ARTICLE TYPE

Hilbert-Schmidt and Sobol sensitivity indices for static and time series Wnt signaling measurements in colorectal cancer† [work in progress]

Shriprakash Sinha*[a]

Ever since the accidental discovery of Wingless [Sharma R.P., Drosophila information service, 1973, 50, p 134], research in the field of Wnt signaling pathway has taken significant strides in wet lab experiments and various cancer clinical trials, augmented by recent developments in advanced computational modeling of the pathway. Information rich gene expression profiles reveal various aspects of the signaling pathway and help in studying different issues simultaneously. Hitherto, not many computational studies exist which incorporate the simultaneous study of these issues. This manuscript • explores the strength of contributing factors in the signaling pathway, • analyzes the existing causal relations among the inter/extracellular factors effecting the pathway based on prior biological knowledge and • investigates the deviations in fold changes in the recently found prevalence of psychophysical laws working in the pathway. To achieve this goal, local and global sensitivity analysis is conducted on the (non)linear responses between the factors obtained from static and time series expression profiles using the density (Hilbert-Schmidt Information Criterion) and variance (Sobol) based sensitivity indices. The results show the advantage of using density based indices over variance based indices mainly due to the former's employment of distance measures & the kernel trick via Reproducing kernel Hilbert space (RKHS) that capture nonlinear relations among various intra/extracellular factors of the pathway in a higher dimensional space. In time series data, using these indices it is now possible to observe where in time, which factors get influenced & contribute to the pathway, as changes in concentration of the other factors are made. This synergy of prior biological knowledge, sensitivity analysis & representations in higher dimensional spaces can facilitate in time based administration of target therapeutic drugs & reveal hidden biological information within colorectal cancer samples. Code has been made available at Google drive on `https://drive.google.com/folderview?id=0B7Kkv8wlhPU-Q2NBZGt1ZERrSVE&usp=sharing`

1 Introduction

1.1 A short review

Sharma[1]'s accidental discovery of the Wingless played a pioneering role in the emergence of a widely expanding research field of the Wnt signaling pathway. A majority of the work has focused on issues related to • the discovery of genetic and epigenetic factors affecting the pathway (Thorstensen *et al.*[2] & Baron and Kneissel[3]), • implications of mutations in the pathway and its dominant role on cancer and other diseases (Clevers[4]), • investigation into the pathway's contribution towards embryo development (Sokol[5]), homeostasis (Pinto *et al.*[6], Zhong *et al.*[7]) and apoptosis (Pećina-Šlaus[8]) and • safety and feasibility of drug design for the Wnt pathway (Kahn[9], Garber[10], Voronkov and Krauss[11], Blagodatski *et al.*[12] & Curtin and Lorenzi[13]). Approximately forty years after the discovery, important strides have been made in the research work involving several wet lab experiments and cancer clinical trials (Kahn[9], Curtin and Lorenzi[13]) which

* *This research was conducted independently. Corresponding Author : shriprakash sinha.*

[a] *Address : 104-Madhurisha Heights Phase 1, Risali, Bhilai - 490006, INDIA; E-mail : sinha.shriprakash@yandex.com*

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

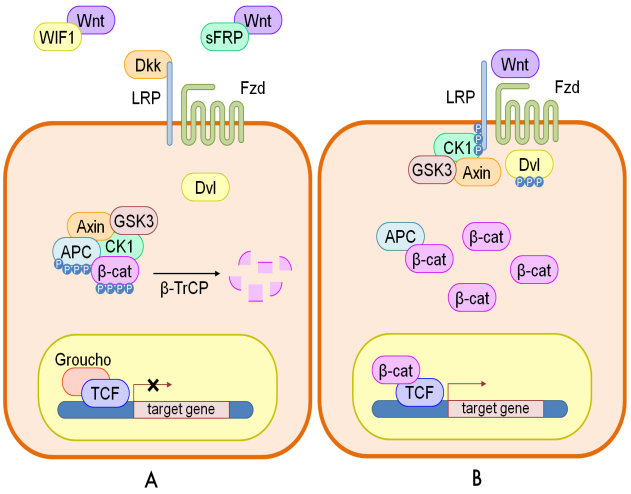

Fig. 1 A cartoon of Wnt signaling pathway contributed by Verhaegh *et al.*[18]. Part (A) represents the destruction of β-*catenin* leading to the inactivation of the Wnt target gene. Part (B) represents activation of Wnt target gene.

have been augmented by the recent developments in the various advanced computational modeling techniques of the pathway. 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[14]) 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[15]).

The work in this paper investigates some of the current aspects of research regarding the pathway via sensitivity analysis while using static (Jiang *et al.*[16]) and time series (Gujral and MacBeath[17]) gene expression data retrieved from colorectal cancer samples.

1.2 Canonical Wnt signaling pathway

Before delving into the problem statement, a brief introduction to the Wnt pathway is given here. From the recent work of Sinha[19], the canonical Wnt signaling pathway is a transduction mechanism that contributes to embryo development and controls homeostatic self renewal in several tissues (Clevers[4]). Somatic mutations in the pathway are known to be associated with cancer in different parts of the human body. Prominent among them is the colorectal cancer case (Gregorieff and Clevers[20]). In a succinct overview, the Wnt signaling pathway works when the Wnt ligand gets attached to the Frizzled(FZD)/LRP coreceptor complex. FZD may interact with the Dishevelled (DVL) causing phosphorylation. It is also thought that Wnts cause phosphorylation of the LRP via casein kinase 1 ($CK1$) and kinase $GSK3$. These developments further lead to attraction of Axin which causes inhibition of the formation of the degradation complex. The degradation complex constitutes of $AXIN$, the β-*catenin* transportation complex APC, $CK1$ and $GSK3$. When the pathway is active the dissolution of the degradation complex leads to stabilization in the concentration of β-*catenin* in the cytoplasm. As β-*catenin* enters into the nucleus it displaces the $GROUCHO$ and binds with transcription cell factor TCF thus instigating transcription of Wnt target genes. $GROUCHO$ acts as lock on TCF and prevents the transcription of target genes which may induce cancer. In cases when the Wnt ligands are not captured by the coreceptor at the cell membrane, $AXIN$ helps in formation of the degradation complex. The degradation complex phosphorylates β-*catenin* which is then recognized by $FBOX/WD$ repeat protein β-$TRCP$. β-$TRCP$ is a component of ubiquitin ligase complex that helps in ubiquitination of β-*catenin* thus marking it for degradation via the proteasome. Cartoons depicting the phenomena of Wnt being inactive and active are shown in figures 1(A) and 1(B), respectively.

2 Problem statement & sensitivity analysis

Succinctly, the endeavour is to address the following issues - • explore the strength of contributing factors in the signaling pathway, • analyse the existing causal relations among the inter/extracellular factors effecting the pathway based on prior biological knowledge and • investigate the recently found prevalence of psychophysical laws working in the pathway in a multi-parameter setting. The issues related to • inference of hidden biological relations among the factors, that are yet to be discovered and • discovery of new causal relations using hypothesis testing, will be addressed in a subsequent manuscript.

In order to address the above issues, sensitivity analysis (SA) is performed on either the datasets or results obtained from biologically inspired causal models. The reason for using these tools of sensitivity analysis is that they help in observing the behaviour of the output and the importance of the contributing input factors via a robust and an easy mathematical framework. In this manuscript both local and global SA methods are used. Where appropriate, a description of the biologically inspired causal models ensues before the analysis of results from these models. The approach taken here is that first a problem will be addressed and then the analysis of results and discussion ensues before working with the next issue.

2.1 Sensitivity analysis

Seminal work by Russian mathematician Sobol'[21] lead to development as well as employment of SA methods to study various complex systems where it was tough to measure the contribution of various input parameters in the behaviour of the output. A re-

cent unpublished review on the global SA methods by Iooss and Lemaître[22] categorically delineates these methods with the following functionality • screening for sorting influential measures (Morris[23] method, Group screening in Moon *et al.*[24] & Dean and Lewis[25], Iterated factorial design in Andres and Hajas[26], Sequential bifurcation in Bettonvil and Kleijnen[27] and Cotter[28] design), • quantitative indicies for measuring the importance of contributing input factors in linear models (Christensen[29], Saltelli *et al.*[30], Helton and Davis[31] and McKay *et al.*[32]) and nonlinear models (Homma and Saltelli[33], Sobol[34], Saltelli[35], Saltelli *et al.*[36], Saltelli *et al.*[37], Cukier *et al.*[38], Saltelli *et al.*[39], & Tarantola *et al.*[40] Saltelli *et al.*[41], Janon *et al.*[42], Owen[43], Tissot and Prieur[44], Da Veiga and Gamboa[45], Archer *et al.*[46], Tarantola *et al.*[47], Saltelli *et al.*[41] and Jansen[48]) and • exploring the model behaviour over a range on input values (Storlie and Helton[49] and Da Veiga *et al.*[50], Li *et al.*[51] and Hajikolaei and Wang[52]). Iooss and Lemaître[22] also provide various criteria in a flowchart for adapting a method or a combination of the methods for sensitivity analysis.

Besides the above Sobol'[21]'s variance based indicies, more recent developments regarding new indicies based on density, derivative and goal-oriented can be found in Borgonovo[53], Sobol and Kucherenko[54] and Fort *et al.*[55], respectively. In a more recent development, Da Veiga[56] propose new class of indicies based on density ratio estimation (Borgonovo[53]) that are special cases of dependence measures. This in turn helps in exploiting measures like distance correlation (Székely *et al.*[57]) and Hilbert-Schmidt independence criterion (Gretton *et al.*[58]) as new sensitivity indicies. The basic framework of these indicies is based on use of Csiszár *et al.*[59] f-divergence, concept of dissimilarity measure and kernel trick Aizerman *et al.*[60]. Finally, Da Veiga[56] propose feature selection as an alternative to screening methods in sensitivity analysis. The main issue with variance based indicies (Sobol'[21]) is that even though they capture importance information regarding the contribution of the input factors, they • do not handle multivariate random variables easily and • are only invariant under linear transformations. In comparison to these variance methods, the newly proposed indicies based on density estimations (Borgonovo[53]) and dependence measures are more robust.

2.2 Relevance in systems biology

Recent efforts in systems biology to understand the importance of various factors apropos output behaviour has gained prominence. Sumner *et al.*[61] compares the use of Sobol'[21] variance based indices versus Morris[23] screening method which uses a One-at-a-time (OAT) approach to analyse the sensitivity of *GSK*3 dynamics to uncertainty in an insulin signaling model. Similar efforts, but on different pathways can be found in Zheng and Rundell[62] and Marino *et al.*[63].

SA provides a way of analyzing various factors taking part in a biological phenomena and deals with the effects of these factors on the output of the biological system under consideration. Usually, the model equations are differential in nature with a set of inputs and the associated set of parameters that guide the output. SA helps in observing how the variance in these parameters and inputs leads to changes in the output behaviour. The goal of this manuscript is not to analyse differential equations and the parameters associated with it. Rather, the aim is to observe which input genotypic factors have greater contribution to observed phenotypic behaviour like a sample being normal or cancerous in both static and time series data. In this process, the effect of fold changes in time is also considered for analysis in the light of the recently observed psychophysical laws acting downstream of the Wnt pathway (Goentoro and Kirschner[64]).

2.3 Sensitivity indices

Given the range of estimators available for testing the sensitivity, it might be useful to list a few which are going to be employed in this research study. Also, a brief introduction into the fundamentals of the derivation of the three main indicies has been provided.

2.3.1 Variance based indices

The variance based indices as proposed by Sobol'[21] prove a theorem that an integrable function can be decomposed into summands of different dimensions. Also, a Monte Carlo algorithm is used to estimate the sensitivity of a function apropos arbitrary group of variables. It is assumed that a model denoted by function $u = f(x)$, $x = (x_1, x_2, ..., x_n)$, is defined in a unit n-dimensional cube $\mathscr{K}^n$ with u as the scalar output. The requirement of the problem is to find the sensitivity of function $f(x)$ with respect to different variables. If $u^* = f(x^*)$ is the required solution, then the sensitivity of u^* apropos x_k is estimated via the partial derivative $(\partial u/\partial x_k)_{x=x^*}$. This approach is the local sensitivity. In global sensitivity, the input $x = x^*$ is not specified. This implies that the model $f(x)$ lies inside the cube and the sensitivity indices are regarded as tools for studying the model instead of the solution. Detailed technical aspects with examples can be found in Homma and Saltelli[33] and Sobol[65].

Let a group of indices $i_1, i_2, ..., i_s$ exist, where $1 \leq i_1 < ... < i_s \leq n$ and $1 \leq s \leq n$. Then the notation for sum over all different groups of indices is -

$$\widehat{\Sigma} T_{i_1,i_2,...,i_s} = \Sigma_{i=1}^{n} T_i + \Sigma_{s=1}^{n} \Sigma_{1 \leq i < j \leq n} T_{i,j} + ... + T_{1,2,...,n} \quad (1)$$

Then the representation of $f(x)$ using equation 1 in the form -

$$\begin{aligned} f(x) &= f_0 + \widehat{\Sigma} f_{i_1,i_2,...,i_s} \quad (2) \\ &= f_0 + \Sigma_i f_i(x_i) + \Sigma_{i<j} f_{i,j}(x_i, x_j) + ... + f_{1,2,...,n}(x_1, x_2, ..., x_n) \end{aligned}$$

is called ANOVA-decomposition from Archer *et al.*[46] or expansion into summands of different dimensions, if f_0 is a constant and integrals of the summands $f_{i_1,i_2,...,i_s}$ with respect to their own variables are zero, i.e,

$$f_0 = \int_{\mathscr{K}^n} f(x)dx \tag{3}$$

$$\int_0^1 f_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s})dx_{i_k} = 0, 1 \leq k \leq s \tag{4}$$

It follows from equation 3 that all summands on the right hand side are orthogonal, i.e if at least one of the indices in $i_1,i_2,...,i_s$ and $j_1,j_2,...,j_l$ is not repeated i.e

$$\int_0^1 f_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s})f_{j_1,j_2,...,j_l}(x_{j_1},x_{j_2},...,x_{j_s})dx = 0 \tag{5}$$

Sobol'[21] proves a theorem stating that there is an existence of a unique expansion of equation 3 for any $f(x)$ integrable in $\mathscr{K}^n$. In brief, this implies that for each of the indices as well as a group of indices, integrating equation 3 yields the following -

$$\int_0^1 ... \int_0^1 f(x)dx/dx_i = f_0 + f_i(x_i) \tag{6}$$

$$\int_0^1 ... \int_0^1 f(x)dx/dx_i dx_j = f_0 + f_i(x_i) + f_j(x_j) + f_{i,j}(x_i,x_j) \tag{7}$$

were, dx/dx_i is $\prod_{\forall k \in \{1,..,n\}; i \notin k} dx_k$ and $dx/dx_i dx_j$ is $\prod_{\forall k \in \{1,..,n\}; i,j \notin k} dx_k$. For higher orders of grouped indices, similar computations follow. The computation of any summand $f_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s})$ is reduced to an integral in the cube $\mathscr{K}^n$. The last summand $f_{1,2,...,n}(x_1,x_2,...,x_n)$ is $f(x) - f_0$ from equation 3. Homma and Saltelli[33] stresses that use of Sobol sensitivity indices does not require evaluation of any $f_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s})$ nor the knowledge of the form of $f(x)$ which might well be represented by a computational model i.e a function whose value is only obtained as the output of a computer program.

Finally, assuming that $f(x)$ is square integrable, i.e $f(x) \in \mathscr{L}_2$, then all of $f_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s}) \in \mathscr{L}_2$. Then the following constants

$$\int_{\mathscr{K}^n} f^2(x)dx - f_0^2 = D \tag{8}$$

$$\int_0^1 ... \int_0^1 f^2_{i_1,i_2,...,i_s}(x_{i_1},x_{i_2},...,x_{i_s})dx_{i_1}dx_{i_2}...dx_{i_s} = D_{i_1,i_2,...,i_s} \tag{9}$$

are termed as variances. Squaring equation 3, integrating over $\mathscr{K}^n$ and using the orthogonality property in equation 5, D evaluates to -

$$D = \widehat{\Sigma} D_{i_1,i_2,...,i_s} \tag{10}$$

Then the global sensitivity estimates is defined as -

$$S_{i_1,i_2,...,i_s} = \frac{D_{i_1,i_2,...,i_s}}{D} \tag{11}$$

It follows from equations 10 and 11 that

$$\widehat{\Sigma} S_{i_1,i_2,...,i_s} = 1 \tag{12}$$

Clearly, all sensitivity indices are non-negative, i.e an index $S_{i_1,i_2,...,i_s} = 0$ if and only if $f_{i_1,i_2,...,i_s} \equiv 0$. The true potential of Sobol indices is observed when variables $x_1,x_2,...,x_n$ are divided into m different groups with $y_1,y_2,...,y_m$ such that $m < n$. Then $f(x) \equiv f(y_1,y_2,...,y_m)$. All properties remain the same for the computation of sensitivity indices with the fact that integration with respect to y_k means integration with respect to all the x_i's in y_k. Details of these computations with examples can be found in Sobol[65]. Variations and improvements over Sobol indices have already been stated in section 2.1.

2.3.2 Density based indices

As discussed before, the issue with variance based methods is the high computational cost incurred due to the number of interactions among the variables. This further requires the use of screening methods to filter out redundant or unwanted factors that might not have significant impact on the output. Recent work by Da Veiga[56] proposes a new class of sensitivity indicies which are a special case of density based indicies Borgonovo[53]. These indicies can handle multivariate variables easily and relies on density ratio estimation. Key points from Da Veiga[56] are mentioned below.

