig (1996).

229 The Bayesian networks work by estimating the posterior probability of the
230 model given the data set. This estimation is usually referred to as the Bayesian
231 score of the model conditioned on the data set. Mathematically, let $\mathcal{S}$ represent
232 the model given the data $\mathcal{D}$ and ξ is the background knowledge. Then according
233 to the Bayes Theorem Bayes and Price (1763):

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

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

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

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

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

Work on biological systems that make use of Bayesian networks can also be found in Friedman et al. (2000), Hartemink et al. (2001), Sachs et al. (2002), Sachs et al. (2005) and Peer et al. (2001). Bayesian networks are good in generating network structures and testing a targeted hypothesis which confine the experimenter to derive causal inferences Brent and Lok (2005). But a major disadvantage of the Bayesian networks is that they rely heavily on the conditional probability distributions which require good sampling of datasets and are computationally intensive. On the other hand, these networks are quite robust to the existence of the unobserved variables and accommodates noisy datasets. They also have the ability to combine heterogeneous data sets that incorporate different modalities.

In Sinha (2014), simple static Bayesian Network models have been developed with an aim to show how • incorporation of heterogeneous data can be done to increase prediction accuracy of test samples • prior biological knowledge can be embedded to model biological phenomena behind the Wnt pathway in colorectal cancer • to test the hypothesis regarding direct correspondence of active state

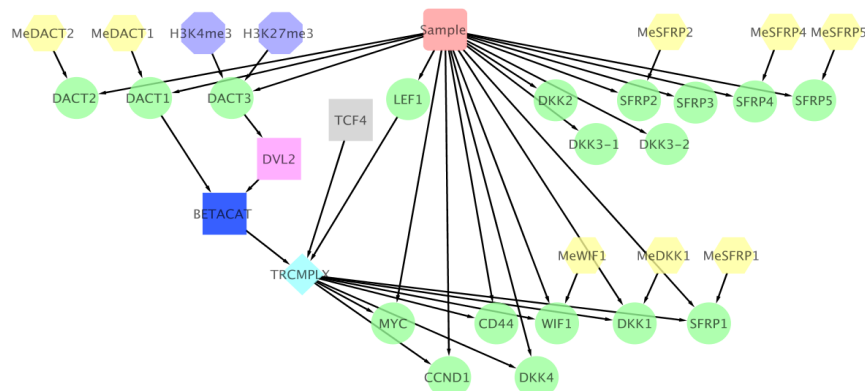


Figure 2: Influence diagram of $\mathcal{M}_{PBK+EI}$ contains partial prior biological knowledge and epigenetic information in the form of methylation and histone modification. Diagram drawn using CYTOSCAPE Shannon et al. (2003). In this model the state of Sample is distinguished from state of *TRCMPLX* that constitutes the Wnt pathway.

of β -catenin based transcription complex and the state of the test sample via segregation of nodes in the directed acyclic graphs of the proposed models and inferences can be made regarding the hidden biological relationships between a particular gene and the β -catenin transcription complex. This work uses MATLAB implemented BN toolbox from Murphy et al. (2001).

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

The NB model Ver assumes that the activation (inactivation) of β -catenin based transcription complex is equivalent to the fact that the sample is cancerous (normal). This assumption needs to be tested and Sinha (2014) proposes a newly improvised models based on prior biological knowledge and epigenetic information regarding the signaling pathway with the assumption that sample prediction may not always mean that the β -catenin based transcription complex is activated. These assumptions are incorporated by inserting another node of *Sample* for which gene expression measurements were available. This is separate from the *TRCMPLX* node that influences a particular set of known genes in the human colorectal cancer. For those genes whose relation with the *TRCMPLX* is currently not known or biologically affirmed, indirect paths through the *Sample* node to the *TRCMPLX* exist, technical aspect of which is described next.

For those factors whose relations were not yet confirmed but known to be involved in the pathway, the causal arcs were segregated via a latent variable introduced into the Bayesian network. The latent variable in the form of *Sample* (see figure 2) is extremely valuable as it connects the factors whose relations have not been confirmed till now, to factors whose influences have been confirmed

in the pathway. Finally, the introduction of latent variable in a causal model opens the avenue to assume the presence of measurements that haven't been recorded. Intuitively, for cancer samples the hidden measurements might be different from those for normal samples. The connectivity of factors through the variable provides an important route to infer biological relations.

Since all gene expressions have been measured from a sample of subjects the expression of genes is conditional on the state of the *Sample*. Here both tumorous and normal cases are present in equal amounts. The transcription factor *TRCMPLX* under investigation is known to operate with the help of interaction between β -catenin with *TCF4* and *LEF1* Waterman (2004), Kriegl et al. (2010). It is also known that the regions in the TSS of *MYC* Yochum (2011), *CCND1* Schmidt-Ott et al. (2007), *CD44* Kanwar et al. (2010), *SFRP1* Caldwell et al. (2006), *WIF1* Reguart et al. (2004), *DKK1* González-Sancho et al. (2004) and *DKK4* Pendas-Franco et al. (2008), Baehs et al. (2009) contain factors that have affinity to β -catenin based *TRCMPLX*. Thus expression of these genes are shown to be influenced by *TRCMPLX*, in figure 2.

Roles of *DKK2* Matsui et al. (2009) and *DKK3* Zitt et al. (2008), Veeck and Dahl (2012) have been observed in colorectal cancer but their transcriptional relation with β -catenin based *TRCMPLX* is not known. Similarly, *SFRP2* is known to be a target of *Pax2* transcription factor and yet it affects the β -catenin Wnt signaling pathway Brophy et al. (2003). Similarly, *SFRP4* Feng Han et al. (2006), Huang et al. (2010) and *SFRP5* Suzuki et al. (2004) are known to have affect on the Wnt pathway but their role with *TRCMPLX* is not well studied. *SFRP3* is known to have a different structure and function with respect to the remaining *SFRPx* gene family Hoang et al. (1996). Also, the role of *DACT2* is found to be conflicting in the Wnt pathway Kivimäe et al. (2011). Thus for all these genes whose expression mostly have an extracellular affect on the pathway and information regarding their influence on β -catenin based *TRCMPLX* node is not available, an indirect connection has been made through the *Sample* node. This connection will be explained at the end of this section.

Lastly, it is known that concentration of *DVL2* (a member of Disheveled family) is inversely regulated by the expression of *DACT3* Jiang et al. (2008). High *DVL2* concentration and suppression of *DACT1* leads to increase in stabilization of β -catenin which is necessary for the Wnt pathway to be active Jiang et al. (2008). But in a recent development Yuan et al. (2012) it has been found that expression of *DACT1* positively regulates β -catenin. Both scenarios need to be checked via inspection of the estimated probability values for β -catenin using the test data. Thus there exists direct causal relations between parent nodes *DACT1* and *DVL2* and child node, β -catenin. Influence of methylation (yellow hexagonal) nodes to their respective gene (green circular) nodes represent the affect of methylation on genes. Influence of histone modifications in *H3K27me3* and *H3K4me3* (blue octagonal) nodes to *DACT3* gene node represents the affect of histone modification on *DACT3*. The β -catenin (blue square) node is influenced by concentration of *DVL2* (depending on the expression state of *DACT3*) and behavior of *DACT1*.

The aforementioned established prior causal biological knowledge is imposed

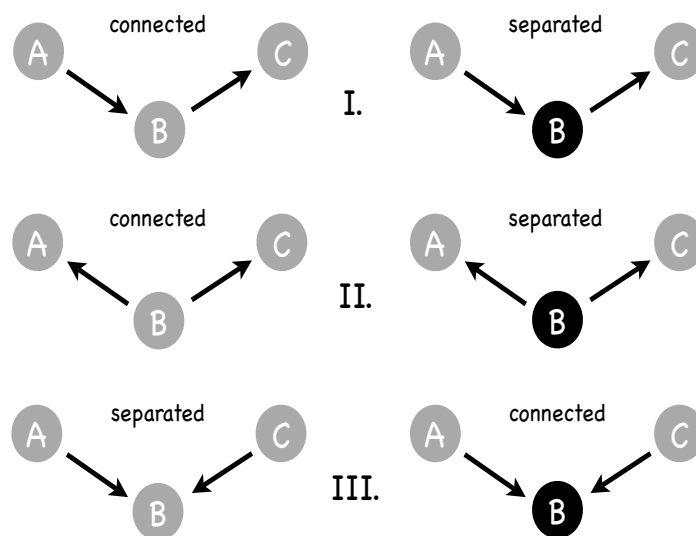


Figure 3: Cases for d-connectivity and d-separation. Black (Gray) circles mean evidence is available (not available) regarding a particular node.

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

In order to understand indirect connections further it is imperative to know about **d-connectivity/separability**. In a BN model this connection is established via the principle of **d-connectivity** which states that nodes are *connected* in a path when there exists no node in the path that has more than one incoming influence edge or there exists nodes in path with more than one incoming influence edge which are observed (i.e evidence regarding such nodes is available) Charniak (1991). Conversely, via principle of **d-separation** nodes are *separated* in a path when there exists nodes in the path that have more than one incoming influence edge or there exists nodes in the path with at most one incoming influence edge which are observed (i.e evidence regarding such nodes is available). Figure 3 represents three different cases of connectivity and separation between nodes $\mathcal{A}$ and $\mathcal{C}$ when the path between them passes through node $\mathcal{B}$. Connectivity or dependency exists between nodes $\mathcal{A}$ and $\mathcal{C}$ when • evidence is not present regarding node $\mathcal{B}$ in the left graphs of I. and II. in figure 3 or • evidence is present regarding node $\mathcal{B}$ in the right graph of III. in figure 3. Conversely, separation or independence exists between nodes $\mathcal{A}$ and $\mathcal{C}$ when • evidence is present regarding node $\mathcal{B}$ in the right graphs of I. and II. in figure 3 or • evidence is not present regarding node $\mathcal{B}$ in the left graph of III. in figure 3.

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

1. *SFRP3*, *Sample*, *SFRP1*, *TRCMPLX*
2. *SFRP3*, *Sample*, *DKK1*, *TRCMPLX*
3. *SFRP3*, *Sample*, *WIF1*, *TRCMPLX*
4. *SFRP3*, *Sample*, *CD44*, *TRCMPLX*
5. *SFRP3*, *Sample*, *DKK4*, *TRCMPLX*
6. *SFRP3*, *Sample*, *CCND1*, *TRCMPLX*
7. *SFRP3*, *Sample*, *MYC*, *TRCMPLX*
8. *SFRP3*, *Sample*, *LEF1*, *TRCMPLX*
9. *SFRP3*, *Sample*, *DACT3*, *DVL2*, β -catenin, *TRCMPLX*
10. *SFRP3*, *Sample*, *DACT1*, β -catenin, *TRCMPLX*

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

2.5. The logarithmic psychophysical law

Masin et al. (2009) states the Weber's law as follows -

Consider a sensation magnitude γ determined by a stimulus magnitude β . Fechner (1860) (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 (1860) (vol 1, p. 65) attributed to the German physiologist Ernst Heinrich Weber the empirical finding Weber (1834) that $\Delta\gamma$ remains constant when the relative stimulus increment $\frac{\Delta\beta}{\beta}$ remains constant, and named this finding Weber's law. Fechner (1860) (vol 2, p. 10) underlined that Weber's law was empirical.

It has been found that Bernoulli's principle Bernoulli (1738) is different from Weber's law Weber (1834) 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. (2009) shows that Weber's law is a special case of Bernoulli's principle and can be derived as follows - Equation 4 depicts the Bernoulli's principle and increment in sensation represented by $\Delta\gamma$ is proportional to change in stimulus represented by $\Delta\beta$.

$$\gamma = b \times \log \frac{\beta}{\alpha} \quad (4)$$

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

$$\Delta\gamma = b \times \log \frac{\beta + \Delta\beta}{\alpha} - b \times \log \frac{\beta}{\alpha} = b \times \log \left(\frac{\beta + \Delta\beta}{\beta} \right) = b \times \log \left(1 + \frac{\Delta\beta}{\beta} \right) \quad (5)$$

Since b is a constant, equation 5 reduces to

$$\Delta\gamma \propto \frac{\Delta\beta}{\beta} \quad (6)$$

where $\propto$ means "is constant when there is constancy of" from Masin et al. (2009). The final equation 6 is a formulation of Weber's laws in wordings and thus Bernoulli's principles imply Weber's law as a special case. Using Fechner (1860) derivation, it is possible to show the relation between Bernoulli's principles and Weber's law. Starting from the last line of equation 5, the following yields the relation.

