Article

Conserved Machine Learning Rankings of Myc Gene Combinations Across Different Sensitivity Methods Connote Existence of Biological Synergy

Shriprakash Sinha *

Posted Date: 23 January 2025

doi: 10.20944/preprints202501.1733.v1

Keywords: MYC; Gene combination; Colorectal Cancer; ETC-1922159; Machine learning

Article

Conserved Machine Learning Rankings of MYC Gene Combinations Across Different Sensitivity Methods Connote Existence of Biological Synergy †

Conserved ML Rankings of MYC Gene Combinations

Shriprakash Sinha

Independent Researcher; 104-Madhurisha Heights Phase 1, Risali, Bhilai-490006, India; sinha.shriprakash@yandex.com

† Aspects of unpublished work were presented in a poster session at the first Wnt Gordon Research Conference, from 6-11 August 2017, held in Stowe, VT 05672, USA.

Abstract: A recent design of a machine learning based search engine was published, that ranks combinations of genes that might be working synergistically in cells in various processes. To demonstrate its efficacy in real life scenario, the data set containing recordings of up/down regulated genes generated from colorectal cancer (CRC) cells treated with PROCN-WNT inhibitor drug ETC-1922159 was taken. The regulation of the genes were recorded individually, but in many cases, it is still not known which higher (≥ 2) order gene combinations might be playing a greater role in CRC. Here, I demonstrate that the rankings assigned to gene combinations at 2^{nd} order, by the search engine are conserved across the different sensitivity methods (and kernels/variants). This conservation points to the possible existence of the synergy between genes at the biological level. To establish the hypothesis I present ranked combinations of v-myc avian myelocytomatosis viral oncogene homolog (MYC), known to encode proteins that play significant role as transcription factors in cancer and target various kinds of genes, thus contributing to regrowth and proliferation. The manuscript identifies experimentally tested combinations of MYC-X in literature (whether in CRC cell or other cancer/ordinary cell). Second, the work reveals machine learning rankings for these MYC-X combinations in ETC-1922159 treated CRC cells. For experimentally established combinations, these rankings bolster confirmatory results. Based on the second step, the work points to new rankings of unknown/untested/unexplored MYC-X combinations that might be working synergistically in CRC cells.

Keywords: MYC; gene combination; colorectal cancer; ETC-1922159; machine learning

1. Significance

A search engine was used to reveal and prioritise gene combinations, by adapting the code from a recently published work. Rankings of combinations were observed to be conserved across the different sensitivity methods (used to estimate the influence of components of a combination). This points to possible existence of synergy between genes at biological level. Presented here are rankings of experimentally established combinations of MYC-X for CRC treated with ETC-1922159 and rankings that are unexplored. These point to efficacy and potential of the search engine. The engine is effective for ranking combinations of any gene of choice.

2. Introduction

2.1. Combinatorial Search Problem and a Possible Solution

A recent design of a machine learning based search engine was published [1], that ranks combinations of genes that might be working synergistically in cells in various processes. To demostrate its efficacy in real life scenario, the data set containing recordings of up/down regulated genes generated from colorectal cancer (CRC) cells treated with PROCN-WNT inhibitor drug ETC-1922159 was

 doi:10.20944/preprints202501.1733.v1

taken [2]. The regulation of the genes were recorded individually, but in many cases, it is still not known which higher (≥ 2) order gene combinations might be playing a greater role in CRC. Here, I demonstrate that the rankings assigned to gene combinations at 2^{nd} order, by the search engine are conserved across the different sensitivity methods (and kernels/variants). This conservation points to the possible existance of the synergy between genes at the biological level. Readers are requested to go through the adaptation of the above mentioned work for gaining deeper insight into the working of the pipeline and its use of published data set generated after administration of ETC-1922159, [3].

2.2. Insight Behind the Work

Across all search engines has the fundamental principle remains the same i.e to capture the pattern available in the data and based on that pattern, rank a list of queries. Different algorithms can be applied, however if the fundamental pattern is captured accurately, then the rankings will remain approximately the same, with slight variations, across the different kinds of search engines used. I use one search engine, however, vary the way the patterns are captured via use of different sensitivity methods. Each sensitivity method uses a different flavour/mathematical formulation to compute the sensitivity indices to estimate the influence of the involved factors. These involved factors are genes that play a role in cell biology, in the above research. The insight is that all methods will capture the sensitivity of the involved factors based on their recorded regulations and the search engine will rank the combination of factors based on these sensitivity indices. Since the role of involved factors are captured properly, the search engine will give appropriate rankings to the combinations, thus capturing which gene combinations might be playing significantly in a biological phenomena. The above work shows rankings for experimentally confirmed combinations as well as unexplored/untested combinations. These rankings are not just numbers. They point to the existence of biological synergy in the form of gene combinations, whether tested in wet lab or unexplored till now. Finally, the findings suggest that the rankings are conserved across the different sensitivity methods used.

