bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ARTICLE TYPE

Prioritizing 2nd and 3rd order interactions via support vector ranking using sensitivity indices on static Wnt measurements †

shriprakash sinha* ◕

It is widely known that the sensitivity analysis plays a major role in computing the strength of the influence of involved factors in any phenomena under investigation. When applied to expression profiles of various intra/extracellular factors that form an integral part of a signaling pathway, the variance and density based analysis yields a range of sensitivity indices for individual as well as various combinations of factors. These combinations denote the higher order interactions among the involved factors that might be of interest in the working mechanism of the pathway. For example, $DACT3$ is known to be a epigenetic regulator of the Wnt pathway in colorectal cancer and subject to histone modifications. But many of the $n^{th} \geq 2$ order interactions of $DACT3$ that might be influential have not been explored/tested. In this work, after estimating the individual effects of factors for a higher order combination, the individual indices are considered as discriminative features. A combination, then is a multivariate feature set in higher order (≥ 2). With an excessively large number of factors involved in the pathway, it is difficult to search for important combinations in a wide search space over different orders. Exploiting the analogy of prioritizing webpages using ranking algorithms, for a particular order, a full set of combinations of interactions can then be prioritized based on these features using a powerful ranking algorithm via support vectors. The computational ranking sheds light on unexplored combinations that can further be investigated using hypothesis testing based on wet lab experiments. Here, the basic framework and results obtained on 2nd and 3rd order interactions for members of family of $DACT$, $SFRP$, DKK (to name a few) in both normal and tumor cases is presented using a static data set.

Significance

The search and wet lab testing of unknown biological hypotheses in the form of combinations of various intra/extracellular factors that are involved in a signaling pathway, costs a lot in terms of time, investment and energy. To reduce this cost of search in a vast combinatorial space, a pipeline has been developed that prioritises these list of combinations so that a biologist can narrow down their investigation. The pipeline uses kernel based sensitivity indices to capture the influence of the factors in a pathway and employs powerful support vector ranking algorithm. The generic workflow and future improvements are bound to cut down the cost for many wet lab experiments and reveal unknown/untested biological hypothesis.

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])

* *Corresponding Author : shriprakash sinha*

◕ *Author worked on this project as an independent researcher. The aspects of the work have been presented in a poster session at the EMBO Wnt 2016 at Brno, Czech Republic. 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/

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

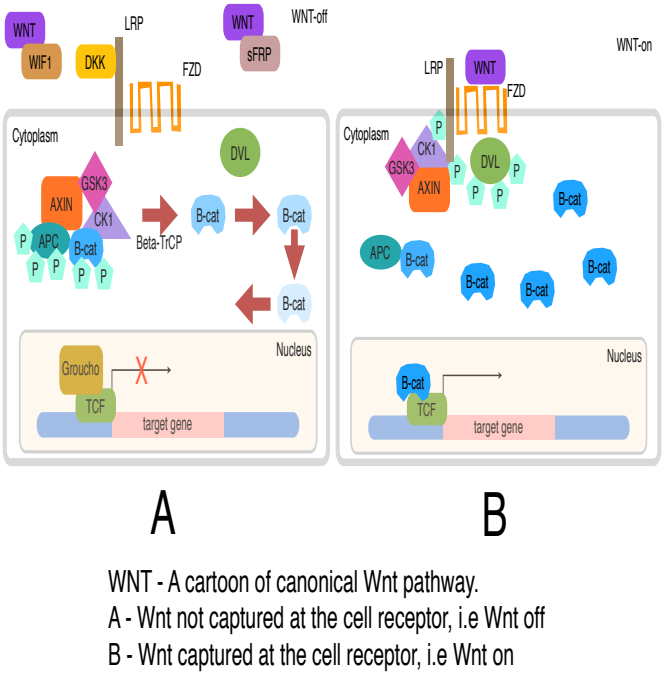


Fig. 1 A cartoon of Wnt signaling pathway. Part (A) represents the destruction of *β-catenin* leading to the inactivation of the Wnt target gene. Part (B) represents activation of Wnt target gene.

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 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]) 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[17], 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[18]). 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 (*CK*1) and kinase *GSK*3. 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*, *CK*1 and *GSK*3. 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

It is widely known that the sensitivity analysis plays a major role in computing the strength of the influence of involved factors in any phenomena under investigation. When applied to expression profiles of various intra/extracellular factors that form an integral part of a signaling pathway, the variance and density based analysis yields a range of sensitivity indices for individual as well as various combinations of factors. These combinations denote the higher order interactions among the involved factors. Computation of higher order interactions is often time consuming but they give a chance to explore the various combinations that might be of interest in the working mechanism of the pathway. For example, in a range of fourth order combinations among the various factors of the Wnt pathway, it would be easy to assess the influence of the destruction complex formed by APC, AXIN, CSKI and GSK3 interaction. Unknown interactions can be further investigated by

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

transforming biological hypothesis regarding these interactions in vitro, in vivo or in silico. But to mine these unknown interactions it is necessary to search a wide space of all combinations of input factors involved in the pathway.

In this work, after estimating the individual effects of factors for a higher order combination, the individual indices are considered as discriminative features. A combination, then is a multivariate feature set in higher order (>2). With an excessively large number of factors involved in the pathway, it is difficult to search for important combinations in a wide search space over different orders. Exploiting the analogy with the issues of prioritizing webpages using ranking algorithms, for a particular order, a full set of combinations of interactions can then be prioritized based on these features using a powerful ranking algorithm via support vectors (Joachims[19]). The computational ranking sheds light on unexplored combinations that can further be investigated using hypothesis testing based on wet lab experiments. In this manuscript both local and global SA methods are used.

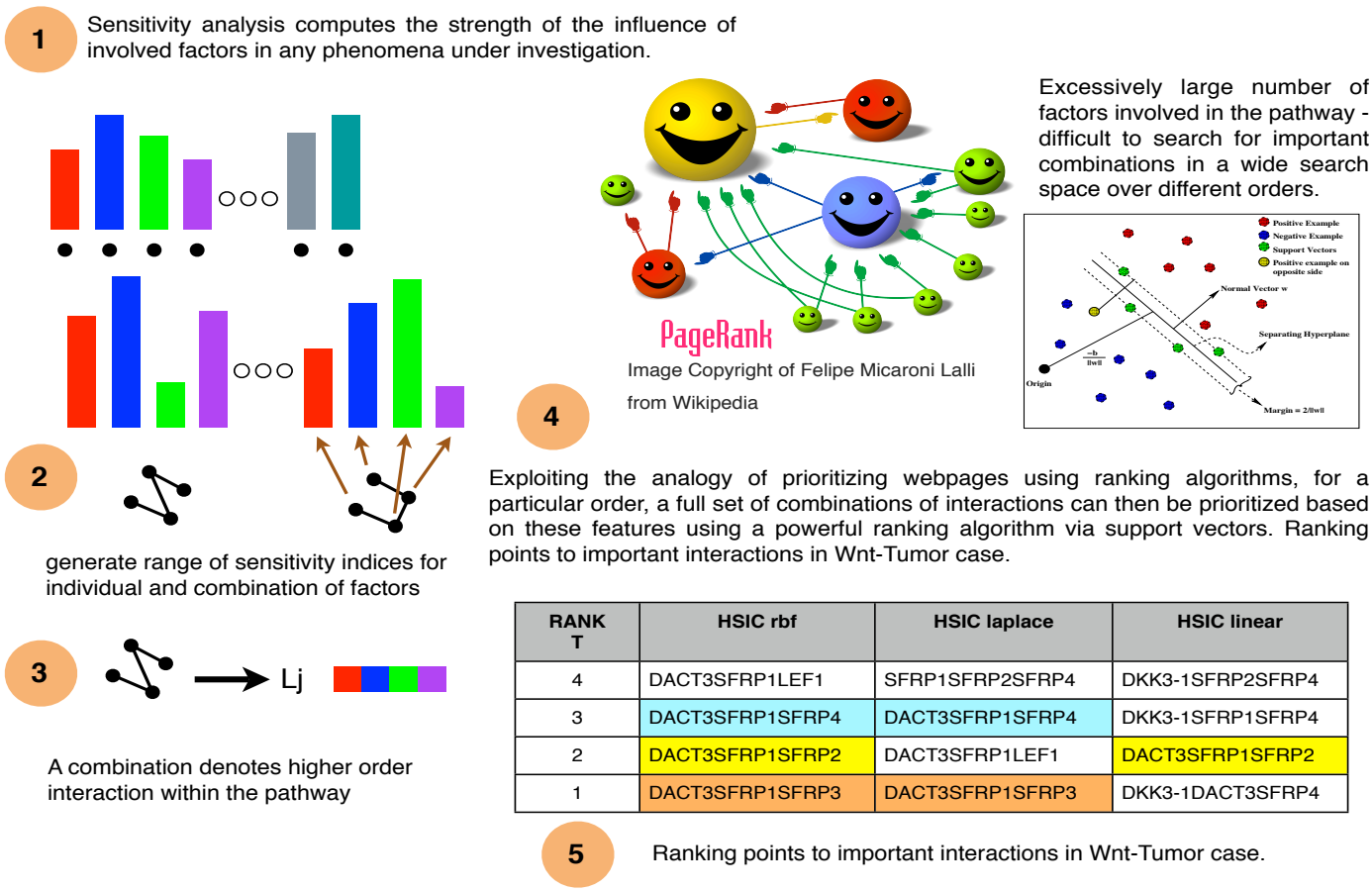
Similar higher order interactions can be computed and prioritized. This gives a chance to examine the biological hypothesis of interest regarding the positive/negative roles of unexplored combinations of the various factors involved in the Wnt pathway, in tumor and normal cases. Using recordings in time, it is possible to find how the prioritization of a specific combination of involved factors is changing in time. This further helps in revealing when an important combination like destruction complex formed by APC, AXIN, CSKI and GSK3 interaction could be influenced via an intervention. Thus a powerful computational mechanism is presented to explore the interactions involved in Wnt pathway. The framework can be used in any other pathway also.

3 Sensitivity analysis

Seminal work by Russian mathematician Sobol'[20] 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 recent unpublished review on the global SA methods by Iooss and Lemaître[21] categorically delineates these methods with the following functionality • screening for sorting influential measures (Morris[22] method, Group screening in Moon *et al.*[23] & Dean and Lewis[24], Iterated factorial design in Andres and Hajas[25], Sequential bifurcation in Bettonvil and Kleijnen[26] and Cotter[27] design), • quantitative indicies for measuring the importance of contributing input factors in linear models (Christensen[28], Saltelli *et al.*[29], Helton and Davis[30] and McKay *et al.*[31]) and nonlinear models (Homma and Saltelli[32], Sobol[33], Saltelli[34], Saltelli *et al.*[35], Saltelli *et al.*[36], Cukier *et al.*[37], Saltelli *et al.*[38], & Tarantola *et al.*[39] Saltelli *et al.*[40], Janon *et al.*[41], Owen[42], Tissot and Prieur[43], Da Veiga and Gamboa[44], Archer *et al.*[45], Tarantola *et al.*[46], Saltelli and Annoni[47] 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[21] also provide various criteria in a flowchart for adapting a method or a combination of the methods for sensitivity analysis. Figure 3 shows the general flow of the mathematical formulation for computing the indices in the variance based Sobol method. The general idea is as follows - A model could be represented as a mathematical function with a multidimensional input vector where each element of a vector is an input factor. This function needs to be defined in a unit dimensional cube. Based on ANOVA decomposition, the function can then be broken down into f_0 and summands of different dimensions, if f_0 is a constant and integral of summands with respect to their own variables is 0. This implies that orthogonality follows in between two functions of different dimensions, if at least one of the variables is not repeated. By applying these properties, it is possible to show that the function can be written into a unique expansion. Next, assuming that the function is square integrable variances can be computed. The ratio of variance of a group of input factors to the variance of the total set of input factors constitute the sensitivity index of a particular group. Detailed derivation is present in the Appendix.

Besides the above Sobol'[20]'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'[20]) 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. Figure 4 shows the general flow of the mathematical formulation for computing the indices in the density based HSIC method. The general idea is as follows - The sensitivity index is actually a distance correlation which incorporates the kernel based Hilbert-Schmidt Information Criterion between two input vectors in higher dimension. The criterion is nothing but the Hilbert-Schmidt norm of cross-covariance operator which generalizes the covariance matrix by representing higher order correlations between the input

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.



RANK T	HSIC rbf	HSIC laplace	HSIC linear
4	DACT3SFRP1LEF1	SFRP1SFRP2SFRP4	DKK3-1SFRP2SFRP4
3	DACT3SFRP1SFRP4	DACT3SFRP1SFRP4	DKK3-1SFRP1SFRP4
2	DACT3SFRP1SFRP2	DACT3SFRP1LEF1	DACT3SFRP1SFRP2
1	DACT3SFRP1SFRP3	DACT3SFRP1SFRP3	DKK3-1DACT3SFRP4

Fig. 2 A graphical view of the general idea behind the current work. (1) Sensitivity indices capture the influence of involved factors in a pathway. (2) Generate sensitivity indices for individual and combinations of involved factors. Note that for a combination, the indices for the involved factors in the combination are generated separately. (3) Vectorize the indices per combination. (4) Rank these combinations based on the sensitivity indices using support vector ranking as web pages are ranked using a ranking algorithm. (5) Obtained is a prioritized list of interactions that could point to important interactions in the pathway in cancer cases.

vectors through nonlinear kernels. For every operator and provided the sum converges, the Hilbert-Schmidt norm is the dot product of the orthonormal bases. For a finite dimensional input vectors, the Hilbert-Schmidt Information Criterion estimator is a trace of product of two kernel matrices (or the Gram matrices) with a centering matrix such that HSIC evalutes to a summation of different kernel values. Detailed derivation is present in the Appendix.

It is this strength of the kernel methods that HSIC is able to capture the deep nonlinearities in the biological data and provide reasonable information regarding the degree of influence of the involved factors within the pathway. Improvements in variance based methods also provide ways to cope with these nonlinearities but do not exploit the available strength of kernel methods. Results in the later sections provide experimental evidence for the same.

3.1 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'[20] variance based indices versus Morris[22] screening method which uses a One-at-a-time (OAT) approach to analyse the sensitivity of $GSK3$ 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

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

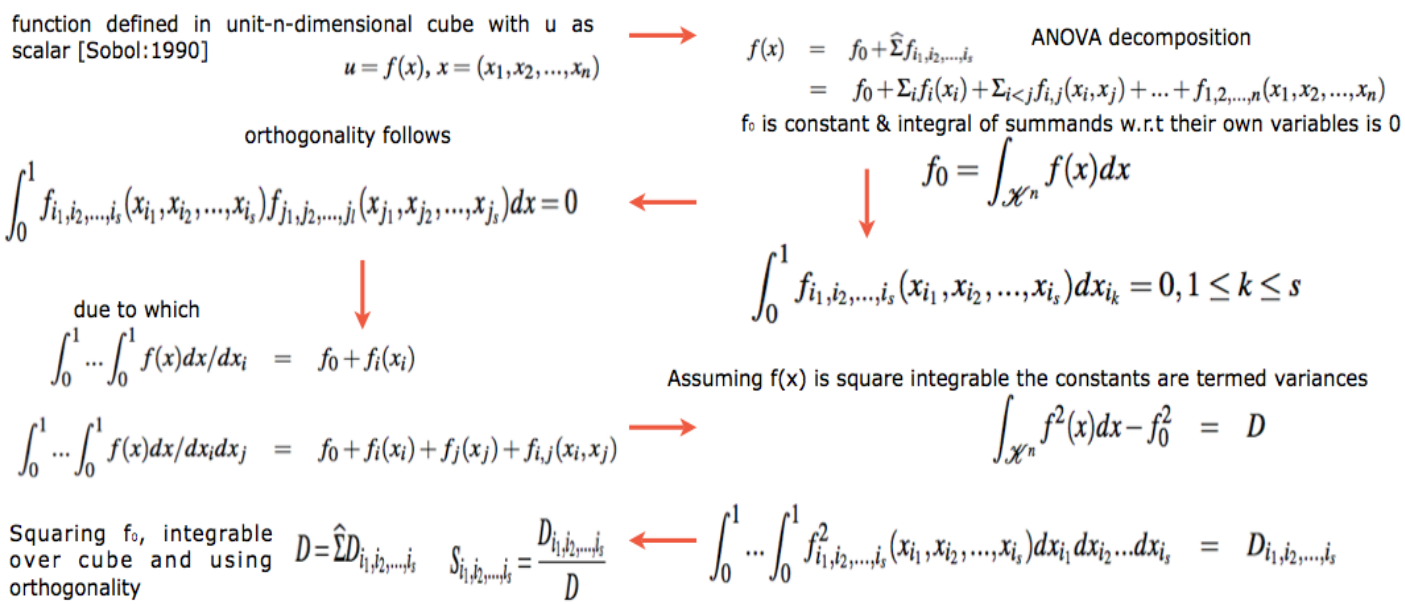


Fig. 3 Computation of variance based sobol sensitivity indices. For detailed notations, see appendix.