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

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

3. Materials and methods

The models purported by Sinha (2014) involving the biological knowledge as well as epigenetic information depicted by $\mathcal{M}_{PBK+EI}$ and biological knowl-

edge 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 (2014).

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. (2008), a training set will contain 46 samples and a test set will contain 2 samples (one of normal and one of tumor). This procedure is repeated for every normal sample which is combined with each of the tumorous sample to form a series of test datasets. In total there will be 576 pairs of test data and 576 instances of training data. Note that for each test sample in a pair, the expression value for a gene is discretized using a threshold computed for that particular gene from the training set. Computation of the threshold has been elucidated in Sinha (2014). 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 $\Pr(g_i = \text{active} | g_k \text{ evidence}) \forall k$ genes. Note that these probabilities are recorded for both normal and tumor test samples.

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

4. Results and discussion on observation 1

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

Let γ be $\Pr(TRCMPLX = \text{active} | \text{all gene evidences})$, 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.

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

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

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

To compute $\Delta\gamma$, the 576 predictions of γ observed at $\beta = 90\%$ is subtracted from the 576 predictions of γ observed at $\beta = 80\%$ and a mean of the deviations recorded. This mean becomes $\Delta\gamma$. The procedure is computed again for different value of β . In this manuscript, the effect of constant and incremental deviations are observed. Tables 1 and 2 represent the deviations for models $\mathcal{M}_{PBK+EI}$ and $\mathcal{M}_{PBK}$, respectively.

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

For the case of constant deviations (figure 4) in model $\mathcal{M}_{PBK+EI}$, it was observed that deviations in activation of $TRCMPLX$ conditional on gene evidences for the tumor test samples showed a logarithmic behaviour and were directly proportional to the negative of both the Weber's and Bernoulli's law. This can be seen by the blue curve almost following the green and cyan curves. For the case of deviations in activation of $TRCMPLX$ conditional on gene evidences for the normal test samples showed an exponential behaviour and were proportional to negative of both the Weber's and Bernoulli's law. Similar behaviour was observed for all the coloured curves in case of incremental deviations as shown in figure 5. The exponential behaviour for activation of $TRCMPLX$ being active conditional on gene evidences correctly supports to the last line of

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

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

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

equation 7 which is the derivation of Weber's law from Bernoulli's equation. It actually point to Fechner's derivation of Weber's law from logarithmic formulation.

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

4.2. Intuition behind the curve behaviour

Lastly, the intuitive idea behind the behaviour of the curves generated from constant deviation in table 1 is as follows. It is expected that $\Pr(TRCMPLX = \text{active}|\text{all gene evidences})$ is low (high) in the case of Normal (Tumor) samples. The change $\Delta\Pr(TRCMPLX = \text{active}|\text{all gene evidences})$ jumps by power of

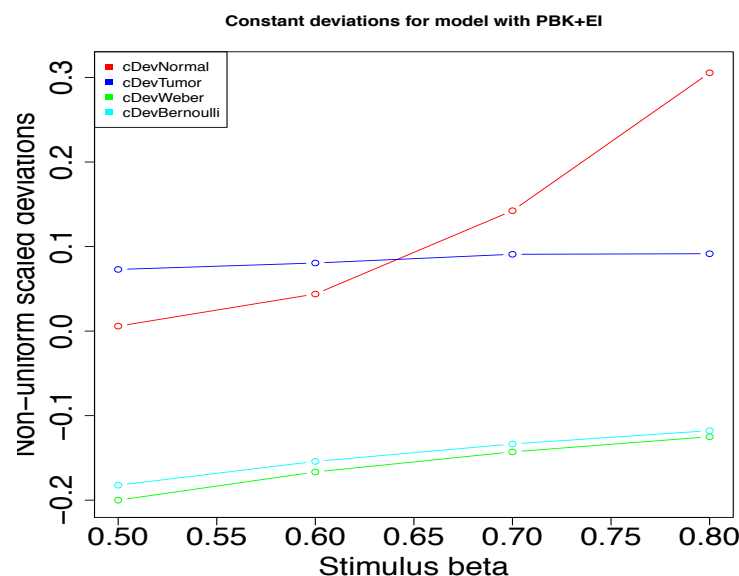


Figure 4: Constant deviations in β i.e. *ETGN* and corresponding deviations in $\Pr(\text{TRCMPLX} = \text{active} | \forall ge_i \text{ evidences})$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled 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.

10 as the β values change from 50% to 90% in Normal cases. It can be observed from the table that there are low deviations in $\Pr(\text{TRCMPLX} = \text{active} | \text{all gene evidences})$ when β is low i.e. the effect of transcription complex is low and high deviations in $\Pr(\text{TRCMPLX} = \text{active} | \text{all gene evidences})$ when β is high i.e. the effect of transcription complex is high. But it should be noted that the deviations still tend to be small. This implies that the *TRCMPLX* is switched off at a constant rate. Thus changes in β leads to exponential curves as in the formulation $\frac{\Delta\beta}{\beta}$, $\Delta\beta \rightarrow 0$ and $\beta \rightarrow \infty$.

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

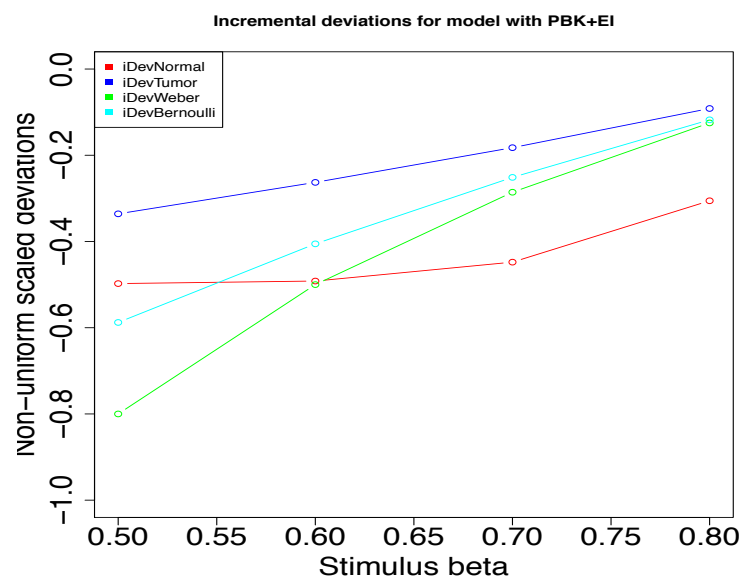


Figure 5: Incremental deviations in β i.e. *ETGN* and corresponding deviations in $\Pr(TRCMPLX = \text{active} | \forall g e_i \text{ evidences})$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations 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.

the logarithmic curve in deviations of prediction of transcription complex.

Finally, these observations present a crude yet important picture regarding the downstream transcriptional behaviour of signaling pathway in case of colorectal cancer. Though the current model does not measure the fold changes in the concentration levels of β -catenin, it can help in measuring the deviations in activity of the transcription complex conditional on the gene evidences by observing the deviations in the strength of belief assigned as priors in the probability tables of the node representing the transcription complex of the network. Thus sensitivity analysis facilitates in observing such natural phenomena at computational level. In context of the work by Goentoro and Kirschner (2009), the presented results are crude in terms of static observations yet they show corresponding behaviour of transcriptional activity in terms of psychophysical laws. Further investigations using dynamic models might reveal more information in comparison to the static models used in Sinha (2014). The observations presented here bolster the existence of behavioural phenomena in terms of logarithmic laws.

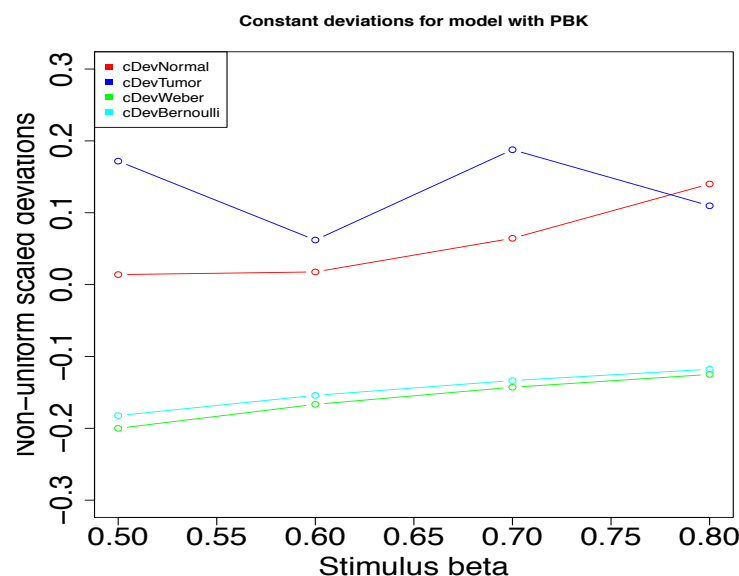


Figure 6: Constant deviations in β i.e. *ETGN* and corresponding deviations in $\Pr(\text{TRCMPLX} = \text{active} | \forall g_i \text{ evidences})$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled 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.

5. Preservation of gene gene interactions

The second part of this study was to find interactions between two genes by observing the conditional probability of activation status of one gene given the evidence of another gene. Let g be a gene. To obtain the results, two steps need to be executed in a serial manner. The first step is to construct gene gene interactions based on the available conditional probabilities denoted by $\Pr(g_i = \text{active/repressed} | g_k \text{ evidence}) \forall k$ genes. The conditional probabilities are inferred using the junction tree algorithm that employs two-pass message passing scheme. Example code and implementations of the same can be found in Murphy et al. (2001). The steps for constructing the gene gene interactions based on these conditional probabilities are documented in the Appendix. The second step is to infer gene gene interaction network based purely on reversible interactions. Note that networks are inferred for gene evidences using normal and tumor test samples separately.

Finally, once the interaction network is ready, the computational empirical estimates for deviations in gene-gene interaction is recorded and observation on the prevalence of psychophysical laws in these interactions is discussed. An important point that needs to be kept in mind is that the inferred interaction network differs based on the choice of the threshold involved (which is a

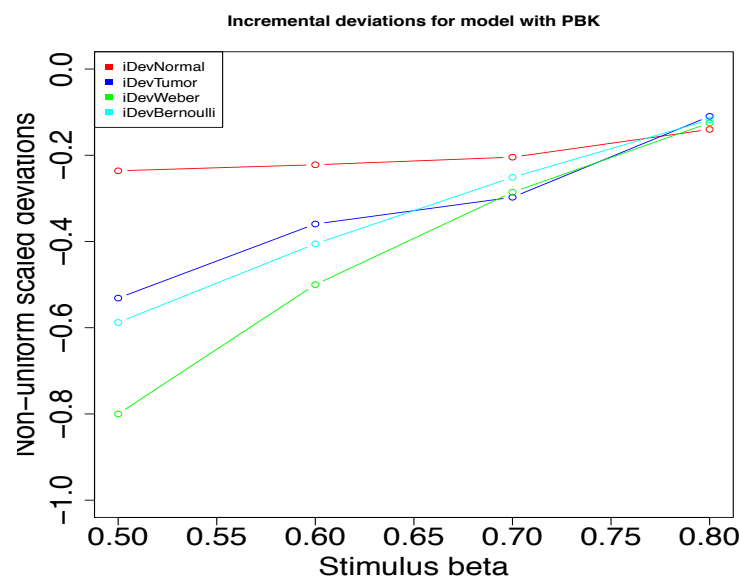


Figure 7: Incremental deviations in β i.e. *ETGN* and corresponding deviations in $\Pr(TRCPLX = \text{active} | \forall g_i \text{ evidences})$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations are also recorded. Note that the plots and the y-axis depict scaled deviations 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.