2.3. PORCN-WNT Inhibitors

The regulation of the Wnt pathway is dependent on the production and secretion of the WNT proteins. Thus, the inhibition of a causal factor like PORCN which contributes to the WNT secretion has been proposed to be a way to interfere with the Wnt cascade, which might result in the growth of tumor. Several groups have been engaged in such studies and known PORCN-WNT inhibitors that have been made available till now are IWP-L6 [4,5], C59 [6], LGK974 [7] and ETC-1922159 [8]. In this study, the focus of the attention is on the implications of the ETC-1922159, after the drug has been administered. The drug is a enantiomer with a nanomolar activity and excellent bioavailability as claimed in [8].

2.4. MYC

c-MYC (MYC) is a gene that encodes for transcription factor. It belongs to the family of MYC genes which includes B-MYC, N-MYC, L-MYC and S-MYC. Mutations in c-MYC have been found to be involved in various kinds of cancers. MYC (or v-myc avian myelocytomatosis viral oncogene homolog) contains a helix-loop-helix (HLH) and leucine zipper (LZ) domain that helps it to bind to interact with DNA and form a complex with MAX [9], respectively. A good introductory review on c-MYC can be found in [10] which covers topics on transcription factors and binding sites, transcriptional properties, MYC affected genes and finally role of MYC in cell cycle, apoptosis [11] and metabolism [12]. A lot of work has been done in MYC at various levels as it is implicated in various types of cancer. Additionally, MYC has been found to have influence on the chromatin structure also [13]. In colon cancer, c-MYC has been found to be highly expressed [14,15]. In colorectal cancer cells treated with ETC-1922159, MYC was found to be down regulated along with other genes. In this study, I use MYC as a choice of gene because it is induced by mitogenic signals and regulates downstream cellular

 doi:10.20944/preprints202501.1733.v1

responses. Overexpressed MYC promotes malignant transformation and modulates expression of a range of genes in experimental systems, however only a few are proven direct targets.

For curating experimentally established combinations of MYC, I use the list tabulated by [16]. They present a large-scale screen for MYC-binding sites in live human cells by identifying genes directly bound by MYC through the consensus E-box element CACGTG. Their strategy was based on the preselection of candidate sites with bioinformatic tools, which was followed by the experimental analysis of a large number of individual sites with a quantitative ChIP assay. They identify 257 genes within the high-affinity group of MYC-targets and assign them to a functional category. The genes qualify as high-affinity targets in all cell lines tested.

3. Tools of Study

Sensitivity analysis and its relevance in systems biology have been covered in a recently published article [17], which forms the foundation for this work. In this work, the sensitivity indices are computed for all factors or combination of factors affecting the pathway. Ranking using support vector machines (SVM) are then employed using these sensitivity indices. The work uses SVM package by [18] in https://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html. I use the adaptation of the above engine to rank 2^{nd} order gene combinations.

3.1. Sensitivity Analysis

For completeness, a part of [17] has been reproduced below.

3.1.1. Variance Based Sobol Method

Seminal work by Russian mathematician [19] 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 [20] categorically delineates these methods with the following functionality • screening for sorting influential measures ([21] method, Group screening in [22,23], Iterated factorial design in [24], Sequential bifurcation design in [25,26]), • quantitative indicies for measuring the importance of contributing input factors in linear models ([27–30]) and nonlinear models ([31–47]) and • exploring the model behaviour over a range on input values ([48–51]). [20] also provide various criteria in a flowchart for adapting a method or a combination of the methods for sensitivity analysis.

The general idea for variance based Sobol method 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.

3.1.2. Density Based HSIC Method

Besides the above [19]'s variance based indicies, more recent developments regarding new indicies based on density, derivative and goal-oriented can be found in [52–54], respectively. In a latest development, [55] propose new class of indicies based on density ratio estimation [52] that are special cases of dependence measures. This in turn helps in exploiting measures like distance correlation [56] and Hilbert-Schmidt independence criterion [57] as new sensitivity indicies. The framework of these indicies is based on use of [58] f-divergence, concept of dissimilarity measure and kernel trick [59].

 doi:10.20944/preprints202501.1733.v1

Finally, [55] propose feature selection as an alternative to screening methods in sensitivity analysis. The main issue with variance based indicies [19] 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 [52] and dependence measures are more robust.

The general idea for density based HSIC method 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 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.

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.