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]).

There are two approaches to sensitivity analysis. The first is the local sensitivity analysis in which if there is a required solution, then the sensitivity of a function apropos a set of variables is estimated via a partial derivative for a fixed point in the input space. In global sensitivity, the input solution is not specified. This implies that the model function lies inside a cube and the sensitivity indices are regarded as tools for studying the model instead of the solution. The general form of g-function (as the model or output variable) is used to test the sensitivity of each of the input factor (i.e expression profile of each of the genes). This is mainly due to its non-linearity, non-monotonicity as well as the capacity to produce analytical sensitivity indices. The g-function takes the form -

$$f(x) = \Pi_{i=1}^{d} \frac{|4 * x_i - 2| + a_i}{1 + a_i} \quad (1)$$

were, d is the total number of dimensions and $a_i \geq 0$ are the indicators of importance of the input variable x_i. Note that lower values of a_i indicate higher importance of x_i. In our formulation, we randomly assign values of $x_i \in [0, 1]$. For the static (time series) data $d = 18$ (factors affecting the pathway). The value of d varies from 2 to 17, depending on the order of the combination one might be interested in. Thus the expression profiles of the various genetic factors in the pathway are considered as input factors and the global analysis conducted. Note that in the predefined dataset, the working of the signaling pathway is governed by a preselected set of genes that affect the pathway. For comparison purpose, the local sensitivity analysis method is also used to study how the individual factor is behaving with respect to the remaining factors while working of the pathway is observed in terms of expression profiles of the various factors.

Finally, in context of Goentoro and Kirschner[64]'s work regarding the recent development of observation of Weber's law working downstream of the pathway, it has been found that the law is governed by the ratio of the deviation in the input and the absolute input value. More importantly, it is these deviations in input that are of significance in studing such a phemomena. The current manuscript explores the sensitivity of deviation in the fold changes between measurements of fold changes at consecutive time points to explore in what duration of time, a particular factor is affecting the pathway in a major way. This has deeper implications in the fact that one is now able to observe when in time an intervention can be made or a gene be perturbed to study the behaviour of the pathway in tumorous cases. Thus sensitivity analysis of deviations in the mathematical formulation of the psychophysical law can lead to insights into the time period based

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

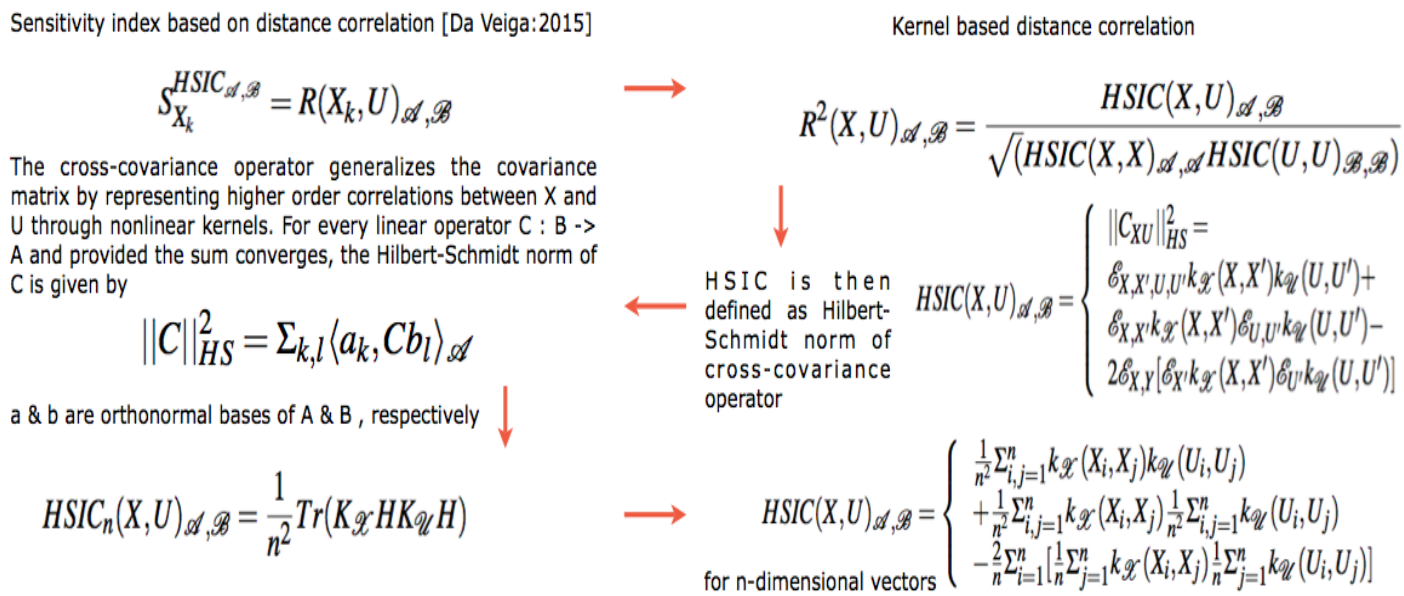


Fig. 4 Computation of density based hsic sensitivity indices. For detailed notations, see appendix.

influence of the involved factors in the pathway. Thus, both global and local anaylsis methods are employed to observe the entire behaviour of the pathway as well as the local behaviour of the input factors with respect to the other factors, respectively, via analysis of fold changes and deviations in fold changes, in time.

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 and the choice of sensitivity packages which are already available in literature, has been described in the Appendix.

4 Ranking Support Vector Machines

Learning to rank is a machine learning approach with the idea that the model is trained to learn how to rank. A good introduction to this work can be found in Li[65]. Existing methods involve pointwise, pairwise and listwise approaches. In all these approaches, Support Vector Machines (SVM) can be employed to rank the required query. SVMs for pointwise approach build various hyperplanes to segregate the data and rank them. Pairwise approach uses ordered pair of objects to classify the objects and then utilize the classifyer to rank the objects. In this approach, the group structure of the ranking is not taken into account. Finally, the listwise ranking approach uses ranking list as instances for learning and prediction. In this case the ranking is straightforward and the group structure of ranking is maintained. Various different designs of SVMs have been developed and the research in this field is still in preliminary stages. In context of the gene expression data set employed in this manuscript, the objects are the genes with their RECORDED EXPRESSION VALUES FOR NORMAL AND TUMOR CASES. Both cases are treated separately.

Note that rankings algorithms have been developed to be employed in the genomic datasets but to the author's awareness, these algorithms do not rank the range of combinations in a wide combinatorial search space in time. Also, they do not take into account the ranking of unexplored biological hypothesis which are assigned to a particular sensitivity value or vector that can be used for prioritization. For example, Kolde *et al.*[66] presents a ranking algorithm that betters existing ranking model based on the assignment of *P*-value. As stated by Kolde *et al.*[66] it detects genes that are ranked consistently better than expected under null hypothesis of uncorrelated inputs and assigns a significance score for each gene. The underlying probabilistic model makes the algorithm parameter free and robust to outliers, noise and errors. Significance scores also provide a rigorous way to keep only the statistically relevant genes in the final list. The proposed work here develops on sensitivity analysis and computes the influences of the factors for a system under investigation. These sensitivity indices give a much realistic view of the biological influence than the proposed *P*-value assignment and the probabilistic model. The manuscript at the current stage does not compare the algorithms as it is a pipeline to investigate and conduct a systems wide study. Instead of using SVM-Ranking it is possible to use other algorithms also, but the author has restricted to the development of the pipeline per se. Finally, the current work tests the effectiveness of the variance based (SOBOL) sensitivity

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

indices apropos the density and kernel based (HSIC) sensitivity indices. Finally, Blondel *et al.*[67] provides a range of comparison for 10 different regression methods and a score to measure the models. Compared to the frame provided in Blondel *et al.*[67], the current pipeline takes into account biological information an converts into sensitivity scores and uses them as discriminative features to provide rankings. Thus the proposed method is algorithm independent.

5 Description of the dataset & design of experiments

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 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, $WIF1$ and $DACT$. Also, expression values of established Wnt pathway target genes like $LEF1$, MYC, $CD44$ and $CCND1$ were recorded per sample.

Note that green (red) represents activation (repression) in the heat maps of data in Jiang *et al.*[16]. 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. 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.

GENERAL ISSUES - • Here the input factors are the gene expression values for both normal and tumor cases in static data. For the case of time series data, the input factors are the fold change (deviations in fold change) expression values of genes at different time points (periods). Also, for the time series data, in the first experiment the analysis of a pair of the fold changes recorded at to different consecutive time points i.e t_i & t_{i+1} is done. In the second experiment, the analysis of a pair of deviations in fold changes recorded at t_i & t_{i+1} and t_{i+1} & t_{i+2}. In this work, in both the static and the time series datasets, the analysis is done to study the entire model/pathway rather than find a particular solution to the model/pathway. Thus global sensitivity analysis is employed. But the local sensitivity methods are used to observe and compare the affect of individual factors via 1st order analysis w.r.t total order analysis (i.e global analysis). In such an experiment, the output is the sensitivity indices of the individual factors participating in the model. This is different from the general trend of observing the sensitivity of parameter values that affect the pathway based on differential equations that model a reaction. Thus the model/pathway is studied as a whole by observing the sensitivities of the individual factors.

Note that the 24 normal and tumor cases are all different from each other. The 18 genes that are used to study in[16] are the input factors and it is unlikely that there will be correlations between different patients. The phenotypic behaviour might be similar at a grander scale. Also, since the sampling number is very small for a network of this scale, large standard deviations can be observed in many results, especially when the Sobol method is used. But this is not the issue with the sampling number. By that analysis, large deviations are not observed in kernel based density methods. The deviations are more because of the fact that the nonlinearities are not captured in an efficient way in the variance based Sobol methods. Due to this, the resulting indicies have high variance in numerical value. For the same number of samplings, the kernel based methods don't show high variance.

The procedure begins with the listing of all C_k^n combinations for k number of genes from a total of n genes. k is ≥ 2 and $\leq (n-1)$. Each of the combination of order k represent a unique set of interaction between the involved genetic factors. While studying the interaction among the various genetic factors using static data, tumor samples are considered separated from normal samples. For the experiments conducted here on a toy example, 20 bootstrap replicates were generated for each of normal and tumor samples without replacement. For each bootstrap replicate, the normal and turmor samples are divided into two different sets of equal size. Next the datasets are combined in a specifed format which go as input as per the requirement of a particular sensitivity analysis method. Thus for each pth combination in C_k^n combinations, the dataset is prepared in the required format from both normal and tumor samples (See .R code in mainscript-1-1.R in google drive and step 3 in figure 5). After the data has been transformed, vectorized programming is employed for density based sensitivity analysis and looping is employed for variance based sensitivity analysis to compute the required sensitivity indices for each of the p combinations. Once the sensitivity indices are generated for each of the pth combination, for every bootstrap replicate in normal and tumor cases, confidence intervals are estimated for each sensitivity index. This procedure is done for different kinds of sensitivity analysis methods.

After the above sensitivity indices have been stored for each of the pth combination, each of the sensitivity analysis method for normal and tumor cases per bootstrap replicates, the next step in the design of experiment is conducted. Here, for a particular k^{th} order of combination, a choice is made regarding the number of sample size (say p), where $2 \leq p \leq noObs-1$ (noObs is the number of observations i.e 20 replicates). Then for all sample sets of order p in C_p^{noObs}, generate training index set of order p and test index set of order $noObs-p$. For each of

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

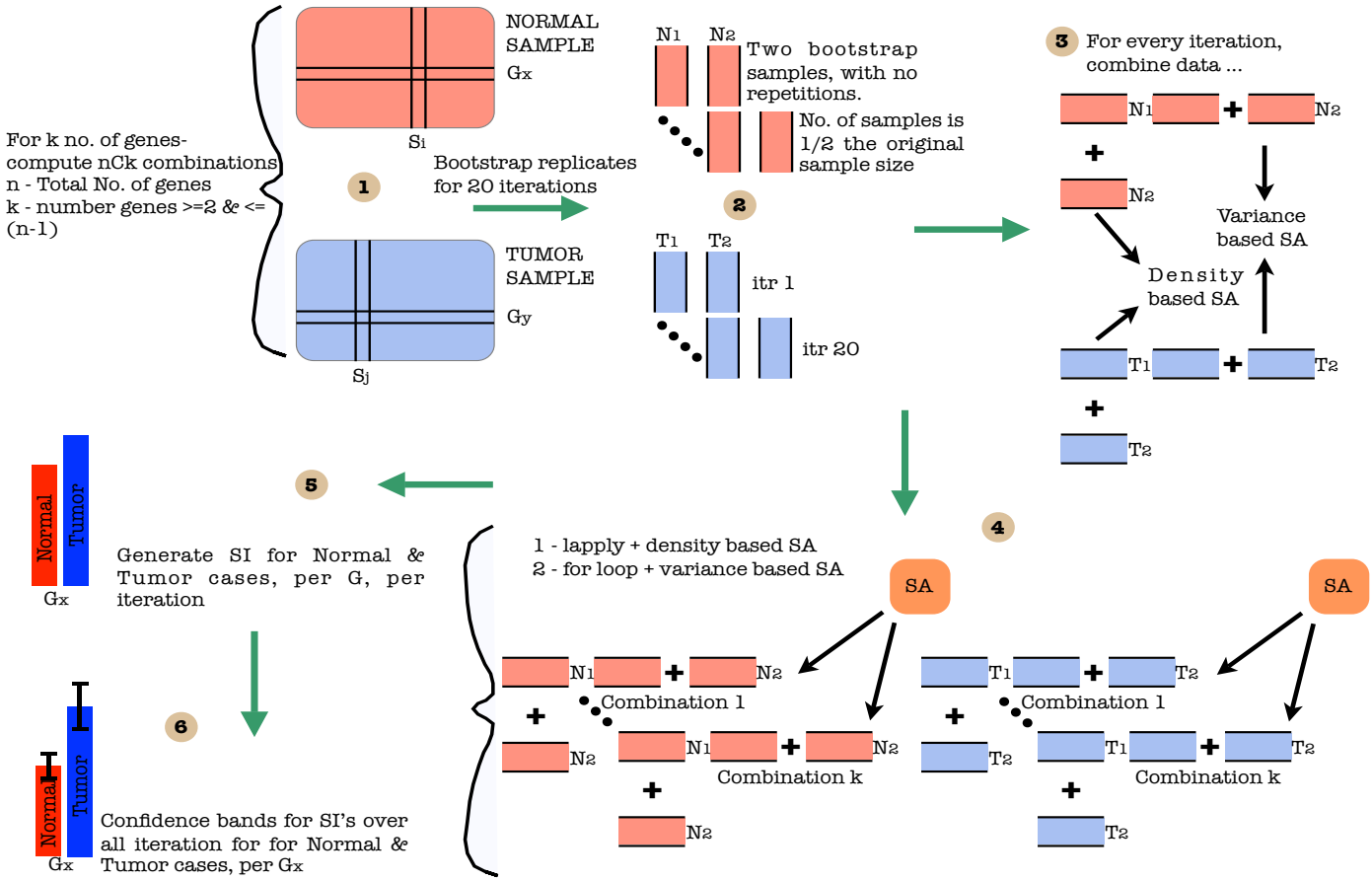


Fig. 5 A cartoon of experimental setup. IMPORTANT NOTE - In this figure, G_x, G_y and G represent a combination. Step - (1) Segregation of data into normal and tumor cases. (2) Further data division per case and bootstrap sampling with no repetitions for different iterations. (3) Assembling bootstrapped data and application of SA methods. (4) Generation of SI's for normal and tumor case per gene per iteration. (5) Generation of averaged SI and confidence bands per case per gene.

the sample set, considering the sensitivity index for each individual factor of a gene combination in the previous step as described in the foregoing paragraph, a training and a test set is generated. Thus an observation in a training and a test set represents a gene combination with sensitivity indices of involved genetic factors as feature values. For a particular gene combination there are p training samples $noObs-p$ test samples. In total there are C_p^{noObs} training sets and corresponding test sets. Next, SVM_{learn}^{Rank} (Joachims[19]) is used to generate a model on default value C value of 20. In the current experiment on toy model C value has not been tunned. The training set helps in the generation of the model as the different gene combinations are numbered in order which are used as rank indices. The model is then used to generate score on the observations in the testing set using the $SVM_{classify}^{Rank}$ (Joachims[19]). Next the scores are averaged across all C_p^{noObs} test samples. The experiment is conducted for normal and tumor samples separately. This procedure is executed for each and every sensitivity analysis method. Finally, for each sensitivity analysis method, for all k^{th} order combinations, the mean across the averaged p scores is computed. This is followed by sorting of these scores along with the rank indices already assigned to the gene combinations. The end result is a sorted order of the gene combinations based on the ranking score learned by the SVM^{Rank} algorithm. These steps are depicted in figure 6.