555 computational issue) but the underlying psychophysical laws remain unchanged
556 (which is a natural phenomena irrespective of the components involved). Thus
557 the while reading the observations on the psychophysical laws, readers must not
558 get confused regarding plots made for interactions from different networks.

559 5.1. Constructing gene-gene interactions

560 Gene interactions are constructed by labeling the inferred conditional proba-
561 bility of activation of g_j given the state of g_i , for all j 's & i 's. Here labels refer to
562 assigning $\langle \rangle$ for an activated gene and $|$ for repressed gene. Thus the following
563 possible combinations can be inferred - $g_j \langle \rangle - \langle \rangle g_i$, $g_j | - | g_i$, $g_j \langle \rangle - | g_i$
564 and $g_j | - \langle \rangle g_i$. Note that all interactions are basically depicting the degree
565 of belief in the state of g_i given or conditional on g_j i.e. $\Pr(g_i | g_j)$. The label
566 related to g_i is derived by discretizing $\Pr(g_i | g_j)$ with respect to a weighted mean
567 or the arbitrary value of 0.5. In any interaction, the label associated with g_j
568 is the evidence and the label associated with g_i is the predicted conditional
569 probability. Thus there will always exist two way interactions corresponding to
570 $\Pr(g_j | g_i)$ and $\Pr(g_i | g_j)$ in a Normal case. Similar interactions can be inferred for
571 the Tumor case. Which interactions to select is based on criteria of reversibility
572 and duplication, which is addressed later. To reiterate a final note regarding the

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

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

interactions - the inferred interactions differ based on the choice of the threshold involved (which is a computational issue) but if prevalent, the underlying psychophysical laws remain unchanged (which is a natural phenomena irrespective of the components involved).

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

- Each gene has an individual threshold that is different from the other as the expression values are different and the discretization used to estimate a particular 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

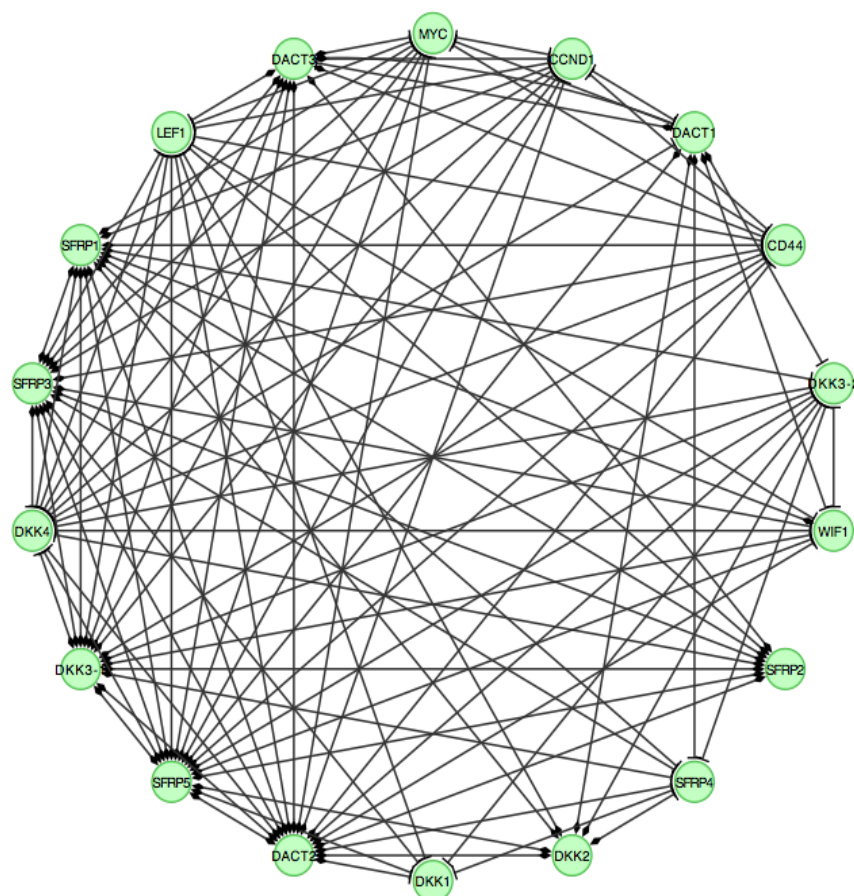


Figure 8: Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = 0.5$. Note that the effect of initialized cpt for *TRCMPLX* is 90% in tumorous case and 10% in normal case. Diamond <> means activation and straight bar | means repression.

588 value of 0.5. • Finally, the weighted mean could reveal interactions between two
589 genes that might be happening at different stages of time. Even though using
590 a static model, capturing such intricate interactions is possible as will be seen
591 later.

592 There is a formulation for weighted means, but the computation of the
593 weighted mean for training samples belonging to Normal and Tumor is done

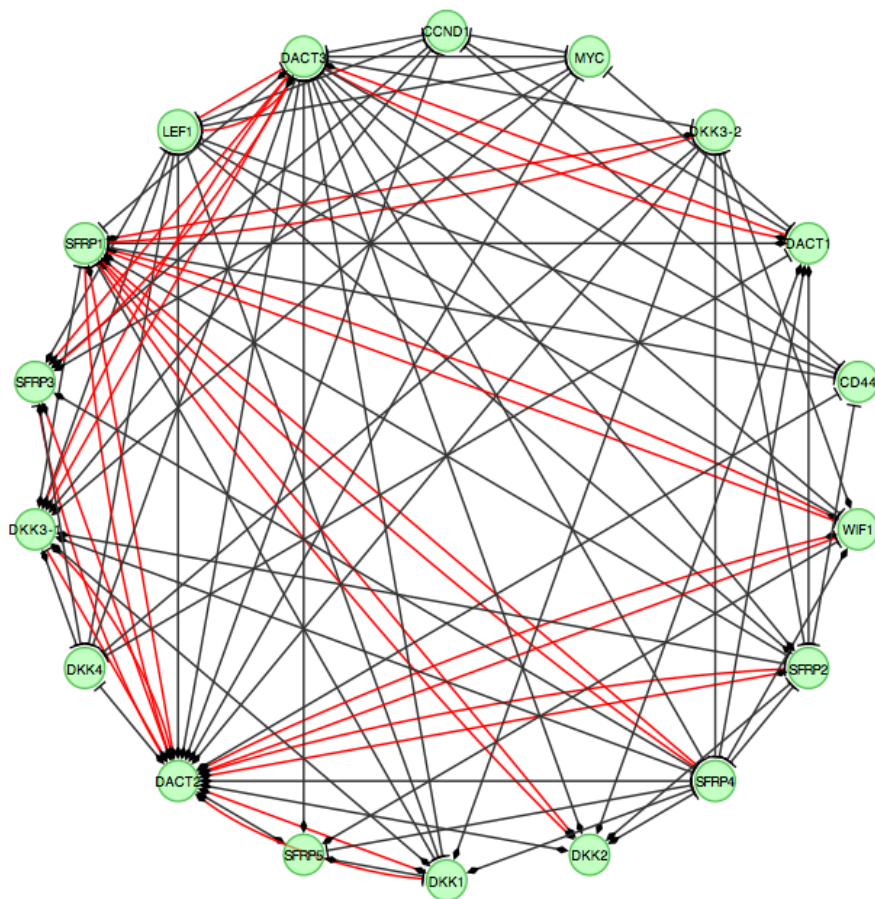


Figure 9: Gene gene interactions for normal case while using $\mathcal{M}_{PBK+EI}$ with $\theta = \theta_N$. Note that the effect of initialized cpt for *TRCMPLX* is 90% in tumorous case and 10% in normal case. Diamond <> means activation and straight bar | means repression.

separately. The separate formulations are given below -

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

were, $n_{1,N}$ is the number of Normal training samples with label 1, $n_{2,N}$ is the number of Normal training samples with label 2, $n_{1,T}$ is the number of Tumor training samples with label 1 and $n_{2,T}$ 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).

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

Table 4: Tabulated gene gene interactions of figure 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.

600 Based on the steps described in Appendix, for each gene a matrix is obtained
601 that shows the statistics of how the status of a gene is affected conditional on
602 the individual evidences of the remaining genes. Also, for each of the i^{th} gene
603 the averaged $\widehat{Pr}_N(g_i|g_k)$ is also stored in vector $PggN$. Same is done for tumor
604 cases. These two vectors are later used to test the veracity of existence of
605 psychophysical laws in gene-gene interaction network. Table 3 represents one
606 such tabulation for gene *SFRP3*. For all runs and all test samples, the following
607 was tabulated in table 3 : aa_N - *SFRP3* is active (a) when a gene is active (a)

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

608 in Normal (N) sample, ar_N - $SFRP3$ is active (a) when a gene is repressed (r)
609 in Normal (N) sample, ra_N - $SFRP3$ is repressed (r) when a gene is active (a)
610 in Normal (N) sample, rr_N - $SFRP3$ is repressed (r) when a gene is repressed
611 (r) in Normal (N) sample, aa_T - $SFRP3$ is active (a) when a gene is active (a)
612 in Tumor (T) sample, ar_T - $SFRP3$ is active (a) when a gene is repressed (r)
613 in Tumor (T) sample, pa_T - $SFRP3$ is repressed (r) when a gene is active (a)
614 in Tumor (T) sample, gg_{IN} - interaction of $SFRP3$ given the gene evidence
615 based on majority voting among aa_N , ar_N , ra_N and rr_N and finally, gg_{IT} -

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 6: Tabulated missing gene gene interactions of figure 8 and 9 using $\mathcal{M}_{PBK+EI}$ obtained in case of Normal samples. Interactions found in Normal samples with 80%, 70%, 60% and 50% effect that are not found with 90% and vice versa have been recorded. Here, the symbols represent the following - <> activation and | 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.

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

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

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

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

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

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

While employing the weighted mean as the threshold to discretize $\Pr(SFRP3 = \text{active}|g_j \text{ evidence})$, the *SFRP3* gene evidences that constitutes the test data are used. See step 5.a.iii in Appendix. Note that the test evidences for *SFRP3* are used for two purpose (1) to discretize $\Pr(SFRP3 = \text{active}|g_j \text{ evidence})$ as discussed above and (2) to compute the probability of activation status of another gene conditional on evidence for *SFRP3*, i.e $\Pr(g_j = \text{active}|SFRP3 \text{ evidence})$. Why to use test evidence or labels to compute weighted mean? Since the test evidence for a gene (i.e the discretized label) has been derived using the median computed on the corresponding training data for the same gene, it absolutely fine to use the discretized test labels to further compute the

662 weighted mean. This is because the median of gene expression is a value which
663 is much higher than the probability value of 1 and cannot be used to discretize a
664 predicted conditional probability value. Also, estimating the density estimates
665 from a small population of gene expression values has its own weakness. To
666 converge on a plausible realistic value the discretized test samples can be used
667 to estimate a weighted mean which represents the summary of the distribution
668 of the discretized values. This weighted mean of *SFRP3* test samples then
669 discretizes $\Pr(SFRP3 = \text{active}|g_j \text{ evidence})$ according to the inherently repre-
670 sented summary. More realistic estimates like kernel density estimates could
671 also be used.

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

686 5.2. Inferring gene-gene interaction network

687 Next, after the construction of gene-gene interactions, it is necessary to infer
688 the network. The inference of the estimated gene-gene interactions network is
689 based on explicitly reversible roles in Normal and Tumor test samples. This
690 means that only those interactions are selected which show the following prop-
691 erty - $g_j \leftrightarrow - \leftrightarrow g_i$ in Normal if and only if $g_j| - |g_i$ in Tumor, $g_j \leftrightarrow -|g_i$
692 in Normal if and only if $g_j| - \leftrightarrow g_i$ in Tumor, $g_j| - \leftrightarrow g_i$ in Normal if and
693 only if $g_j \leftrightarrow -|g_i$ in Tumor and finally, $g_j| - |g_i$ in Normal if and only if
694 $g_j \leftrightarrow - \leftrightarrow g_i$. This restricts the network to only reversible gene-gene inter-
695 actions in Normal and Tumor cases. Note that an interaction $g_j \mathcal{IR} g_i$ ($g_i \mathcal{IR} g_j$)
696 is depicted by $\Pr(g_i|g_j)$ ($\Pr(g_j|g_i)$).