3.1.3. Sensitivity Package in R

The sensitivity package by [60] was used to develop the search engine pipeline. The current research uses the Hilbert Schmidt Independence Criterion (HSIC) and SOBOL method, implemented in the sensitivity package mentioned above. I use three different kernels under the HSIC method namely, • laplace, • linear and • rbf. For SOBOL method, I use • sobol-2002, and • sobol-jansen. Each of these variants or kernels have been implemented in the sensitivity package and option has been provided in the search engine code to generate the rankings for a particular gene, using a choice of a kernel/variant at a time. Technical details about the variants and kernels can be found in references cited in [60].

4. Static Data by [2]

Data used in this research work was released in a publication by [2]. The ETC-1922159 was released in Singapore in July 2015 under the flagship of the Agency for Science, Technology and Research (A*STAR) and Duke-National University of Singapore Graduate Medical School (Duke-NUS). Note that the ETC-1922159 data show numerical point measurements that is as [2] quote - "List of differentially expressed genes identified at three days after the start of ETC-159 treatment of colorectal tumors. Log2 fold-changes between untreated (vehicle, VEH) and ETC-159 treated (ETC) tumors are reported." The numerical point measurements of differentially expressed genes were recorded using the following formulation of fold changes in equation 1 (see [61–63]).

$$\log_2 \frac{VEH_{avg}}{ETC_{avg}} \tag{1}$$

5. Methodology

5.1. Revealing Higher Order Biological Hypotheses via Sensitivity Analysis and Insilico Ranking Algorithm

In the trial experiments on ETC-1922159 [2], a list of genes (2500±) have been reported to be up and down regulated after the drug treatment and a time buffer of 3 days. Some of the transcript levels of these genes have been recorded and the experimental design is explained elaborately in the

 doi:10.20944/preprints202501.1733.v1

same manuscript. In the list are also available unknown or uncharacterised proteins that have been recorded after the drug was administered. These have been marked as "- -" in the list (Note - In this manuscript these uncharacterised proteins have been marked as "XXM", were $M = 1, 2, 3, ...$). The aim of this work is to reveal unknown/unexplored/untested biological hypotheses that form higher order combinations. For example, it is known that the combinations of WNT-FZD or RSPO-LGR-RNF play significant roles in the Wnt pathway. But the $n \geq 2, 3, ...$-order combinations out of $N(> n)$ genes forms a vast combinatorial search forest that is extremely tough to investigate due to the humongous amount of combinations. Currently, a major problem in biology is to cherry pick the combinations based on expert advice, literature survey or random choices to investigate a particular combinatorial hypothesis. The current work aims to reveal these unknown/unexplored/untested combinations by prioritising these combinations using a potent support vector ranking algorithm [18]. This cuts down the cost in time/energy/investment for any investigation concerning a biological hypothesis in a vast search space.

The pipleline works by computing sensitivity indicies for each of these combinations and then vectorising these indices to connote and form discriminative feature vector for each combination. The ranking algorithm is then applied to a set of combinations/sensitivity index vectors and a ranking score is generated. Sorting these scores leads to prioritization of the combinations. Note that these combinations are now ranked and give the biologists a chance to narrow down their focus on crucial biological hypotheses in the form of combinations which the biologists might want to test. Analogous to the webpage search engine, where the click of a button for a few key-words leads to a ranked list of web links, the pipeline uses sensitivity indices as an indicator of the strength of the influence of factors or their combinations, as a criteria to rank the combinations.

5.2. Design for Static Data from [2]

The procedure begins with the listing of all C_k^n combinations for k number of genes from a total of n genes. Here n can be the choice of the biologist. 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. Note that the ETC-1922159 data show numerical point measurements that is as [2] quote "List of differentially expressed genes identified at three days after the start of ETC-159 treatment of colorectal tumors. Log2 fold-changes between untreated (vehicle, VEH) and ETC-159 treated (ETC) tumors are reported." Since the sensitivity analysis methods require a sample for a particular observation, a steep gaussian distribution was generated with a jitter (noise) added to the deviation from the reported point measurement of 0.005. In this experiment, the distribution contained 10 measurements (including the point of measurement under consideration). This is repeated for each point of measurement.

To have an averaged ranking, the experiment was designed to run for 50 iterations. In each iteration, 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 p^{th} combination in C_k^n combinations, the dataset is prepared in a required format (See .R code in mainscript-2-2.R). Details of formatting the data have not been presented in the article to maintain the fluidity and brevity. Interested readers can find examples of formating the data in the sensitivity analysis package in R. 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.