Note that the following is the order in which the files should be executed in R, in order, for obtaining the desired results (Note that the code will not be explained here) - • use source("mainScript-1-1.R") with arguments for Static data • use source("Combine-Static-files.R"), if computing indices separately via previous file, • source("Store-Results-S.R"), • source("SVMRank-Results-S.R"), again this needs to be done separately for different kinds of SA methods and • source("Sort-n-

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

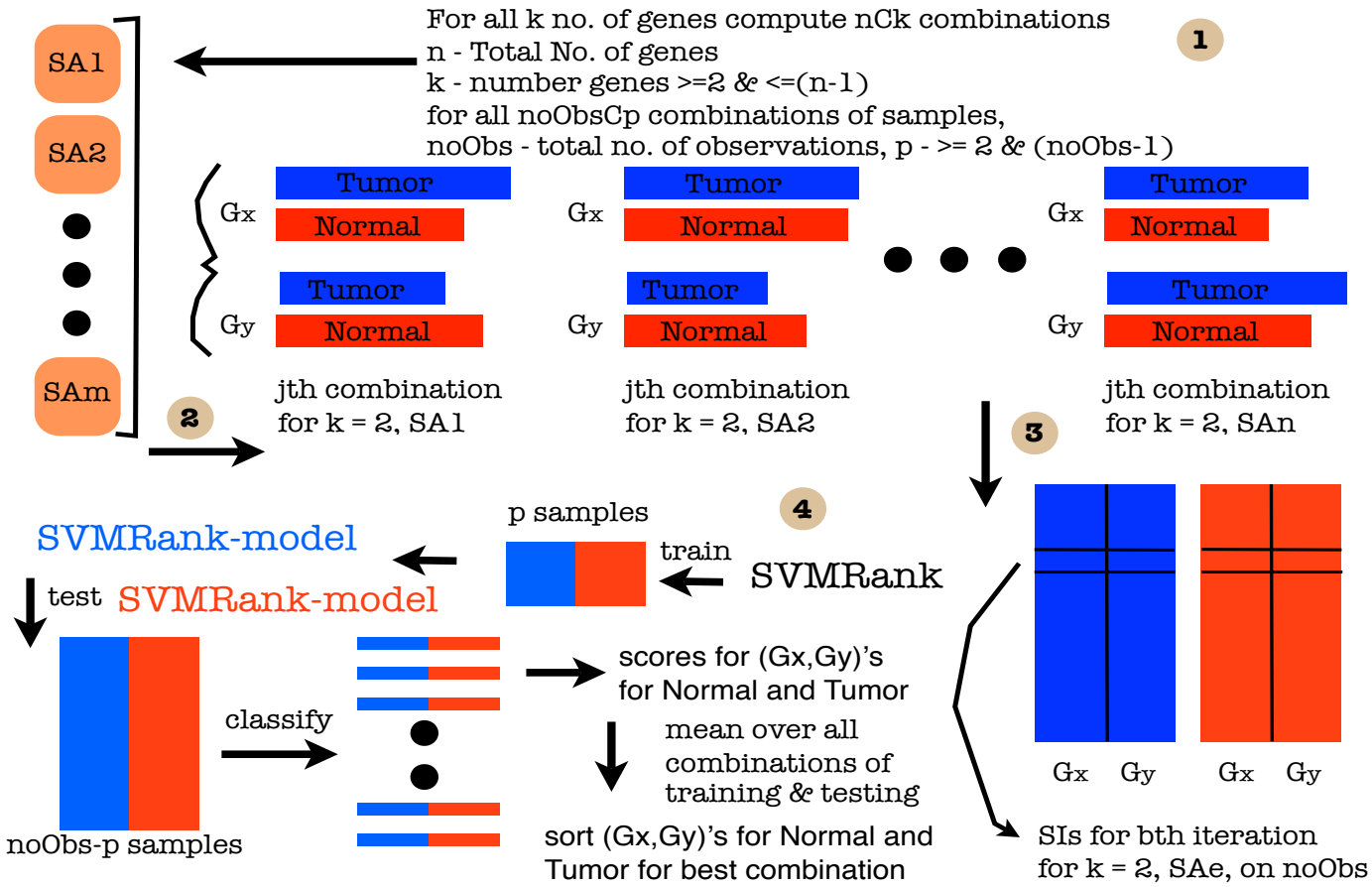


Fig. 6 A cartoon of experimental setup. IMPORTANT NOTE - In this figure, G_x, G_y and G represent separate genes. Step - (1) Assembling p training indices and $noObs-p$ testing indices for every p^{th} order of samples in C_p^{noObs}. Thus there are a total of C_p^{noObs} training and corresponding test sets. (2) For every SA, combine (say for $k=2$, i.e interaction level 2) SI's of genetic factors for normal and tumor separately, for each observation in training and test data. (3) For $noObs=20$ different replicates, per SA_e and a particular combination of $< G_x, G_y >$ in normal and tumor, a matrix of observations is consturcted. (4) Using indices in (1) SVM_{learn}^{Rank} is employed on p training data to generate a model. This model is used to generate a ranking score on the test data via $SVM_{classify}^{Rank}$. These score are averaged over C_p^{noObs} test data sets. Further, mean of scores over $noObs-p$ test replicates per $< G_x, G_y >$ are computed and finally the combinations are ranked based on sorting for each of normal and training set.

Plot-S.R") to sort the interactions.

6 Results and Discussion

Initial results on prioritization of 2nd order interactions learnt from support vector ranking using these sensitivity indices were plotted for visualization. For a training sample size of 4 and test sample size of 16, from a total of 20 bootstraps for each of the interactions, in normal and tumor cases, and the sensitivity index using $Fdiv-Chi2$, the computed sensitivity indices were sorted in increasing order of value and are shown in figure 7 and 8, respectively. The $y-axis$ represents the sensitivity indices and the $x-axis$ represents the values of the different factors involved under investigation (here 2nd order combinations). Since the indices have been sorted based on the ranking score, the combinations are now ordered according to the strength of their significance. The combinations with low significance or sensitivity index is placed towards the left most end and vice versa. Visualizations for other indices were also generated and are relegated to the appendix. Since various indices have been tested, it would be important to see how good the rankings tally across the different sensitivity methods. To tally the rankings visually, the top and the bottom 10 combinations were tabulated in tables 1, 2, 5, 3 and 4. The top and the bottom 10 rankings in figures 7 and 8 can be found in the second columns of tables 1 and 2.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

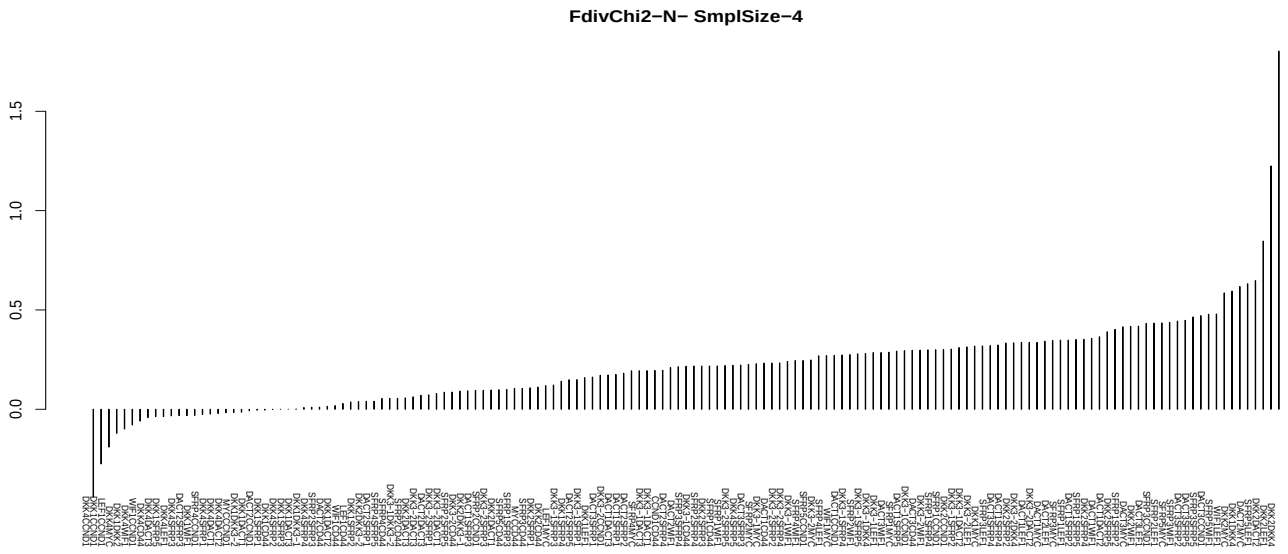


Fig. 7 FdivChi2; Training sample size - 4; Test sample size - 16; Case - Normal

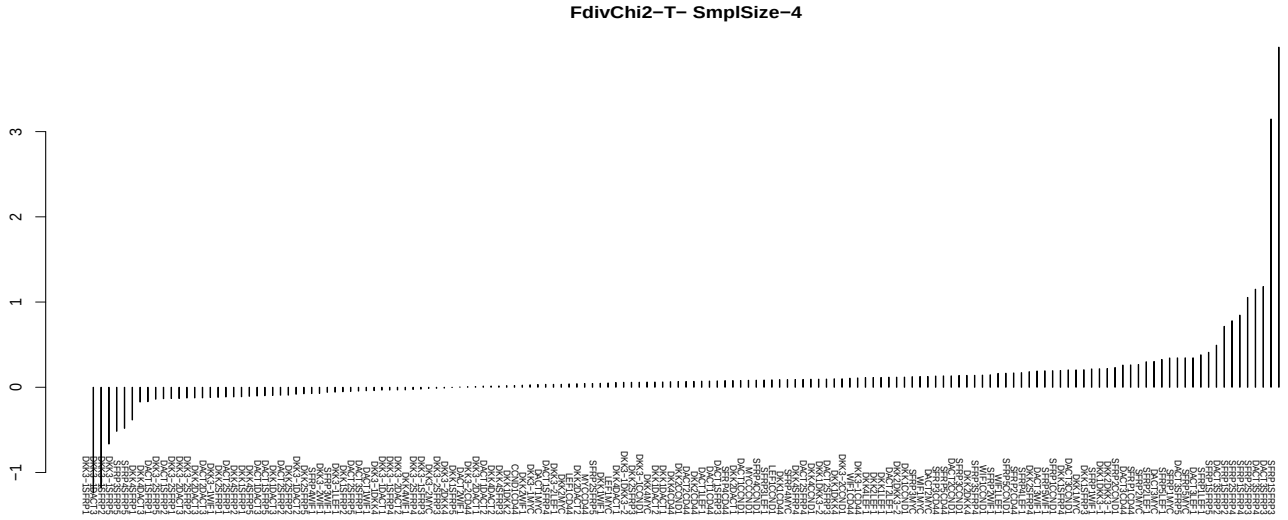


Fig. 8 FdivChi2; Training sample size - 4; Test sample size - 16; Case - Tumor

6.1 2nd Order interactions

In tumor cases (table 1), for most of the sensitvity indices we find that *SFRP*1-*SFRP*3 has the highest priority in the this data. This is followed by various others interactions like *DACT*3-*SFRP*3 and *DACT*3-*SFRP*4. For other kinds of the interactions we find varying kinds of rankings. It is important to note that the rankings are relative but they do give a glimpse of the importance of the interactions in a vast combinatorial space. These rankings are of importance in a particular phenomena of investigation. A particular colour across the columns, more specifically the sensitivity methods, represent the relative positioning of a particular interaction. The different colours in a row of a column do not

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Tumor Case - Density based methods

Rank	Fdiv				HSIC	
	High priority interactions					
	chi2	hellinger	kl	tv	rbf	laplace
10	SFRP1LEF1	SFRP1MYC	SFRP5LEF1	SFRP5LEF1	SFRP2SFRP4	SFRP1MYC
9	SFRP1SFRP5	DACT3LEF1	DACT3SFRP5	DACT3SFRP5	SFRP2SFRP3	SFRP5LEF1
8	DACT3SFRP2	SFRP1SFRP2	SFRP5MYC	SFRP5MYC	DACT3LEF1	DACT3MYC
7	SFRP1SFRP2	SFRP1LEF1	SFRP1MYC	SFRP1MYC	SFRP1LEF1	SFRP2LEF1
6	SFRP2SFRP4	SFRP2SFRP4	DACT3SFRP4	DACT3SFRP4	SFRP1SFRP4	DACT3LEF1
5	SFRP1SFRP4	SFRP2SFRP3	DACT3MYC	DACT3MYC	DACT3SFRP4	SFRP1WIF1
4	SFRP2SFRP3	SFRP1SFRP4	DACT3LEF1	DACT3LEF1	SFRP1SFRP2	DACT3WIF1
3	DACT3SFRP4	DACT3SFRP4	SFRP1LEF1	SFRP1LEF1	DACT3SFRP2	DACT3SFRP3
2	DACT3SFRP3	DACT3SFRP3	DACT3SFRP3	DACT3SFRP3	DACT3SFRP3	SFRP1LEF1
1	SFRP1SFRP3	SFRP1SFRP3	SFRP1SFRP3	SFRP1SFRP3	SFRP1SFRP3	SFRP1SFRP3
	Low priority interactions					
153	DKK3-1SFRP1	DKK3-1SFRP1	DKK3-1SFRP1	DKK3-1SFRP1	DKK4DACT3	DKK4DACT3
152	DKK3-1DACT3	DKK3-1DACT3	DKK3-1DACT3	DKK3-1DACT3	DACT2DACT3	DKK3-1SFRP5
151	DKK3-1SFRP2	DKK3-1SFRP2	DKK3-1SFRP2	DKK3-1SFRP2	DKK3-1SFRP5	DKK4SFRP1
150	DKK3-1SFRP5	DKK3-1SFRP5	DKK3-1SFRP5	DKK3-1SFRP5	DKK1DACT3	DACT2DACT3
149	SFRP3SFRP5	SFRP3SFRP5	SFRP3SFRP5	SFRP3SFRP5	DKK1SFRP1	DKK3-1SFRP2
148	SFRP4SFRP5	SFRP4SFRP5	DACT1SFRP1	DACT1SFRP1	DACT1SFRP1	DACT2SFRP1
147	DKK4SFRP1	DACT1SFRP1	SFRP4SFRP5	SFRP4SFRP5	DKK3-2DACT3	DACT2SFRP1
146	DKK4DACT3	DACT1SFRP2	DACT1DACT3	DACT1DACT3	DKK2SFRP1	DKK4SFRP2
145	DACT1SFRP1	DACT1DACT3	DACT1SFRP5	DACT1SFRP5	DACT2SFRP1	DACT1SFRP1
144	DKK3-2SFRP5	DKK4SFRP1	DKK2DACT3	DKK2DACT3	DKK4SFRP1	DKK2SFRP1

Table 1 Ranking of second order interactions in Tumor case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.

mean any thing. What has been found is that $DACT3$ is implicated in the Wnt β-catenin signaling in colorectal cancer through epigenetic modifications and is a histone modification therapeutic target Jiang *et al.*[16]. As a confirmatory result, the pipeline indicates the priority that $DACT3$ gets along with the members of the $SFRP$ family in the tumor. This high priority also indicates the efficacy of the designed pipeline. Combinations of the members of the $SFRP$ family also show at the high priority list. More specifically $SFRP1$-$SFRP3$, $SFRP2$-$SFRP3$ and $SFRP1$-$SFRP2$.

On the contrary, when we look at the set of low priority interactions, there is a range of combinations involving the family of DKK and members of $DACT$, $SFRP$. Surpurisingly, specific interactions of within family members also show up at this stage. This is specially seen for $DACT1$-$DACT3$, $DACT2$-$DACT3$, $SFRP3$-$SFRP5$ and $SFRP4$-$SFRP5$. These indicate the non-effectiveness of the 2nd order interactions among a family of protein in the tumor case. Based on these rankings it might be possible to narrow down the search for important interactions in the vast combinatorial search space. We also generated the similar rankings for the interactions between the same set of factors in normal case. These are represented in table 2.