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

705 Next, duplicate interactions are removed from the network for normal sam-
706 ples. This is repeated for the network based on tumor samples also. This is

Deviation study for *SFRP3* and *MYC* for normal case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1 + \frac{\Delta\beta}{\beta})$	$\Delta\gamma$ $\Pr(SFRP3 MYC)$	$\Delta\gamma$ $\Pr(MYC SFRP3)$
0.8	0.1	0.125	0.117783	0.003014287	0.00324456
0.7	0.1	0.1428571	0.1335314	0.002766111	0.00324456
0.6	0.1	0.1666667	0.1541507	0.002504868	0.00324456
0.5	0.1	0.2	0.1823216	0.002228110	0.00324456
0.8	0.1	0.125	0.117783	0.010513376	0.01297824
0.7	0.2	0.2857143	0.2513144	0.007499089	0.00973368
0.6	0.3	0.5	0.4054651	0.004732978	0.00648912
0.5	0.4	0.8	0.5877867	0.002228110	0.00324456

Table 7: Deviation study for $\Pr(SFRP3|MYC)$ and $\Pr(MYC|SFRP3)$ for normal case

achieved by removing one of the interactions from the following pairs ($g_j <>$
 $- <> g_i$ and $g_i <> - <> g_j$), ($g_j <> -|g_i$ and $g_i| - <> g_j$), ($g_j| - <> g_i$
and $g_i <> -|g_j$) and ($g_j| - |g_i$ and $g_i| - |g_j$). This process is done to remove
redundant interactions that are recorded via steps mentioned in construction of
gene-gene interaction network. Figure 8 shows one such network after complete
network construction, interaction labeling, consideration of reversible interactions
and removal of duplicate interactions using Normal test samples with
ETGN of 90% in $\mathcal{M}_{PBK+EI}$. For the case of Tumor test samples with ETGN
90% in $\mathcal{M}_{PBK+EI}$, only the reversal of interactions need to be done. Table
4 and 5 represents these interactions in figures 8 and 9 in a tabulated form,
respectively.

Finally, different networks were generated by varying the effect of *TRCMPLX*
(ETGN) and compared for the normal test samples. Table 6 represents the dif-
ferent interactions that were preserved in network from ETGN 90% with respect
to networks obtained from ETGN with values of 80%, 70%, 60% and 50%. It
was found that most of the genetic interactions depicted in figures 8 and 9
were found to be preserved across the different variations in ETGN as shown
in table 6. 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.

6. Results and discussion on observations 2 & 3

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

In the previous section, it was found that some of the interactions remain
preserved as there was change in the affect of transcription complex. The first

Deviation study for *SFRP3* and *MYC* for tumor case

β	$\Delta\beta$	$\frac{\Delta\beta}{\beta}$	$\log(1 + \frac{\Delta\beta}{\beta})$	$\Delta\gamma$ $\Pr(SFRP5 MYC)$	$\Delta\gamma$ $\Pr(MYC SFRP5)$
0.8	0.1	0.125	0.117783	-0.006463410	0.000000e+00
0.7	0.1	0.1428571	0.1335314	-0.006967724	5.551115e-17
0.6	0.1	0.1666667	0.1541507	-0.007515486	-5.551115e-17
0.5	0.1	0.2	0.1823216	-0.008112496	0.000000e+00
0.8	0.1	0.125	0.117783	-0.029059115	0.000000e+00
0.7	0.2	0.2857143	0.2513144	-0.022595705	0.000000e+00
0.6	0.3	0.5	0.4054651	-0.015627982	-5.551115e-17
0.5	0.4	0.8	0.5877867	-0.008112496	0.000000e+00

Table 8: Deviation study for $\Pr(SFRP3|MYC)$ and $\Pr(MYC|SFRP3)$ for tumor case

observation of this work was that deviations in the activity of the transcription complex followed a logarithmic-power psychophysical law. The manifestation of these laws at transcriptional levels can be attributed to the fold changes in *β -catenin* levels and the prevalence of Weber's law observed by Goentoro and Kirschner (2009). In this perspective, it would be interesting to observe if these laws are prevalent among the gene-gene interactions in the network or not.

Case: $\langle \rangle - |$ or $| - \langle \rangle$ with $\theta = \theta_N$

In Sinha (2014), the unknown behaviour of *SFRP3* in the Wnt pathway has been revealed slightly using computational causal inference. In figure 8, *SFRP3* shows preservation in the network and its interaction with other genetic factors involved in the model proposed in Sinha (2014) has been depicted. In one such paired interaction between *SFRP3* and *MYC*, *SFRP3* showed activation (repression) and *MYC* showed repression (activation) in normal (tumor) samples. As the change in the effect of transcription complex was induced by changing the initially assigned cpt values for *TRCMPLX* node, the deviations in the prediction of the gene-gene interaction network was observed to follow the logarithmic-power law crudely. What this means is that deviations or fold changes might also be prevalent at the gene-gene interaction level due to the upstream fold changes in *β -catenin* that induces transcriptional activity. More specifically, the deviation in the joint interaction that is represented by the degree of belief via the conditional probability of status of one gene given the evidence or activation status regarding another gene i.e $\Pr(g_i|g_j)$, is influenced by the fold changes upstream of the pathway and thus exhibit similar psychophysical laws.

Table 7 and 8 show these deviations in the prediction of the interactions for both the normal and the tumor cases. The tables show how deviations are affected when the changes in the effect of the transcription complex are done at constant and incremental rate. To summarize the results in these tables, graphs

were plotted in figures 10 for $\Pr(SFRP3|MYC)$ (constant deviations), 11 for $\Pr(MYC|SFRP3)$ (constant deviations), 12 for $\Pr(SFRP3|MYC)$ (incremental deviations) and 11 for $\Pr(MYC|SFRP3)$ (incremental deviations).

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

Similar behaviour was observed in the case of incremental deviations as depicted in figures 12 and 13. Analysis of the behaviour of other gene-gene interactions showing $\leftrightarrow -|$ or $|- \leftrightarrow$ can be observed in a similar way and can be produced by executing the R code in `Weber_Fechner_law.r` provided in Google drive <https://drive.google.com/folderview?id=0B7Kkv8w1hPU-T05wTTNodWNydjA&usp=sharing>. Note that plots need manual axis and title adjustments. Some of the plot results has been compressed in the zip file titled `Results-2015.zip`.

Case: $|-|$ or $\leftrightarrow - \leftrightarrow$ with $\theta = \theta_N$

Again, as pointed out in Sinha (2014), the unknown behaviour of $SFRP2$ in the Wnt pathway has been captured using computational causal inference. In figure 9, $SFRP2$ shows preservation in the network and its interaction with other genetic factors involved in the model proposed in Sinha (2014) has been depicted. In one such paired interaction between $SFRP2$ and $CD44$, both showed repression (activation) in normal (tumor) samples. As the change in the effect of transcription complex was induced via sensitizing the initially assigned cpt values, the deviations in the prediction of the gene-gene interaction network was observed to follow the logarithmic-power law crudely.

Table 9 and 10 show these deviations in the prediction of the interactions for both the normal and the tumor cases. The tables show how deviations are affected when the changes in the effect of the transcription complex are done at constant and incremental level. To summarize the results in these tables, graphs were plotted in figures 14 for $\Pr(SFRP2|CD44)$ (constant deviations), 15 for $\Pr(CD44|SFRP2)$ (constant deviations), 16 for $\Pr(SFRP2|CD44)$ (incremental deviations) and 15 for $\Pr(CD44|SFRP2)$ (incremental deviations).

Considering figure 14, when deviations are constant in both Weber and Bernoulli formulation, the deviations in the prediction of $\Pr(SFRP2|CD44)$ is observed to be logarithmic in the normal samples (apropos the Weber and Bernoulli deviations represented by green and cyan curves). Deviation in pre-

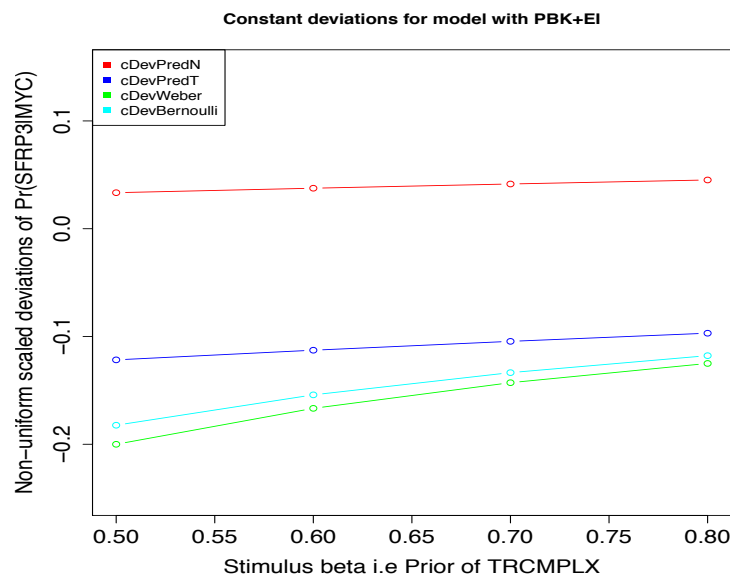


Figure 10: Constant deviations in β i.e ETGN and corresponding deviations in $\Pr(SFRP3|MYC)$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The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

808 dictions are depicted by the red (blue) curves for normal (tumor) samples.
809 Such a behaviour is not observed for $\Pr(CD44|SFRP2)$ as is depicted in figure
810 15. Even though $\Pr(CD44|SFRP2)$ was computationally estimated through a
811 model, the interaction for $CD44$ given $SFRP2$ was not observed in both normal
812 and tumor cases while the interaction for $SFRP2$ given $CD44$ was observed to
813 be reversible. This points to a crucial fact that the interactions interpreted from
814 conditional probabilities are not always two sided. Thus the interpretation for
815 $\Pr(g_i|g_j)$ is investigated in both directions as $g_i\mathcal{IR}g_j$ and $g_j\mathcal{IR}g_i$ to get a full
816 picture. Not that the results are wrong, but all angles of interpretations need
817 to be investigated to get the picture between any two genes. Similar behaviour
818 was observed in the case of incremental deviations as depicted in figures 16 and
819 17. Note that graph for incremental deviation in $\Pr(CD44|SFRP2)$ is just a
820 cumulative effect and does not state anything about the logarithmic law.

821 Finally, note that the predicted conditional probability a gene i given ev-
822 idence for gene j does not change but the inferred gene-gene interactions do
823 change depending on the choice of the threshold. These changes are depicted
824 in the figures 8 (table 4) and 9 (table 5). Dual interactions were inferred using
825 the weighted mean as a discretization factor, as is shown next. These are dual

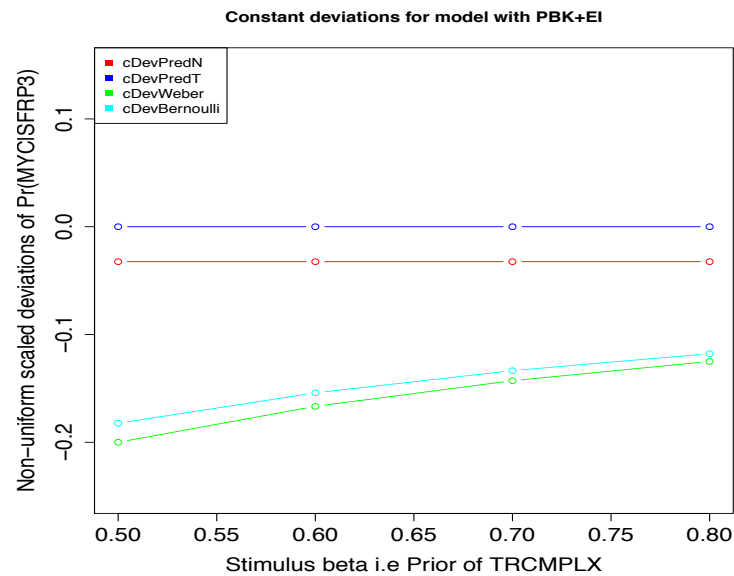


Figure 11: Same as figure 10 but for $\Pr(MYC|SFRP3)$.