After the above sensitivity indices have been stored for each of the p^{th} combination, for a chosen sensitivity analysis method, the next step in the design of experiment is conducted. Here, the indices are averaged per combination to have a mean index value. These index values form the discriminative features for a particular combination. For a k^{th} order combination, a vector of k elements or indices forms a feature vector. Thus for C_k^n combinations there will be C_k^n vectors, each containing k elements. Next, SVM_{learn}^{Rank} [18] is used to generate a model on default value C value of 20. In the current

 doi:10.20944/preprints202501.1733.v1

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 $SVM^{Rank}_{classify}$ [18]. 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.

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("extractETCdata.R") • use source("mainScript-2-2.R") • use source("SVMRank-Results-S-mean.R").

6. Results & Discussion

6.1. How to Interpret the Ranking?

In each of the sections below, one will find two tables.

The first table lists the rankings of a particular gene combination based on these kernels. Based on majority voting, a combination is decided to be low ranked or high ranked. So, for example, if majority rankings point to low numerical value (i.e below the half way mark of approximately 2500 gene combinations), then the combination is possibly not highly ranked in colorectal cancer cells AFTER the ETC-1922159 treatment. Looking at it in another way, this low ranking suggests that the combination might have been up regulated in colorectal cancer cells BEFORE the ETC-1922159 treatment. This points to the inference that the combination of the two genes/proteins was working synergistically in colorectal cancer, while being up regulated and before ETC-1922159 treatment. The ETC-1922159 administration had caused a down regulation of genes in colorectal cancer cells and what is available as data by [2] points to down regulated recordings of genes/proteins taken individually.

The second table uses the majority voting mentioned above, to filter out which combinations need to be further tests in the wet lab. These combinations recorded in the second table are inferences/pointers to existence of possible synergistic combinations that might be working in the cell, in a particular scenario (here colorectal cancer cells). Additionally, one will see two inferences - • based on the experimentally tested and established synergies in any other pathological/normal cell, if recorded in colorectal cancer cells treated with ETC-1922159, these combinations will ranked by the engine appropriately (or note - there might be a possibility that the experimentally tested combination established in a different scenario, might not get an approprate rank by the search engine in the colorectal cancer cells treated with ETC-1922159.) and • based on the cues from previous point, there will be combinations ranked by the engine, that point to new synergies that have not been explored/tested in wet lab.

Further, in a list of approximately 2500 genes that were up/down regulated after ETC-1922159 treatment, for the second order combinations there will be 2499 combinations. The engine generates the ranking for all these 2499 combinations. However, it is not possible to report the ranking of all 2499 combinations in a single article, for a particular gene under investigation. The full set of rankings reveal a prioritized list of new combinations that emerge as plausible biological hypotheses that might be working synergistically in colorectal cancer cells. These require further tests. For transperancy and reproducibility, one can download the code of search engine in R language and run it on the data made available by [2], to get a full list of some 2499, 2^{nd} order combinations for a particular gene of choice. Higher order combinations can also be generated using this engine.

Finally, we also see how the rankings behave across the different sensitivity methods and how they are conserved across the same. This conservation points to existence of biological synergy between the components of a combination (i.e either experimentally established or is unexplored/untested).

6.2. Conserved Rankings of Experimentally Established and Unexplored/Untested, MYC - X Combinations

NOTE - X denotes a particular gene/protein of interest.

 doi:10.20944/preprints202501.1733.v1

[16] divide the MYC target genes into diverse functional categories. I use the same categories (the below sections) and present rankings of those combinations that were recorded in CRC cells treated with ETC-1922159. Further, rankings of unexplored/untested combinations for genes that might be related to family of X are also presented. Note that I present rankings of combinations of genes that were down regulated after the ETC-1922159 treatment. Also, bold **MYC - X** means the target genes mentioned in [16] were also found/recorded in CRC treated with ETC-1922159. Conserved rankings for these bolster confirmatory experimental results.

6.2.1. Adehesion / Matrix / Tissue Remodeling (AMT)

Under this functional category in [16], the following MYC-target genes are reported - integrin subunit beta 1 (ITGB1), integrin subunit alpha 6 (ITGA6), collagen type IV alpha (1/2) chain (COL4A-1/2), serpin family E member 1 (SERPINE1) and acid phosphatase 5, tartrate resistant (ACP5).

Table 1 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 2 generated from analysis of the ranks in Table 1. The Table 1 shows rankings w.r.t MYC. For genes related to Integrins (i.e ITGA9, ITGB3BP and ITGAE); Collagen (i.e COL9A3); Serpin family (i.e SERPINF2); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Collagen (i.e COL18A1 and COL27A1); Acid phosphatase (i.e ACP1 and ACP6); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 1 graphically, in Table 2.

Table 1. 2^{nd} order interaction ranking between MYC VS AMT members.