Considering the similar notation in previous section, $f : \mathscr{R}^n \rightarrow \mathscr{R}$ $(u = f(x))$ is assumed to be continuous. It is also assumed that X_k has a known distribution and are independent. Baucells and Borgonovo[66] state that a function which measures the similarity between the distribution of U and that of $U|X_k$ can define the impact of X_k on U. Thus the impact is defined as -

$$S_{X_k} = \mathscr{E}(d(U,U|X_k)) \tag{13}$$

were $d(\cdot,\cdot)$ is a dissimilarity measure between two random variables. Here d can take various forms as long as it satisfies the criteria of a dissimilarity measure. Csiszár *et al.*[59]'s f-divergence between U and $U|X_k$ when all input random variables are considered to be absolutely continuous with respect to Lebesgue measure on $\mathscr{R}$ is formulated as -

$$d_F(U||U|X_k) = \int_{\mathscr{R}} F(\frac{p_U(u)}{p_{U|X_k}(u)})p_{U|X_k}(u)du \tag{14}$$

were F is a convex function such that $F(1) = 0$ and p_U and $p_{U|X_k}$ are the probability distribution functions of U and $U|X_k$. Standard choices of F include Kullback-Leibler divergence $F(t) = -\log_e(t)$,

Hellinger distance $(\sqrt{t}-1)^2$, Total variation distance $F(t)=|t-1|$, Pearson χ^2 divergence $F(t)=t^2-1$ and Neyman χ^2 divergence $F(t)=(1-t^2)/t$. Substituting equation 14 in equation 13, gives the following sensitivity index -

$$\begin{aligned} S^F_{X_k} &= \int_{\mathscr{R}} d_F(U||U|X_k)p_{X_k}(x)dx \\ &= \int_{\mathscr{R}}\int_{\mathscr{R}} F(\frac{p_U(u)}{p_{U|X_k}(u)})p_{U|X_k}(u)p_{X_k}(x)dxdu \\ &= \int_{\mathscr{R}^2} F(\frac{p_U(u)p_{X_k}(x)}{p_{U|X_k}(u)p_{X_k}(x)})p_{U|X_k}(u)p_{X_k}(x)dxdu \\ &= \int_{\mathscr{R}^2} F(\frac{p_U(u)p_{X_k}(x)}{p_{X_k,U}(x,u)})p_{X_k,U}(x,u)dxdu \end{aligned} \tag{15}$$

were p_{X_k} and $p_{X_k,Y}$ are the probability distribution functions of X_k and (X_k,U), respectively. Csiszár *et al.*[59] f-divergences imply that these indices are positive and equate to 0 when U and X_k are independent. Also, given the formulation of $S^F_{X_k}$, it is invariant under any smooth and uniquely invertible transformation of the variables X_k and U (Kraskov *et al.*[67]). This has an advantage over Sobol sensitivity indices which are invariant under linear transformations.

By substituting the different formulations of F in equation 15, Da Veiga[56]'s work claims to be the first in establishing the link that previously proposed sensitivity indices are actually special cases of more general indices defined through Csiszár *et al.*[59]'s f-divergence. Then equation 15 changes to estimation of ratio between the joint density of (X_k,U) and the marginals, i.e -

$$S^F_{X_k} = \int_{\mathscr{R}^2} F(\frac{1}{r(x,u)})p_{X_k,U}(x,u)dxdu = \mathscr{E}_{(X_k,U)}F(\frac{1}{r(X_k,U)}) \tag{16}$$

were, $r(x,y) = (p_{X_k,U}(x,u))/(p_U(u)p_{X_k}(x))$. Multivariate extensions of the same are also possible under the same formulation.

Finally, given two random vectors $X \in \mathscr{R}^p$ and $Y \in \mathscr{R}^q$, the dependence measure quantifies the dependence between X and Y with the property that the measure equates to 0 if and only if X and Y are independent. These measures carry deep links (Sejdinovic *et al.*[68]) with distances between embeddings of distributions to reproducing kernel Hilbert spaces (RHKS) and here the related Hilbert-Schmidt independence criterion (HSIC by Gretton *et al.*[58]) is explained.

In a very brief manner from an extremely simple introduction by Daumé III[69] - "We first defined a field, which is a space that supports the usual operations of addition, subtraction, multiplication and division. We imposed an ordering on the field and described what it means for a field to be complete. We then defined vector spaces over fields, which are spaces that interact in a friendly way with their associated fields. We defined complete vector spaces and extended them to Banach spaces by adding a norm. Banach spaces were then extended to Hilbert spaces with the addition of a dot product." Mathematically, a Hilbert space $\mathscr{H}$ with elements $r,s \in \mathscr{H}$ has dot product $\langle r,s\rangle_{\mathscr{H}}$ and $r\cdot s$. When $\mathscr{H}$ is a vector space over a field $\mathscr{F}$, then the dot product is an element in $\mathscr{F}$. The product $\langle r,s\rangle_{\mathscr{H}}$ follows the below mentioned properties when $r,s,t \in \mathscr{H}$ and for all $a \in \mathscr{F}$ -

- Associative : $(ar)\cdot s = a(r\cdot s)$
- Commutative : $r\cdot s = s\cdot r$
- Distributive : $r\cdot(s+t) = r\cdot s + r\cdot t$

Given a complete vector space $\mathscr{V}$ with a dot product $\langle\cdot,\cdot\rangle$, the norm on $\mathscr{V}$ defined by $||r||_{\mathscr{V}} = \sqrt{(\langle r,r\rangle)}$ makes this space into a Banach space and therefore into a full Hilbert space.

A reproducing kernel Hilbert space (RKHS) builds on a Hilbert space $\mathscr{H}$ and requires all Dirac evaluation functionals in $\mathscr{H}$ are bounded and continuous (on implies the other). Assuming $\mathscr{H}$ is the $\mathscr{L}_2$ space of functions from X to $\mathscr{R}$ for some measurable X. For an element $x \in X$, a Dirac evaluation functional at x is a functional $\delta_x \in \mathscr{H}$ such that $\delta_x(g) = g(x)$. For the case of real numbers, x is a vector and g a function which maps from this vector space to $\mathscr{R}$. Then δ_x is simply a function which maps g to the value g has at x. Thus, δ_x is a function from $(\mathscr{R}^n \mapsto \mathscr{R})$ into $\mathscr{R}$.

The requirement of Dirac evaluation functions basically means (via the Riesz[70] representation theorem) if ϕ is a bounded linear functional (conditions satisfied by the Dirac evaluation functionals) on a Hilbert space $\mathscr{H}$, then there is a unique vector ℓ in $\mathscr{H}$ such that $\phi g = \langle g,\ell\rangle_{\mathscr{H}}$ for all $\ell \in \mathscr{H}$. Translating this theorem back into Dirac evaluation functionals, for each δ_x there is a unique vector k_x in $\mathscr{H}$ such that $\delta_x g = g(x) = \langle g,k_x\rangle_{\mathscr{H}}$. The reproducing kernel K for $\mathscr{H}$ is then defined as : $K(x,x') = \langle k_x,k_{x'}\rangle$, were k_x and $k_{x'}$ are unique representatives of δ_x and $\delta_{x'}$. The main property of interest is $\langle g,K(x,x')\rangle_{\mathscr{H}} = g(x')$. Furthermore, k_x is defined to be a function $y \mapsto K(x,y)$ and thus the reproducibility is given by $\langle K(x,\cdot),K(y,\cdot)\rangle_{\mathscr{H}} = K(x,y)$.

Basically, the distance measures between two vectors represent the degree of closeness among them. This degree of closeness is computed on the basis of the discriminative patterns inherent in the vectors. Since these patterns are used implicitly in the distance metric, a question that arises is, how to use these distance metric for decoding purposes?

The kernel formulation as proposed by Aizerman *et al.*[60], is a solution to our problem mentioned above. For simplicity, we consider the labels of examples as binary in nature. Let $\mathbf{x}_i \in \mathscr{R}^n$, be the set of n feature values with corresponding category of the example label (y_i) in data set $\mathscr{D}$. Then the data points can be mapped to a higher dimensional space $\mathscr{H}$ by the transformation ϕ:

$$\phi : \mathbf{x}_i \in \mathscr{R}^n \mapsto \phi(\mathbf{x}_i) \in \mathscr{H} \tag{17}$$

This $\mathscr{H}$ is the *Hilbert Space* which is a strict inner product space, along with the property of completeness as well as separability. The inner product formulation of a space helps in discriminating the location of a data point w.r.t a separating hyperplane in $\mathscr{H}$. This is achieved by the evaluation of the inner product between the normal vector representing the hyperplane along with the vectorial representation of a data point in $\mathscr{H}$. Thus, the idea behind equation(17) is that even if the data points are nonlinearly clustered in space $\mathscr{R}^n$, the transformation spreads the data points into $\mathscr{H}$, such that they can be linearly separated in its range in $\mathscr{H}$.

Often, the evaluation of dot product in higher dimensional spaces is computationally expensive. To avoid incurring this cost, the concept of kernels in employed. The trick is to formulate kernel functions that depend on a pair of data points in the space $\mathscr{R}^n$, under the assumption that its evaluation is equivalent to a dot product in the higher dimensional space. This is given as:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = < \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) > \tag{18}$$

Two advantages become immediately apparent. First, the evaluation of such kernel functions in lower dimensional space is computationally less expensive than evaluating the dot product in higher dimensional space. Secondly, it relieves the burden of searching an appropriate transformation that may map the data points in $\mathscr{R}^n$ to $\mathscr{H}$. Instead, all computations regarding discrimination of location of data points in higher dimensional space involves evaluation of the kernel functions in lower dimension. The matrix containing these kernel evaluations is referred to as the *kernel* matrix. With a cell in the kernel matrix containing a kernel evaluation between a pair of data points, the kernel matrix is square in nature.

As an example in practical applications, once the kernel has been computed, a pattern analysis algorithm uses the kernel function to evaluate and predict the nature of the new example using the general formula:

$$\begin{aligned} f(\mathbf{z}) &= < \mathbf{w}, \phi(\mathbf{z}) > + b \\ &= < \sum_{i=1}^{N} \alpha_i \times y_i \times \phi(\mathbf{x}_i), \phi(\mathbf{z}) > + b \\ &= \sum_{i=1}^{N} \alpha_i \times y_i \times < \phi(\mathbf{x}_i), \phi(\mathbf{z}) > + b \\ &= \sum_{i=1}^{N} \alpha_i \times y_i \times \kappa(\mathbf{x}_i, \mathbf{z}) + b \end{aligned} \tag{19}$$

where $\mathbf{w}$ defines the hyperplane as some linear combination of training basis vectors, $\mathbf{z}$ is the test data point, y_i the class label for training point $\mathbf{x}_i$, α_i and b are the constants. Various transformations to the kernel function can be employed, based on the properties a kernel must satisfy. Interested readers are referred to Taylor and Cristianini[71] for description of these properties in detail.

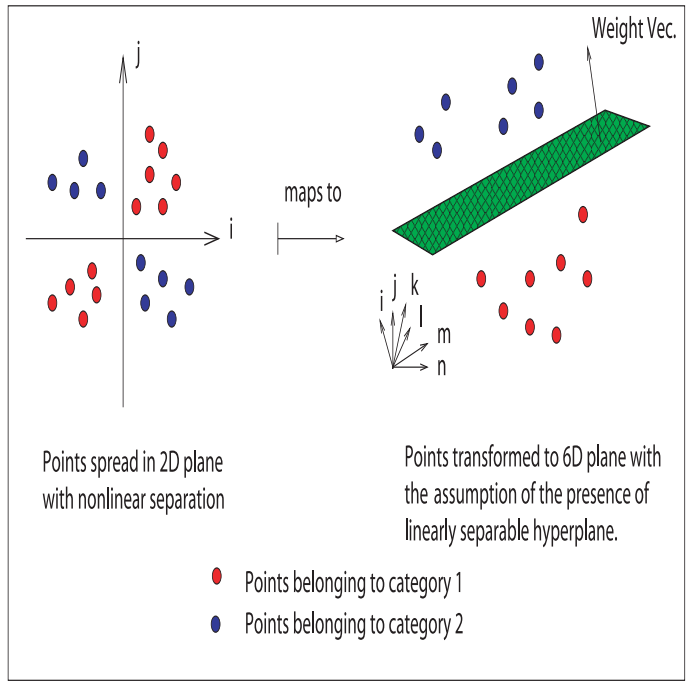

Fig. 2 A geometrical interpretation of mapping nonlinearly separable data into higher dimensional space where it is assumed to be linearly separable, subject to the holding of dot product.

The Hilbert-Schmidt independence criterion (HSIC) proposed by Gretton *et al.*[58] is based on kernel approach for finding dependences and on cross-covariance operators in RKHS. Let $X \in \mathscr{X}$ have a distribution P_X and consider a RKHS $\mathscr{A}$ of functions $\mathscr{X} \to \mathscr{R}$ with kernel $k_{\mathscr{X}}$ and dot product $\langle \cdot, \cdot \rangle_{\mathscr{A}}$. Similarly, Let $U \in \mathscr{Y}$ have a distribution P_Y and consider a RKHS $\mathscr{B}$ of functions $\mathscr{U} \to \mathscr{R}$ with kernel $k_{\mathscr{B}}$ and dot product $\langle \cdot, \cdot \rangle_{\mathscr{B}}$. Then the cross-covariance operator $C_{X,U}$ associated with the joint distribution P_{XU} of (X, U) is the linear operator $\mathscr{B} \to \mathscr{A}$ defined for every $a \in \mathscr{A}$ and $b \in \mathscr{B}$ as -

$$\langle a, C_{XU} b \rangle_{\mathscr{A}} = \mathscr{E}_{XU}[a(X), b(U)] - \mathscr{E}_X a(X) \mathscr{E}_U b(U) \tag{20}$$

The cross-covariance operator generalizes the covariance matrix by representing higher order correlations between X and U through nonlinear kernels. For every linear operator $C : \mathscr{B} \to \mathscr{A}$ and provided the sum converges, the Hilbert-Schmidt norm of C is given by -

$$||C||^2_{HS} = \Sigma_{k,l} \langle a_k, C b_l \rangle_{\mathscr{A}} \tag{21}$$

were a_k and b_l are orthonormal bases of $\mathscr{A}$ and $\mathscr{B}$, respectively. The HSIC criterion is then defined as the Hilbert-Schmidt norm of cross-covariance operator -

$$HSIC(X,U)_{\mathscr{A},\mathscr{B}} = \begin{cases} ||C_{XU}||^2_{HS} = \\ \mathscr{E}_{X,X',U,U'} k_{\mathscr{X}}(X,X') k_{\mathscr{U}}(U,U') + \\ \mathscr{E}_{X,X'} k_{\mathscr{X}}(X,X') \mathscr{E}_{U,U'} k_{\mathscr{U}}(U,U') - \\ 2\mathscr{E}_{X,Y}[\mathscr{E}_{X'} k_{\mathscr{X}}(X,X') \mathscr{E}_{U'} k_{\mathscr{U}}(U,U')] \end{cases} \tag{22}$$

were the equality in terms of kernels is proved in Gretton *et al.* [58]. Finally, assuming (X_i, U_i) $(i = 1, 2, ..., n)$ is a sample of the random vector (X, U) and denote $K_{\mathscr{X}}$ and $K_{\mathscr{U}}$ the Gram matrices with entries $K_{\mathscr{X}}(i,j) = k_{\mathscr{X}}(X_i, X_j)$ and $K_{\mathscr{U}}(i,j) = k_{\mathscr{U}}(U_i, U_j)$. Gretton *et al.* [58] proposes the following estimator for $HSIC_n(X,U)_{\mathscr{A},\mathscr{B}}$ -

$$HSIC_n(X,U)_{\mathscr{A},\mathscr{B}} = \frac{1}{n^2} Tr(K_{\mathscr{X}} H K_{\mathscr{U}} H) \tag{23}$$

were H is the centering matrix such that $H(i,j) = \delta_{i,j} - \frac{1}{n}$. Then $HSIC_n(X,U)_{\mathscr{A},\mathscr{B}}$ can be expressed as -

$$HSIC(X,U)_{\mathscr{A},\mathscr{B}} = \begin{cases} \frac{1}{n^2} \Sigma^n_{i,j=1} k_{\mathscr{X}}(X_i, X_j) k_{\mathscr{U}}(U_i, U_j) \\ + \frac{1}{n^2} \Sigma^n_{i,j=1} k_{\mathscr{X}}(X_i, X_j) \frac{1}{n^2} \Sigma^n_{i,j=1} k_{\mathscr{U}}(U_i, U_j) \\ - \frac{2}{n} \Sigma^n_{i=1} [\frac{1}{n} \Sigma^n_{j=1} k_{\mathscr{X}}(X_i, X_j) \frac{1}{n} \Sigma^n_{j=1} k_{\mathscr{U}}(U_i, U_j)] \end{cases} \tag{24}$$

Finally, Da Veiga [56] proposes the sensitivity index based on distance correlation as -

$$S^{HSIC_{\mathscr{A},\mathscr{B}}}_{X_k} = R(X_k, U)_{\mathscr{A},\mathscr{B}} \tag{25}$$

were the kernel based distance correlation is given by -

$$R^2(X,U)_{\mathscr{A},\mathscr{B}} = \frac{HSIC(X,U)_{\mathscr{A},\mathscr{B}}}{\sqrt{(HSIC(X,X)_{\mathscr{A},\mathscr{A}} HSIC(U,U)_{\mathscr{B},\mathscr{B}})}} \tag{26}$$

were kernels inducing $\mathscr{A}$ and $\mathscr{B}$ are to be chosen within a universal class of kernels. Similar multivariate formulation for equation 23 are possible.

2.3.3 Choice of sensitivity indices

The SENSITIVITY PACKAGE (Faivre *et al.* [72] and Iooss and Lemaître [22]) in R langauge provides a range of functions to compute the indices and the following indices will be taken into account for addressing the posed questions in this manuscript.