In comparison to the tables 1 and 2, which take into account the density based sensitivity methods, we also generated the rankings in tumor and normal cases using variance based sensitivity indices (table 5 in appendix). What we find is that there are interactions that get high priority in both the tumor and normal case. For example, the cases of $DKK3-1$-$WIF1$ and $DKK3-2$-$LEF1$, show up with high priority in both the normal and tumor case. This indicates the inability of the variance based sensitivity indices to capture crucial influence of the participating factors or combination of the factors, properly. Clearly, what we find is that the density based methods show superior selection and capture of improtant nonlinear biological information to provide proper indices that can be ranked later on.

6.2 3rd Order Interactions

We now turn our attention to the 3rd order combinations which open a deeper and a greater space of investigation as compared to the space of the second order combinations. Tables 3 and 4 show the rankings for the 3rd order interactions in tumor and normal cases, respectively. A major finding is that $DACT3$ combination with members of the family of $SFRP$ show very high ranking in tumor cases. Note that we could only generate the rankings for the different kernels in the HSIC method as other density based

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Normal Case - Density based methods

Rank	Fdiv				HSIC	
	High priority interactions					
	chi2	hellinger	kl	tv	rbf	laplace
10	DACT3CCND1	DKK2WIF1	DKK2DKK4	DACT3MYC	DKK2SFRP5	DACT3SFRP1
9	SFRP5WIF1	DKK2SFRP5	DKK2LEF1	DKK2WIF1	DKK2WIF1	DACT3CCND1
8	WIF1LEF1	DACT3SFRP2	DKK2DACT2	DACT3SFRP2	SFRP5LEF1	DKK2DACT2
7	DKK2MYC	SFRP5LEF1	DKK2MYC	SFRP5LEF1	DACT3SFRP5	DKK4SFRP5
6	DKK1DKK4	DACT3CCND1	DACT3LEF1	DACT3CCND1	DACT3CCND1	SFRP5CCND1
5	DACT2MYC	DKK1MYC	DACT3CCND1	DACT3LEF1	SFRP5CCND1	DACT3SFRP5
4	SFRP5LEF1	DKK2MYC	SFRP5LEF1	DKK2MYC	DKK2LEF1	DKK2SFRP5
3	DKK2DACT2	DKK2DACT2	DACT3SFRP2	DKK2DACT2	DKK1MYC	DKK3-1CD44
2	DKK2LEF1	DKK2LEF1	DKK2WIF1	DKK2LEF1	DKK2DACT2	DKK2DKK4
1	DKK2DKK4	DKK2DKK4	DACT3MYC	DKK2DKK4	DKK2DKK4	DKK1MYC
	Low priority interactions					
153	DKK4CCND1	DKK4LEF1	DKK4CD44	DKK4CD44	DKK1DKK3-2	DKK4LEF1
152	DKK1CCND1	DKK1DKK3-2	DKK4DACT3	DKK4DACT3	DKK4SFRP3	DKK1DKK3-2
151	LEF1CCND1	DKK4SFRP1	DKK4MYC	DKK4MYC	DKK4SFRP1	DKK1WIF1
150	DKK4MYC	DKK4SFRP3	DKK4SFRP3	DKK4SFRP3	DKK4LEF1	DKK4SFRP1
149	DKK1DKK2	DKK4DACT1	DKK1DKK2	DKK1DKK2	DKK1SFRP3	DKK3-2CD44
148	DKK4WIF1	DKK4DACT3	DKK4DACT1	DKK4DACT1	DACT2SFRP3	DKK3-2SFRP3
147	WIF1CCND1	DKK4CD44	DKK4SFRP1	DKK4SFRP1	DKK4DACT1	DKK3-2DACT3
146	DKK4CD44	DKK1SFRP1	LEF1CCND1	LEF1CCND1	DKK1WIF1	DKK3-2DACT1
145	DKK4DACT3	DKK1SFRP3	DKK1DKK3-2	DKK1DKK3-2	DKK1SFRP1	DKK3-2SFRP4
144	DKK1SFRP5	DACT2SFRP3	DACT2SFRP3	DACT2SFRP3	WIF1CD44	DKK3-2SFRP1

Table 2 Ranking of second order interactions in Normal case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.

methods were computationally intensive and costly in time. On the other hand the combinations of the family of *DKK* with other factors showed very low rankings in the tumor case, thus indicating their non participation or down regulated participation in the tumor cases. The most beautiful aspect of these rankings is that one can specifically see which members of a family are having potent role in the tumor and which ones are not. For example, *DACT*3 plays very important role along with various members of the *SFRP* family but the combination *DACT*3-*SFRP*3-*SFRP*5 shows very low participation in tumor case. This potential of the search engine to specifically rank the different combinations gives the biologists deep insights into the interactions of member of a single family of protiens also. A similar ranking profile was generated for the normal case.

CODE AVAILABILITY Code has been made available on Google drive at `https://drive.google.com/folderview?id=0B7Kkv8wlhPU-V1Fkd1dMSTd5ak0&usp=sharing`

7 Conclusions

A workflow has been presented that can prioritize the entire range of interactions among the constituent or subgroup of intra/extracellular factors affecting the pathway by using powerful algorithm of support vector ranking on interactions that have sensitivity indices of the involved factors as features. These sensitivity indices compute the influences of the factors on the pathway and represent nonlinear biological relations among the factors that are captured using kernel methods. SVM ranking then scores the testing data which can be sorted to find the highly prioritized interactions that need further investigation. Using this efficient workflow, it is possible to analyse any combination of involved factors in a signaling pathway. Here, we show confirmatory results for members of family of *DACT*, *SFRP* (*DKK*) to have highly prioritized role in tumor (normal) cases, in a single experimental setup. These preliminary results show the efficacy of the pipeline in providing a ranked list of interactions which the biologists can narrow down to in a vast serach space of combinations, thus cutting down the cost of terms of time, investment and energy. Additionally, the pipeline also generates unknown/untested biological hypothesis reagrding the combinations involved in the pathway in both tumor and normal cases. This opens up the potential for the biologists to work on recommendations from this pipeline in a vast combinatorial forest of combinations, when many a times, it is not possible to know where to start the search.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	Tumor Case		
Ranking	HSIC		
	rbf	laplace	linear
	High priority interactions		
10	SFRP1SFRP2LEF1	DACT3SFRP1WIF1	DKK3-1DACT1SFRP4
9	SFRP1SFRP2SFRP4	DACT3SFRP1MYC	DKK3-1SFRP2SFRP3
8	DACT3SFRP2LEF1	DACT3SFRP5LEF1	DACT3SFRP1SFRP4
7	SFRP1SFRP2SFRP3	DACT3SFRP2LEF1	DKK3-1SFRP1SFRP3
6	DACT3SFRP2SFRP4	DACT3SFRP1SFRP2	DKK3-1DACT3SFRP3
5	DACT3SFRP1LEF1	DACT3SFRP2SFRP4	DACT3SFRP2SFRP4
4	DACT3SFRP2SFRP3	SFRP1SFRP2SFRP4	DKK3-1SFRP2SFRP4
3	DACT3SFRP1SFRP4	DACT3SFRP1SFRP4	DACT3SFRP1SFRP2
2	DACT3SFRP1SFRP2	DACT3SFRP1LEF1	DKK3-1SFRP1SFRP4
1	DACT3SFRP1SFRP3	DACT3SFRP1SFRP3	DKK3-1DACT3SFRP4
	Low priority interactions		
816	DKK1DKK3-1SFRP2	SFRP1SFRP3SFRP5	DKK1DKK4WIF1
815	DACT3SFRP3SFRP5	DACT3SFRP4SFRP5	DKK1DKK2DACT3
814	SFRP1SFRP3SFRP5	SFRP2SFRP4SFRP5	DKK2DACT2WIF1
813	DKK1DKK4SFRP1	SFRP1SFRP4SFRP5	DKK1DKK4SFRP1
812	SFRP2SFRP3SFRP5	SFRP2SFRP3SFRP5	DKK2DKK4SFRP1
811	DKK1DKK3-1SFRP5	DKK1DKK3-1SFRP5	DKK1DKK4LEF1
810	DKK2DKK3-1SFRP5	DKK1DKK3-1SFRP2	DKK1DACT2LEF1
809	DKK1DKK4DACT3	DACT3SFRP3SFRP5	DKK1DKK4SFRP2
808	DKK2DKK4SFRP1	DKK1DKK3-2SFRP1	DKK1DKK2WIF1
807	SFRP1SFRP4SFRP5	DKK1DACT1SFRP2	DACT3SFRP2WIF1

Table 3 Ranking of third order interactions in Tumor case using variance based sensitivity indices. Here 1 has high priority and 816 has low priority.

8 Acknowledgement

The author thanks Mr. Prabhat Sinha and Mrs. Rita Sinha for financially supporting the project.

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 S. Sinha, *Integr. Biol.*, 2014, **6**, 1034–1048.
18 A. Gregorieff and H. Clevers, *Genes & development*, 2005, **19**, 877–890.
19 T. Joachims, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 2006, pp. 217–226.
20 I. M. Sobol', *Matematicheskoe Modelirovanie*, 1990, **2**, 112–118.
21 B. Iooss and P. Lemaître, *arXiv preprint arXiv:1404.2405*, 2014.
22 M. D. Morris, *Technometrics*, 1991, **33**, 161–174.
23 H. Moon, A. M. Dean and T. J. Santner, *Technometrics*, 2012, **54**, 376–387.
24 A. Dean and S. Lewis, *Screening: methods for experimentation in industry, drug discovery, and genetics*, Springer Science & Business Media, 2006.
25 T. H. Andres and W. C. Hajas, 1993.
26 B. Bettonvil and J. P. Kleijnen, *European Journal of Operational Research*, 1997, **96**, 180–194.
27 S. C. Cotter, *Biometrika*, 1979, **66**, 317–320.
28 R. Christensen, *Linear models for multivariate, time series, and spatial data*, Springer Science & Business Media, 1991.
29 A. Saltelli, K. Chan and E. Scott, *Willey, New York*, 2000.
30 J. C. Helton and F. J. Davis, *Reliability Engineering & System Safety*, 2003, **81**, 23–69.
31 M. D. McKay, R. J. Beckman and W. J. Conover, *Technometrics*, 1979, **21**, 239–245.
32 T. Homma and A. Saltelli, *Reliability Engineering & System Safety*, 1996, **52**, 1–17.
33 I. M. Sobol, *Mathematics and computers in simulation*, 2001, **55**, 271–280.
34 A. Saltelli, *Computer Physics Communications*, 2002, **145**, 280–297.
35 A. Saltelli, M. Ratto, S. Tarantola and F. Campolongo, *Chemical reviews*, 2005,

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

	Normal Case		
Ranking	HSIC		
	rbf	laplace	linear
	High priority interactions		
10	DKK2WIF1MYC	DACT3SFRP5MYC	DKK4SFRP4CCND1
9	DKK2DKK4WIF1	DKK3-1DACT3CCND1	DKK1DKK4MYC
8	DKK1LEF1MYC	DKK2DACT2LEF1	SFRP5WIF1CCND1
7	DKK3-1LEF1CCND1	DACT3LEF1CCND1	WIF1LEF1CCND1
6	DKK2DKK4MYC	DACT3SFRP2SFRP5	DKK4SFRP2CCND1
5	DKK2DACT2LEF1	DACT3WIF1CCND1	DKK1SFRP2CCND1
4	DACT3WIF1CCND1	DACT3SFRP5CCND1	DACT2LEF1CCND1
3	DACT3LEF1CCND1	DACT3SFRP2CCND1	DKK4LEF1CCND1
2	DKK2DKK4DACT2	DKK2DKK4DACT2	SFRP2LEF1CCND1
1	DKK2DKK4LEF1	DKK2DKK4LEF1	DKK1LEF1CCND1
	Low priority interactions		
816	DKK1DKK2SFRP5	DKK4SFRP1LEF1	DKK4DACT2LEF1
815	DKK1DKK2DKK4	DKK1DKK3-2SFRP5	DKK4WIF1LEF1
814	DKK4SFRP1LEF1	DKK1DKK3-2SFRP2	DKK1DKK2LEF1
813	DKK4WIF1LEF1	DKK1DKK3-2SFRP4	DKK1DKK2DKK4
812	DKK4SFRP3CCND1	DKK1DKK3-2DACT2	DACT2WIF1LEF1
811	DKK4DACT3SFRP1	DKK1DKK3-2CCND1	DKK1DKK2WIF1
810	DKK4SFRP1WIF1	DKK4SFRP1MYC	DACT2DACT3SFRP5
809	DKK4SFRP1SFRP2	DKK4DACT3SFRP1	DKK1DKK2DACT2
808	DKK4SFRP3LEF1	DKK1SFRP3SFRP4	DKK4DACT3SFRP5
807	DKK1SFRP3CCND1	DKK4SFRP1WIF1	DKK2DACT3CCND1

Table 4 Ranking of third order interactions in Normal case using variance based sensitivity indices. Here 1 has high priority and 816 has low priority.

105, 2811–2828.

36 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.

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

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

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

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

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

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

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

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

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

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

47 A. Saltelli and P. Annoni, *Environmental Modelling & Software*, 2010, **25**, 1508–1517.

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 H. Li, *IEICE TRANS. INF. & SYST.*, 2011, **E94-D**, year.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

66 R. Kolde, S. Laur, P. Adler and J. Vilo, *Bioinformatics*, 2012, **28**, 573–580.

67 M. Blondel, A. Onogi, H. Iwata and N. Ueda, *PloS one*, 2015, **10**, e0128570.

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

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

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

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

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

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

74 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.

75 S. Sinha, *MS Thesis*, 2004.

76 I. K.-U. Bletzinger, *Basic Mathematics*, 2002.

77 C. J. Burges, *Data mining and knowledge discovery*, 1998, **2**, 121–167.

78 N. Cristianini and J. Shawe-Taylor, *An introduction to support vector machines and other kernel-based learning methods*, Cambridge university press, 2000.

79 B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT press, 2002.

80 V. N. Vapnik and V. Vapnik, *Statistical learning theory*, Wiley New York, 1998, vol. 1.

81 B. E. Boser, I. M. Guyon and V. N. Vapnik, Proceedings of the fifth annual workshop on Computational learning theory, 1992, pp. 144–152.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Appendix

9 Sensitivity indices

9.1 Variance based indices

The variance based indices as proposed by Sobol'[20] 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 $u = f(x)$, $x = (x_1, x_2, \ldots, x_n)$, is defined in a unit n-dimensional cube $\mathcal{K}^n$ 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[32] and ? .