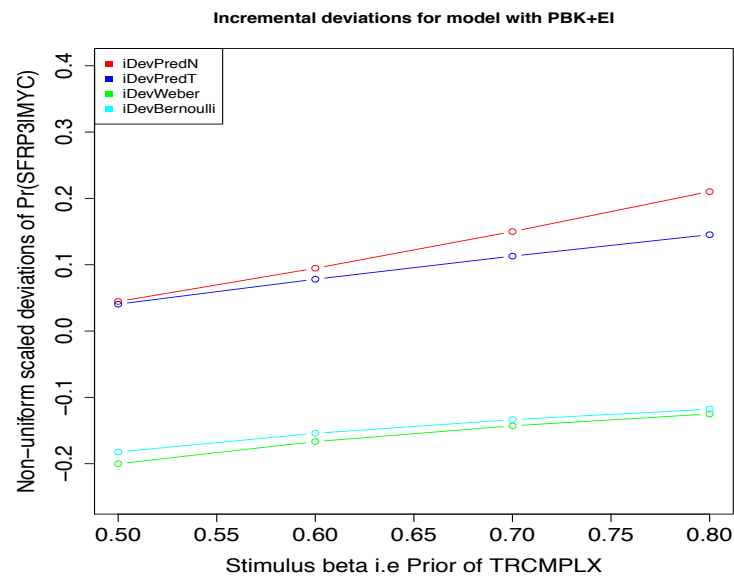


Figure 12: Same as figure 10 but for $\Pr(SFRP3|MYC)$. Instead of constant deviations, incremental deviations are represented.

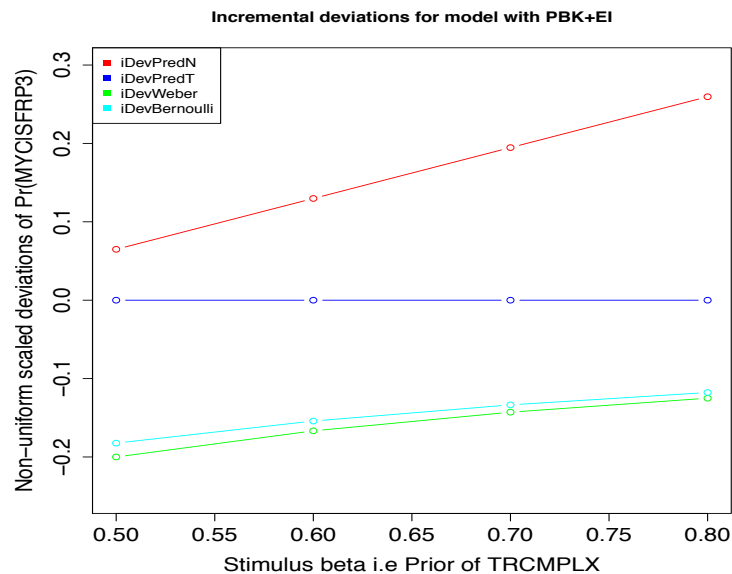


Figure 13: Same as figure 10 but for $\Pr(MYC|SFRP3)$. Instead of constant deviations, incremental deviations are represented.

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

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

Table 9: Deviation study for $\Pr(SFRP2|CD44)$ and $\Pr(CD44|SFRP2)$ for normal case

interactions are marked in red colour in figure 9.

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

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

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

Table 10: Deviation study for $\Pr(SFRP2|CD44)$ and $\Pr(CD44|SFRP2)$ for tumor case

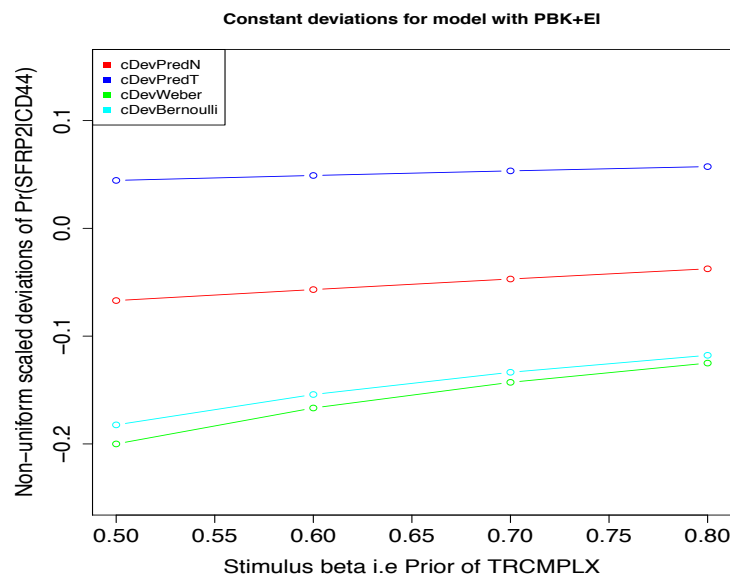


Figure 14: Constant deviations in β i.e ETGN and corresponding deviations in $\Pr(SFRP2|CD44)$ for both normal and tumor test samples. Corresponding Weber and Bernoulli deviations were also recorded. Note that the plots and the y-axis depict scaled deviations to visually analyse the observations. The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

832 expressed in normal sample might reverse activity at some stage or the other
833 (an vice versa). Here, one such interaction is discussed in detail. Interpretations

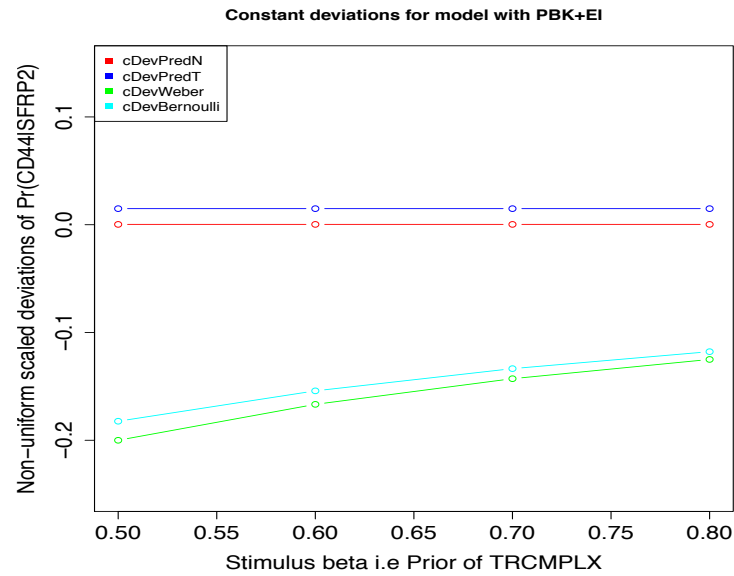


Figure 15: Same as figure 14 but for $\Pr(CD44|SFRP2)$.

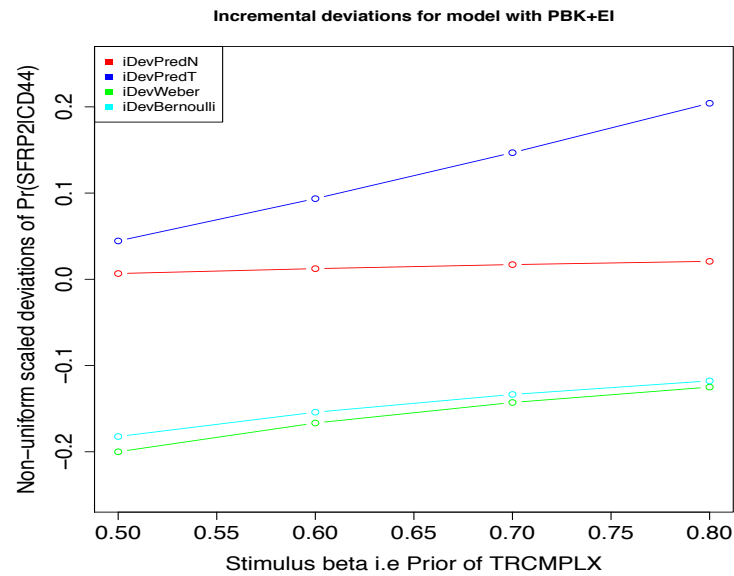


Figure 16: Same as figure 14 but for $\Pr(SFRP2|CD44)$. Instead of constant deviations, incremental deviations are represented.

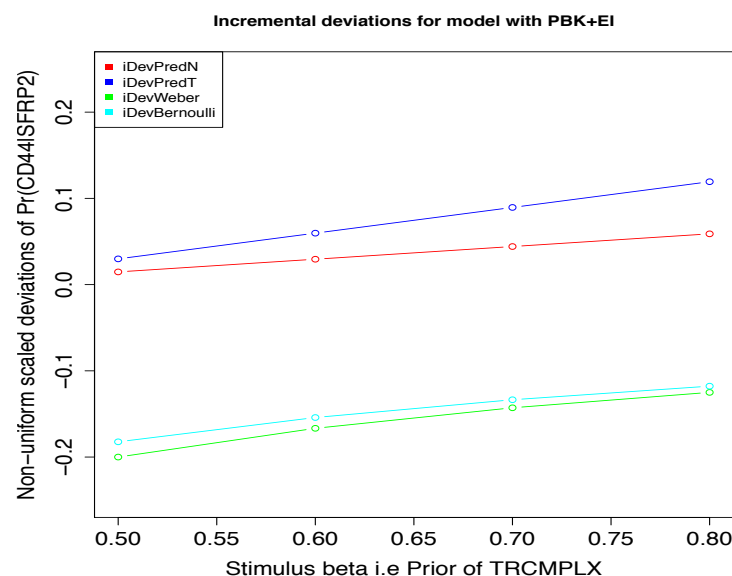


Figure 17: Same as figure 15 but for $\Pr(CD44|SFRP2)$. Instead of constant deviations, incremental deviations are represented.