1. sensiFdiv - conducts a density-based sensitivity analysis where the impact of an input variable is defined in terms of dissimilarity between the original output density function and the output density function when the input variable is fixed. The dissimilarity between density functions is measured with Csiszar f-divergences. Estimation is performed through kernel density estimation and the function kde of the package ks. (Borgonovo [53], Da Veiga [56])

2. sensiHSIC - conducts a sensitivity analysis where the impact of an input variable is defined in terms of the distance between the input/output joint probability distribution and the product of their marginals when they are embedded in a Reproducing Kernel Hilbert Space (RKHS). This distance corresponds to HSIC proposed by Gretton *et al.* [58] and serves as a dependence measure between random variables.

3. soboljansen - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) $\times$ n model evaluations. These are called the Jansen estimators. (Jansen [48] and Saltelli *et al.* [41])

4. sobol2002 - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) $\times$n model evaluations. These are called the Saltelli estimators. This estimator suffers from a conditioning problem when estimating the variances behind the indices computations. This can seriously affect the Sobol indices estimates in case of largely non-centered output. To avoid this effect, you have to center the model output before applying "sobol2002". Functions "soboljansen" and "sobolmartinez" do not suffer from this problem. (Saltelli [35])

5. sobol2007 - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) $\times$ n model evaluations. These are called the Mauntz estimators. (Saltelli *et al.* [41])

6. sobolmartinez - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices using correlation coefficients-based formulas, at a total cost of (p + 2) $\times$ n model evaluations. These are called the Martinez estimators.

7. sobol - implements the Monte Carlo estimation of the Sobol sensitivity indices. Allows the estimation of the indices of the variance decomposition up to a given order, at a total cost of (N + 1) $\times$ n where N is the number of indices to estimate. (Sobol' [21])

3 Description of the dataset & design of experiments

STATIC DATA - A simple static dataset containing expression values measured for a few genes known to have important role in human colorectal cancer cases has been taken from Jiang *et al.* [16]. Most of the expression values recorded are for genes that play a role in Wnt signaling pathway at an extracellular level and are known to

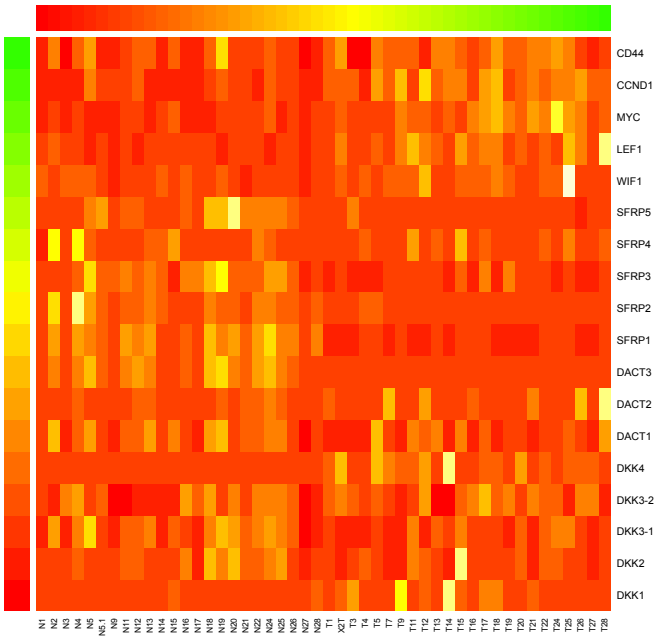

Fig. 3 Heat map for gene expression values for each of the 24 normal mucosa and 24 human colorectal tumor cases from Jiang *et al.*[16]

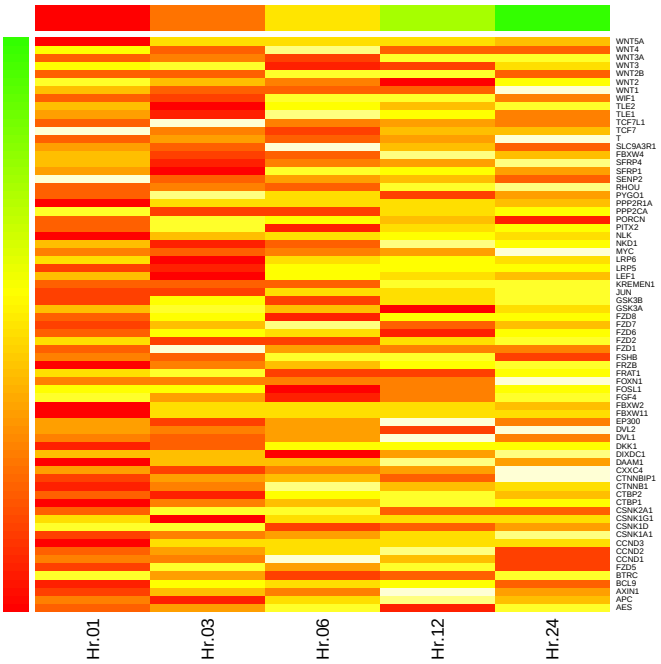

Fig. 4 Heat map for gene expression values for 6 time points from Gujral and MacBeath[17]

have inhibitory affect on the Wnt pathway due to epigenetic factors. For each of the 24 normal mucosa and 24 human colorectal tumor cases, gene expression values were recorded for 14 genes belonging to the family of *SFRP*, *DKK*, *WIF*1 and *DACT*. Also, expression values of established Wnt pathway target genes like *LEF*1, *MYC*, *CD*44 and *CCND*1 were recorded per sample.

TIME SERIES DATA - Contrary to the static data described above, Gujral and MacBeath[17] presents a bigger set of 71 Wnt-related gene expression values for 6 different times points over a range of 24-hour period using qPCR. The changes represent the fold-change in the expression levels of genes in 200 ng/mL *WNT*3*A*-stimulated HEK 293 cells in time relative to their levels in unstimulated, serum-starved cells at 0-hour. The data are the means of three biological replicates. Only genes whose mean transcript levels changed by more than two-fold at one or more time points were considered significant. Positive (negative) numbers represent up (down) -regulation.

Note that green (red) represents activation (repression) in the heat maps of data in Jiang *et al.*[16] and Gujral and MacBeath[17]. Figures 3 and 4 represent the heat maps for the static and time series data respectively.

DESIGN OF EXPERIMENTS - The reported results will be based on scaled as well as unscaled datasets. For the static data, only the scaled results are reported. This is mainly due to the fact that the measurements vary in a wide range and due to this there is often an error in the computed estimated of these indices. The data for time series does not vary in a wide range and thus the results are reported for both the scaled and the non scaled versions. Total sensitivity indices and 1st order indices will be used for sensitivity analysis. For addressing a biological question with known prior knowledge, the order of indices might be increased. While studying the interaction among the various genetic factors using static data, tumor samples are considered separated from normal samples. Bootstrapping without replicates on a smaller sample number is employed to generate estimates of indices which are then averaged. This takes into account the variance in the data and generates confidence bands for the indices. For the case of time series data, interactions among the contributing factors are studied by comparing (1) pairs of fold-changes at single time points and (2) pairs of deviations in fold changes between pairs of time points. Generation of distribution around measurements at single time points with added noise is done to estimate the indices.

4 Static data

To measure the strength of the contributing factors in the static dataset by Jiang *et al.*[16], 1st order and total sensitivity indices were generated. For each of the expression values of the genes recorded in the normal and tumor cases, the computation of the indices was done using bootstrapped samples in three different experiments each with a sample size of 8, 16 and 24, respectively.

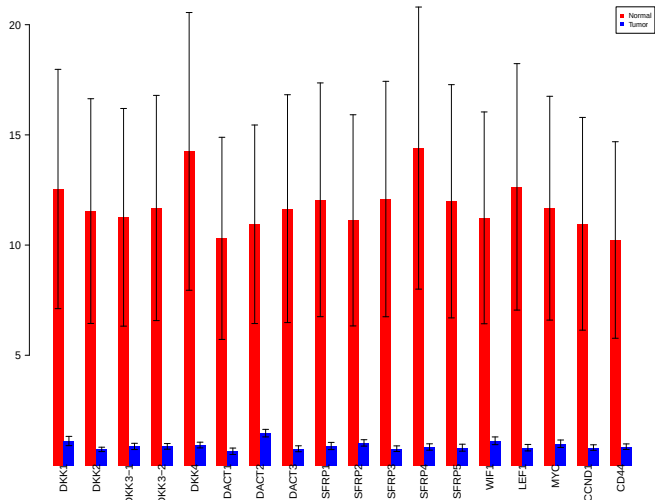

Fig. 5 sensiFdiv indices using Total Variation distance. Red - indices for normal. Blue - indices for tumor.

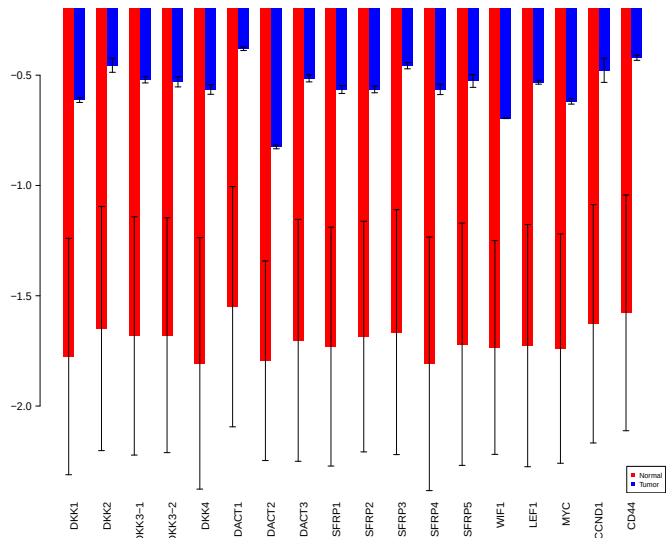

Fig. 6 sensiFdiv indices for Kullback-Leibler divergence. Red - indices for normal. Blue - indices for tumor.

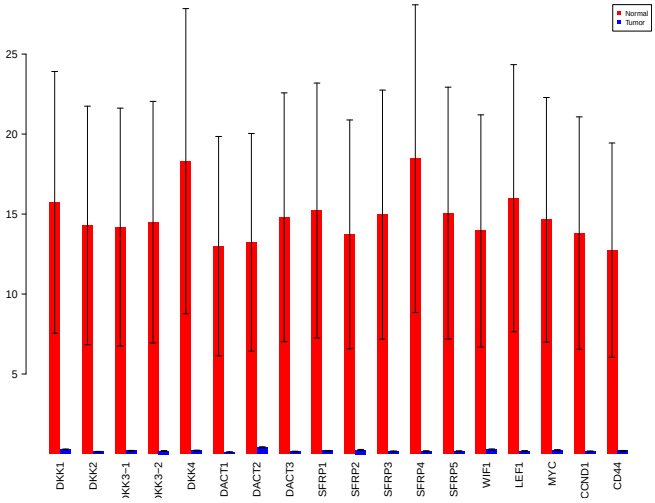

Fig. 7 sensiFdiv indices for Hellinger distance. Red - indices for normal. Blue - indices for tumor.

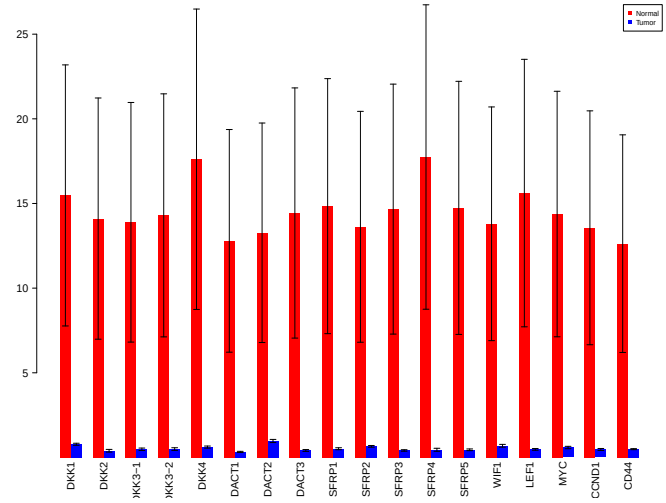

Fig. 8 sensiFdiv indices for Pearson χ^2 distance. Red - indices for normal. Blue - indices for tumor.

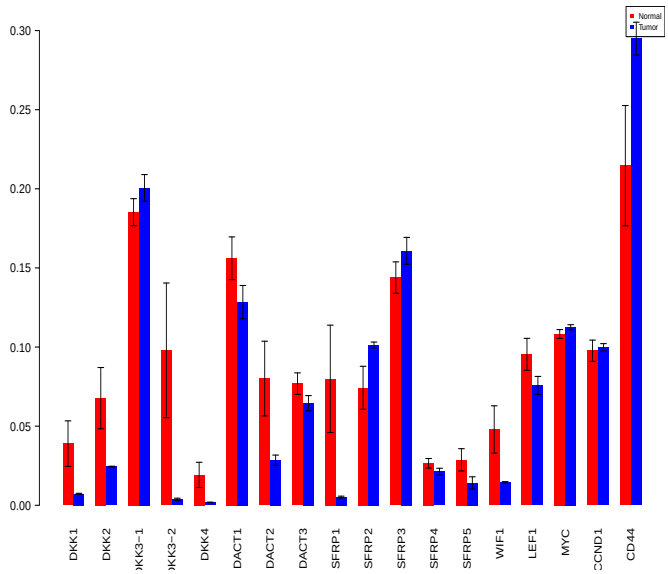

Fig. 9 sensiHSIC indices for linear kernel. Red - indices for normal. Blue - indices for tumor.

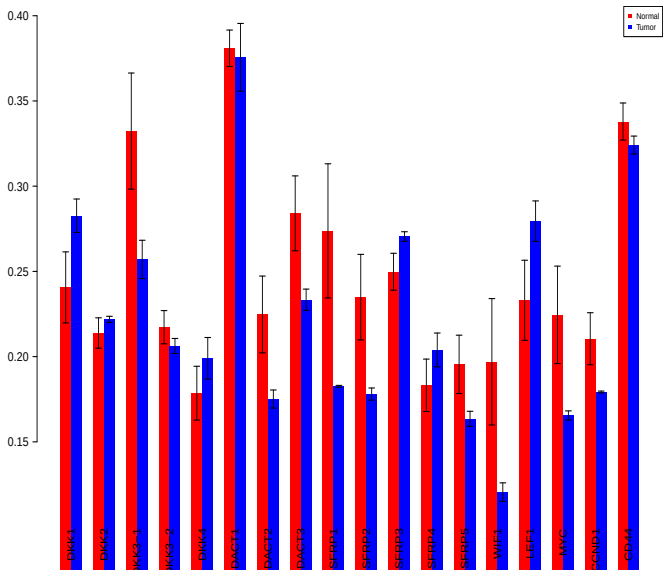

Fig. 10 sensiHSIC indices for laplace kernel. Red - indices for normal. Blue - indices for tumor.

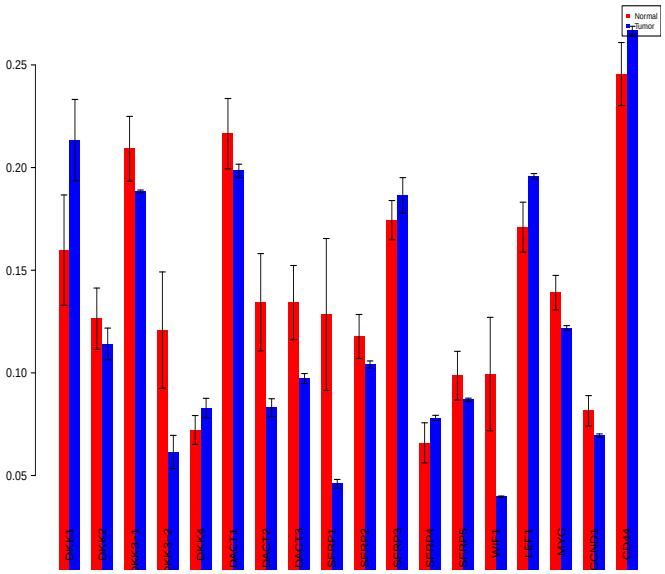

Fig. 11 sensiHSIC indices for rbf kernel. Red - indices for normal. Blue - indices for tumor.

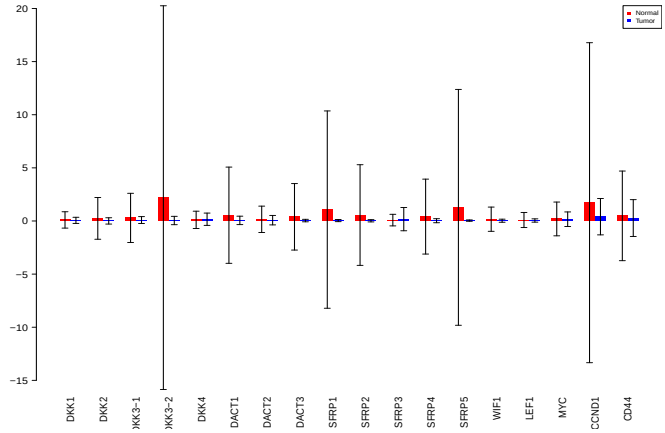

Fig. 12 Sobol 2002 first order indices. Red - indices for normal. Blue - indices for tumor.