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

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

Then the representation of $f(x)$ using equation 2 in the form -

$$\begin{aligned} f(x) &= f_0 + \widehat{\Sigma} f_{i_1,i_2,\ldots,i_s} \\ &= f_0 + \Sigma_i f_i(x_i) + \Sigma_{i<j} f_{i,j}(x_i, x_j) + \ldots + f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n) \end{aligned} \quad (3)$$

is called ANOVA-decomposition from Archer *et al.*[45] or expansion into summands of different dimensions, if f_0 is a constant and integrals of the summands $f_{i_1,i_2,\ldots,i_s}$ with respect to their own variables are zero, i.e,

$$f_0 = \int_{\mathcal{K}^n} f(x) dx \quad (4)$$

$$\int_0^1 f_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) dx_{i_k} = 0, 1 \le k \le s \quad (5)$$

It follows from equation 4 that all summands on the right hand side are orthogonal, i.e if at least one of the indices in $i_1, i_2, \ldots, i_s$ and $j_1, j_2, \ldots, j_l$ is not repeated i.e

$$\int_0^1 f_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) f_{j_1,j_2,\ldots,j_l}(x_{j_1}, x_{j_2}, \ldots, x_{j_s}) dx = 0 \quad (6)$$

Sobol'[20] proves a theorem stating that there is an existence of a unique expansion of equation 4 for any $f(x)$ integrable in $\mathcal{K}^n$. In brief, this implies that for each of the indices as well as a group of indices, integrating equation 4 yields the following -

$$\int_0^1 \ldots \int_0^1 f(x) dx/dx_i = f_0 + f_i(x_i) \quad (7)$$

$$\int_0^1 \ldots \int_0^1 f(x) dx/dx_i dx_j = f_0 + f_i(x_i) + f_j(x_j) + f_{i,j}(x_i, x_j) \quad (8)$$

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 $f_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s})$ is reduced to an integral in the cube $\mathcal{K}^n$. The last summand $f_{1,2,\ldots,n}(x_1, x_2, \ldots, x_n)$ is $f(x) - f_0$ from equation 4. Homma and Saltelli[32] stresses that use of Sobol sensitivity indices does not require evaluation of any $f_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s})$ 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 $f(x) \in \mathcal{L}_2$, then all of $f_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) \in \mathcal{L}_2$. Then the following constants

$$\int_{\mathcal{K}^n} f^2(x) dx - f_0^2 = D \quad (9)$$

$$\int_0^1 \ldots \int_0^1 f^2_{i_1,i_2,\ldots,i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) dx_{i_1} dx_{i_2} \ldots dx_{i_s} = D_{i_1,i_2,\ldots,i_s} \quad (10)$$

are termed as variances. Squaring equation 4, integrating over $\mathcal{K}^n$ and using the orthogonality property in equation 6, D evaluates to -

$$D = \widehat{\Sigma} D_{i_1,i_2,\ldots,i_s} \quad (11)$$

Then the global sensitivity estimates is defined as -

$$S_{i_1,i_2,\ldots,i_s} = \frac{D_{i_1,i_2,\ldots,i_s}}{D} \quad (12)$$

It follows from equations 11 and 12 that

$$\widehat{\Sigma} S_{i_1,i_2,\ldots,i_s} = 1 \quad (13)$$

Clearly, all sensitivity indices are non-negative, i.e an index $S_{i_1,i_2,\ldots,i_s} = 0$ if and only if $f_{i_1,i_2,\ldots,i_s} \equiv 0$. The true potential of Sobol indices is observed when variables $x_1, x_2, \ldots, x_n$ are divided into m different groups with $y_1, y_2, \ldots, y_m$ such that $m < n$. Then $f(x) \equiv f(y_1, y_2, \ldots, y_m)$. 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 ? . Variations and improvements over Sobol indices

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Variance Based Method

	Tumor Case			Normal Case		
	High priority interactions					
Rank	2007	jansen	martinez	2007	jansen	martinez
10	DACT3WIF1	SFRP3LEF1	SFRP3LEF1	DKK3-2CD44	SFRP3LEF1	DKK3-1DACT2
9	SFRP1LEF1	DKK2LEF1	DKK3-1MYC	SFRP2LEF1	DKK2LEF1	SFRP3WIF1
8	DACT3LEF1	DACT1LEF1	DKK3-1DACT2	DKK3-2WIF1	DACT1LEF1	SFRP3LEF1
7	SFRP4LEF1	DKK3-1WIF1	DKK2LEF1	DKK3-2DACT2	DKK2DKK4	DACT1LEF1
6	DKK3-1WIF1	DKK2DKK4	DACT1LEF1	WIF1MYC	SFRP4LEF1	DKK2DKK4
5	DACT1LEF1	DKK3-2DKK4	SFRP5LEF1	DKK4MYC	DKK3-2DKK4	SFRP5LEF1
4	DKK3-2LEF1	SFRP4LEF1	DKK3-1LEF1	DKK3-2DKK4	DKK3-1DKK4	DKK2LEF1
3	SFRP4WIF1	DKK3-2LEF1	DKK3-1WIF1	DKK3-2MYC	DKK3-1WIF1	DKK3-1LEF1
2	DACT2LEF1	DKK3-1DKK4	DKK2DKK4	LEF1MYC	DKK3-2LEF1	DKK3-1WIF1
1	SFRP5LEF1	SFRP5LEF1	DKK3-1DKK4	DKK3-2LEF1	SFRP5LEF1	DKK3-1DKK4
	Low priority interactions					
153	DKK4MYC	LEF1CCND1	DKK4SFRP3	DKK1DKK4	LEF1CCND1	DKK4SFRP3
152	DKK4DACT2	DKK4MYC	DKK1DKK3-1	DKK1LEF1	DKK1SFRP5	DKK1DKK3-1
151	DKK1MYC	DKK1SFRP5	DKK1SFRP3	DKK4LEF1	DKK4MYC	DKK4DACT1
150	LEF1MYC	DKK4SFRP4	DKK4DACT1	DKK1WIF1	DKK4SFRP3	DKK1SFRP3
149	DKK4CCND1	DKK4SFRP3	DKK4SFRP4	WIF1CD44	DKK4SFRP4	DKK4SFRP4
148	DKK1DACT2	DKK4SFRP5	DKK4CD44	DKK1DKK3-2	DKK1DKK3-2	LEF1CCND1
147	DKK4SFRP5	DKK4DACT1	DKK1DKK2	DACT2LEF1	DKK4SFRP5	DKK4CD44
146	DACT2MYC	DKK4CCND1	DKK1DACT1	DKK4SFRP2	DKK1CCND1	DKK1DKK2
145	LEF1CCND1	DKK1DKK3-2	DKK4MYC	DKK1SFRP2	DKK4DACT1	DKK1DACT1
144	DKK1CCND1	DKK1CCND1	LEF1CCND1	DKK4CD44	DKK4CCND1	DKK1SFRP5

Table 5 Ranking of second order interactions in Tumor & Normal case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.

have already been stated in section 3.

9.2 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, $f : \mathcal{R}^n \rightarrow \mathcal{R}$ ($u = f(x)$) is assumed to be continuous. It is also assumed that X_k has a known distribution and are independent. Baucells and Borgonovo[68] 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 -

$$S_{X_k} = \mathcal{E}(d(U, U|X_k)) \tag{14}$$

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 $\mathcal{R}$ is formulated as -

$$d_F(U||U|X_k) = \int_{\mathcal{R}} F\left(\frac{p_U(u)}{p_{U|X_k}(u)}\right) p_{U|X_k}(u)du \tag{15}$$

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 15 in

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

equation 14, gives the following sensitivity index -

$$\begin{aligned} S^F_{X_k} &= \int_{\mathcal{R}} d_F(U||U|X_k)p_{X_k}(x)dx \\ &= \int_{\mathcal{R}}\int_{\mathcal{R}} F(\frac{p_U(u)}{p_{U|X_k}(u)})p_{U|X_k}(u)p_{X_k}(x)dxdu \\ &= \int_{\mathcal{R}^2} F(\frac{p_U(u)p_{X_k}(x)}{p_{U|X_k}(u)p_{X_k}(x)})p_{U|X_k}(u)p_{X_k}(x)dxdu \\ &= \int_{\mathcal{R}^2} F(\frac{p_U(u)p_{X_k}(x)}{p_{X_k,U}(x,u)})p_{X_k,U}(x,u)dxdu \end{aligned} \tag{16}$$

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.*[69]). This has an advantage over Sobol sensitivity indices which are invariant under linear transformations.

By substituting the different formulations of F in equation 16, 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 16 changes to estimation of ratio between the joint density of (X_k, U) and the marginals, i.e -

$$S^F_{X_k} = \int_{\mathcal{R}^2} F(\frac{1}{r(x,u)})p_{X_k,U}(x,u)dxdu = \mathcal{E}_{(X_k,U)}F(\frac{1}{r(X_k,U)}) \tag{17}$$

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 $X \in \mathcal{R}^p$ and $Y \in \mathcal{R}^q$, 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.*[70]) 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[71] - "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 $\mathcal{H}$ with elements $r, s \in \mathcal{H}$ has dot product $\langle r, s\rangle_{\mathcal{H}}$ and $r \cdot s$. When $\mathcal{H}$ is a vector space over a field $\mathcal{F}$, then the dot product is an element in $\mathcal{F}$. The product $\langle r, s\rangle_{\mathcal{H}}$ follows the below mentioned properties when $r, s, t \in \mathcal{H}$ and for all $a \in \mathcal{F}$ -

- 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 $\mathcal{V}$ with a dot product $\langle\cdot,\cdot\rangle$, the norm on $\mathcal{V}$ defined by $||r||_{\mathcal{V}} = \sqrt{(\langle r, r\rangle)}$ 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 $\mathcal{H}$ and requires all Dirac evaluation functionals in $\mathcal{H}$ are bounded and continuous (on implies the other). Assuming $\mathcal{H}$ is the $\mathcal{L}_2$ space of functions from X to $\mathcal{R}$ for some measurable X. For an element $x \in X$, a Dirac evaluation functional at x is a functional $\delta_x \in \mathcal{H}$ 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 $\mathcal{R}$. Then δ_x is simply a function which maps g to the value g has at x. Thus, δ_x is a function from $(\mathcal{R}^n \mapsto \mathcal{R})$ into $\mathcal{R}$.

The requirement of Dirac evaluation functions basically means (via the Riesz[72] representation theorem) if ϕ is a bounded linear functional (conditions satisfied by the Dirac evaluation functionals) on a Hilbert space $\mathcal{H}$, then there is a unique vector ℓ in $\mathcal{H}$ such that $\phi g = \langle g, \ell\rangle_{\mathcal{H}}$ for all $\ell \in \mathcal{H}$. Translating this theorem back into Dirac evaluation functionals, for each δ_x there is a unique vector k_x in $\mathcal{H}$ such that $\delta_x g = g(x) = \langle g, k_x\rangle_{\mathcal{H}}$. The reproducing kernel K for $\mathcal{H}$ 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 $\langle g, K(x, x')\rangle_{\mathcal{H}} = g(x')$. Furthermore, k_x is defined to be a function $y \mapsto K(x, y)$ and thus the reproducibility is given by $\langle K(x,\cdot), K(y,\cdot)\rangle_{\mathcal{H}} = K(x, y)$.

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 $\mathbf{x}_i \in \mathcal{R}^n$, be the set of n feature values with corresponding category of the example label (y_i) in data set $\mathcal{D}$. Then the data points can be mapped to a higher dimensional space $\mathcal{H}$ by the transformation ϕ:

$$\phi : \mathbf{x}_i \in \mathcal{R}^n \mapsto \phi(\mathbf{x}_i) \in \mathcal{H} \tag{18}$$

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This $\mathcal{H}$ 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 $\mathcal{H}$. 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 $\mathcal{H}$. Thus, the idea behind equation(18) is that even if the data points are nonlinearly clustered in space $\mathcal{R}^n$, the transformation spreads the data points into $\mathcal{H}$, such that they can be linearly separated in its range in $\mathcal{H}$.

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 $\mathcal{R}^n$, under the assumption that its evaluation is equivalent to a dot product in the higher dimensional space. This is given as:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = < \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) > \tag{19}$$

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 $\mathcal{R}^n$ to $\mathcal{H}$. 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:

$$\begin{aligned} f(\mathbf{z}) &= < \mathbf{w}, \phi(\mathbf{z}) > + b \\ &= < \sum_{i=1}^{N} \alpha_i \times y_i \times \phi(\mathbf{x}_i), \phi(\mathbf{z}) > + b \\ &= \sum_{i=1}^{N} \alpha_i \times y_i \times < \phi(\mathbf{x}_i), \phi(\mathbf{z}) > + b \\ &= \sum_{i=1}^{N} \alpha_i \times y_i \times \kappa(\mathbf{x}_i, \mathbf{z}) + b \end{aligned} \tag{20}$$

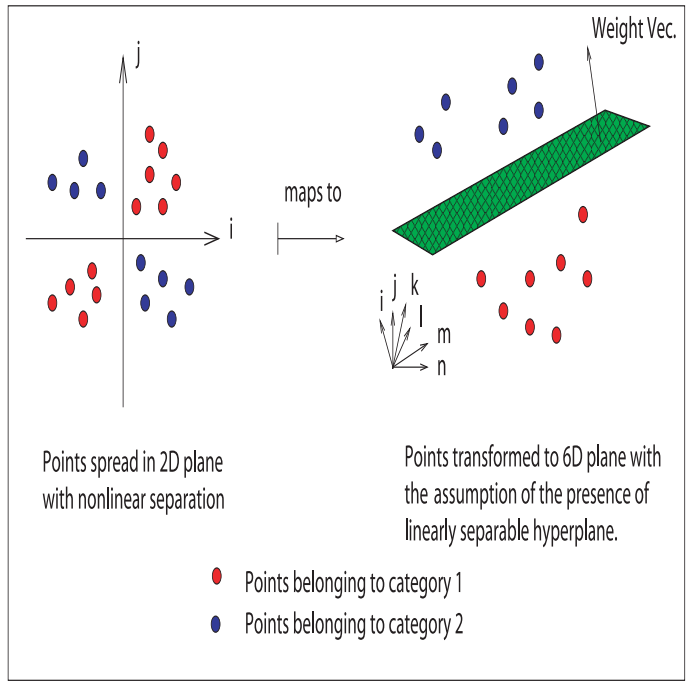


Fig. 9 A geometrical interpretation of mapping nonlinearly separable data into higher dimensional space where it is assumed to be linearly separable, subject to the holding of dot product.

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[73] for description of these properties in detail.

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 $X \in \mathcal{X}$ have a distribution P_X and consider a RKHS $\mathcal{A}$ of functions $\mathcal{X} \rightarrow \mathcal{R}$ with kernel $k_{\mathcal{X}}$ and dot product $\langle \cdot, \cdot \rangle_{\mathcal{A}}$. Similarly, Let $U \in \mathcal{Y}$ have a distribution P_Y and consider a RKHS $\mathcal{B}$ of functions $\mathcal{U} \rightarrow \mathcal{R}$ with kernel $k_{\mathcal{B}}$ and dot product $\langle \cdot, \cdot \rangle_{\mathcal{B}}$. Then the cross-covariance operator $C_{X,U}$ associated with the joint distribution P_{XU} of (X, U) is the linear operator $\mathcal{B} \rightarrow \mathcal{A}$ defined for every $a \in \mathcal{A}$ and $b \in \mathcal{B}$ as -

$$\langle a, C_{XU} b \rangle_{\mathcal{A}} = \mathcal{E}_{XU}[a(X), b(U)] - \mathcal{E}_X a(X) \mathcal{E}_U b(U) \tag{21}$$

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

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

is given by -

$$||C||_{HS}^2 = \Sigma_{k,l}\langle a_k, Cb_l\rangle_{\mathcal{A}} \tag{22}$$

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

$$HSIC(X,U)_{\mathcal{A},\mathcal{B}} = \begin{cases} ||C_{XU}||_{HS}^2 = \\ \mathcal{E}_{X,X',U,U'}k_{\mathcal{X}}(X,X')k_{\mathcal{U}}(U,U')+ \\ \mathcal{E}_{X,X'}k_{\mathcal{X}}(X,X')\mathcal{E}_{U,U'}k_{\mathcal{U}}(U,U')- \\ 2\mathcal{E}_{X,Y}[\mathcal{E}_{X'}k_{\mathcal{X}}(X,X')\mathcal{E}_{U'}k_{\mathcal{U}}(U,U')] \end{cases} \tag{23}$$

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

$$HSIC_n(X,U)_{\mathcal{A},\mathcal{B}} = \frac{1}{n^2}Tr(K_{\mathcal{X}}HK_{\mathcal{U}}H) \tag{24}$$

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

$$HSIC(X,U)_{\mathcal{A},\mathcal{B}} = \begin{cases} \frac{1}{n^2}\Sigma_{i,j=1}^n k_{\mathcal{X}}(X_i,X_j)k_{\mathcal{U}}(U_i,U_j) \\ +\frac{1}{n^2}\Sigma_{i,j=1}^n k_{\mathcal{X}}(X_i,X_j)\frac{1}{n^2}\Sigma_{i,j=1}^n k_{\mathcal{U}}(U_i,U_j) \\ -\frac{2}{n}\Sigma_{i=1}^n[\frac{1}{n}\Sigma_{j=1}^n k_{\mathcal{X}}(X_i,X_j)\frac{1}{n}\Sigma_{j=1}^n k_{\mathcal{U}}(U_i,U_j)] \end{cases} \tag{25}$$

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

$$S_{X_k}^{HSIC_{\mathcal{A},\mathcal{B}}} = R(X_k, U)_{\mathcal{A},\mathcal{B}} \tag{26}$$

were the kernel based distance correlation is given by -

$$R^2(X,U)_{\mathcal{A},\mathcal{B}} = \frac{HSIC(X,U)_{\mathcal{A},\mathcal{B}}}{\sqrt{(}HSIC(X,X)_{\mathcal{A},\mathcal{A}}HSIC(U,U)_{\mathcal{B},\mathcal{B}})} \tag{27}$$

were kernels inducing $\mathcal{A}$ and $\mathcal{B}$ are to be chosen within a universal class of kernels. Similar multivariate formulation for equation 24 are possible.