RANKING MYC TARGETS IN AMT

MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Integrin					
ITGA9 - MYC	292	841	34	292	1284
ITGB3BP - MYC	451	96	380	370	405
ITGAE - MYC	1421	1615	1352	1557	669
Collagen					
COL9A3 - MYC	343	1087	210	700	987
COL18A1 - MYC	2412	2042	2692	314	301
COL27A1 - MYC	2602	1870	2522	1205	1716
Serpin family					
SERPINF2 - MYC	2513	2109	2599	1870	1993
Acid phosphatase					
ACP1 - MYC	1615	1867	1516	2031	2029
ACP6 - MYC	2122	1860	2152	2274	2466

 doi:10.20944/preprints202501.1733.v1

Table 2. 2^{nd} order combinatorial hypotheses between MYC and AMT members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
AMT members	synergy BEFORE drug treatment w.r.t
ITG-A9/B3BP/AE	MYC
COL9A3	MYC
SERPINF2	MYC
AMT members	synergy AFTER drug treatment w.r.t
COL-18A1/27A1	MYC
ACP-1/6	MYC

6.2.2. Ligand / Receptor (LR)

Under this functional category in [16], the following MYC-target genes are reported - fibroblast growth factor receptor 4 (FGFR4), G protein-coupled receptor 4 (GPR4), Interleukin (IL-2/11RA/13), cholinergic receptor nicotinic beta 1 subunit (CHRNB1), NOTCH4 and transforming growth factor beta (TGFB-1/2/3).

Table 3 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 4 generated from analysis of the ranks in Table 3. The Table 3 shows rankings w.r.t MYC. For genes related to Fibroblast growth factor receptor (i.e FGFR4); G protein-coupled receptor (i.e GPR63 and GPRC5B); Interleukin (i.e IL33, IL17RB and IL17D); Cholinergic receptors nicotinic subunits (i.e CHRNA5); NOTCH (i.e NOTCH1) and Transforming growth factor beta (i.e TGFBR3) all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to G protein-coupled receptor (i.e GPR68, GPR137C and GPR19); Interleukin (i.e IL17RD and IL1RL2); NOTCH (i.e NOTCH4) and Transforming growth factor beta (i.e TGFB1 and TGFBRAP1); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 3 graphically, in Table 4.

 doi:10.20944/preprints202501.1733.v1

Table 3. 2^{nd} order interaction ranking between MYC VS LR members.

RANKING MYC TARGETS IN LR					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Fibroblast growth ... factor receptor					
FGFR4 - MYC	1191	681	1396	1541	1372
G protein-coupled receptor					
GPR63 - MYC	38	990	30	125	1125
GPRC5B - MYC	1194	1097	1234	1821	1539
GPR68 - MYC	2402	2638	2497	762	1157
GPR137C - MYC	2680	2443	2661	358	1460
GPR19 - MYC	2722	2264	2717	206	418
Interleukin					
IL33 - MYC	46	163	224	90	395
IL17RB - MYC	270	501	314	1030	866
IL17D - MYC	1019	1199	851	531	1353
IL17RD - MYC	1683	2175	1842	2390	2278
IL1RL2 - MYC	1727	2522	1897	1526	1831
Cholinergic receptors ... nicotinic subunits					
CHRNA5 - MYC	65	164	93	92	471
NOTCH					
NOTCH1 - MYC	723	571	388	1034	220
NOTCH4 - MYC	2670	2684	2663	779	247
Transforming growth ... factor beta					
TGFB1 - MYC	1208	2091	1124	1903	2372
TGFBR3 - MYC	933	1042	897	937	266
TGFBRAP1 - MYC	1126	1651	956	1837	1891

Table 4. 2^{nd} order combinatorial hypotheses between MYC and LR members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
LR members	synergy BEFORE drug treatment w.r.t
FGFR4	MYC
GPR-63/C5B	MYC
IL-33/17RB/17D	MYC
CHRNA5	MYC
NOTCH1	MYC
TGFB-R3	MYC
LR members	synergy AFTER drug treatment w.r.t
GPR-68/137C/19	MYC
ACP-1/6	MYC
IL-17RD/1RL2	MYC
NOTCH4	MYC
TGFB-1/RAP1	MYC

6.2.3. Structural (STR)

Under this functional category in [16], the following MYC-target genes are reported - erythrocyte membrane protein band 4.2 (EPB42), laminin subunit beta 2 (LAMB2), lamin A/C (LMNA) and stathmin 1/oncoprotein 18 (STMN1).

 doi:10.20944/preprints202501.1733.v1

Table 5 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 6 generated from analysis of the ranks in Table 5. The Table 5 shows rankings w.r.t MYC. For genes related to Erythrocyte membrane protein band (EPB41L4A); Lamin (i.e LMNB1 and LMNB2) and Stathmin 1/oncoprotein 18 (i.e STMN1); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Laminin subunit beta (i.e LAMB1); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 5 graphically, in Table 6.