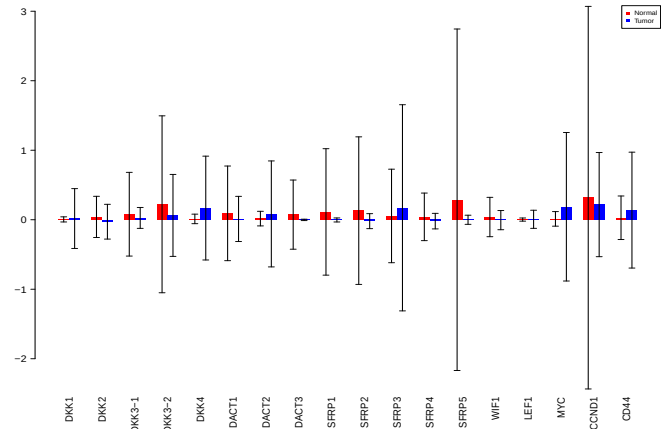

Fig. 13 Sobol 2007 first order indices. Red - indices for normal. Blue - indices for tumor.

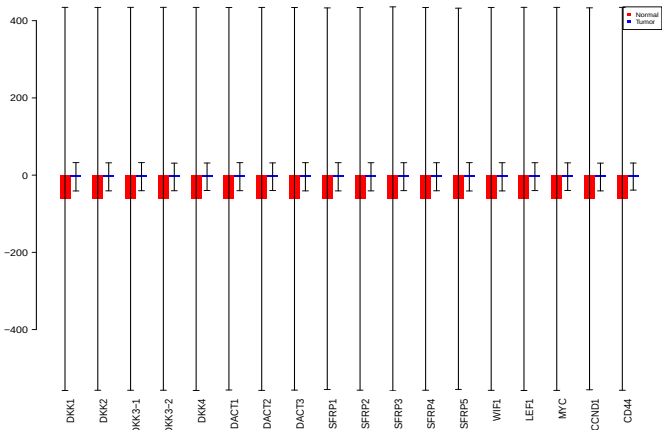

Fig. 14 Sobol jansen first order indices. Red - indices for normal. Blue - indices for tumor.

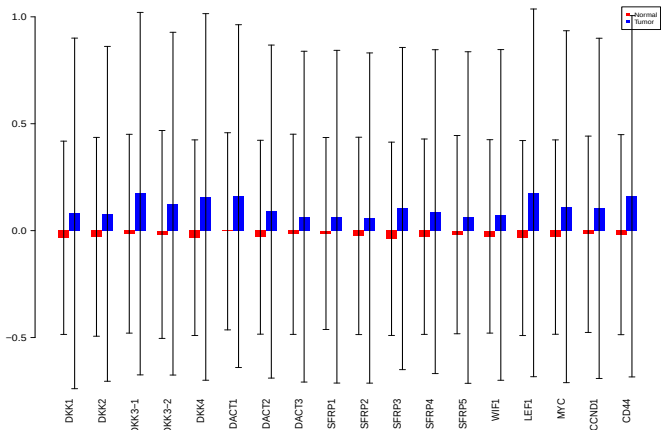

Fig. 15 Sobol martinez first order indices. Red - indices for normal. Blue - indices for tumor.

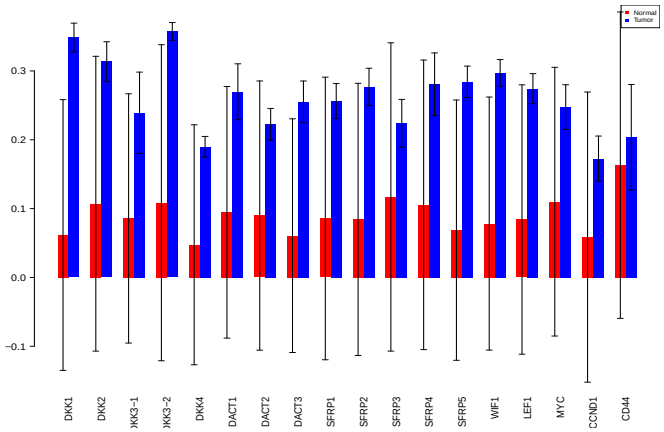

Fig. 16 Sobol first order indices. Red - indices for normal. Blue - indices for tumor.

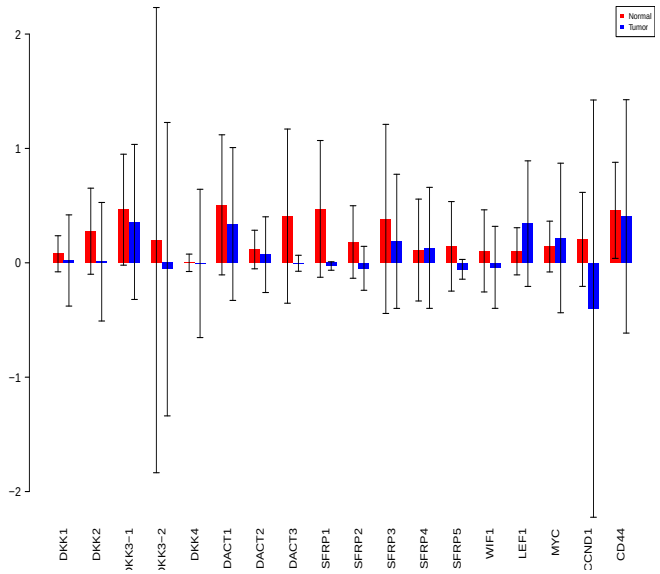

Fig. 17 Sobol 2002 total order indices. Red - indices for normal. Blue - indices for tumor.

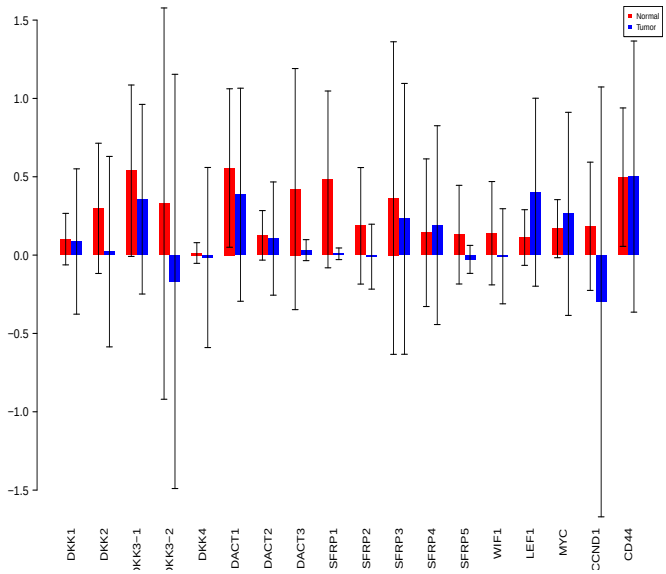

Fig. 18 Sobol 2007 total order indices. Red - indices for normal. Blue - indices for tumor.

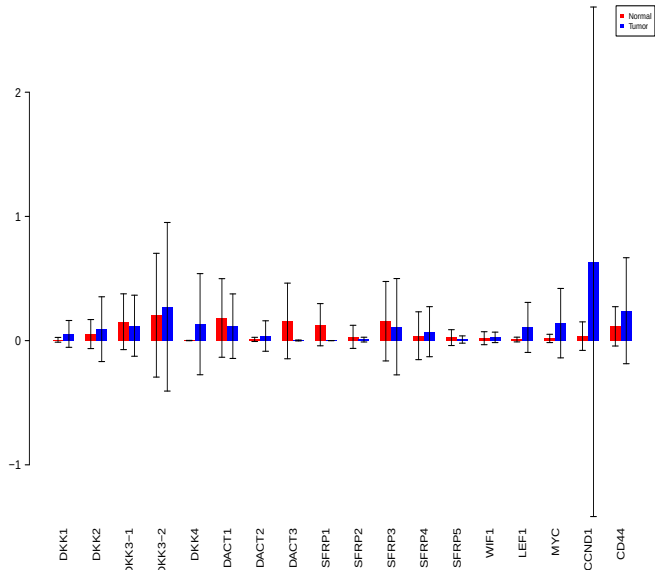

Fig. 19 Sobol jansen total order indices. Red - indices for normal. Blue - indices for tumor.

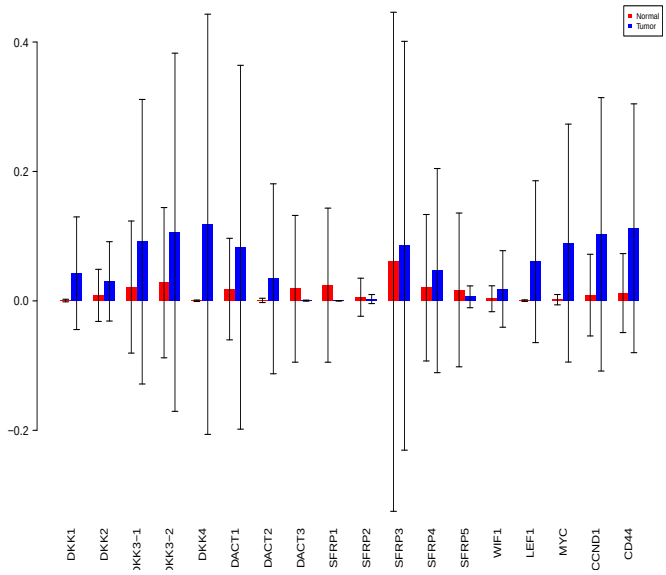

Fig. 20 Sobol martinez total order indices. Red - indices for normal. Blue - indices for tumor.

With only 24 samples in total, 20 bootstraps were generated for each set and the results were generated. From these replicates, the mean of the indices is reported along with the 0.95% confidence bands.

Using the sensiFdiv, all indices are computed as positive and those nearing to zero indicate the contribution of a factor as independent from the behaviour under consideration. Here, while comparing the indices of the gene expression values for normal and tumor cases, it was found that most of the involved intra/extracellular factors had some degree of contribution in the normal case and almost negligible contribution in the tumor case (see figures 5, 7 and 8). Apart from the negative reading for the KL divergence 6 the interpretations remain the same. This implies that the basic Csiszár *et al.*[59] f-divergence based indices might not capture the intrinsic genotypic effects for the normal and the tumorous cases. From the biological perspective, these graphs do not help in interpreting the strength of the contributions in normal and tumor cases. One might rank the indices for relative contributions, but this might not shed enough light on the how the factors are behaving in normal and tumor cases.

A more powerful way to analyse the contributions is the newly proposed HSIC based measures by Da Veiga[56]. These distances use the kernel trick which can capture intrinsic properties inherent in the recorded measurements by transforming the data into a higher dimensional space. Using these distances in sensiHSIC, it was found that the contributions of the various factors in the normal and the tumor cases vary drastically. This is shown in figures 9, 10 and 11. The laplace and the rbf kernels give more reliable sensitivity estimates for the involved factors than the linear kernel. Studying the results in figures 6 and 7 of Sinha[19] based on prior biological knowledge encoded in the Bayesian network models along with the indices of aforementioned figures, it can be found that indices of the family of $DACT-1/2/3$ show higher (lower) sensitivity in the normal (tumor) case where the activation (repression) happens. Again, of the $DACT$ family, $DACT-1$ has greater influence than $DACT-3$ (than $DACT-2$) based on the values of the sensitivity indices. These indices indicate the dependence of a factor on the output of the model characterized by the signaling being active (passive) in the normal (tumor) cases. 0(1) mean no (full) dependence of the output on the input factor. The laplace and the rbf kernels were found to give more consistent results than the linear kernel and the following description discusses the results from these kernels. For the $SFRP$ family $SFRP-1/2/5$ show higher (lower) sensitivity in normal (tumor) case where the activation (repression) happens (see figures 10 and 11). For $SFRP-3/4$ the influence is higher (lower) in the tumor (normal) case. In all the three types of kernels, $WIF1$, MYC and $CCND1$ show stronger (weaker) influence of repression (activation) in the normal (tumor) case (see figures 10 and 11). $CD44$ showed variable influence while observing the normal and tumor cases. Sinha[19] could not derive proper inferences for $LEF1$ but the sensitivity indices indicate that the influence of $LEF1$ in tumor samples to be higher than in normal samples. This points to the $LEF1$'s major role in tumor cases. Finally, for the family of DKK, $DKK1$ and $DKK3-2$ show similar behaviour of expression (repression) in normal (tumor cases) (see Sinha[19]). For the former, the prominence of the influence is shown in the higher (lower) sensitivity for tumor (normal) case. For the latter higher (lower) sensitivity was recorded for normal (tumor) case. This implies that the latter has more influential role in normal while the former has more influential role in tumor case. $DKK3-1$ was found to be expressed (repressed) in normal (tumor) and its dominant role is prominent from the higher bar sensitivity bar for normal than the tumor. Similar behavior of $DKK2$ was inferred by Sinha[19] but the sensitivity indices point to varied results and thus a conclusion cannot be drawn. Note that greater the value of the sensitivity index, greater is an input factor's contribution to the model.

The first order indices generated by sobol functions implemented in sobol2002 (figure 12), sobol2007 (figure 13), soboljansen (figure 14), sobolmartinez (figure 15) and sobol (figure 16) do not point to significant dependencies of the input factors. This can be attributed to the fact that there are less number of samples that help in the estimation of the sensitivity indicies. Finally, the total order indices need to be investigated in the context of the first order indices. It can be observed , sobol2002 (figure 17) and sobol2007 (figure 17) give much better estimates than soboljansen (figure 19) and sobolmartinez (figure 20). Most importantly, it is the former two that closely match with the sensitivity indices estimated using the HSIC distance measures. Interpretations from sobol2002 (figure 17) and sobol2007 (figure 18) are the same as those described above using the laplace and the rbf kernels from density based HSIC measure.

In summary, the sensitivity indices confirm the inferred results in Sinha[19] but do not help in inferring the causal relations using the static data. In combination with the results obtained from the Bayesian network models in Sinha[19] it is possible to study the effect of the input factors for the pathway in both normal and tumor cases. The results of sensitivity indices indicate how much these factors influence the pathway in normal and tumor cases. Again, not all indices reveal important information. So users must be cautious of results and see which measures reveal information that are close to already established or computationally estimated biological facts. Here the density based sensitivity indices captured information more precisely than the variance based indices (except for the total order indices from sobol2002/7 which gave similar results as sensiHSIC). This is attributed to the analytical strength provided by the distance measures using the kernel trick via RKHS that capture nonlinear relations in higher dimensional space, more precisely. Finally, in a recent unpublished work by

De Lozzo and Marrel[73], it has been validated that the HSIC indices prove to be more sensitive to the global behaviour than the Sobol indices.

5 Time series data

Next, the analysis of the time series data is addressed using the sensitivity indices. There are two experiments that have been performed. First is related to the analysis of a pair of the fold changes recorded at to different consecutive time points i.e t_i & t_{i+1}. The second is related to the analysis of a pair of deviations in fold changes recorded at t_i & t_{i+1} and t_{i+1} & t_{i+2}. The former compares the measurements in time while the latter takes into account the deviations that happens in time. For each measurement at a time point a normal distribution was generated with original recorded value as the mean, a standard deviation of 0.005 and an added noise in the form of jitter (see function jitter in R langauge). For the time measurements of each of the genes recorded in Gujral and MacBeath[17] an analysis of the sensitivity indices for both the scaled and the non-scaled data was done. Here the analysis for non-scaled data is presented. The reason for not presenting the scaled data is that the sample measurements did not vary drastically as found in the case of static data which caused troubles in the estimation of indices earlier. Another reason for not reporting the results on the scaled data is that the non-scaled ones present raw sensitive information which might be lost in scaling via normalization. Note that the authors use self organizing maps (SOM) to cluster data and use correlational analysis to derive their conclusions. In this work, the idea of clustering is abandoned and sensitivity indices are estimated for recorded factors participating in the pathway. Also the simple correlational analysis is dropped in favour of highly analytical kernel based distance measures.

Also, in a recent development, Goentoro and Kirschner[64] point to two findings namely, • the robust fold changes of β-*catenin* and • the transcriptional machinery of the Wnt pathway depends on the fold changes in β-*catenin* instead of absolute levels of the same and some gene transcription networks must respond to fold changes in signals according to the Weber's law in sensory physiology. The second study also carries a weight in the fact that due to the study of the deviations in the fold changes it is possible to check if the recently observed and reported natural psychophysical laws in the signaling pathway hold true or not. Finally, using the sensitivity indicies an effort is made to confirm the existing biological causal relations.

5.1 Analysis of fold changes at different time points

Lets begin with the gene $WNT3A$ as changes in its concentration lead to recording of the measurements of the different genes by Gujral and MacBeath[17]. Of the list of genes recorded, the indices of the those which are influenced by the concentration of $WNT3A$ are analysed. Next based on these confirmations and patterns of indices over time, conclusions for other enlisted genes are drawn. For the former list, the following genes $FZD1$, $FZD2$, $LEF1$, $TCF7$, $TCF7L1$, $LRP6$, $DVL1$, $SFRP1$, $SFRP4$, $CTBP1$, $CTBP2$, $PORCN$, $GSK3\beta$, MYC, APC and $CTNNB1$ are considered.

Figures 21 and 22 represent the indices computed over time. Columns represent the different kinds of indices computed while the rows show the respective genes. Each graph contains the sensitivity index computed at a particular time point (represented by a coloured bar). It should be observed from the aforementioned figures that the variants of the Sobol first order (FO) and the total order (TO) indices computed under different formulations were not very informative. This can be seen in graphs were some indices are negative and at some places the behaviour across time and genes remain the same. In contrast to this, the indices generated via the original Sobol function (under the column Sobol-SBL) as well as the sensiHSIC were found to be more reliable. Again, the rbf and laplace kernels under the HSIC formulations showed similar behaviour in comparison to the use of the linear kernel.