9.3 Choice of sensitivity indices

The SENSITIVITY PACKAGE (Faivre *et al.*[74] and Iooss and Lemaître[21]) 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) × n model evaluations. These are called the Jansen estimators. (Jansen[48] and Saltelli *et al.*[40])

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) ×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[34])

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) × n model evaluations. These are called the Mauntz estimators. (Saltelli and Annoni[47])

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) × 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) × n where N is the number of indices to estimate. (Sobol'[20])

10 Optimization and Support Vector Machines

Aspects of SVMs from Sinha[75] are reproduced for completeness.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

10.1 Optimization Problems

10.1.1 Introduction

The main focus in this section is optimization problems, the concept of Lagrange multipliers and KKT conditions, which will be later used to explain the details about the SVMs.

10.1.2 Mathematical Formulation

Optimization problems arise in almost every area of engineering. The goal is to achieve an almost perfect and efficient result, while carrying out certain procedures of optimization. Our main source of reference on this topic derives from Bletzinger[76]. We will be using notations used in Bletzinger[76]. In mathematical terms the general form of optimization problem can be represented as :

$$\begin{aligned} \text{minimize} \quad &: \quad f(\mathbf{x});\ \mathbf{x} \in \mathbf{R}^n \\ \text{such that} \quad &: \quad g_j(\mathbf{x}) \leq 0;\ \text{j} = 1,.....,\text{p} \\ &: \quad h_j(\mathbf{x}) = 0;\ \text{j} = 1,.....,\text{q}. \end{aligned} \tag{28}$$

where f, g_j and h_j are the objective function, equality constraints and inequality constraints. Generally, the number of constraints is less than the number of variables used to formulate the optimization problem. For a problem to be linear, both the constraints and the objective function need to be linear. Quadratic problems require only the objective function to be quadratic, while the constraints remain linear in formulation. Besides these, if any one of the functions is nonlinear, then the problem becomes nonlinear in nature. A graphical view of the types of the problems can be seen in fig. 10).

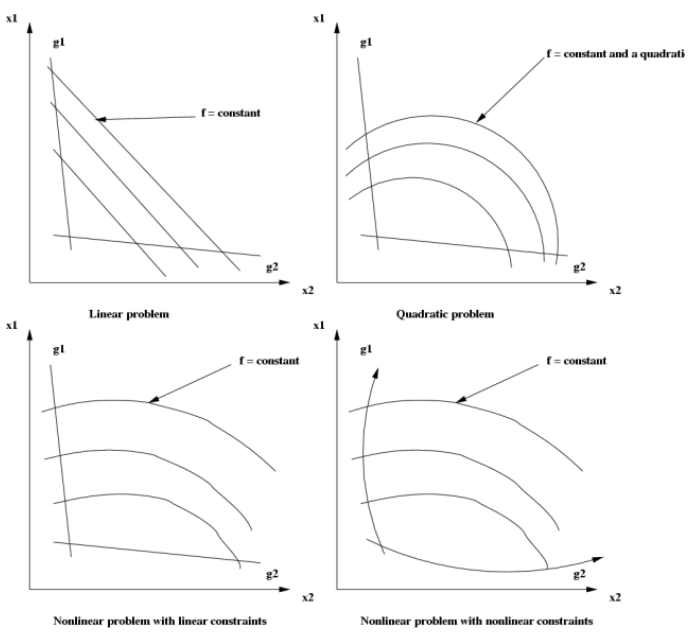


Fig. 10 Kinds of optimization problems.

10.1.3 Lagrange Multipliers

In unconstrained optimization problems, where the first order derivatives are assumed continuous, the solution is found by solving:

$$\nabla_x f = \frac{\partial f}{\partial x_i} = 0;\ \text{i} = 1,.....,\text{n}. \tag{29}$$

where f is a function of x. Since most of the optimization problems are constrained, the concept of Lagrange multipliers is introduced in order to solve the problem. Thus, the Lagrangian formulation, for Eqn. 28 becomes:

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{j=1}^{p} \lambda_j g_j(\mathbf{x}) + \sum_{j=1}^{q} \mu_j h_j(\mathbf{x}) \tag{30}$$

where L is the Lagrangian, λ and μ are the vectors of the Lagrange multipliers for inequality and equality constraints, respectively.

Next comes the solving of the Lagrangian. We try to derive a solution in terms of variables used and show that the final solution achieved by Equ. 28 and Eqn. 30 remains the same. For the sake of derivation, we assume that each of the vectors x, λ and μ have a single element and also there exists a single optimal solution. We will then generalize the solution to vectors containing various elements. Let x^*, λ^* and μ^* be the optimal solution for the Lagrangian. Let $x^!$ be the optimal solution for $f(x)$. To begin with, our Lagrangian has the form:

$$L(x, \lambda, \mu) = f(x) + \lambda g(x) + \mu h(x) \tag{31}$$

Derivation:

- **Step 1:** Differentiate the Lagrangian in Eqn 31 w.r.t x and equate it to zero.

$$\frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} + \mu \frac{\partial h}{\partial x} = 0 \tag{32}$$

- **Step 2:** Find x in terms of λ and μ, such that $x = x(\lambda, \mu)$.

- **Step 3:** Differentiate the Lagrangian in Eqn. 31 w.r.t λ and equate it to zero.

$$\frac{\partial L}{\partial \lambda} = g(x) = 0 \tag{33}$$

- **Step 4:** Differentiate the Lagrangian in Eqn 31 w.r.t μ and equate it to zero.

$$\frac{\partial L}{\partial \mu} = h(x) = 0 \tag{34}$$

- **Step 5:** Substitute $x(\lambda, \mu)$ in Equ. 33 and Eqn. 34 to get two equations in two unknowns λ and μ and solve to get

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

the optimal values.

$$g(x(\lambda,\mu)) = 0 \tag{35}$$

$$h(x(\lambda,\mu)) = 0 \tag{36}$$

Let λ^*, μ^* be the solution. Substituting these in $x = x(\lambda,\mu)$, we get x^*.

- **Step 6:** Combining Eqn. 33 and Eqn. 34 in Eqn. 31, along with λ^*, μ^* and x^*, we have:

$$L(x^*,\lambda^*,\mu^*) = f(x^*) + \lambda^* g(x^*) + \mu^* h(x^*) = f(x^*) \tag{37}$$

Since, it is assumed that there exist only one optimal solution we have:

$$\begin{aligned} L(x^*,\lambda^*,\mu^*) &= f(x^*) = f(x^!) \\ x^* &= x^! \end{aligned} \tag{38}$$

Lastly, since $g(x)$ in Eqn. 33 is a inequality constraint, we have:

$$\begin{aligned} \lambda g(x) = 0 \\ \lambda \geq 0 \end{aligned} \tag{39}$$

10.1.4 Dual Functions

For sake of simplicity, let us for a moment ignore the equality constraint. Then the Lagrangian becomes:

$$L(x,\lambda,\mu) = f(x) + \lambda g(x). \tag{40}$$

It is sometimes easy to transform the Lagrangian into a simpler form, in order to find an optimal solution. We can represent the Lagrangian as a Dual function in such a manner that the optimal solution defined as minimum of $L(x,\lambda^*)$ w.r.t x where $\lambda = \lambda^*$, can be represented as the maximum of dual function $D(\lambda)$ w.r.t λ. For a given λ, the dual is evaluated by finding the minimum of $L(x,\lambda)$ w.r.t x. Thus to find the optimal point we evaluate:

$$\max_{\lambda} D(\lambda) = \min_{x} L(x,\lambda^*) \tag{41}$$

So the basic steps to solve the dual problem are as follows: **Step 1:** Minimize $L(x,\lambda)$ w.r.t x, and find x in terms of λ. **Step 2:** Substitute $x(\lambda)$ in L s.t. $D(\lambda) = L(x(\lambda),\lambda)$. **Step 3:** Maximize $D(\lambda)$ w.r.t λ.

10.1.5 Karush Kuhn Tucker Conditions

The derivation in the last part (Eqn. 31 to Eqn. 39)gives us a set of equations that need to be evaluated along with the consideration of constraints present. These set of equations and constraints in terms of the Lagrangian, form the Karush Kuhn Tucker Conditions. We give here the generalized KKT conditions and explain the necessary details.

$$\begin{array}{lll} \frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{p} \lambda_j \frac{\partial g_j}{\partial x_i} + \sum_{j=1}^{p} \mu_j \frac{\partial h_j}{\partial x_i} = 0 & : & i = 1,\ldots,n \\ \frac{\partial L}{\partial \lambda_j} = g_j(\mathbf{x}) = 0 & : & j = 1,\ldots,p \\ \frac{\partial L}{\partial \mu_j} = h_j(\mathbf{x}) = 0 & : & j = 1,\ldots,q \\ \lambda_j g_j = 0 & : & j = 1,\ldots,p \\ \lambda_j \geq 0 & : & j = 1,\ldots,p \end{array} \tag{42}$$

where L is Eqn. 30.

The KKT conditions specify a few points which are as follows:

1. The first line states that the linear combination of objective and constraint gradients vanishes.

2. A prerequisite of the KKT conditions is that the gradients of the constraints must be continuous (evident from second and third lines in Eqn. 42).

3. The last two lines in Eqn. 42 state that at optimum either the constraints are active or the constrains are inactive.

10.2 Support Vector Machines

Armed with the knowledge of optimization problems and concept of Lagrange multipliers, we now delve into the workings of support vector machines. Burges[77] provides a good introduction to SVMs and is our main reference. Interested readers should refer to Cristianini and Shawe-Taylor[78], Schölkopf and Smola[79] and Vapnik and Vapnik[80] for detailed references.

10.2.1 Separable Case

Let us suppose that we are presented with a data set that is linearly separable. We assume that there are m examples of data in the format $\{\mathbf{x}_i, y_i\}$, s.t. $\mathbf{x}_i \in \mathbf{R}^n$; $i = 1,\ldots,m$, where $y_i \in \{-1,1\}$ is the corresponding true label of $\mathbf{x}_i$. We also suppose there is an existence of a linear hyperplane in the n dimensional space that separates the positively labeled data from the negatively labeled data. Let this separating hyperplane be given by

$$\mathbf{w}\cdot\mathbf{x} + b = 0. \tag{43}$$

where, $\mathbf{w}$ is the normal vector $\perp$ to the hyperplane and $|b|/||\mathbf{w}||$ is the shortest perpendicular distance of the hyperplane to the origin. $||\mathbf{w}||$ is the Euclidean norm of $\mathbf{w}$. The *margin* of a hyperplane is then defined as the minimum of the distance of the positively and negatively labeled examples, to the hyperplane. For the linear case, the SVM searches for the hyperplane with largest margin. We now have three conditions, based on the location of

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

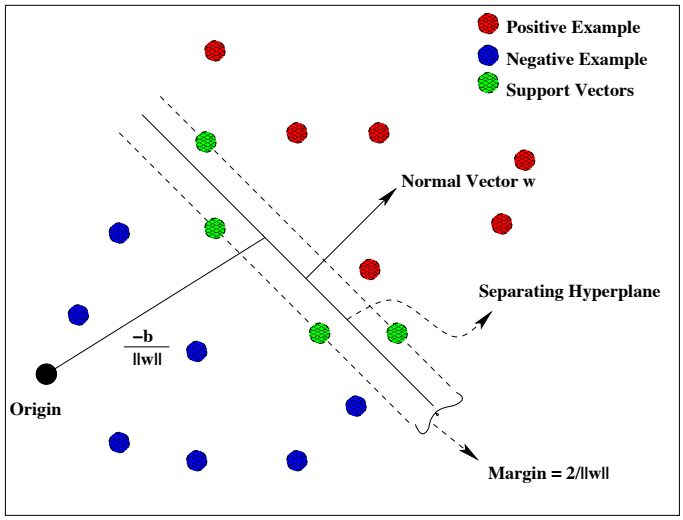


Fig. 11 Linear hyperplane for separable data. Adapted from Burges (A Tutorial on Support Vector Machines)

an example $\mathbf{x}_i$ w.r.t the hyperplane:

$$\mathbf{w}\cdot\mathbf{x}_i + b = 0 \quad : \quad \text{example lying on the hyperplane.} \tag{44}$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \geq +1 \quad : \quad \text{positively labeled example.} \tag{45}$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \leq -1 \quad : \quad \text{negatively labeled example.} \tag{46}$$

Combining the equality and the two inequalities we have:

$$y_i(\mathbf{w}\cdot\mathbf{x}_i + b) - 1 \geq 0 \tag{47}$$

Since the SVMs search for the largest margin, we now try to find a mathematical expression of the margin. Considering the examples that satisfy equality in Eqn. 45, the distance of the closest positive example can be expressed as $|1-b|/||\mathbf{w}||$. Similarly, considering the negative examples that satisfy equality in Eqn. 46, the distance of the closest negative example can be expressed as $|-1-b|/||\mathbf{w}||$. On summation of the two shortest distances, we get the margin of the hyperplane as $2/||\mathbf{w}||$. Since the labels are $\{-1, 1\}$, no example lies inside the hyperplanes representing the margin in this case. Taking into account that the SVM searches for the largest margin, we can say that it can be achieved by minimizing $||\mathbf{w}||^2$, subject to the constraints in Eqn. 47. Examples lying on the hyperplanes of the margins are termed *support vectors*, as their removal would change the margin and thus the solution. Figure 11 represents the conceptual points about separating hyperplanes.

10.3 Lagrangian Representation: Separable Case

Clearly, the previous paragraph shows that finding the margin is a problem of optimization as the goal is to minimize $||\mathbf{w}||^2$ subject to constraints in Eqn. 47. Employing the ideas of Chapter 3, the Lagrangian for the above problem, is:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{m} \alpha_i y_i(\mathbf{x}_i\cdot\mathbf{w} + b) + \sum_{i=1}^{m} \alpha_i \tag{48}$$

where $\frac{1}{2}||\mathbf{w}||^2$ is the objective function, α is the Lagrangian multiplier and the Eqn. 47 is the inequality constraint. Since the minimization of the objective function is required, we employ the ideas of the derivation of KKT conditions (Eqn. 42) to Eqn. 48. In short, we would require the $L(\mathbf{w}, b, \alpha)$ to be minimized w.r.t $\mathbf{w}$ and b and also require its derivative w.r.t all α_i's to vanish. Thus the KKT conditions take the form:

$$\begin{aligned}
&\tfrac{\partial L}{\partial w_j} = w_j - \textstyle\sum_i \alpha_i y_i x_{ij} = 0 && : \quad \text{j = 1,....,n} \\
&\tfrac{\partial L}{\partial b} = -\textstyle\sum_{i=1}^{m} \alpha_i y_i = 0 && : \quad \text{i = 1,....,m} \\
&\tfrac{\partial L}{\partial \alpha_i} = -\textstyle\sum_{i=1}^{m} y_i(\mathbf{x}_i\cdot\mathbf{w} + b) + \sum_{i=1}^{m} 1 = 0 && : \quad \text{i = 1,....,m} \\
&\alpha_i(y_i(\mathbf{x}_i\cdot\mathbf{w} + b) - 1) = 0 && : \quad \text{i = 1,....,m} \\
&\alpha_i \geq 0 && : \quad \text{i = 1,....,m}
\end{aligned} \tag{49}$$

Thus solving the SVMs is equivalent to solving the KKT conditions. While $\mathbf{w}$ is determined by the training set, b can be found by solving the penultimate equation in Eqn. 49 for which $\alpha_i \neq 0$. Also note that examples that have $\alpha_i \neq 0$ form the set of support vectors.

The dual problem for the same Lagrangian is:

$$D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i\cdot\mathbf{x}_j \tag{50}$$

Solving for Eqn. 50 requires maximization of D w.r.t α_i, subject to second line of Eqn. 49 and positivity of α_i, with the solution given by first line of Eqn. 49.

To classify or predict the label of a new example $\mathbf{x}_{new}$, the SVM has to evaluate $(\mathbf{x}_{new}\cdot\mathbf{w} + b)$ and check the sign of the evaluated value. A positive sign would lead to assignment of a $+1$ label and a negative sign to -1.

10.4 Nonseparable Case

For many classification problems, the data present is nonseparable. To extend the idea to nonseparable case, some amount of cost is added, which takes care of particular cases of examples. This is achieved by introducing slack in the constraints Eqn. 45 and Eqn. 46 (Burges[77], Vapnik and Vapnik[80]). The equations

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

then becomes

$$\mathbf{w}\cdot\mathbf{x}_i + b \geq 1-\xi_i \quad : \quad \text{positively labeled example.} \quad (51)$$

$$\mathbf{w}\cdot\mathbf{x}_i + b \leq -1+\xi_i \quad : \quad \text{negatively labeled example.} \quad (52)$$

For an error to occur, the ξ_i value must exceed unity. To take care of the cost of errors, a penalty is introduced which changes the objective function from $||\mathbf{w}||^2/2$ to $||\mathbf{w}||^2/2 + C(\sum_i \xi_i)^k$. Thus $\sum_i \xi_i$ represents the upper bound on the training error. For quadratic problems, k can be 1 or 2.