Table 5. 2^{nd} order interaction ranking between MYC VS STR members.

RANKING MYC TARGETS IN STR					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Erythrocyte membrane ... protein band					
EPB41L4A - MYC	834	1385	638	858	1254
Laminin subunit beta					
LAMB1 - MYC	2119	1776	2277	2512	2441
Lamin					
LMNB1 - MYC	231	58	171	75	646
LMNB2 - MYC	1007	166	906	823	694
Stathmin 1/oncoprotein 18					
STMN1 - MYC	109	114	139	110	805

Table 6. 2^{nd} order combinatorial hypotheses between MYC and STR members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
STR members	synergy BEFORE drug treatment w.r.t
EPB41L4A	MYC
LMNB-1/2	MYC
STMN1	MYC
STR members	synergy AFTER drug treatment w.r.t
LAMB1	MYC

6.2.4. Channels / Components (CC)

Under this functional category in [16], the following MYC-target genes are reported - solute carrier family 4 member 2 (SLC4A2).

Table 7 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 8 generated from analysis of the ranks in Table 7. The Table 7 shows rankings w.r.t MYC. For genes related to Solute carrier family 4 (SLC4A7); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

One can also interpret the results of the Table 7 graphically, in Table 8.

 doi:10.20944/preprints202501.1733.v1

Table 7. 2^{nd} order interaction ranking between MYC VS CC members.

RANKING MYC TARGETS IN CC					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Solute carrier family 4 SLC4A7 - MYC	1159	282	796	541	1314

Table 8. 2^{nd} order combinatorial hypotheses between MYC and CC members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
CC members	synergy BEFORE drug treatment w.r.t
SLC4A7	MYC

6.2.5. Chaperone / Protein Folding (CPF)

Under this functional category in [16], the following MYC-target genes are reported - heat shock protein family (HSPA8, HSPE1, HSPD1 and HSPCAL3).

Table 9 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 10 generated from analysis of the ranks in Table 9. The Table 9 shows rankings w.r.t MYC. For genes related to Heat shock protein family (i.e HSPB6, HSPE1, HSPD1, HSPA4L and HSPA9); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Heat shock protein family (i.e HSPA4); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 9 graphically, in Table 10.

Table 9. 2^{nd} order interaction ranking between MYC VS CPF members.

RANKING MYC TARGETS IN CPF					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Heat shock ... protein family					
HSPB6 - MYC	193	223	111	330	645
HSPE1 - MYC	440	1041	467	1134	1095
HSPD1 - MYC	747	487	784	1023	654
HSPA4L - MYC	1085	1268	542	944	1684
HSPA9 - MYC	1481	1488	1454	2160	1719
HSPA4 - MYC	1705	768	1541	1160	1612

Table 10. 2^{nd} order combinatorial hypotheses between MYC and CPF members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
CPF members	synergy BEFORE drug treatment w.r.t
HSP-B6/E1/D1/A4L/A9	MYC
CPF members	synergy AFTER drug treatment w.r.t
HSPA4	MYC

 doi:10.20944/preprints202501.1733.v1

6.2.6. Translation / Ribosomal Protein (TRP)

Under this functional category in [16], the following MYC-target genes are reported - eukaryotic translation initiation factor (EIF-4A1/4E/5A2); Poly(A)-binding protein (PABP); ribosomal protein S (RPS-19/6) and ribosomal protein L (RPL-13A/19/22/27A).

Table 11 and shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 13 generated from analysis of the ranks in Table 11 and 12. The Table 11 shows rankings w.r.t MYC. For genes related to Eukaryotic translation initiation factor (i.e EIF2B3 and EIF2D); Poly(A)-binding protein (i.e PABPC1L); and Ribosomal protein S (i.e RPSA, RPS2, RPS2P46, RPS3A, RPS18, RPS9 and RPS23) all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Eukaryotic translation initiation factor (i.e EIF3L, EIF2B1, EIF2AK4, EIF4B, EIF3E, EIF3F, EIF2B5 and EIF4EBP1); Poly(A)-binding protein (i.e PABPC4) and Ribosomal protein S (i.e RPS5, RPS11, RPS27, RPS4X, RPS3, RPS6KL1, RPS27A, RPS14, RPS21, RPS24, RPS16, RPS19, RPS20, RPS13, RPS7, RPS12, RPS8, RPS29, RPS25 and RPS23); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

The Table 12 shows rankings w.r.t MYC. For genes related to Ribosomal protein L (i.e RPL22, RPL5, RPL7L1, RPL21, RPL10A, RPL15, RPL19, RPL41, RPL23, RPL13A, RPL4 and RPL3) all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Ribosomal protein L (i.e RPL24, RPL39, RPL30, RPL27A, RPL35A, RPL7, RPL7A, RPL38, RPL26, RPL23A, RPL14, RPL6, RPL11, RPL27, RPL12, RPL37A, RPL39L, RPL37, RPL36A, RPL31, RPL9, RPL35, RPL18A, RPL32, RPL34 and RPL13); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Tables 11 and 12 graphically, in Table 13.

 doi:10.20944/preprints202501.1733.v1

Table 11. 2^{nd} order interaction ranking between MYC VS TRP members.