Gujral and MacBeath[17] simulate the serum starved HEK293 cells with 200 ng/mL of $WNT3A$ at different lengths of time. After the first hour (t_1), (under HSIC-rbf/laplace) it was observed that the sensitivity of $WNT3A$ was low (red bar). The maximum contribution of $WNT3A$ can be recoreded after the 12th stimulation. But due to increased stimulation by $WNT3A$ later on, there is an increased sensitivity of FZD-1/2 as well as $LRP6$. The FZD or the frizzled family of 7-transmember protein (Ueno *et al.*[74]) works in tandem with LRP-5/6 as binding parameters for the Wnt ligands to initiate the Wnt signaling. Consistent with the findings of Holcombe *et al.*[75] and Planutis *et al.*[76], $FZD1$ was found to be expressed. But there is a fair decrease in the contribution of the same in the next two time frames i.e after 3rd and the 6th hour. The maximum contribution of $FZD1$ is found after the $WNT3A$ simulation at 12th hour. This probably points to repetitive involvement of $FZD1$ after a certain period of time to initiate the working of the signaling pathway. $FZD2$ showed increasing significance in contribution after the first two time frames. The contribution drops significantly after the 3rd simulation and gradually increases in the next two time frames. The repetitive behaviour is similar to $FZD1$, yet it's role is not well studied as it appears to bind to both $WNT3A$ which promotes $Wnt/beta\text{-}catenin$ signaling and $WNT5A$ which inhibits it as shown by Sato *et al.*[77], respectively.

Klapholz-Brown *et al.*[78] and Yokoyama *et al.*[79] show that there is increased β-*catenin* due to $WNT3A$ stimulation which is depicted by the increased sensitivity of $CTNNB1$ expression in one of the above mentioned figures. MYC (i.e $c-MYC$) is known to be over expressed in colorectal cancer cases mainly due to the activation of $TCF-4$ transcription factor via intra nuclear bind-

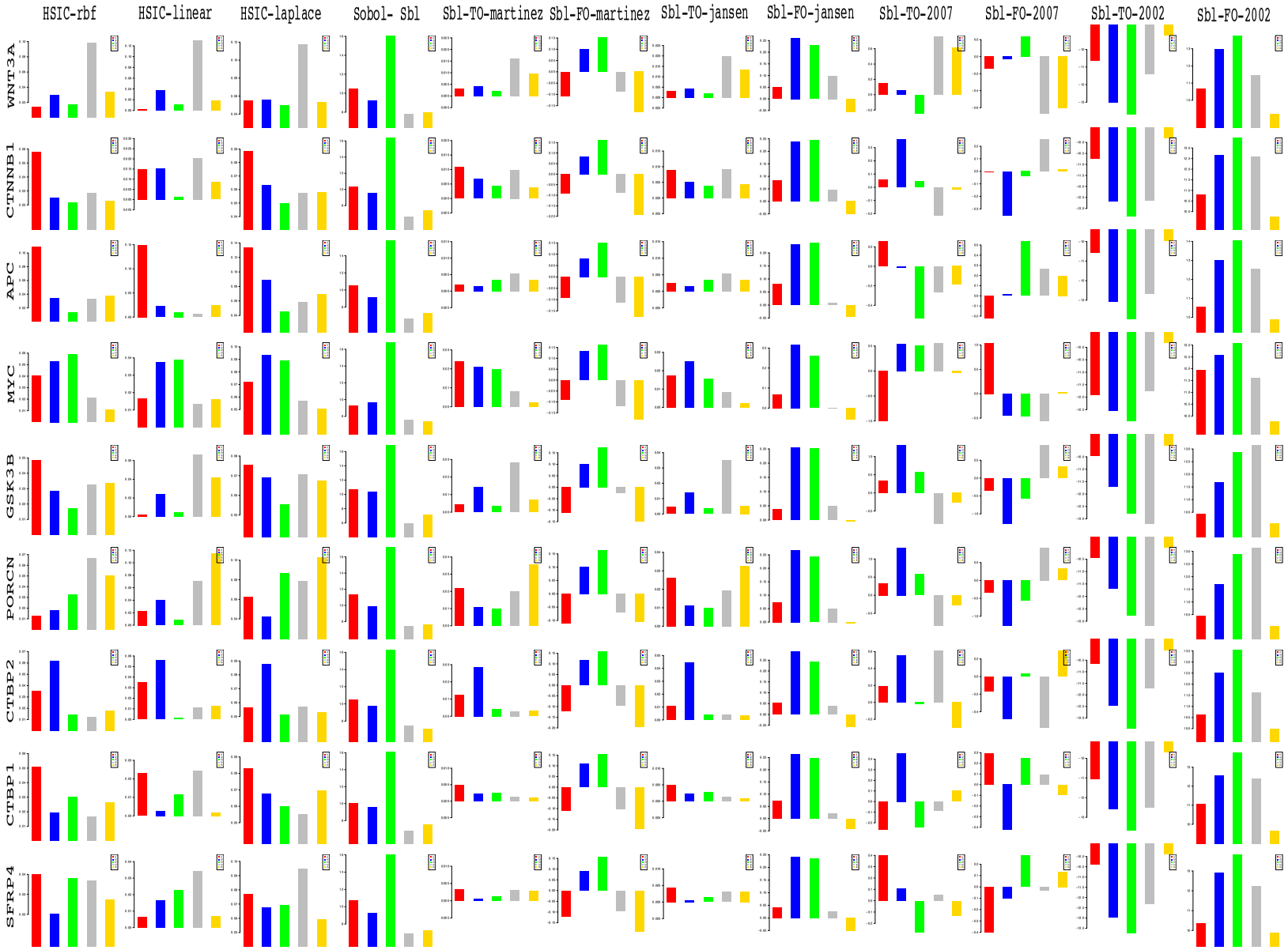

Fig. 21 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene. For each graph, the bars represent sensitivity indices computed at t1 (red), t2 (blue), t3 (green), t4 (gray) and t5 (yellow). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

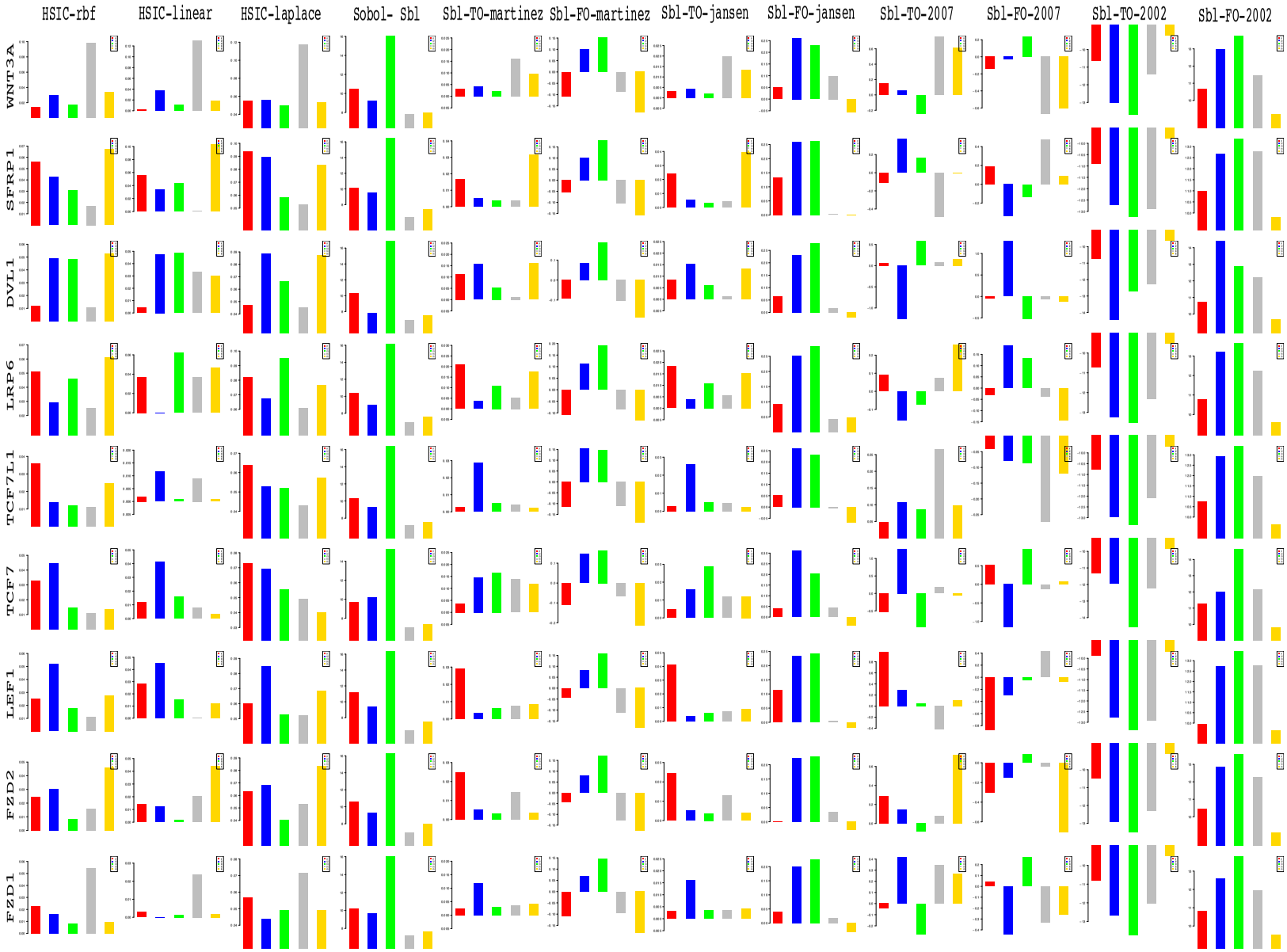

Fig. 22 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene. For each graph, the bars represent sensitivity indices computed at t1 (red), t2 (blue), t3 (green), t4 (gray) and t5 (yellow). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

ing of β-*catenin* (He *et al.*[80]), either by *APC* mutations (Korinek *et al.*[81]) or β-*catenin* mutations (Morin *et al.*[82]). The sensitivity of *MYC* increased monotonically but after the 6th hour it dropped significantly. Probably *MYC* does not play important role at later stages. As found in Hino *et al.*[83] and You *et al.*[84], *DVL* family interacts with the frizzled *FDZ* members leading to disassembly of the β-*catenin* destruction complex and subsequent translocation of β-*catenin* to the nucleus. Development on *DVL* family have been extensively recorded in González-Sancho *et al.*[85] and [86], and significance of *DVL*1 in Taiwanese colorectal cancer in Huang *et al.*[87]. *DVL*1 shows a marked increase in sensitivity as the concentration of the *WNT*3*A* increases in time. This is supported by the fact that ligand binding at the membrane leads to formation of complex including *DVL*1, *FZD* and *AXIN*.

Negative regulators like *SFRP*4 were found to have lower sensitivity as *WNT*3*A* concentration increases, but remained constant for most period. Meanwhile the significance of Wnt antagonist *SFRP*1 (Galli *et al.*[88], Suzuki *et al.*[89] and Caldwell *et al.*[90]) decreases over the period as the concentration of *WNT*3*A* increases. Chinnadurai[91] reviews the co-repressor ability of the *CTBP* family, while Hamada and Bienz[92] shows *CTBP* as a binding factor that interacts with *APC* thus lowering the availability of free nuclear β-*catenin*. This interaction is further confirmed in the recent research work by Schneikert *et al.*[93]. As shown by Yokoyama *et al.*[79] *CTBP*1 showed increased sensitivity with increased stimulation of *WNT*3*A* in the first hour. The latter stages show a decreased contribution of *CTBP*1 as the concentration of *WNT*3*A* was increased. This is in line with what Gujral and MacBeath[17] show in their manuscript and indicate the lowering of the co-repressor effect of *CTBP* at later stages. On the other hand, *CTBP*2 showed reverse behaviour of sensitivity in comparison to *CTBP*1 across different time points. Increased significance of *CTBP*2 was observed in the first two time frames, i.e after 1st and 3rd hour of stimulation, followed by lower contribution to the pathway at the latter stages. In both cases, the diminishing co-repressive nature of *CTBP* in time is observed. Contrary to these finding, recent results in Patel *et al.*[94] suggest that both *CTBP*1 and *CTBP*2 are up-regulated in colon cancer stem cells.

PORCN showed less sensitivity in the initial stages than in final stages indicating its importance in the contribution to Wnt secretion which is necessary for signaling (Willert and Nusse[95]). The sensitivity of *GSK*3β and *APC* decreased in time indicating the lowering of its significance in later stages due to no formation of the degradation complex. Activity of *TCF* gains greater prominence in the first and the second time frames after the initial *WNT*3*A* stimulation. This is in conjugation with the pattern showed by *CTBP*2. Regarding *TCF*7*L*2, the activity is observed to be maximum during the first time frame with decrease in the contribution in the later time frames.

Now, the analysis for the remaining 55 genes ensues. The estimated sensitivity indices for these genes are depicted in figures 23 and 24. Due to the above mentioned reasons regarding the issues related to the Sobol indicies the results presented in these two figures are from sensiHSIC and first order Sobol. *AES* follows the similar pattern of contribution as *WNT*3*A* contribution with high peaks at the end of the 1st and the 6th hour. But there is a reversal in the affect of *AES* after the 12th and 24th hour. This implies that in later stages *AES* is not a valuable contributor which is not so in the case of *WNT*3*A*. Similar behaviour can be found for *AXIN*1. In contrast to this, *BCL*9 shows a reverse behaviour in the contribution for the first three time frames. This indicates its maximum effect during the 3rd hour of simulation with *WNT*3*A*.

+++WORK IN PROGRESS+++
Now the indicies for other genes are also considered.

5.2 The logarithmic psychophysical law

In a recent development, Goentoro and Kirschner[64] point to two findings namely, • the robust fold changes of β-catenin and • the transcriptional machinery of the Wnt pathway depends on the fold changes in β-catenin instead of absolute levels of the same and some gene transcription networks must respond to fold changes in signals according to the Weber's law in sensory physiology. In an unpublished work by Sinha[96], preliminary analysis of results in Sinha[19] shows that the variation in predictive behaviour of β-catenin based transcription complex conditional on gene evidences follows power and logarithmic psychophysical law crudely, implying deviations in output are proportional to increasing function of deviations in input and showing constancy for higher values of input. This relates to the work of Adler *et al.*[97] on power and logarithmic law albeit at a coarse level. A description of these laws ensues before the analysis of the results.

Masin *et al.*[98] states the Weber's law as follows - Consider a sensation magnitude γ determined by a stimulus magnitude β. Fechner[99] (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[99] (vol 1, p. 65) attributed to the German physiologist Ernst Heinrich Weber the empirical finding Weber[100] that $\Delta\gamma$ remains constant when the relative stimulus increment $\frac{\Delta\beta}{\beta}$ remains constant, and named this finding Weber's law. Fechner[99] (vol 2, p. 10) underlined that Weber's law was empirical. ■

It has been found that Bernoulli's principle (Bernoulli[101]) is different from Weber's law (Weber[100]) 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.*[98] shows that Weber's law is a special case of Bernoulli's principle and can be derived as follows - Equation 27 depicts the Bernoulli's principle and increment in sensation represented by $\Delta\gamma$ is proportional to change in

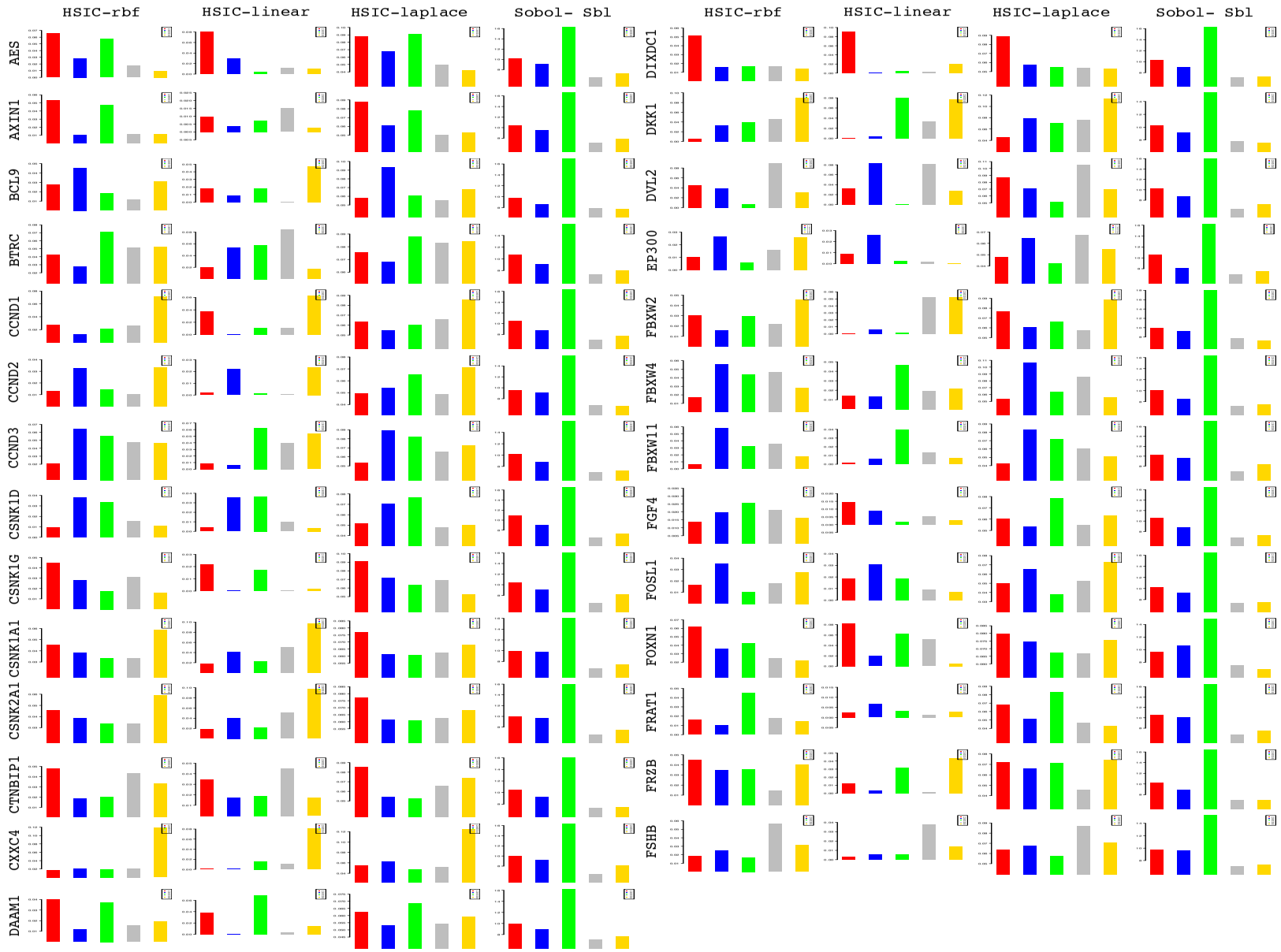

Fig. 23 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene. For each graph, the bars represent sensitivity indices computed at t1 (red), t2 (blue), t3 (green), t4 (gray) and t5 (yellow). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