10.5 Lagrangian Representation: Nonseparable Case

Since the formulation of the Lagrangian and its dual follow the same procedure, as mentioned before, we only mention the equations. The Lagrangian for nonlinear nonseparable case is:

$$\begin{aligned} L(\mathbf{w}, b, \alpha, \xi) &= \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i\{y_i(\mathbf{x}_i\cdot\mathbf{w}+b) \\ &\quad -1+\xi_i\} - \sum_{i=1}^{m}\mu_i\xi_i \end{aligned} \quad (53)$$

The corresponding KKT conditions are:

$$\begin{array}{lll} \frac{\partial L}{\partial w_j} = w_j - \sum_i \alpha_i y_i x_{ij} = 0 & : & j = 1,\ldots,n \\ \frac{\partial L}{\partial b} = -\sum_{i=1}^{m}\alpha_i y_i = 0 & : & i = 1,\ldots,m \\ \frac{\partial L}{\partial \alpha_i} = y_i(\mathbf{x}_i\cdot\mathbf{w}+b) - 1 + \xi_i = 0 & : & i = 1,\ldots,m \\ \frac{\partial L}{\partial \xi_i} = C - \alpha_i - \mu_i = 0 & : & i = 1,\ldots,m \\ \alpha_i\{y_i(\mathbf{x}_i\cdot\mathbf{w}+b) - 1 + \xi_i\} = 0 & : & i = 1,\ldots,m \\ \mu_i\xi_i = 0 & : & i = 1,\ldots,m \\ \alpha_i \geq 0 & : & i = 1,\ldots,m \\ \xi_i \geq 0 & : & i = 1,\ldots,m \\ \mu_i \geq 0 & : & i = 1,\ldots,m \end{array} \quad (54)$$

The dual formulation $k = 1$ for the Lagrangian just discussed is:

$$D = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j \mathbf{x}_i\cdot\mathbf{x}_j \quad (55)$$

All the previous conditions remain same, except that the Lagrangian multiplier α_i now has a upper bound of value C. The solution for the dual is given by $\mathbf{w} = \sum_{i=1}^{N_s}\alpha_i y_i \mathbf{x}_i$. N_s is the number of support vectors. Figure 12 depicts the nonseparable case.

10.6 Kernels and Space Dimensionality Transformation

The above cases were for linear separating hyperplanes. In order to generalize for nonlinear cases, Boser et.al[81] employed the idea of Aizerman[60] as follows; Since the Dual in Eqn. 55 and

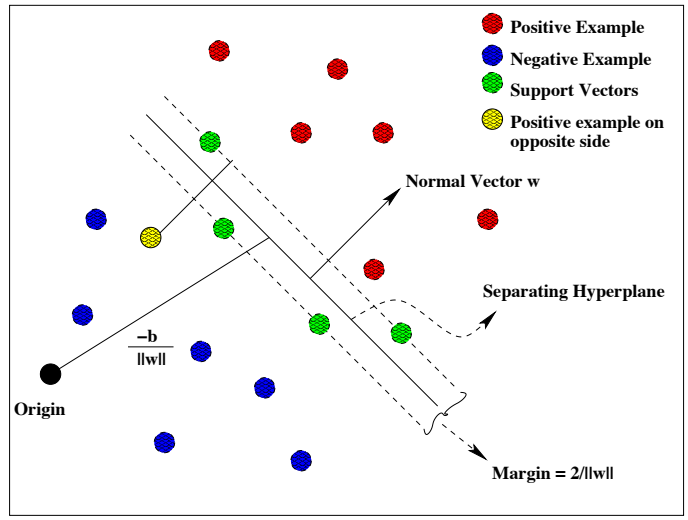


Fig. 12 Linear hyperplane for nonseparable data. Adapted from Burges (A Tutorial on Support Vector Machines)

its corresponding constraint equations employ the dot product of the examples, $\mathbf{x}_i\cdot\mathbf{x}_j$, it was proposed to map the data in a higher dimensional space using a function ϕ s.t. the algorithm would depend only on dot products in the higher space. Next, the existence of a function called *kernel*, dependent on $\mathbf{x}_i$ and $\mathbf{x}_j$, was assumed s.t. the value reported by the kernel was equal to the value resulting from the dot product in the higher space. A mathematical representation of the above concept is -

$$\phi : \mathbf{R}^n \mapsto H \quad (56)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j) \quad (57)$$

where H is a higher dimensional space.

This technique drastically reduces the amount of work required while dealing with nonlinear separating hyperplanes, concerning search for appropriate ϕ. Instead, one only works with $K(\mathbf{x}_i, \mathbf{x}_j)$, in place of $\mathbf{x}_i\cdot\mathbf{x}_j$. For classification purpose, where the sign of the function $(\mathbf{x}_{new}\cdot\mathbf{w}+b)$ is evaluated, the formulation employing kernels become:

$$\begin{aligned} f(\mathbf{x}_{new}) &= (\mathbf{w}\cdot\mathbf{x}_{new}+b) \\ &= \sum_{i=1}^{N_s}\alpha_i y_i \phi(\mathbf{s}_i)\cdot\phi(\mathbf{x}_{new}) + b \\ &= \sum_{i=1}^{N_s}\alpha_i y_i K(\mathbf{s}_i, \mathbf{x}_{new}) + b \end{aligned} \quad (58)$$

where $\mathbf{s}_i$ are the support vectors.

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

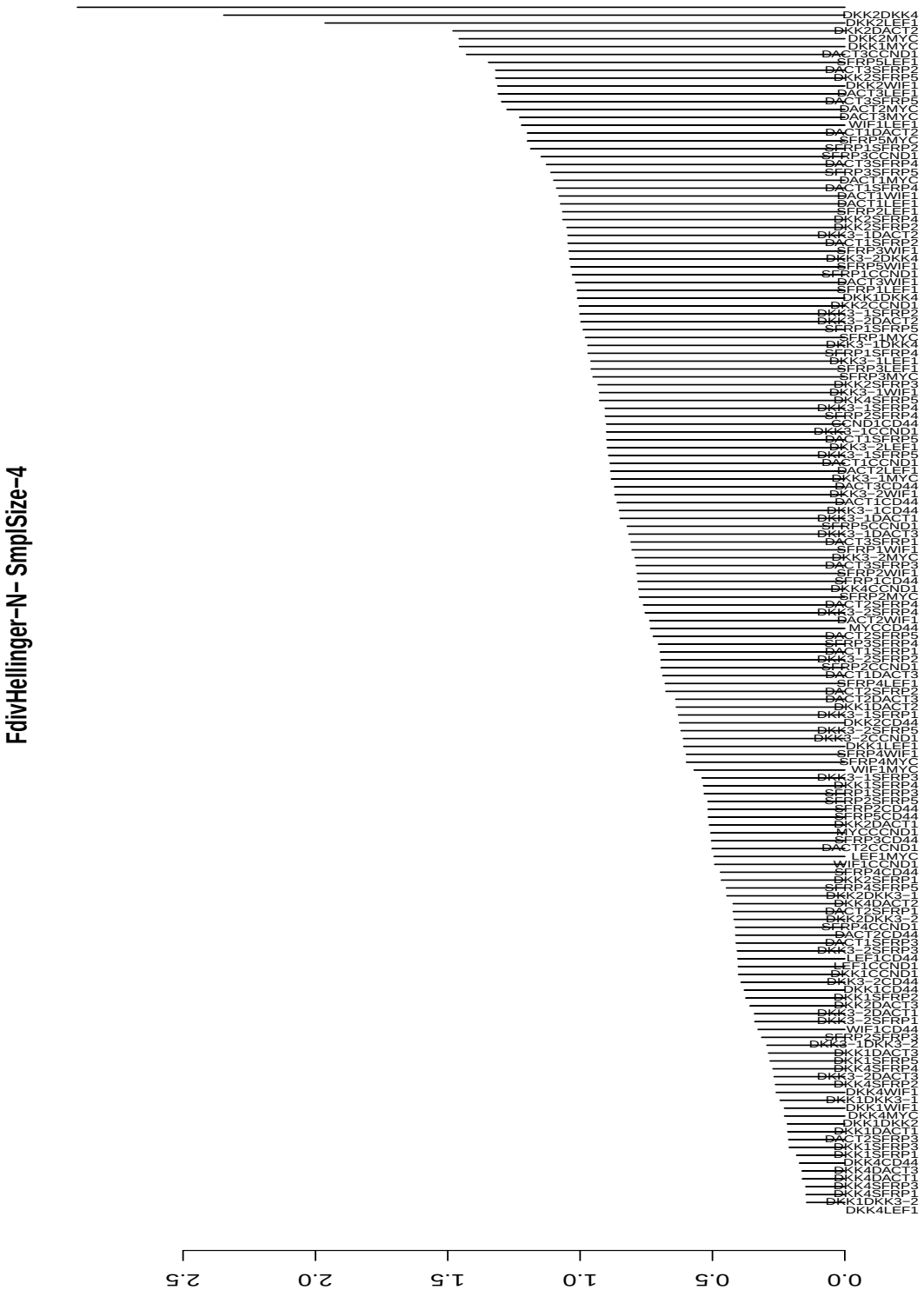


Fig. 13 FdivHellinger; Training sample size - 4; Test sample size - 16; Case - Normal

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

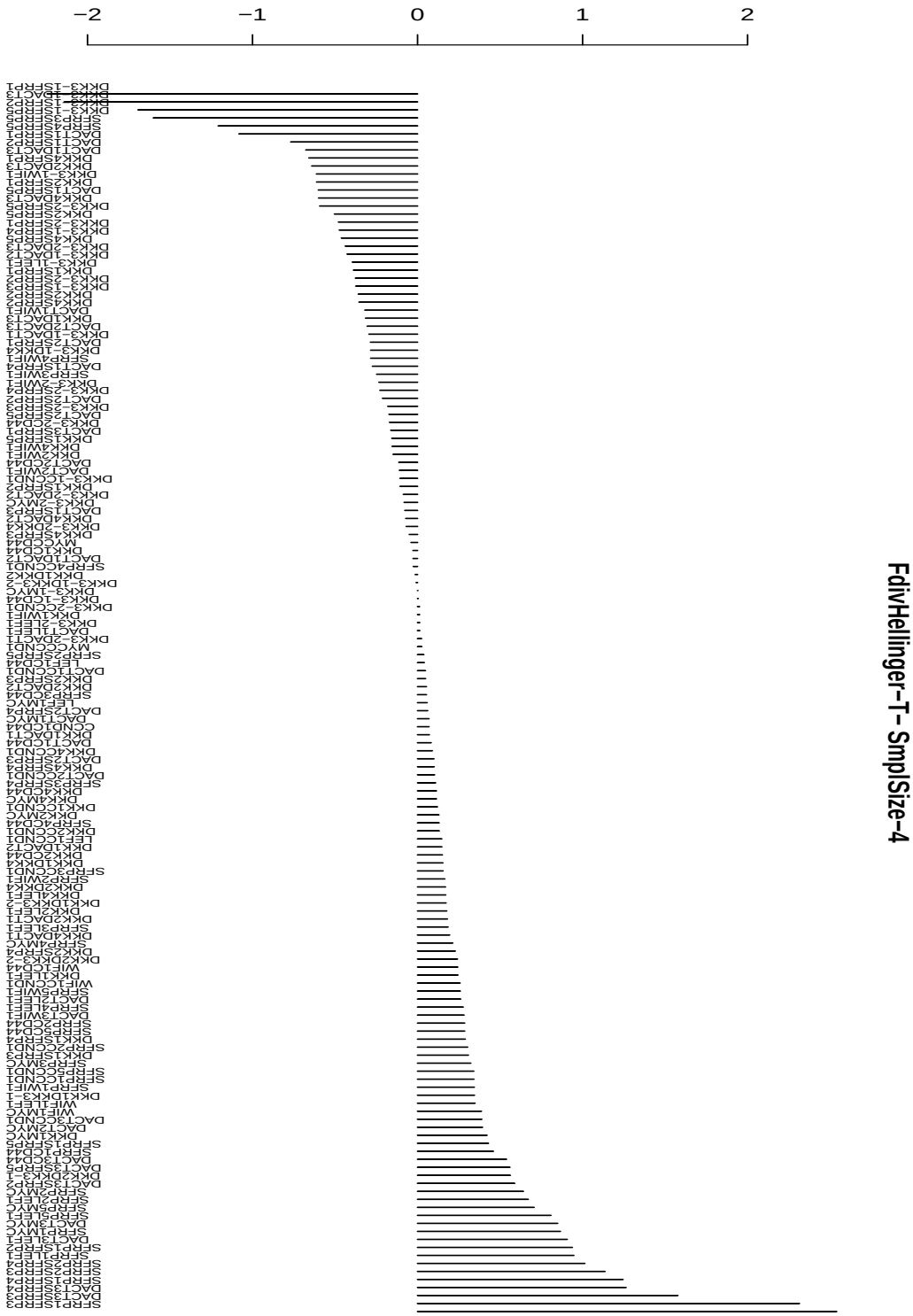


Fig. 14 FdivHellinger; Training sample size - 4; Test sample size - 16; Case - Tumor