RANKING MYC TARGETS IN TRP					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Eukaryotic translation ...					
initiation factor					
EIF2B3 - MYC	564	871	479	1294	258
EIF2D - MYC	953	1166	985	1353	1174
EIF3L - MYC	1045	1672	1300	2476	2622
EIF2B1 - MYC	1666	1679	1882	1492	2000
EIF2AK4 - MYC	1687	2013	1409	2522	2370
EIF4B - MYC	1737	1744	1799	2615	2565
EIF3E - MYC	1808	1716	1715	1707	1538
EIF3F - MYC	2135	1769	2105	1989	1645
EIF2B5 - MYC	2175	1669	2337	1957	1974
EIF4EBP1 - MYC	2210	2149	2151	1948	86
Poly(A)-binding protein					
PABPC4 - MYC	2288	2226	2301	2040	2073
PABPC1L - MYC	1274	242	1410	1207	2047
Ribosomal protein S					
RPSA - MYC	465	1439	669	1193	779
RPS2 - MYC	800	611	554	1296	1246
RPS5 - MYC	1017	2057	792	2620	2634
RPS2P46 - MYC	1067	875	1081	1185	1097
RPS3A - MYC	1095	1202	1047	1179	1403
RPS18 - MYC	1249	1158	1412	2172	1838
RPS9 - MYC	1317	2441	1217	2634	5
RPS11 - MYC	1528	1874	1697	2256	1955
RPS27 - MYC	1573	2059	1750	2133	2311
RPS4X - MYC	1703	1167	1523	1797	2267
RPS3 - MYC	1757	1329	1903	1824	1960
RPS6KL1 - MYC	1763	2336	1411	2538	2624
RPS27A - MYC	1812	1000	1905	2186	1981
RPS14 - MYC	1886	1751	1836	2473	2404
RPS21 - MYC	1923	2238	1920	2671	2695
RPS24 - MYC	1938	1484	2104	2146	2354
RPS16 - MYC	1975	2296	1828	2284	2455
RPS19 - MYC	1995	2224	1947	2519	2249
RPS20 - MYC	2022	1118	1978	2116	2429
RPS13 - MYC	2048	1162	2230	2511	2590
RPS7 - MYC	2058	1001	2191	2465	12
RPS12 - MYC	2089	553	1898	2093	2101
RPS8 - MYC	2102	1470	1963	2155	2080
RPS29 - MYC	2112	1446	2080	2599	2600
RPS25 - MYC	2165	1255	2003	2152	2147
RPS23 - MYC	924	400	2056	1258	198

 doi:10.20944/preprints202501.1733.v1

Table 12. 2^{nd} order interaction ranking between MYC VS TRP members.

RANKING MYC TARGETS IN TRP					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Ribosomal protein L					
RPL22 - MYC	621	799	1185	1201	1467
RPL5 - MYC	709	932	926	1426	1564
RPL7L1 - MYC	1204	1423	1519	2533	2559
RPL21 - MYC	1224	1263	1435	1737	124
RPL10A - MYC	1230	1512	1357	1958	1366
RPL24 - MYC	1285	1864	1032	2091	1996
RPL15 - MYC	1388	1478	1618	2471	2450
RPL19 - MYC	1434	1245	1135	2376	2388
RPL41 - MYC	1455	1240	1080	749	1648
RPL23 - MYC	1509	1524	1631	1556	1872
RPL13A - MYC	1513	1102	1481	2701	2741
RPL4 - MYC	1516	590	1657	1491	1301
RPL39 - MYC	1557	1818	1508	2311	2414
RPL30 - MYC	1567	1609	1690	2173	2169
RPL27A - MYC	1673	2090	1660	2568	2510
RPL35A - MYC	1682	1743	1632	2610	2628
RPL7 - MYC	1744	1364	1615	1784	1408
RPL7A - MYC	1770	1333	1891	2223	2059
RPL38 - MYC	1799	1835	1655	1940	2052
RPL26 - MYC	1841	2017	1862	2023	2077
RPL23A - MYC	1888	1168	1940	1814	54
RPL14 - MYC	1892	720	1975	2254	2386
RPL6 - MYC	1965	734	1766	1976	1794
RPL11 - MYC	1972	954	2049	2396	2572
RPL27 - MYC	1979	1945	2076	2354	1953
RPL12 - MYC	1990	1148	1984	2692	2705
RPL3 - MYC	2007	1497	1866	1561	1111
RPL37A - MYC	2033	1378	2083	2708	2738
RPL39L - MYC	2038	2079	2048	2625	2692
RPL37 - MYC	2061	2306	1982	2699	2675
RPL36A - MYC	2071	2038	1998	2185	1753
RPL31 - MYC	2194	1690	2355	2698	2743
RPL9 - MYC	2203	2001	2260	2713	2677
RPL35 - MYC	2222	2124	2275	2372	2152
RPL18A - MYC	2236	1814	2153	2220	108
RPL32 - MYC	2345	1538	2309	1942	2028
RPL34 - MYC	2538	1405	2478	2739	2719
RPL13 - MYC	2545	2168	2439	2399	2174