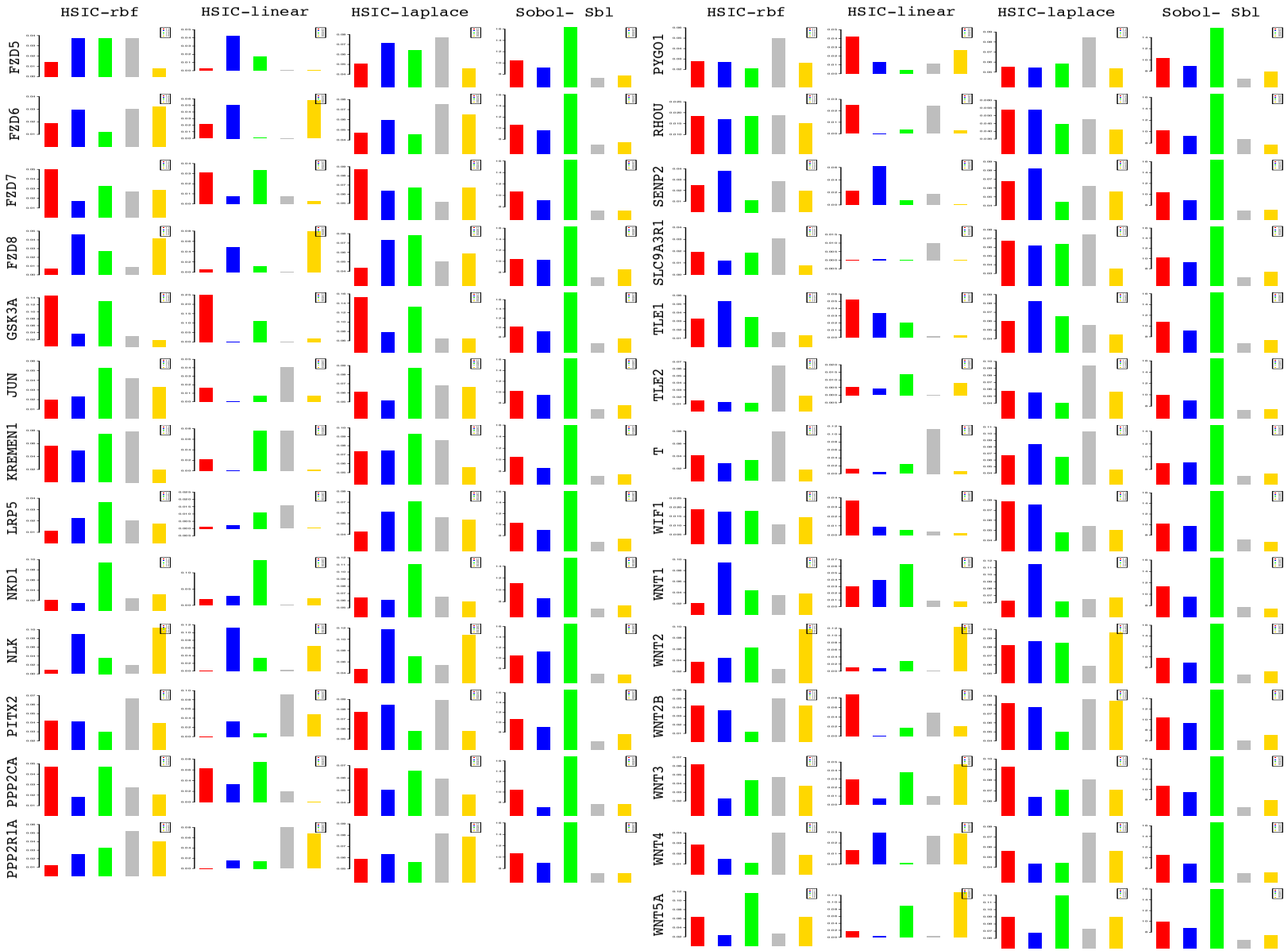

Fig. 24 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene. For each graph, the bars represent sensitivity indices computed at t1 (red), t2 (blue), t3 (green), t4 (gray) and t5 (yellow). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

stimulus represented by $\Delta\beta$.

$$\gamma = b \times \log \frac{\beta}{\alpha} \tag{27}$$

were b is a constant and α is a threshold. To evaluate the increment, the following equation 28 and the ensuing simplification gives -

$$\begin{aligned} \Delta\gamma &= b \times \log \frac{\beta + \Delta\beta}{\alpha} - b \times \log \frac{\beta}{\alpha} \\ &= b \times \log(\frac{\beta + \Delta\beta}{\beta}) \\ &= b \times \log(1 + \frac{\Delta\beta}{\beta}) \end{aligned} \tag{28}$$

Since b is a constant, equation 28 reduces to

$$\Delta\gamma \circ \frac{\Delta\beta}{\beta} \tag{29}$$

were $\circ$ means "is constant when there is constancy of" from Masin *et al.* [98]. The final equation 29 is a formulation of Weber's laws in wordings and thus Bernoulli's principles imply Weber's law as a special case. Using Fechner [99] derivation, it is possible to show the relation between Bernoulli's principles and Weber's law. Starting from the last line of equation 28, the following steps yield the relation -

$$\begin{aligned} e^{\Delta\gamma} &= e^{b \times \log(1+\frac{\Delta\beta}{\beta})} \\ k_p &= e^{\log(1+\frac{\Delta\beta}{\beta})^b}; \text{ were } k_p = e^{\Delta\gamma} \\ k_p &= (1 + \frac{\Delta\beta}{\beta})^b; \text{ since } e^{\log(x)} = x \\ \sqrt[b]{k_p} &= 1 + \frac{\Delta\beta}{\beta} \\ k_q - 1 &= \frac{\Delta\beta}{\beta}; \text{ were } \sqrt[b]{k_p} = k_q \\ k_r &= \frac{\Delta\beta}{\beta}; \text{ the weber's law s.t. } k_r = \sqrt[b]{e^{\Delta\gamma}} - 1 \end{aligned} \tag{30}$$

Equation 29 holds true given the last line of equation 30. By observation, it is important to note that the deviation Δ in the stimulus β plays a crucial role in the above depicted formulations. In the current study, instead of computing the sensitivity of the laws for each involved factor, the sensitivity of the deviations in the fold changes of each factor is taken into account. This is done in order to study the affect of deviations in fold changes in time as concentrations of $WNT3A$ changes at a constant rate. Without loss of generality, it was observed over time that most involved factors had sensitivity indices or strength of contributions, parts or whole of whose graphs follow a convex or a concave curvature. These are usually represented by either an exponentially increasing or decreasing curve or nonlinear curves. This points towards the fact that with increasing changes in stimulated concentration of $WNT3A$ the deviations in fold changes of an involved factor behave either in an increasing or decreasing fashion. Thus deviations in fold changes of various involved factors does affect the working of the signaling pathway over time. Finally, these deviations approximately capture the difference in fold changes recorded between two time frames and are thus a measure of how much the involvement of a factor affects the pathway due to these differences. This measure of involvement is depicted via the estimated sensitivity indices. The study of deviations in fold changes might help in deciding when a therapeutic drug could be administered in time. Details on the contributions of each involved factor are discussed below.

5.3 Analysis of deviations in fold changes

In comparison to the contributions estimated via the sensitivity indices using fold changes at different times separately, this section analyzes the contributions due to deviations in the fold changes recorded between two time points i.e t_i & t_{i+1}. These analysis are also a way to test the efficacy of deviations in fold changes versus the absolute levels that have been stressed upon in Goentoro and Kirschner [64]. As with the analysis of the fold changes at different time points, the estimates obtained using the rbf, linear and laplacian kernels in the HSIC based sensitivity analysis have been used here. Of these, the rbf and laplacian kernels give almost similar results.

■ WNT3A - Figure 25 shows the profile of mRNA expression levels of $WNT3A$ after external stimulation. There is a series of $(+-++)$ deviations in the fold change recordings at different time points. A repetitive behaviour is observed in the contribution of the deviations in fold changes for $WNT3A$ estimated via the sensitivity indices. For intervals in t_1, t_3 and t_6 there is an increase in the significance of the contribution of $WNT3A$ in figure 31 (see the first two bars for $<t_1,t_3>$ & $<t_3,t_6>$), even though in the first three time frames levels of $WNT3A$ are shown to be down-regulated (see figure 25). This behaviour is again repeated in figure 31 for intervals t_6, t_{12} and t_{24} (see the next two bars for $<t_6,t_{12}>$ & $<t_{12},t_{24}>$). In both cases one finds an increase in the contribution of the deviation in the fold change. Comparing the contribution of levels of fold changes in figure 21 were it was found that there is a dip in contribution of $WNT3A$ after t_3 and then a further increase in the contribution at a latter time frame, one finds that the deviations in fold changes involving

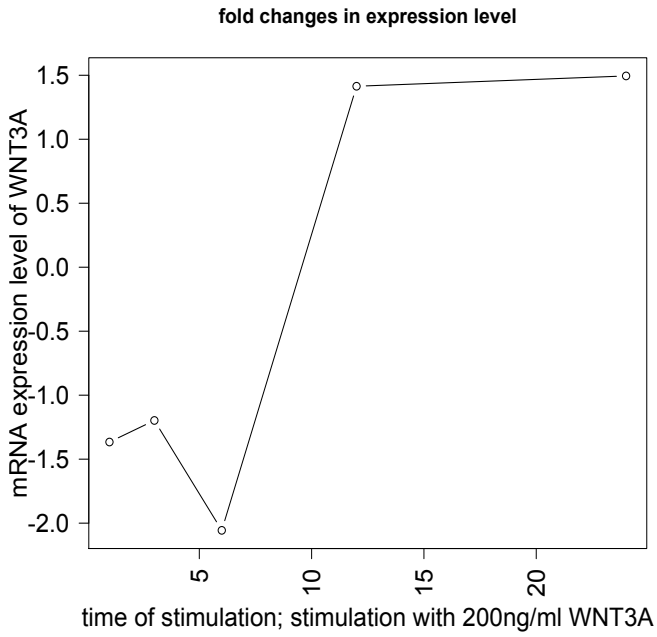

Fig. 25 mRNA expression levels of WNT3A at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

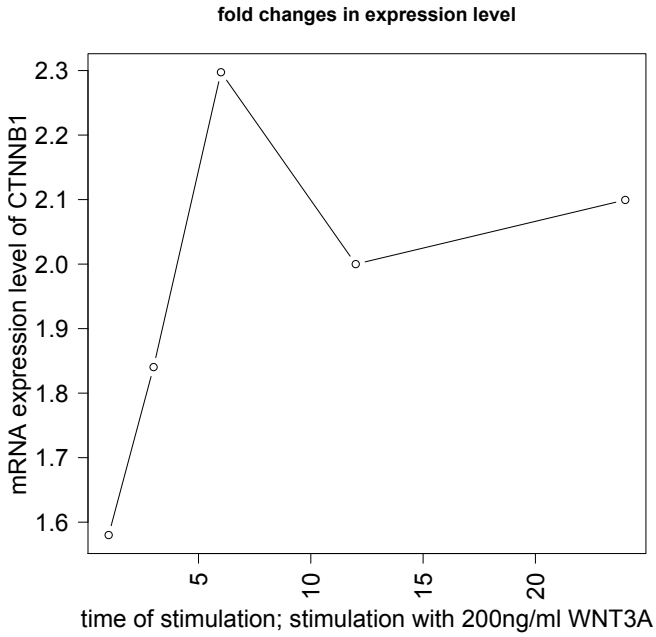

Fig. 26 mRNA expression levels of CTNNB1 at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

$<t_1,t_3>$ & $<t_3,t_6>$ have higher significance than the deviations in fold changes involving $<t_6,t_{12}>$ & $<t_{12},t_{24}>$. It can be noted that even in the deregulated state from $<t_1$, t_3 and $t_6>$ the deviations are minimal and the contributions are significantly high. In case of the regulated states from $<t_6$, t_{12} and $t_{24}>$ the deviation is extremely high between the first two time frames and low in the next two. This results in greater significant contribution in the latter deviation than the former deviation. Thus when deviations are low and the fold changes over time do not vary much, the contributions of the involved factor to the signaling pathway is expected to be high and vice versa. This points to the fact that low variations in fold changes over time have a stablizing influence of $WNT3A$ rather than abrupt high variations in fold changes that might not have the same influence. Thus measurments of deviations in fold changes provide greater support for studying the affect of $WNT3A$ over time.

■ CTNNB1 - Figure 26 shows the profile of mRNA expression levels of $CTNNB1$ after external stimulation. There is a series of $(++-+)$ deviations in the fold change recordings at different time points. An initial increase in the influence of $CTNNB1$ is observed from $<t_1,t_3>$ to $<t_3,t_6>$ (first two bars in figure 31) followed by a gradual decrease of influence from $<t_3,t_6>$ to $<t_6,t_{12}>$ to $<t_{12},t_{24}>$ (last three bars in figure 31). This is observed even though there is an up-regulation in levels of $CTNNB1$ with a slight dip at t_{12} (see figure 26). In comparison to the contribution of levels of fold changes in figure 21 were it was found that there is a gradual decrease in the influence of $CTNNB1$ till t_6 and then a further increase in the contribution at latter time frames, one finds that the deviations in fold changes involving $<t_3,t_6>$ have the highest significance with an almost exponential decrease in the deviations in fold changes involving $<t_6,t_{12}>$ & $<t_{12},t_{24}>$. Even though in a regulated state the influence of deviations in the fold changes indicate a different scenario altogether in comparison to influences of fold changes at distinct time frames. This might point to the fact that the affect of $CTNNB1$ is maximum during $<t_3,t_6>$ in comparison to other stages even after constantly increasing external stimulation with $WNT3A$ at different time points. Exponential decrease in the influence in the deviations in latter time frames points to the ineffectiveness of $CTNNB1$ in the pathway at later stages. Finally, in contrast to the behaviour of influence of $WNT3A$ in the foregoing paragraph, $CTNNB1$ showed higher (lower) influence for greater (lesser) deviations in fold changes.

■ APC - Figure 27 shows the profile of mRNA expression levels of APC after external stimulation. The profile of the deviations of APC in a deregulated state show the following $(-++-)$ pattern. While the $CTNNB1$ expression profile shows non-monotonic increase in levels of fold changes in upregulated state, APC expression profile shows a nonlinear behaviour in levels of fold changes in down regulated state. The significance of deviation in fold

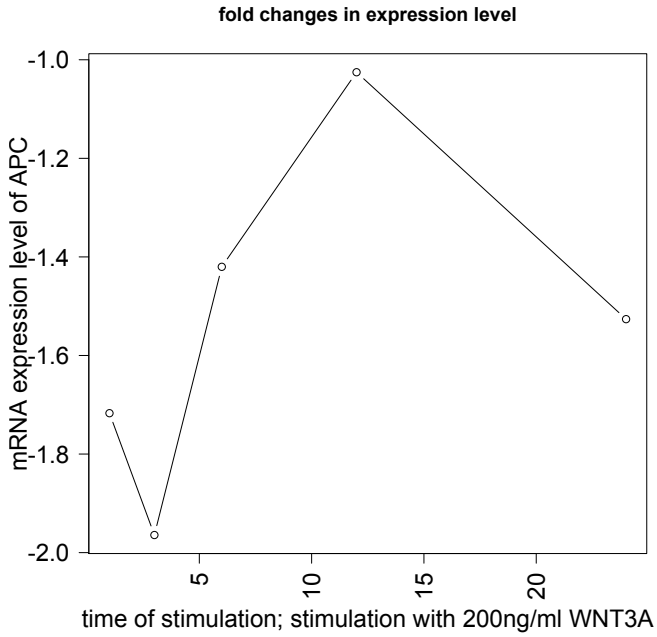

Fig. 27 mRNA expression levels of APC at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

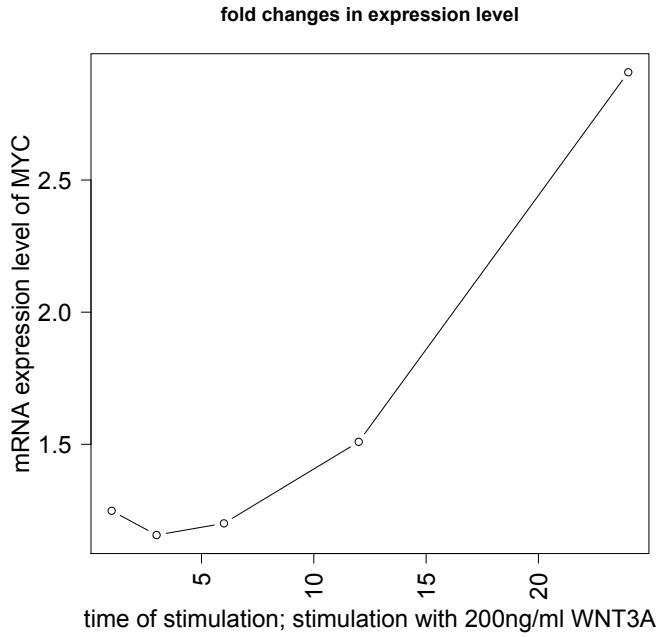

Fig. 28 mRNA expression levels of MYC at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

changes for *APC* is maximum during $<t_3,t_6>$ when the down-regulation is weakened. Further weaking of the downregulation during $<t_6,t_{12}>$ does not have much significance. This attenuation in significance of deviations in fold change might support the fact that *APC*'s weaking in downregulation amplifies shutting down of the Wnt pathway after the intial strong downregulation (where Wnt activity is high). This is corroborated by the finding of Gujral and MacBeath[17] which observes the initial (later) positive (negative) feedback that strenghtens (weakens) the Wnt pathway activity. An initial increase in the influence of *APC* is observed from $<t_1,t_3>$ to $<t_3,t_6>$ (first two bars in figure 31) followed by a gradual decrease of influence from $<t_3,t_6>$ to $<t_6,t_{12}>$ to $<t_{12},t_{24}>$ (last three bars in figure 31). This is observed even though there is an down-regulation in levels of *APC* with slight weaking at t_6, t_{12} and t_{24} in comparison to recordings at other time frames (see figure 27). In comparison to the contribution of levels of fold changes in figure 21 where it was found that there is a gradual decrease in the influence of *APC* till t_6 and then a further increase in the contribution at latter time frames, one finds that the deviations in fold changes involving $<t_3,t_6>$ have the highest significance with an almost exponential decrease in the deviations in fold changes involving $<t_6,t_{12}>$ & $<t_{12},t_{24}>$.