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 arbitrary threshold of 0.5, the receiver operator curves (ROC) along with its corresponding area under the curve (AUC) are plotted. The ROCs are plotted using the discretized predicted values and the discretized labels obtained using the thresholds (computed from the training data) on the test data. The ROC graphs and their respective AUC values indicate how the predictions on the test data behaved under different values assigned to the TRCMPLX while training. Ideally, high values of AUC and steepness in ROC curve indicate good quality results. Finally, 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

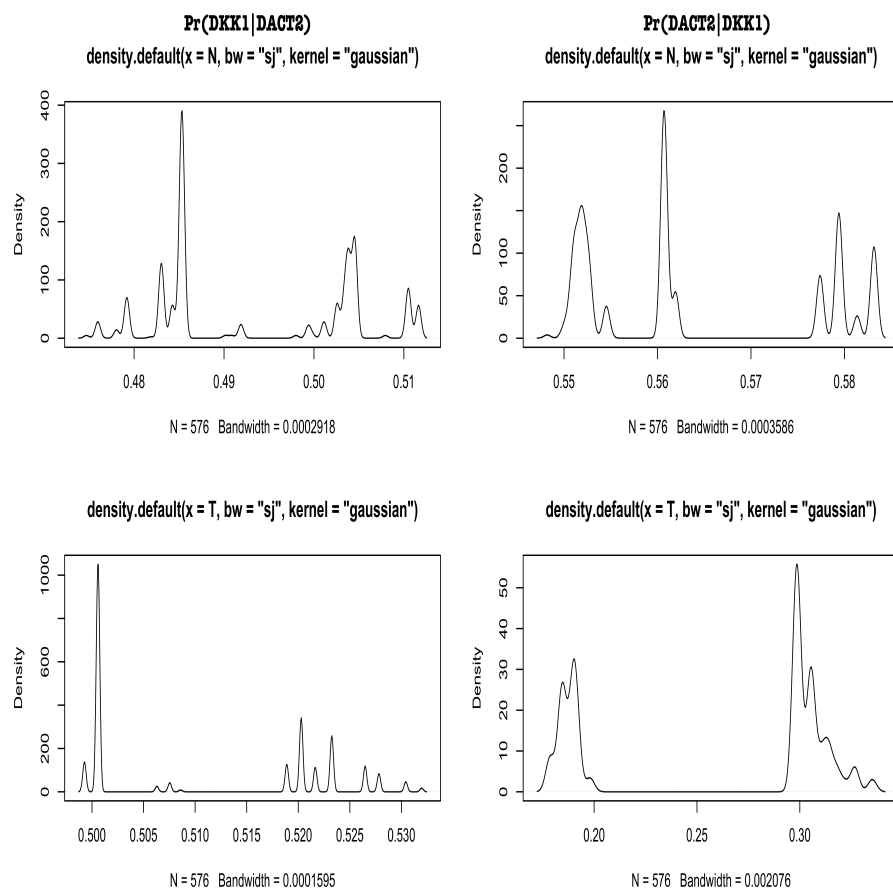


Figure 18: Kernel density estimates for predicted $\Pr(DKK1|DACT2)$ and $\Pr(DACT2|DKK1)$ in Normal and Tumor cases. Gaussian kernel is used for smoothing the density estimate. The bandwidth of the kernel is selected using the pilot estimation of derivative as proposed by Sheather and Jones (1991) and implemented in *R* programming language.

threshold value of 0.5. Plots are made using functions from the PRROC package provided by Grau et al. (2015).

Interaction between DKK1 and DACT2 using $\theta \in \{\theta_N, \theta_T\}$ - Dual interactions $DACT2 \leftrightarrow - \leftrightarrow DKK1$ and $DKK1| - \leftrightarrow DACT2$ ($DACT2| - |DKK1$ and $DKK1 \leftrightarrow -|DACT2$) in normal (tumor) sample were found as depicted in figure 9. Figure 18 shows the kernel density estimate of the predicted conditional probabilities for both normal and tumor test cases. Using the weighted mean of the discretized values of the test samples (discretization done using median estimated from the training data as mentioned before), the predicted $\Pr(DKK1|DACT2)$ and $\Pr(DACT1|DKK1)$ are classified as active

or passive. It might be useful to note that instead of using 0.5 as an arbitrary value, the weighted mean captures the distribution of labels in a much more realistic manner and helps infer interactions among the factors in the Wnt pathway.

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

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

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 following computations in equation 10 -

$$\begin{aligned}\theta_N^{DACT2} &= \frac{1 \times n_{1,N} + 2 \times n_{2,N}}{(1+2) \times (n_{1,N} + n_{2,N})} = \frac{1 \times 360 + 2 \times 216}{3 \times 576} = 0.4583333 \\ \theta_T^{DACT2} &= \frac{1 \times n_{1,T} + 2 \times n_{2,T}}{(1+2) \times (n_{1,T} + n_{2,T})} = \frac{1 \times 216 + 2 \times 360}{3 \times 576} = 0.5416667 \quad (10)\end{aligned}$$

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

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

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

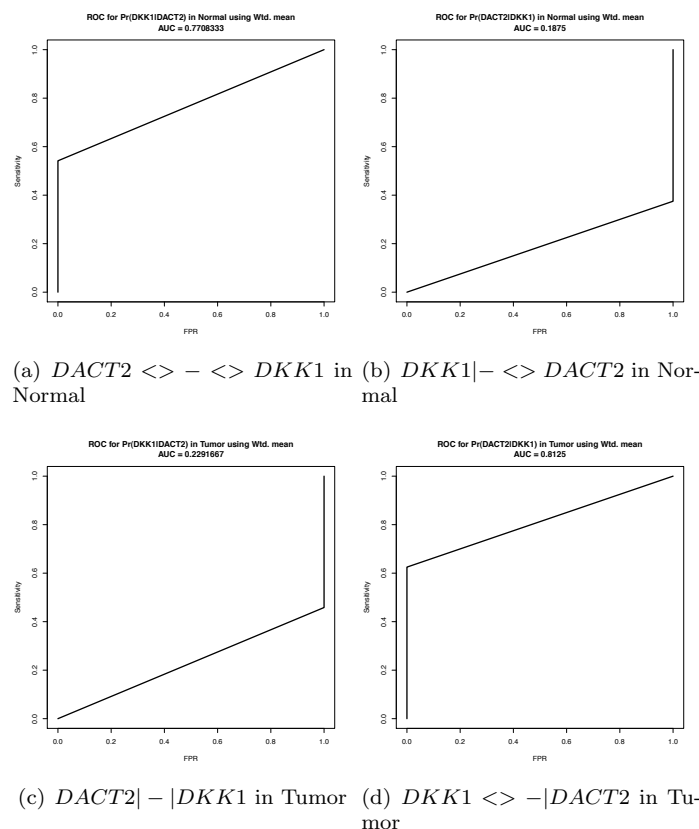


Figure 19: Column wise ROCs for $\Pr(DKK1|DACT2)$ (1st column) and $\Pr(DACT2|DKK1)$ (2nd column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using weighted mean for Normal case and weighted mean for Tumor case. Respective AUC values for the ROC curves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

899 In figure 19 column wise the ROCs for $\Pr(DKK1|DACT2)$ (1st column) and
900 $\Pr(DACT2|DKK1)$ (2nd column) have been plotted with ETGN value for the
901 90%. Row wise the plots depict the curves generated using weighted mean
902 for Normal case and weighted mean for Tumor case. It can be seen that us-
903 ing the weighted mean, the subfigure 19(a) and 19(d) convey a good guess
904 regarding the type of interaction prevailing in normal and tumor case. Thus
905 $DACT2 \lessgtr - \lessgtr DKK1$ i.e $\Pr(DKK1|DACT2)$ is highly favoured in Nor-
906 mal case while $DKK1 \lessgtr -|DACT2$ i.e $\Pr(DACT2|DKK1)$ is highly favoured
907 in Tumor case. Why this is so is because the normal cases show better results
908 in terms of prediction in comparison to the tumor cases. This points to the
909 fact that the interaction $DACT2 \lessgtr - \lessgtr DKK1$ is strongly supported in
910 the normal case in comparison to $DACT2| - |DKK1$ which is weakly sup-
911 ported 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 $DKK1| - <> DACT2$ counterpart in the normal case. Taken together, the dual interactions do exist but with different strengths of belief as shown conditional probability values. The curves in subfigure 19(b) and 19(c) indicate a bad guess and thus do not support the interactions $DKK1| - <> DACT2$ i.e $\Pr(DACT2|DKK1)$ in Normal case and $DACT2| - |DKK1$ i.e $\Pr(DKK1|DACT2)$ in Tumor case.

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

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

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

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

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

Comparing figures 20 and 19, it is clear that the later gives a better guess in terms of the interpretation of the interaction obtained by discretizing the kernel density estimates of inferred conditional probabilities. To evaluate the statistical significance of the predicted probabilities, the values of the KS test

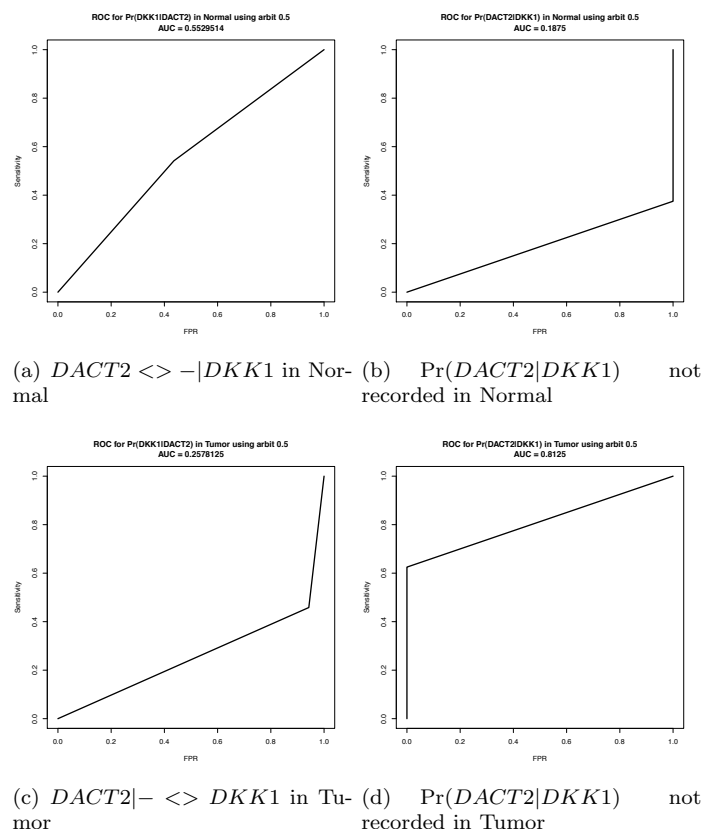


Figure 20: Column wise ROCs for $\Pr(DKK1|DACT2)$ (1^{st} column) and $\Pr(DACT2|DKK1)$ (2^{nd} column) have been plotted with ETGN value for the 90%. Row wise the plots depict the curves generated using arbit value of 0.5 for Normal case and arbit value of 0.5 for Tumor case. Respective AUC values for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

are tabulated and analyzed. Table 11 represents the computed values. The first four rows show the existing significance between the predictions for which the ROC curves have been 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 interpretations can be derived and respective measures can be plotted from the in silico observations.

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 11: 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.

7. Caveats

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

983 8. Future directions

984 In context of the above observations, dynamic models might reveal greater in-
 985 formation regarding the psychophysical laws. Work by Goentoro and Kirschner
 986 (2009) employs sensitivity analysis methods to reveal such laws by tuning sin-
 987 gle parameters. There might be a few ways to measure fold change in single
 988 an multi parameter settings. Future work might involve deeper study of the
 989 phenomena based on multi-parameter setting in a dynamic bayesian network
 990 model. If one incorporates nodes in between two time snapshots of β -catenin
 991 concentration in a dynamic bayesian network, one might be able to measure
 992 the changes at different phases of the signaling pathway. For example, in figure
 993 21 a set of nodes measuring the different concentrations of β -catenin (say N)
 994 are depicted. In a dynamic bayesian network, the previous concentration at t
 995 is connected to the next concentration at $t + 1$. Also, to measure the effect of
 996 difference (ΔN), a change in concentration can be measured. Computations
 997 regarding fold change (ΔN) could then be estimated as posterior probabilities
 998 given the two concentrations, which the Bayesian networks can easily handle.
 999 In case more parameters need to be involved (say the effect of Wnt and APC
 1000 together), nodes might be added as shown below. Then the fold change is
 1001 conditional on $N(t + 1)$, $N(t + 2)$, ΔWnt and ΔAPC and is estimated as
 1002 $\Pr(\Delta N(t + 1) | N(t + 1), N(t + 2), \Delta Wnt, \Delta APC)$.

1003 Regarding sensitivity analysis, in nonlinear problems, it might be useful to
 1004 use Sobol' (1990) indices to estimate the sensitivity of the parameters. These
 1005 indices are a way to estimate the changes in a multiparameter setting thus
 1006 helping one to conduct global sensitivity analysis instead of local sensitivity
 1007 analysis Glen and Isaacs (2012). Finally, with respect to the robustness of the
 1008 gene-gene interaction network, the current work employs a very simple algorithm
 1009 to construct the network and infer preserved interactions across the range of
 1010 values set for a particular parameter. This helps in eliminating interactions
 1011 that do not contribute enough biological information in the pathway or are non
 1012 existant and require further analysis by integration of more data. Work in these
 1013 lines would require incorporation of bigger datasets.