FdivHellinger−T− SmplSize−4

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

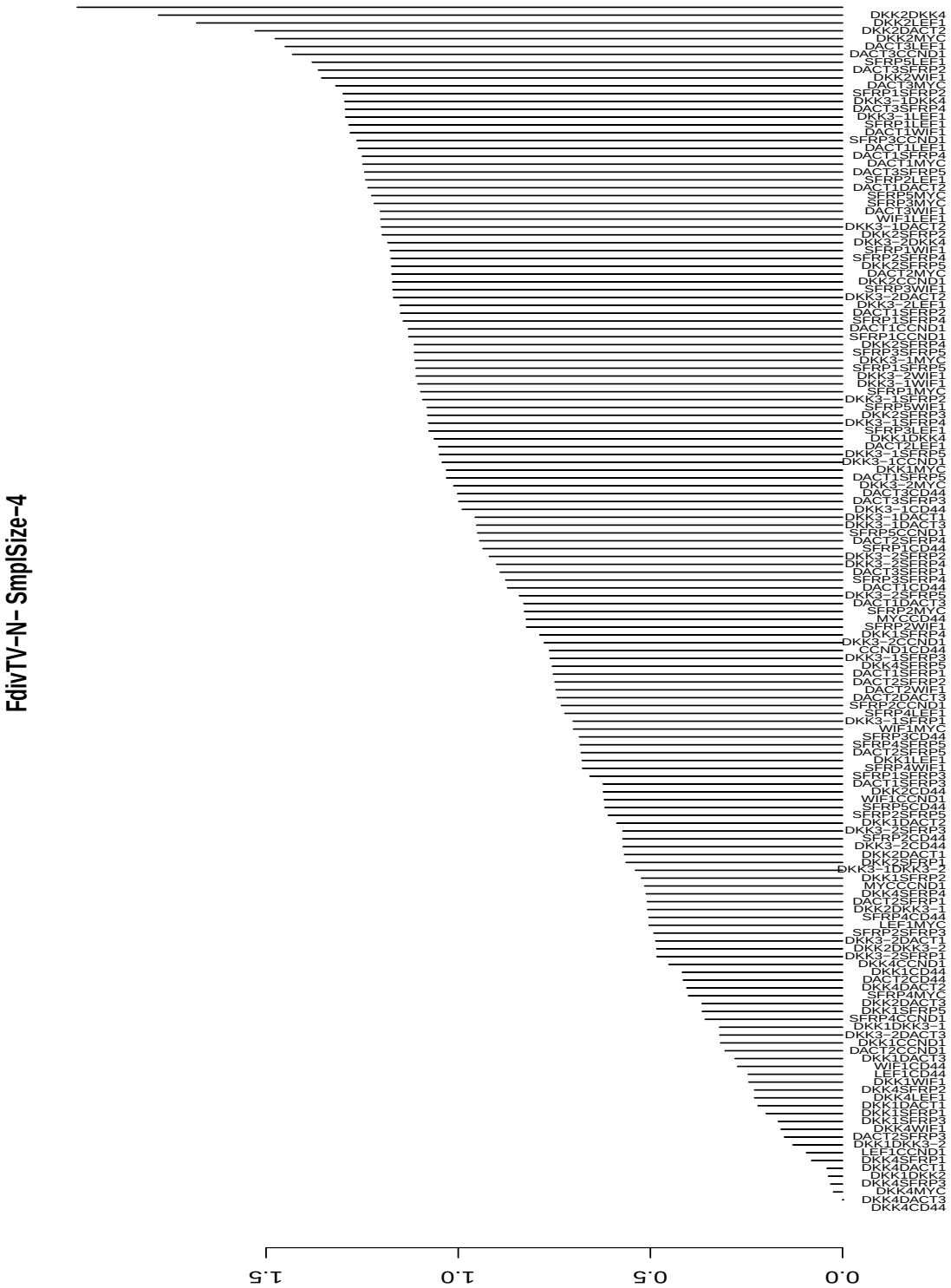


Fig. 15 FdivTV; Training sample size - 4; Test sample size - 16; Case - Normal

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

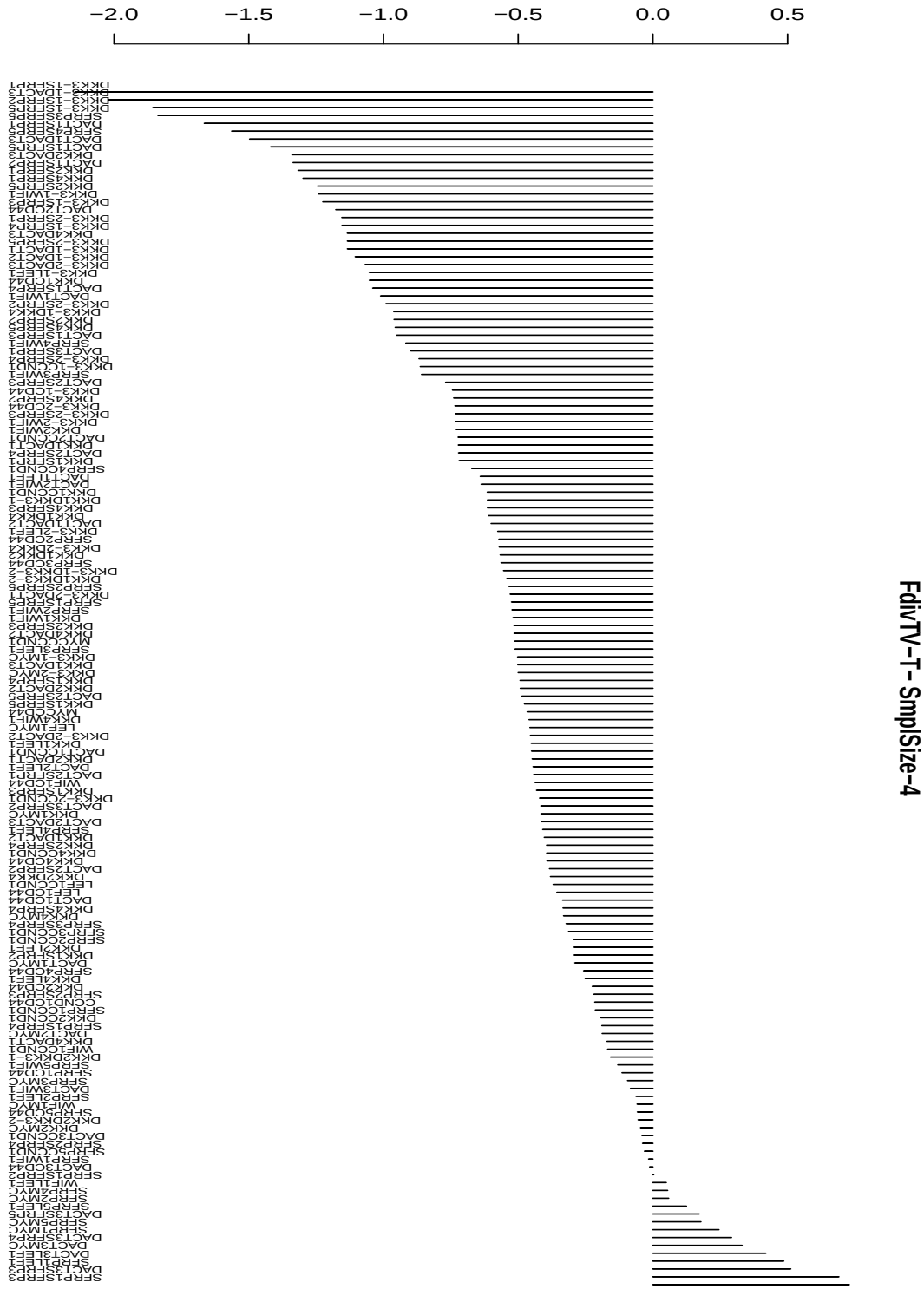


Fig. 16 FdivTV; Training sample size - 4; Test sample size - 16; Case - Tumor

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

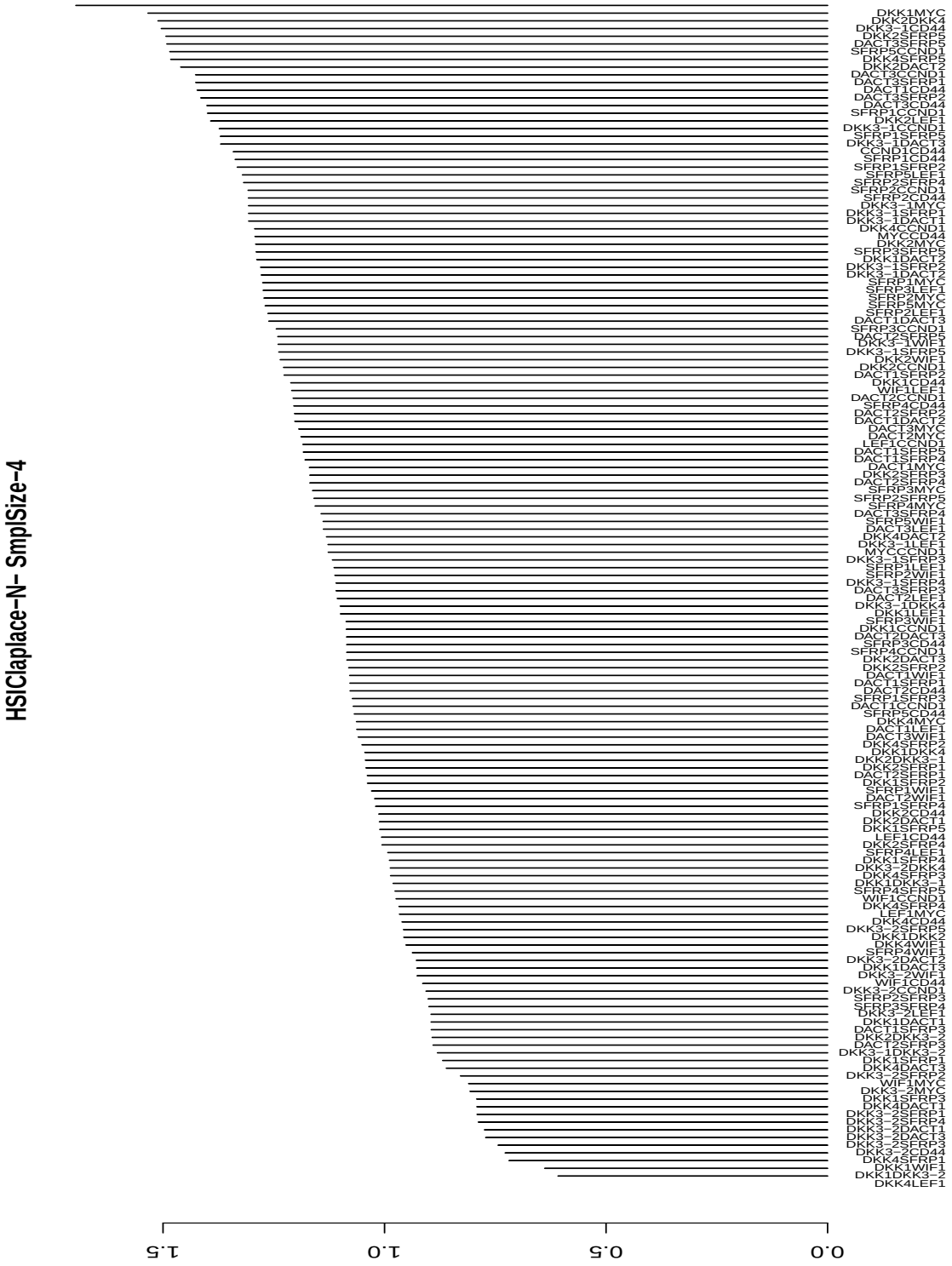


Fig. 17 HSIClaplace; Training sample size - 4; Test sample size - 16; Case - Normal

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

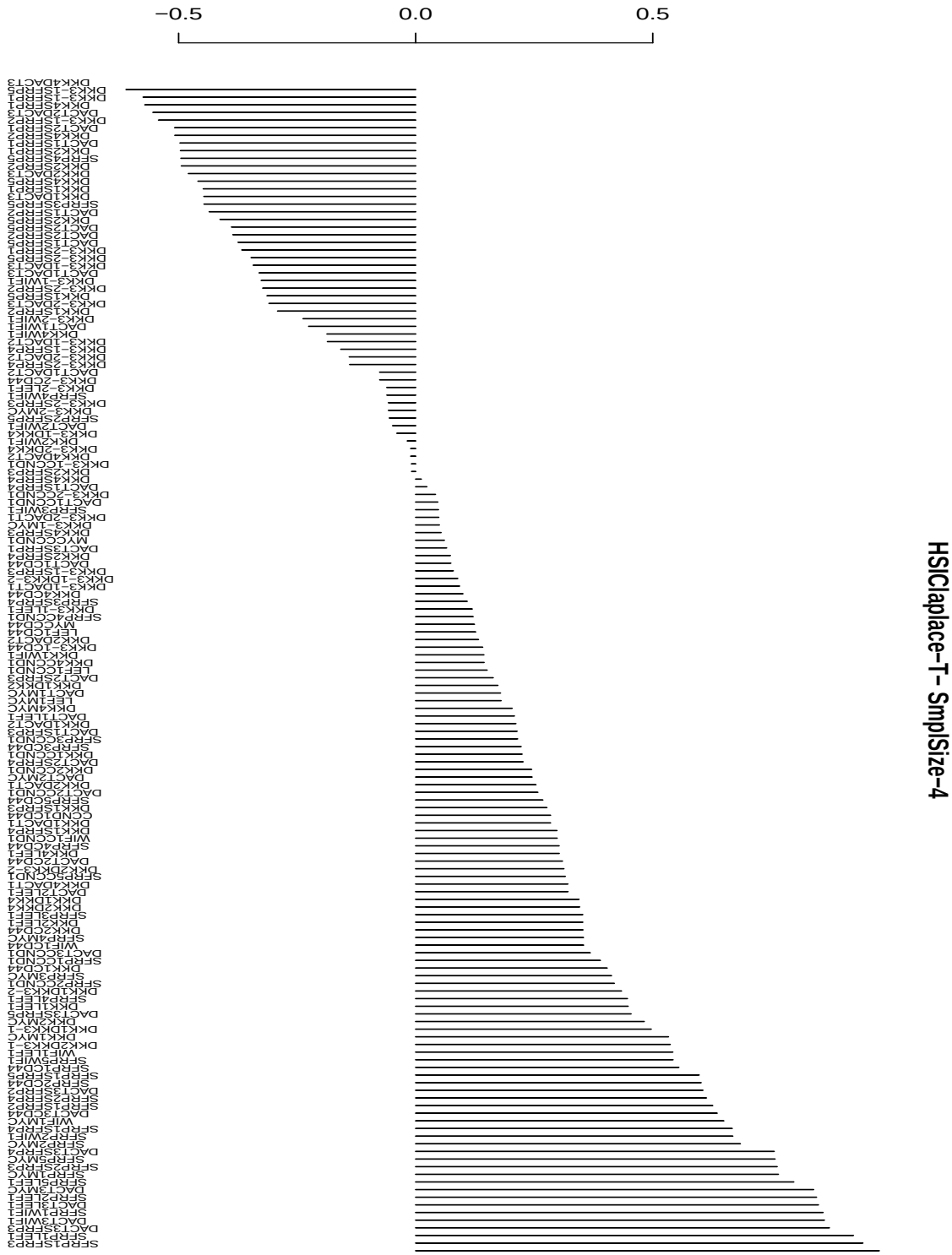


Fig. 18 HSIClaplace; Training sample size - 4; Test sample size - 16; Case - Tumor

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Fig. 19 HSICrbf; Training sample size - 4; Test sample size - 16; Case - Normal

bioRxiv preprint doi: https://doi.org/10.1101/059469; this version posted August 8, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

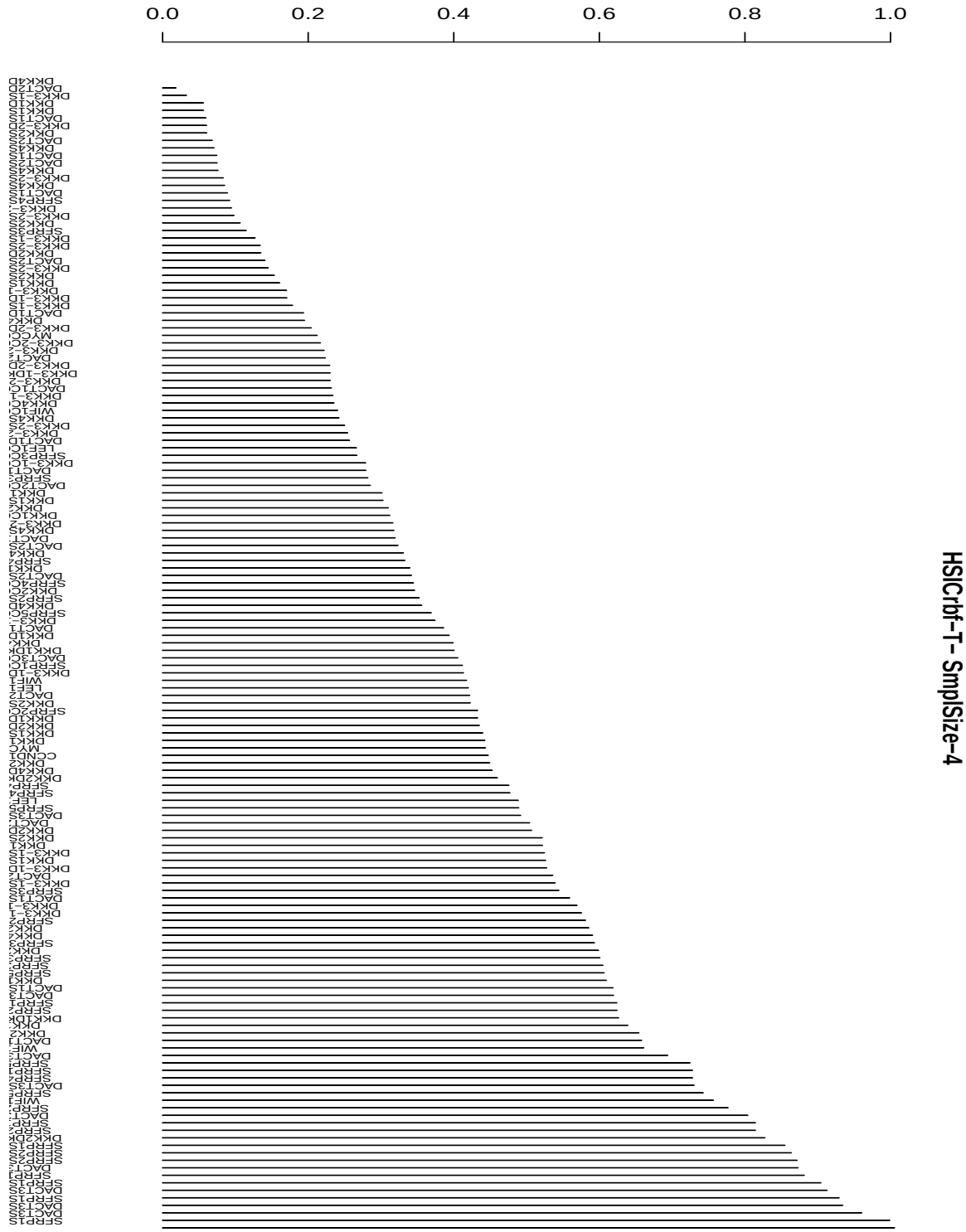


Fig. 20 HSICrbf; Training sample size - 4; Test sample size - 16; Case - Tumor