■ MYC - Figure 28 shows the profile of mRNA expression levels of *MYC* after external stimulation. The profile of the deviations in fold changes of *MYC* in an up-regulated state show the following $(-+++)$ pattern. After an initial dip in the up-regulation at t_6 there is an exponential increase in the fold changes of *MYC* as time progresses. While figure 21 shows an increasing sensitivity of *MYC* for the first three time frames, later up-regulated state of *MYC* due to increasing *WNT3A* stimulations do not hold much significance. In contrast, it is not possible to observe a pattern in the sensitivity of deviations in fold changes for *MYC* except for the fact that the maximum contribution of deviation in fold change is observed for the period of $< t_6, t_{12} >$. This is a period when *MYC*'s significance in the pathway is maximum.

■ GSK3B - Figure 29 shows the profile of mRNA expression levels of $GSK3\beta$ after external stimulation. The profile of the deviations in fold changes of $GSK3\beta$ in an varied regulated state show the following $(-+++)$ pattern. After an initial up-regulation at t_3 there is down regulation at t_6 before which up-regulation follows for latter stages. It is widely known that *WNT* stimulation leads to inhibition of $GSK3\beta$. In contrast to this regard $GSK3\beta$ shows a up-regulated levels at t_3, t_{12} and t_{24}. The author is currently unaware of why this contasting behaviour is exhibited. While work by Gujral and MacBeath[17] does not shed light on this aspect, contrasting models of inhibitions for *GSK3* has been recently proposed in Metcalfe and Bienz[102] which might support this behaviour. Figure 21 shows an decreasing sensitivity of $GSK3\beta$ for the first two time frames, after which there is an increasing sensitivity for the next three time frames. Comparing this with plots

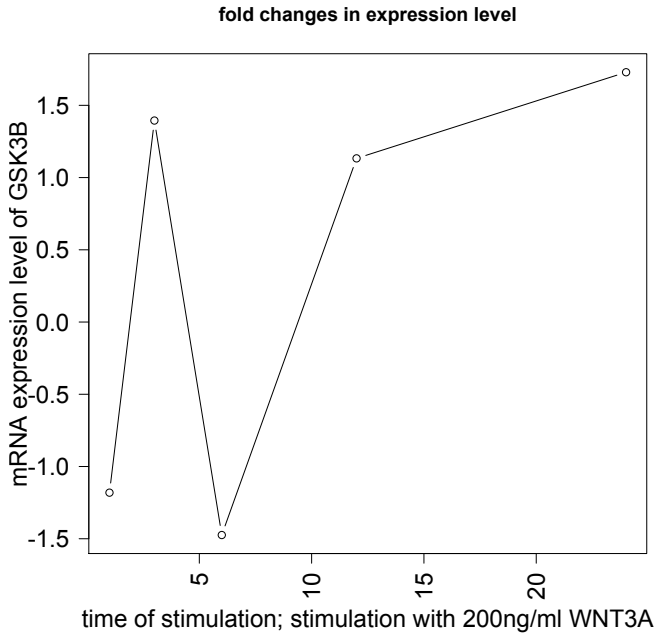

Fig. 29 mRNA expression levels of $GSK3\beta$ at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

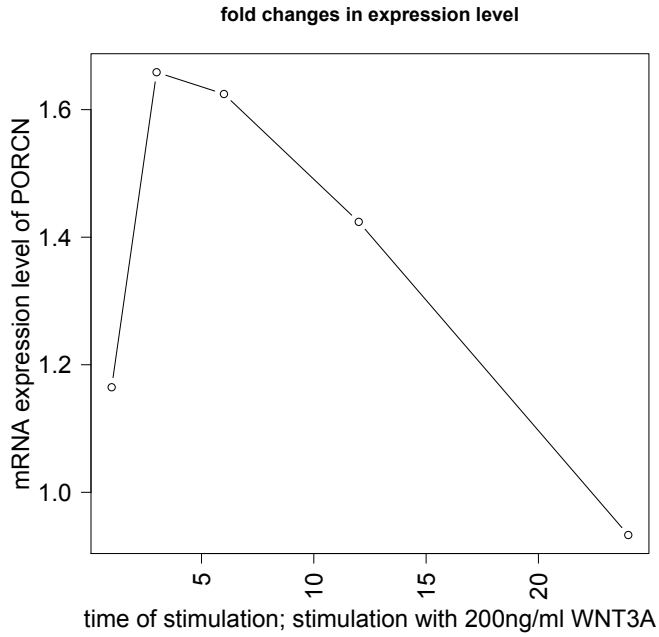

Fig. 30 mRNA expression levels of $GSK3\beta$ at 1st, 3rd, 6th, 12th and 24th hour from Gujral and MacBeath[17].

in figure 31 it is found that there is greater significance of deviations in fold changes of $GSK3\beta$ during later stages of $<t_6,t_{12}>$ and $<t_{12},t_{24}>$. This apparant contradictory behaviour needs to be explained further.

■ PORCN - Figure 30 shows the profile of mRNA expression levels of *PORCN* after external stimulation. The profile of the deviations in fold changes of *PORCN* in an up regulated state show the following $(+---)$ pattern. After an initial hike in up-regulation at t_3 there is continuous decrease in the up regulation. *PORCN* is known to help in the secretion of the Wnt ligands that later on help in the instigation of the signaling activity. Sustanined stimulation by $WNT3A$ over a period of time might lead to decrease in the up regulation of *PORCN* which helps in Wnt secretion. Graph for *PORCN* in figure 21 shows increasing significance of the influence of *PORCN* as time passes, even though there is lower regulation of the same at later stages (fig 30). The highly significant influence of lower regulation at later stages indicates the lessened effectiveness of *PORCN* due to sustained $WNT3A$ stimulation that might have suppressed the functionality of secretion carried out via *PORCN*. Contrary to this, the influences of the deviations in the fold changes over time show the reverse behaviour. The maximum influence is during the first two time frames of $<t_1,t_3>$ and this influence of deviations decreases at later stages. This points to the fact that the deviations in the fold changes at intial stage has greater significance in the pathway than the deviations at later stages.

+++WORK IN PROGRESS+++

6 General Conclusions

Local and global sensitivity analysis on static and time series measurements in Wnt signaling pathway for colorectal cancer is done. Density based Hilbert-Schmidt Information Criterion indices outperformed the variance based Sobol indices. This is attributed to the employment of distance measures & the kernel trick via Reproducing kernel Hilbert space (RKHS) that captures nonlinear relations among various intra/extracellular factors of the pathway in a higher dimensional space. The gained advantage is confirmed on the inferred results obtained via a Bayesian network model based on prior biological knowledge and static gene expression data. In time series data, using these indices it is now possible to observe where in time and to what degree a factor gets influenced & contributes to the pathway, as changes in concentration of the another factor is made. This facilitates in time based administration of target therapeutic drugs & reveal hidden biological information within colorectal cancer samples.

7 Acknowledgement

The author thanks Mrs. Rita Sinha and Mr. Prabhat Sinha for supporting him financially on this project.

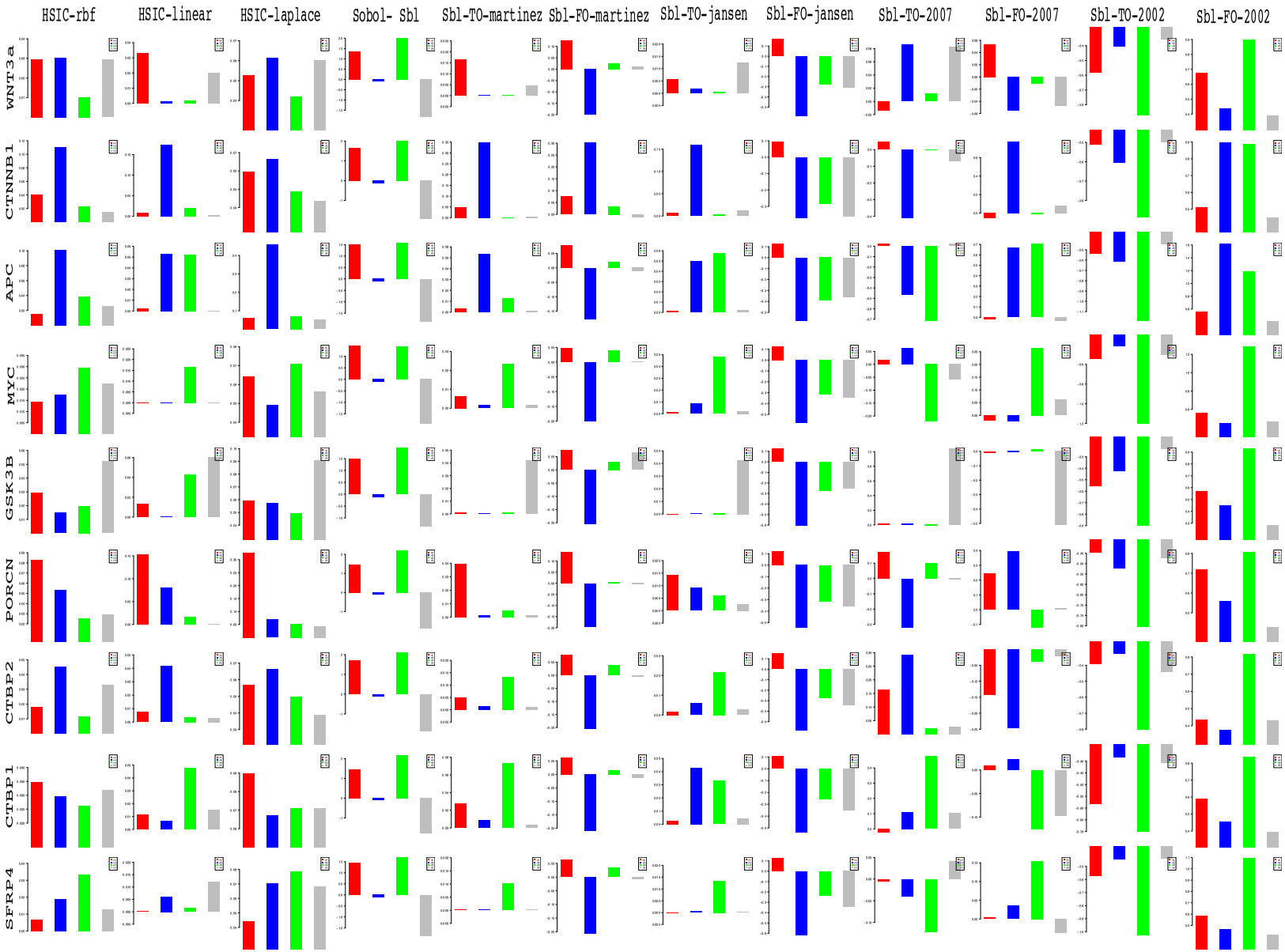

Fig. 31 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene on deviations in fold change. For each graph, the bars represent sensitivity indices computed at <t1,t2> (red), <t2,t3> (blue), <t3,t4> (green) and <t4,t5> (gray). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

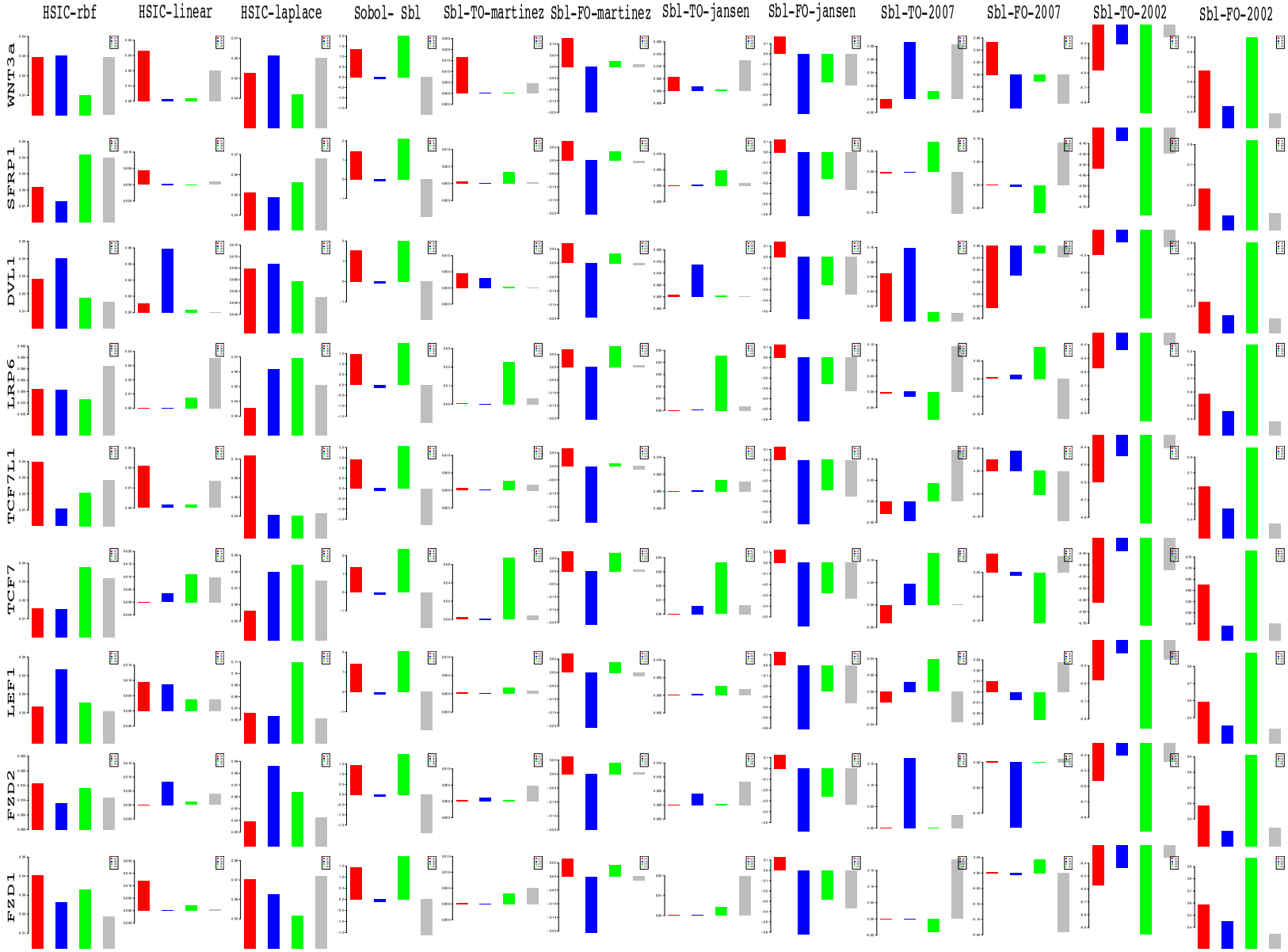

Fig. 32 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene on deviations in fold change. For each graph, the bars represent sensitivity indices computed at <t1,t2> (red), <t2,t3> (blue), <t3,t4> (green) and <t4,t5> (gray). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