1014 9. Conclusions

1015 In this preliminary work via sensitivity analysis, the variation in predictive
 1016 behaviour of β -catenin based transcription complex conditional on gene evi-
 1017 dences is shown to follow power-logarithmic psychophysical law crudely. This
 1018 implies deviations in output are proportional to increasing function of devia-
 1019 tions in the input and show constancy for higher values of input. This points
 1020 towards stability in the behaviour of transcriptional activity downstream of the
 1021 Wnt pathway. As a further development, computational analysis shows that
 1022 the preserved gene-gene interactions are also subject to these power-logarithmic
 1023 psychophysical laws. The prevalence of these laws is reported for interaction
 1024 between elements of pairs of ($SFRP3$, MYC), ($SFRP2$, $CD44$) and ($DKK1$,
 1025 $DACT2$). As a precursor to the analysis of these laws at interaction level,

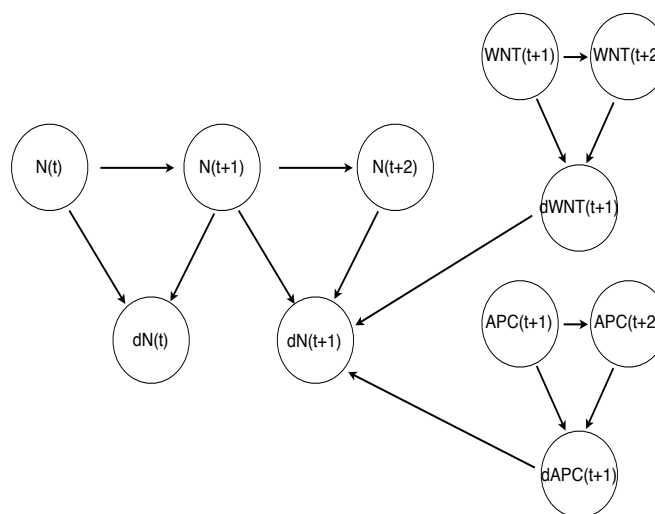


Figure 21: A schematic diagram of a dynamic bayesian network model that might help study the fold change and the logarithmic psychophysical laws behind the changes.

the biologically inspired epigenetically influenced computational causal models were used to infer gene-gene interaction from conditional probabilities of individual gene activation given the status of another gene activation. In relation of colorectal cancer cases, it is now possible to infer the type of interaction that might be happening among the genes at a pair wise level using BN models and further wet lab studies can be developed to investigate the inferred prevalence of power-logarithmic psychophysical laws at interaction level within the pathway. To assert the fact, in a recent development via wet lab experiments by Olsman and Goentoro (2016), it has been confirmed that there are existence of sensors that behave in a logarithmic fashion. The wet lab work by Olsman and Goentoro (2016) supports the earlier proposed crude postulates based on computational sensitivity analysis of this manuscript regarding the existence of logarithmic behaviour in the signaling pathways. It also signifies the importance of systems biology approach where in silico experiments combined with in vivo/in vitro experiments have the power to explore the deeper mechanisms of a signaling pathway.

Acknowledgement

Thanks to - • Dr. Uri Alon from Weizmann Institute of Science, Israel, for lively and critical discussion on behaviour of curves at the International Conference on Systems Biology of Human Disease (2015), held at German Cancer

Research Center in Heidelberg, Germany. • Dr. Silja Renooij for providing useful feedback and editing the manuscript. • the Royal Society of Chemistry (RSC) for giving permission to reproduce parts of material in S. Sinha, Integr. Biol., 2014, DOI: 10.1039/C4IB00124A. • all anonymous reviewers who have helped in refining this manuscript.

Declaration of Interest

No conflict of interest.

Appendix

9.1. Steps for construction of gene gene interaction networks

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

1. $\forall$ 576 runs iterated by a counter j
 - (a) append x_N with the vector whose elements are $\Pr(g_i = \text{active} | g_k \text{ evidence}) \forall k$ genes in the j^{th} 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 $\Pr(g_i = \text{active} | g_k \text{ evidence}) \forall k$ genes in the j^{th} 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 j^{th} 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 j^{th} 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 x_N and x_T to obtain averaged $\widehat{\Pr}_N(g_i | g_k)$ and $\widehat{\Pr}_T(g_i | g_k) \forall k$ gene evidences and $\forall i$ genes. Note $k, i \in 1, \dots, n$ if n is the total number of genes.
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 i^{th} one
 - (a) if($k \neq i$)
 - i. assign variables $tmpaa_N, tmpar_N, tmprr_N, tmpaa_T, tmpar_T, tmprr_T$ to 0.
 - ii. assign threshold values θ to either a fixed value (say 0.5) or a weighted mean.

```

1085   iii. if assigning a weighted mean, compute the threshold  $\theta_N$  as the
1086         weighted mean of the labels of the test data i.e evidences for the
1087          $i^{th}$  gene, in the case of Normal samples (top formula in equation
1088         8). Similarly, compute the threshold  $\theta_T$  as the weighted mean of
1089         the labels of the test data i.e evidences for the  $i^{th}$  gene, in the
1090         case of Tumor samples (bottom formula in equation 8).
1091   iv.  $\forall$  576 runs iterated by a counter  $l$ 
1092         A. if( $ge_N[l,k] == 1$  and  $x_N[l,k] < \theta$ ) increment  $tmprrr_N$  by 1
1093         B. else if( $ge_N[l,k] == 1$  and  $x_N[l,k] \geq \theta$ ) increment  $tmpar_N$ 
1094             by 1
1095         C. else if( $ge_N[l,k] == 2$  and  $x_N[l,k] < \theta$ ) increment  $tmppra_N$  by
1096             1
1097         D. else if( $ge_N[l,k] == 2$  and  $x_N[l,k] \geq \theta$ ) increment  $tmpaa_N$ 
1098             by 1
1099         E. if( $ge_T[l,k] == 1$  and  $x_T[l,k] < \theta$ ) increment  $tmprrr_T$  by 1
1100         F. else if( $ge_T[l,k] == 1$  and  $x_T[l,k] \geq \theta$ ) increment  $tmpar_T$ 
1101             by 1
1102         G. else if( $ge_T[l,k] == 2$  and  $x_T[l,k] < \theta$ ) increment  $tmppra_T$  by
1103             1
1104         H. else if( $ge_T[l,k] == 2$  and  $x_T[l,k] \geq \theta$ ) increment  $tmpaa_T$ 
1105             by 1
1106   v. Comment - store results
1107   vi. append  $ge$  with  $g_k$ ,  $rr_N$  with  $tmprrr_N$ ,  $ar_N$  with  $tmpar_N$ ,  $ra_N$ 
1108       with  $tmppra_N$ ,  $aa_N$  with  $tmpaa_N$ ,  $rr_T$  with  $tmprrr_T$ ,  $ar_T$  with
1109        $tmpar_T$ ,  $ra_T$  with  $tmppra_T$  and  $aa_T$  with  $tmpaa_T$ 
1110   (b) store the variables in the previous step to a data frame (say in R
1111       language) to a variable  $stats$ .
1112   (c) Comment - 1 means aa, 2 means ar, 3 means ra, 4 means rr
1113   (d) assign variables  $gg_{IN}$  and  $gg_{IT}$  as empty vector  $\emptyset$ 
1114   (e)  $\forall j$  gene except the  $i^{th}$  one under consideration
1115       i. find the index  $idx_N$  in stats that corresponds to 1 or 2 or 3 or 4
1116       ii. if( $idx_N == 1$ ) append  $gg_{IN}$  with interaction string  $stats\&g_j <>$ 
1117            $- <> g_i$ 
1118       iii. else if( $idx_N == 2$ ) append  $gg_{IN}$  with interaction string  $stats\&g_j|- <>$ 
1119            $g_i$ 
1120       iv. else if( $idx_N == 3$ ) append  $gg_{IN}$  with interaction string  $stats\&g_j <>$ 
1121            $-|g_i$ 
1122       v. else if( $idx_N == 4$ ) append  $gg_{IN}$  with interaction string  $stats\&g_j|-$ 
1123            $|g_i$ 
1124       vi. find the index  $idx_N$  in stats that corresponds to 1 or 2 or 3 or 4
1125       vii. if( $idx_T == 1$ ) append  $gg_{IT}$  with interaction string  $stats\&g_j <>$ 
1126            $- <> g_i$ 
1127       viii. else if( $idx_T == 2$ ) append  $gg_{IT}$  with interaction string  $stats\&g_j|- <>$ 
1128            $g_i$ 
1129       ix. else if( $idx_T == 3$ ) append  $gg_{IT}$  with interaction string  $stats\&g_j <>$ 
1130            $-|g_i$ 

```

```

1131         x. else if( $idx_T == 4$ ) append  $gg_{IT}$  with interaction string  $stats\{g_j|-$ 
1132            $|g_i$ 
1133         (f) assign  $stats\{gg_{IN}$  with  $gg_{IN}$ 
1134         (g) assign  $stats\{gg_{IT}$  with  $gg_{IT}$ 
1135         (h) Comment -  $i^{th}$  gene influenced
1136         (i)  $gg_I[[i]] < - \text{list}(ig = g_i, stats = stats, PggN = PggN, PggT =$ 
1137            $PggT)$ 

```

1138 References

- 1139 , .
- 1140 Adler, M., Mayo, A., Alon, U., 2014. Logarithmic and power law input-output relations in sensory
1141 systems with fold-change detection. *PLoS computational biology* 10, e1003781.
- 1142 Baehs, S., Herbst, A., Thieme, S., Perschl, C., Behrens, A., Scheel, S., Jung, A., Brabletz, T.,
1143 Göke, B., Blum, H., et al., 2009. Dickkopf-4 is frequently down-regulated and inhibits growth of
1144 colorectal cancer cells. *Cancer letters* 276, 152–159.
- 1145 Baron, R., Kneissel, M., 2013. Wnt signaling in bone homeostasis and disease: from human muta-
1146 tions to treatments. *Nature medicine* 19, 179–192.
- 1147 Bayes, M., Price, M., 1763. An essay towards solving a problem in the doctrine of chances. *Philo-
1148 sophical Transactions* 53, 370.
- 1149 Bernoulli, D., 1738. Specimen theoriae novae de mensura sortis. *Commentarii Academiae Scien-
1150 tiarum Imperialis Petropolitanae* 5, 175–192.
- 1151 Blagodatski, A., Poteryaev, D., Katanaev, V., 2014. Targeting the wnt pathways for therapies. *Mol
1152 Cell Ther* 2, 28.
- 1153 Brent, R., Lok, L., 2005. A fishing buddy for hypothesis generators. *Science* 308, 504.
- 1154 Brophy, P., Lang, K., Dressler, G., 2003. The secreted frizzled related protein 2 (sfrp2) gene is a
1155 target of the pax2 transcription factor. *Journal of Biological Chemistry* 278, 52401–52405.
- 1156 Caldwell, G., Jones, C., Taniere, P., Warrack, R., Soon, Y., Matthews, G., Morton, D., 2006. The
1157 wnt antagonist sfrp1 is downregulated in premalignant large bowel adenomas. *British journal of
1158 cancer* 94, 922–927.
- 1159 Charniak, E., 1991. Bayesian networks without tears. *AI magazine* 12, 50.
- 1160 Clevers, H., 2006. Wnt/ $[\beta]$ -catenin signaling in development and disease. *Cell* 127, 469–480.
- 1161 Costello, J., Plass, C., 2001. Methylation matters. *Journal of medical genetics* 38, 285–303.
- 1162 Curtin, J.C., Lorenzi, M.V., 2010. Drug discovery approaches to target wnt signaling in cancer stem
1163 cells. *Oncotarget* 1, 552.
- 1164 Das, P., Singal, R., 2004. Dna methylation and cancer. *Journal of Clinical Oncology* 22, 4632–4642.
- 1165 Fechner, G.T., 1860. *Elemente der Psychophysik* (2 vols). Breitkopf and Hartel.
- 1166 Feng Han, Q., Zhao, W., Bentel, J., Shearwood, A., Zeps, N., Joseph, D., Iacopetta, B., Dharmara-
1167 jan, A., 2006. Expression of sfrp-4 and β -catenin in human colorectal carcinoma. *Cancer letters*
1168 231, 129–137.
- 1169 Friedman, N., Linial, M., Nachman, I., Pe’er, D., 2000. Using bayesian networks to analyze expres-
1170 sion data. *Journal of computational biology* 7, 601–620.