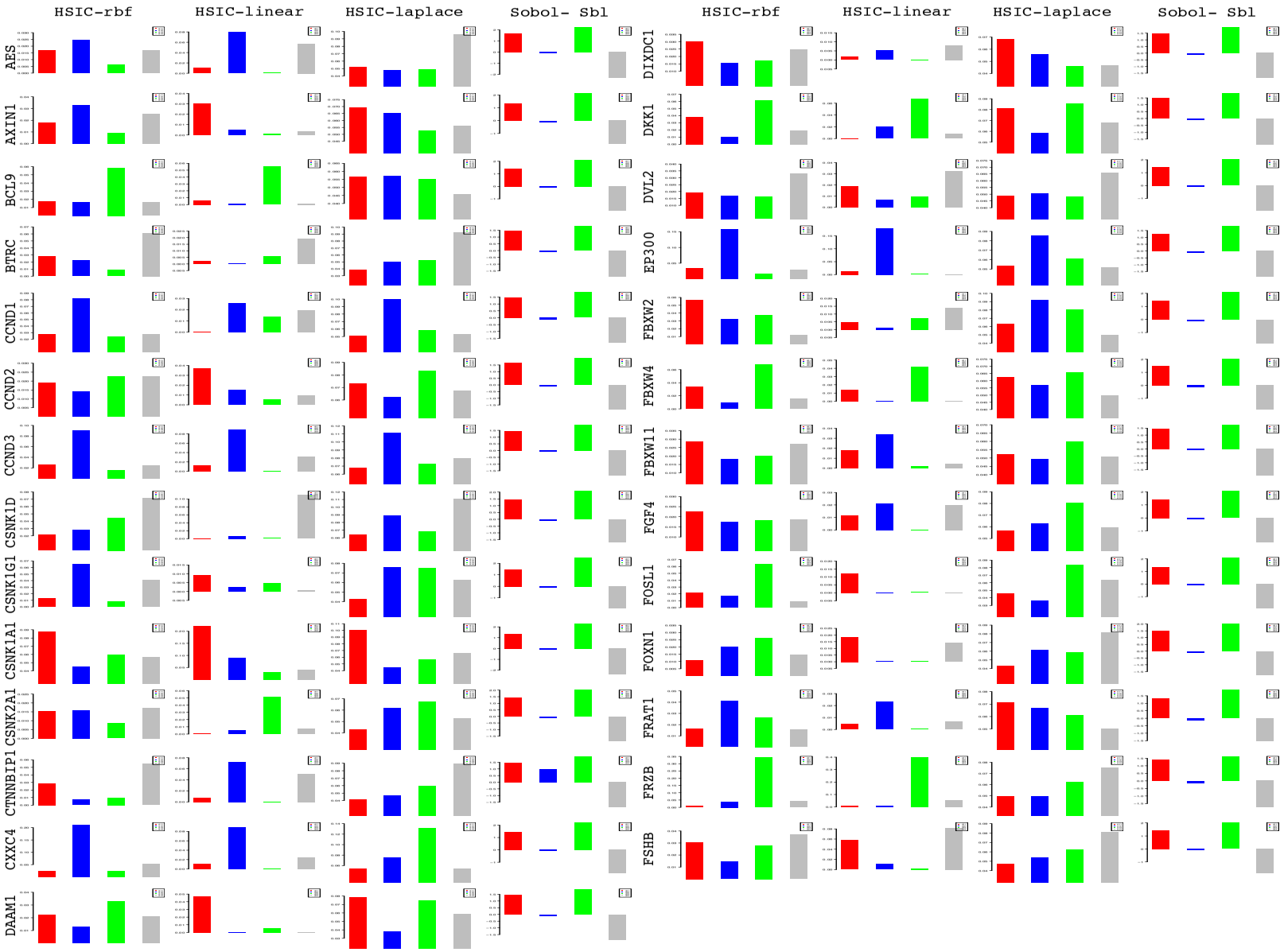

Fig. 33 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene on deviations in fold change. For each graph, the bars represent sensitivity indices computed at <t1,t2> (red), <t2,t3> (blue), <t3,t4> (green) and <t4,t5> (gray). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

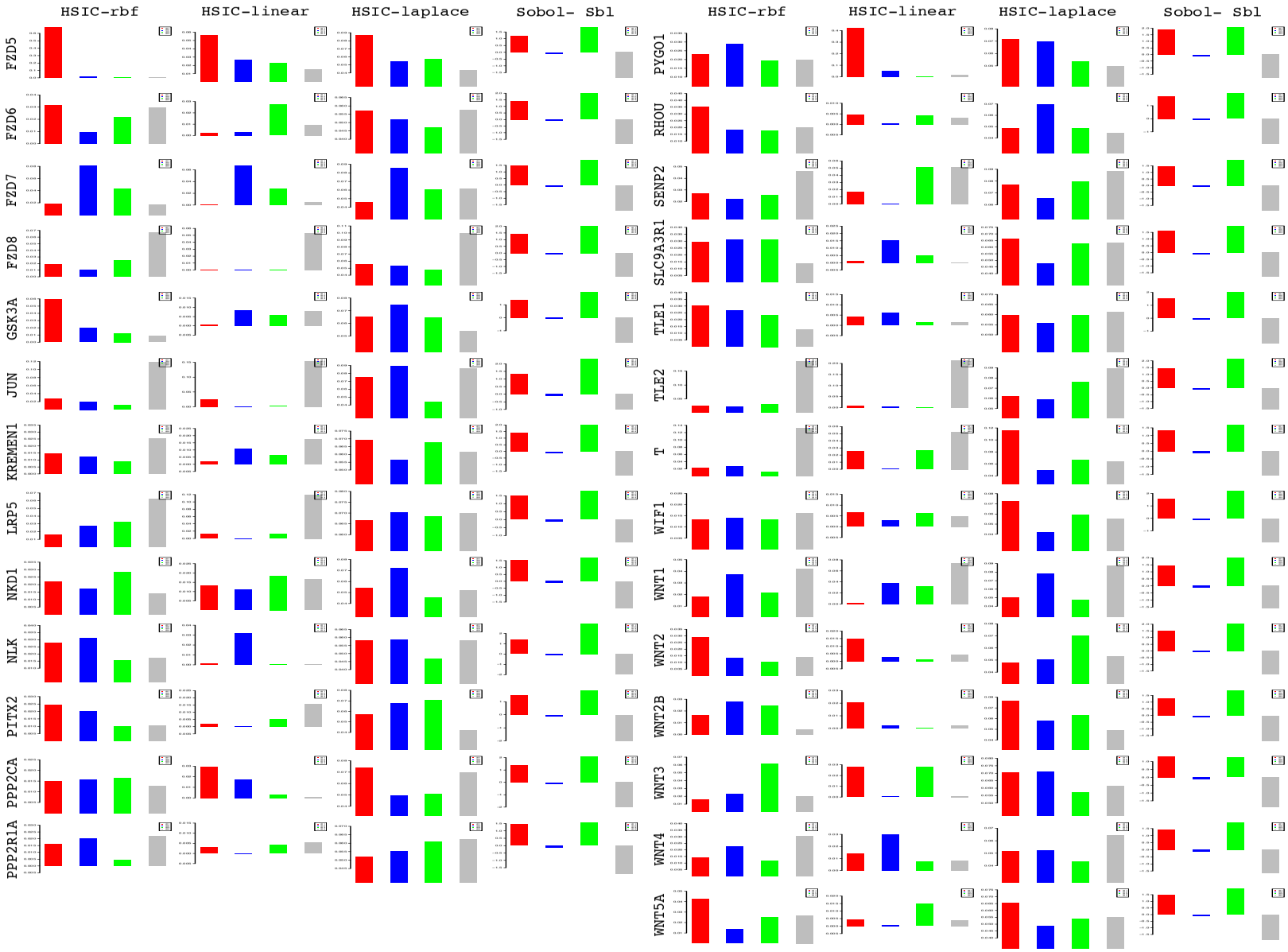

Fig. 34 Column wise - methods to estimate sensitivity indices. Row wise - sensitivity indicies for each gene on deviations in fold change. For each graph, the bars represent sensitivity indices computed at <t1,t2> (red), <t2,t3> (blue), <t3,t4> (green) and <t4,t5> (gray). Indices were computed using non scaled time series data. TO - total order; FO - first order; SBL - Sobol

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

16 X. Jiang, J. Tan, J. Li, S. Kivimäe, X. Yang, L. Zhuang, P. L. Lee, M. T. Chan, L. W. Stanton, E. T. Liu *et al.*, *Cancer cell*, 2008, **13**, 529–541.

17 T. S. Gujral and G. MacBeath, *PloS one*, 2010, **5**, e10024.

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

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

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

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

22 B. Iooss and P. Lemaître, *arXiv preprint arXiv:1404.2405*, 2014.

23 M. D. Morris, *Technometrics*, 1991, **33**, 161–174.

24 H. Moon, A. M. Dean and T. J. Santner, *Technometrics*, 2012, **54**, 376–387.

25 A. Dean and S. Lewis, *Screening: methods for experimentation in industry, drug discovery, and genetics*, Springer Science & Business Media, 2006.

26 T. H. Andres and W. C. Hajas, 1993.

27 B. Bettonvil and J. P. Kleijnen, *European Journal of Operational Research*, 1997, **96**, 180–194.

28 S. C. Cotter, *Biometrika*, 1979, **66**, 317–320.

29 R. Christensen, *Linear models for multivariate, time series, and spatial data*, Springer Science & Business Media, 1991.

30 A. Saltelli, K. Chan and E. Scott, *Willey, New York*, 2000.

31 J. C. Helton and F. J. Davis, *Reliability Engineering & System Safety*, 2003, **81**, 23–69.

32 M. D. McKay, R. J. Beckman and W. J. Conover, *Technometrics*, 1979, **21**, 239–245.

33 T. Homma and A. Saltelli, *Reliability Engineering & System Safety*, 1996, **52**, 1–17.

34 I. M. Sobol, *Mathematics and computers in simulation*, 2001, **55**, 271–280.

35 A. Saltelli, *Computer Physics Communications*, 2002, **145**, 280–297.

36 A. Saltelli, M. Ratto, S. Tarantola and F. Campolongo, *Chemical reviews*, 2005, **105**, 2811–2828.

37 A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana and S. Tarantola, *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.

38 R. Cukier, C. Fortuin, K. E. Shuler, A. Petschek and J. Schaibly, *The Journal of Chemical Physics*, 1973, **59**, 3873–3878.

39 A. Saltelli, S. Tarantola and K.-S. Chan, *Technometrics*, 1999, **41**, 39–56.

40 S. Tarantola, D. Gatelli and T. A. Mara, *Reliability Engineering & System Safety*, 2006, **91**, 717–727.

41 A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto and S. Tarantola, *Computer Physics Communications*, 2010, **181**, 259–270.

42 A. Janon, T. Klein, A. Lagnoux, M. Nodet and C. Prieur, *ESAIM: Probability and Statistics*, 2014, **18**, 342–364.

43 A. B. Owen, *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 2013, **23**, 11.

44 J.-Y. Tissot and C. Prieur, *Reliability Engineering & System Safety*, 2012, **107**, 205–213.

45 S. Da Veiga and F. Gamboa, *Journal of Nonparametric Statistics*, 2013, **25**, 573–595.

46 G. Archer, A. Saltelli and I. Sobol, *Journal of Statistical Computation and Simulation*, 1997, **58**, 99–120.

47 S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz *et al.*, *Reliability Engineering & System Safety*, 2007, **92**, 957–960.

48 M. J. Jansen, *Computer Physics Communications*, 1999, **117**, 35–43.

49 C. B. Storlie and J. C. Helton, *Reliability Engineering & System Safety*, 2008, **93**, 28–54.

50 S. Da Veiga, F. Wahl and F. Gamboa, *Technometrics*, 2009, **51**, 452–463.

51 G. Li, C. Rosenthal and H. Rabitz, *The Journal of Physical Chemistry A*, 2001, **105**, 7765–7777.

52 K. H. Hajikolaei and G. G. Wang, *Journal of Mechanical Design*, 2014, **136**, 011003.

53 E. Borgonovo, *Reliability Engineering & System Safety*, 2007, **92**, 771–784.

54 I. Sobol and S. Kucherenko, *Mathematics and Computers in Simulation*, 2009, **79**, 3009–3017.

55 J.-C. Fort, T. Klein and N. Rachdi, *arXiv preprint arXiv:1305.2329*, 2013.

56 S. Da Veiga, *Journal of Statistical Computation and Simulation*, 2015, **85**, 1283–1305.

57 G. J. Székely, M. L. Rizzo, N. K. Bakirov *et al.*, *The Annals of Statistics*, 2007, **35**, 2769–2794.

58 A. Gretton, O. Bousquet, A. Smola and B. Schölkopf, Algorithmic learning theory, 2005, pp. 63–77.

59 I. Csiszár *et al.*, *Studia Sci. Math. Hungar.*, 1967, **2**, 299–318.

60 M. Aizerman, E. Braverman and L. Rozonoer, *Automation and Remote Control*, 1964, **25**, 821–837.

61 T. Sumner, E. Shephard and I. Bogle, *Journal of The Royal Society Interface*, 2012, **9**, 2156–2166.

62 Y. Zheng and A. Rundell, *IEE Proceedings-Systems Biology*, 2006, **153**, 201–211.

63 S. Marino, I. B. Hogue, C. J. Ray and D. E. Kirschner, *Journal of theoretical biology*, 2008, **254**, 178–196.

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

65 S. Sobol, IM andKucherenko, *Wilmott Magazine*, 2–7.

66 M. Baucells and E. Borgonovo, *Management Science*, 2013, **59**, 2536–2549.

67 A. Kraskov, H. Stögbauer and P. Grassberger, *Physical review E*, 2004, **69**, 066138.

68 D. Sejdinovic, B. Sriperumbudur, A. Gretton, K. Fukumizu *et al.*, *The Annals of Statistics*, 2013, **41**, 2263–2291.

69 H. Daumé III, *From zero to reproducing kernel hilbert spaces in twelve pages or less*, 2004.

70 F. Riesz, *CR Acad. Sci. Paris*, 1907, **144**, 1409–1411.

71 J. S. Taylor and N. Cristianini, *Properties of Kernels*, Cambridge University Press, 2004.

72 R. Faivre, B. Iooss, S. Mahévas, D. Makowski and H. Monod, *Analyse de sensibilité et exploration de modèles: application aux sciences de la nature et de l'environnement*, Editions Quae, 2013.

73 M. De Lozzo and A. Marrel, *arXiv preprint arXiv:1412.1414*, 2014.

74 K. Ueno, H. Hirata, Y. Hinoda and R. Dahiya, *International Journal of Cancer*, 2013, **132**, 1731–1740.

75 R. Holcombe, J. Marsh, M. Waterman, F. Lin, T. Milovanovic and T. Truong, *Molecular Pathology*, 2002, **55**, 220.

76 K. Planutis, M. Planutiene, A. V. Nguyen, M. P. Moyer and R. F. Holcombe, *J Transl Med*, 2013, **11**, 10–1186.

77 A. Sato, H. Yamamoto, H. Sakane, H. Koyama and A. Kikuchi, *The EMBO journal*, 2010, **29**, 41–54.

78 Z. Klapholz-Brown, G. G. Walmsley, Y. M. Nusse, R. Nusse and P. O. Brown, *PloS one*, 2007, **2**, e945.

79 N. Yokoyama, D. Yin and C. C. Malbon, *Journal of molecular signaling*, 2007, **2**, 11.

80 T.-C. He, A. B. Sparks, C. Rago, H. Hermeking, L. Zawel, L. T. da Costa, P. J. Morin, B. Vogelstein and K. W. Kinzler, *Science*, 1998, **281**, 1509–1512.

81 V. Korinek, N. Barker, P. J. Morin, D. van Wichen, R. de Weger, K. W. Kinzler, B. Vogelstein and H. Clevers, *Science*, 1997, **275**, 1784–1787.

82 P. J. Morin, A. B. Sparks, V. Korinek, N. Barker, H. Clevers, B. Vogelstein and K. W. Kinzler, *Science*, 1997, **275**, 1787–1790.

83 S.-i. Hino, T. Michiue, M. Asashima and A. Kikuchi, *Journal of Biological Chemistry*, 2003, **278**, 14066–14073.

84 X. J. You, P. J. Bryant, F. Jurnak and R. F. Holcombe, *Oncology reports*, 2007, **18**, 691–694.

85 J. M. González-Sancho, K. R. Brennan, L. A. Castelo-Soccio and A. M. Brown, *Molecular and cellular biology*, 2004, **24**, 4757–4768.

86 C. Gao and Y.-G. Chen, *Cellular signalling*, 2010, **22**, 717–727.

87 M.-Y. Huang, L.-C. Yen, H.-C. Liu, P.-P. Liu, F.-Y. Chung, T.-N. Wang, J.-Y. Wang and S.-R. Lin, *International journal of molecular sciences*, 2013, **14**, 20492–20507.

88 L. M. Galli, T. Barnes, T. Cheng, L. Acosta, A. Anglade, K. Willert, R. Nusse and L. W. Burrus, *Developmental dynamics*, 2006, **235**, 681–690.

89 H. Suzuki, D. N. Watkins, K.-W. Jair, K. E. Schuebel, S. D. Markowitz, W. D. Chen, T. P. Pretlow, B. Yang, Y. Akiyama, M. van Engeland *et al.*, *Nature genetics*, 2004, **36**, 417–422.

90 G. M. Caldwell, C. Jones, K. Gensberg, S. Jan, R. G. Hardy, P. Byrd, S. Chughtai, Y. Wallis, G. M. Matthews and D. G. Morton, *Cancer research*, 2004, **64**, 883–888.

91 G. Chinnadurai, *Molecular cell*, 2002, **9**, 213–224.

92 F. Hamada and M. Bienz, *Developmental cell*, 2004, **7**, 677–685.

93 J. Schneikert, K. Brauburger and J. Behrens, *Human molecular genetics*, 2011, **20**, 3554–3564.

94 J. Patel, S. Baranwal, I. M. Love, N. J. Patel, S. R. Grossman and B. B. Patel, *Cell Cycle*, 2014, **13**, 3506–3518.

95 K. Willert and R. Nusse, *Cold Spring Harbor perspectives in biology*, 2012, **4**, a007864.

96 S. Sinha, *bioRxiv*, 2015, 015834.

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

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

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

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

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

102 C. Metcalfe and M. Bienz, *Journal of Cell Science*, 2011, **124**, 3537–3544.