Garber, K., 2009. Drugging the wnt pathway: problems and progress. *Journal of the National Cancer Institute* 101, 548–550.

Gardner, T., Faith, J., 2005. Reverse-engineering transcription control networks. *Physics of Life Reviews* 2, 65–88.

Glen, G., Isaacs, K., 2012. Estimating sobol sensitivity indices using correlations. *Environmental Modelling & Software* 37, 157–166.

Goentoro, L., Kirschner, M.W., 2009. Evidence that fold-change, and not absolute level, of β -catenin dictates wnt signaling. *Molecular Cell* 36, 872–884.

González-Sancho, J., Aguilera, O., García, J., Pendás-Franco, N., Peña, C., Cal, S., de Herreros, A., Bonilla, F., Muñoz, A., 2004. The wnt antagonist dickkopf-1 gene is a downstream target of β -catenin/tcf and is downregulated in human colon cancer. *Oncogene* 24, 1098–1103.

Grau, J., Grosse, I., Keilwagen, J., 2015. Pprcc: computing and visualizing precision-recall and receiver operating characteristic curves in r. *Bioinformatics* , btv153.

Gregorieff, A., Clevers, H., 2005. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes & development* 19, 877–890.

Hartemink, A., Gifford, D., Jaakkola, T., Young, R., 2001. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks, in: *Pac. Symp. Biocomput*, Citeseer. pp. 422–433.

Hoang, B., Moos, M., Vukicevic, S., Luyten, F., 1996. Primary structure and tissue distribution of frzb, a novel protein related to drosophila frizzled, suggest a role in skeletal morphogenesis. *Journal of Biological Chemistry* 271, 26131–26137.

Huang, D., Yu, B., Deng, Y., Sheng, W., Peng, Z., Qin, W., Du, X., 2010. Sfrp4 was overexpressed in colorectal carcinoma. *Journal of cancer research and clinical oncology* 136, 395–401.

Issa, J., 2007. Dna methylation as a therapeutic target in cancer. *Clinical Cancer Research* 13, 1634–1637.

Jiang, X., Tan, J., Li, J., Kivimäe, S., Yang, X., Zhuang, L., Lee, P., Chan, M., Stanton, L., Liu, E., Cheyette, B., Yu, Q., 2008. Dact3 is an epigenetic regulator of wnt/ β -catenin signaling in colorectal cancer and is a therapeutic target of histone modifications. *Cancer cell* 13, 529–541.

Kahn, M., 2014. Can we safely target the wnt pathway? *Nature Reviews Drug Discovery* 13, 513–532.

Kanwar, S., Yu, Y., Nautiyal, J., Patel, B., Majumdar, A., 2010. The wnt/ β -catenin pathway regulates growth and maintenance of colonospheres. *Molecular cancer* 9, 212.

Kivimäe, S., Yang, X., Cheyette, B., 2011. All dact (dapper/frodo) scaffold proteins dimerize and exhibit conserved interactions with vangl, dvl, and serine/threonine kinases. *BMC biochemistry* 12, 33.

Kriegel, L., Horst, D., Reiche, J., Engel, J., Kirchner, T., Jung, A., et al., 2010. Lef-1 and tcf4 expression correlate inversely with survival in colorectal cancer. *J. Transl. Med* 8, 123.

Masin, S.C., Zudini, V., Antonelli, M., 2009. Early alternative derivations of fechner’s law. *Journal of History of the Behavioral Sciences* 45, 56–65. doi:10.1002/jhbs.20349.

Matsui, A., Yamaguchi, T., Maekawa, S., Miyazaki, C., Takano, S., Uetake, T., Inoue, T., Otaka, M., Otsuka, H., Sato, T., et al., 2009. Dickkopf-4 and -2 genes are upregulated in human colorectal cancer. *Cancer science* 100, 1923–1930.

Murphy, K., et al., 2001. The bayes net toolbox for matlab. *Computing science and statistics* 33, 1024–1034.

1215 Needham, C.J., Bradford, J.R., Bulpitt, A.J., Westhead, D.R., 2007. A primer on learning in
1216 bayesian networks for computational biology. *PLoS computational biology* 3, e129.

1217 Niehrs, C., 2006. Function and biological roles of the dickkopf family of wnt modulators. *Oncogene*
1218 25, 7469–7481.

1219 Olsman, N., Goentoro, L., 2016. Allosteric proteins as logarithmic sensors. *Proceedings of the*
1220 *National Academy of Sciences*, 201601791.

1221 Pearl, J., 1988. Probabilistic reasoning in intelligent systems: networks of plausible inference.
1222 Morgan Kaufmann.

1223 Pearl, J., 2000. Causality: models, reasoning, and inference. Cambridge Univ Pr.

1224 Pećina-Šlaus, N., 2010. Wnt signal transduction pathway and apoptosis: a review. *Cancer Cell*
1225 *International* 10, 1–5.

1226 Pendas-Franco, N., Garcia, J., Pena, C., Valle, N., Palmer, H., Heinäniemi, M., Carlberg, C.,
1227 Jimenez, B., Bonilla, F., Munoz, A., et al., 2008. Dickkopf-4 is induced by tcf/ β -catenin and up-
1228 regulated in human colon cancer, promotes tumour cell invasion and angiogenesis and is repressed
1229 by 1α , 25-dihydroxyvitamin d3. *Oncogene* 27, 4467–4477.

1230 Peterson, C., Laniel, M., et al., 2004. Histones and histone modifications. *Current Biology* 14,
1231 546–551.

1232 Peer, D., Regev, A., Elidan, G., Friedman, N., 2001. Inferring subnetworks from perturbed expres-
1233 sion profiles. *Bioinformatics* 17, S215.

1234 Pinto, D., Gregorieff, A., Begthel, H., Clevers, H., 2003. Canonical wnt signals are essential for
1235 homeostasis of the intestinal epithelium. *Genes & development* 17, 1709–1713.

1236 Rao, T.P., Kühl, M., 2010. An updated overview on wnt signaling pathways a prelude for more.
1237 *Circulation research* 106, 1798–1806.

1238 Reguart, N., He, B., Xu, Z., You, L., Lee, A., Mazieres, J., Mikami, I., Batra, S., Rosell, R., Mc-
1239 Cormick, F., et al., 2004. Cloning and characterization of the promoter of human wnt inhibitory
1240 factor-1. *Biochemical and biophysical research communications* 323, 229–234.

1241 Roehrig, S., 1996. Probabilistic inference and path analysis. *Decision Support Systems* 16, 55–66.

1242 Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., Lauffenburger, D., 2002. Bayesian network approach
1243 to cell signaling pathway modeling. *Sci. STKE* 148, 38–42.

1244 Sachs, K., Perez, O., Pe’er, D., Lauffenburger, D., Nolan, G., 2005. Causal protein-signaling net-
1245 works derived from multiparameter single-cell data. *Science* 308, 523.

1246 Sato, H., Suzuki, H., Toyota, M., Nojima, M., Maruyama, R., Sasaki, S., Takagi, H., Sogabe, Y.,
1247 Sasaki, Y., Idogawa, M., Sonoda, T., Mori, M., Imai, K., Tokino, T., Shinomura, Y., 2007.
1248 Frequent epigenetic inactivation of dickkopf family genes in human gastrointestinal tumors. *Carcinogenesis* 28, 2459–2466.

1250 Schmidt-Ott, K., Masckauchan, T., Chen, X., Hirsh, B., Sarkar, A., Yang, J., Paragas, N., Wallace,
1251 V., Dufort, D., Pavlidis, P., et al., 2007. β -catenin/tcf/lef controls a differentiation-associated
1252 transcriptional program in renal epithelial progenitors. *Development* 134, 3177–3190.

1253 Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., Amin, N., Schwikowski,
1254 B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular
1255 interaction networks. *Genome research* 13, 2498–2504.

1256 Sharma, R., 1973. Wingless a new mutant in drosophila melanogaster. *Drosophila information*
1257 *service* 50, 134–134.

1258 Sheather, S.J., Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel
1259 density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 683–690.

1260 Sinha, S., 2014. Integration of prior biological knowledge and epigenetic information enhances
1261 the prediction accuracy of the bayesian wnt pathway. *Integr. Biol.* 6, 1034–1048. doi:10.1039/
1262 c4ib000124a.

1263 Sobol', I.M., 1990. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe*
1264 *Modelirovanie* 2, 112–118.

1265 Sokol, S., 2011. *Wnt Signaling in Embryonic Development*. volume 17. Elsevier.

1266 Strahl, B., Allis, C., 2000. The language of covalent histone modifications. *Nature* 403, 41–45.

1267 Sun, J.Z., Wang, G.I., Goyal, V.K., Varshney, L.R., 2012. A framework for bayesian optimality of
1268 psychophysical laws. *Journal of Mathematical Psychology* 56, 495–501.

1269 Suzuki, H., Watkins, D., Jair, K., Schuebel, K., Markowitz, S., Chen, W., Pretlow, T., Yang,
1270 B., Akiyama, Y., Van Engeland, M., et al., 2004. Epigenetic inactivation of sfrp genes allows
1271 constitutive wnt signaling in colorectal cancer. *Nature genetics* 36, 417–422.

1272 Taniguchi, H., Yamamoto, H., Hirata, T., Miyamoto, N., Oki, M., Noshio, K., Adachi, Y., Endo,
1273 T., Imai, K., Shinomura, Y., 2005. Frequent epigenetic inactivation of wnt inhibitory factor-1 in
1274 human gastrointestinal cancers. *Oncogene* 24, 7946–7952.

1275 Thorstensen, L., Lind, G.E., Løvig, T., Diep, C.B., Meling, G.I., Rognum, T.O., Lothe, R.A., 2005.
1276 Genetic and epigenetic changes of components affecting the wnt pathway in colorectal carcinomas
1277 stratified by microsatellite instability. *Neoplasia* 7, 99–108.

1278 Veeck, J., Dahl, E., 2012. Targeting the wnt pathway in cancer: the emerging role of dickkopf-3.
1279 *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* , 18–28.

1280 Voronkov, A., Krauss, S., 2012. Wnt/beta-catenin signaling and small molecule inhibitors. *Current*
1281 *pharmaceutical design* 19, 634.

1282 Waterman, M., 2004. Lymphoid enhancer factor/t cell factor expression in colorectal cancer. *Cancer*
1283 *and Metastasis Reviews* 23, 41–52.

1284 Weber, E.H., 1834. De pulsu resorptione, auditu et tactu. *Annotationes anatomicae et physiologicae*.

1285 Yochum, G., 2011. Multiple wnt/ β -catenin responsive enhancers align with the myc promoter
1286 through long-range chromatin loops. *PloS one* 6, e18966.

1287 Yu, J., Virshup, D.M., 2014. Updating the wnt pathways. *Bioscience reports* 34, 593–607.

1288 Yuan, G., Wang, C., Ma, C., Chen, N., Tian, Q., Zhang, T., Fu, W., 2012. Oncogenic function of
1289 dact1 in colon cancer through the regulation of β -catenin. *PloS one* 7, e34004.

1290 Zhong, Z., Ethen, N.J., Williams, B.O., 2014. Wnt signaling in bone development and homeostasis.
1291 *Wiley Interdisciplinary Reviews: Developmental Biology* 3, 489–500.

1292 Zitt, M., Untergasser, G., Amberger, A., Moser, P., Stadlmann, S., Zitt, M., Müller, H., Mühlmann,
1293 G., Perathoner, A., Margreiter, R., et al., 2008. Dickkopf-3 as a new potential marker for
1294 neoangiogenesis in colorectal cancer: expression in cancer tissue and adjacent non-cancerous
1295 tissue. *Disease Markers* 24, 101–109.