 doi:10.20944/preprints202501.1733.v1

Table 13. 2^{nd} order combinatorial hypotheses between MYC and TRP members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
TRP members	synergy BEFORE drug treatment w.r.t
EIF-2B3/2D	MYC
PABPC1L	MYC
RPS-A/2/2P46/3A/18/9/23	MYC
RPL-22/5/7L1/21/10A/15/19/41/23/13A/4/3	MYC
TRP members	synergy AFTER drug treatment w.r.t
EIF-3L/2B1/2AK4/4B/3E/3F/2B5/4EBP1	MYC
PABPC4	MYC
RPS-5/11/27/4X/3/6KL1/27A/14/21/24/16/... 19/20/13/7/12/8/29/25/23	MYC
RPL-24/39/30/27A/35A/7/7A/38/26/23A/14/6/11/... 27/12/37A/39L/37/36A/31/9/35/18A/32/34/13	MYC

6.2.7. Vesicle Protein / Trafficking (VPT)

Under this functional category in [16], the following MYC-target genes are reported - peroxisomal biogenesis factor (PEX3).

Table 14 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 15 generated from analysis of the ranks in Table 14. The Table 14 shows rankings w.r.t MYC. For genes related to peroxisomal biogenesis factor (i.e PEX5 and PEX6); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to peroxisomal biogenesis factor (i.e PEX3); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 14 graphically, in Table 15.

Table 14. 2^{nd} order interaction ranking between MYC VS VPT members.

RANKING MYC TARGETS IN VPT					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
peroxisomal biogenesis factor					
PEX5 - MYC	868	1142	856	1341	877
PEX3 - MYC	1527	2491	1845	2079	2004
PEX6 - MYC	2471	2463	2436	1456	1863

Table 15. 2^{nd} order combinatorial hypotheses between MYC and VPT members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
VPT members	synergy BEFORE drug treatment w.r.t
PEX-5/6	MYC
VPT members	synergy AFTER drug treatment w.r.t
PEX3	MYC

 doi:10.20944/preprints202501.1733.v1

6.2.8. Carbohydrate (CH)

Under this functional category in [16], the following MYC-target genes are reported - aldehyde dehydrogenases (ALDH2); solute carrier family 2 facilitated glucose transporter (SLC2A4) and aldo-keto reductase family (AKR1A).

Table 16 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 17 generated from analysis of the ranks in Table 16. The Table 16 shows rankings w.r.t MYC. For genes related to aldehyde dehydrogenases (i.e ALDH1B1, ALDH7A1 and ALDH5A1); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to aldehyde dehydrogenases (i.e ALDH3A2, ALDH9A1 and ALDH3A1); solute carrier family 2 facilitated glucose transporter (i.e SLC2A11) and aldo-keto reductase family (i.e AKR1C4); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 16 graphically, in Table 17.

Table 16. 2^{nd} order interaction ranking between MYC VS CH members.

RANKING MYC TARGETS IN CH					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
aldehyde dehydrogenases					
ALDH1B1 - MYC	340	522	516	89	544
ALDH7A1 - MYC	508	1070	205	310	806
ALDH5A1 - MYC	966	418	905	1014	1596
ALDH3A2 - MYC	1719	1872	1363	2076	2085
ALDH9A1 - MYC	1967	1727	1902	2227	2424
ALDH3A1 - MYC	2424	2165	2504	627	643
solute carrier family 2 ... facilitated glucose transporter					
SLC2A11 - MYC	2490	2499	2565	1324	1231
aldo-keto reductase family					
AKR1C4 - MYC	2447	2212	2343	1663	2021

Table 17. 2^{nd} order combinatorial hypotheses between MYC and CH members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
CH members	synergy BEFORE drug treatment w.r.t
ALDH-1B1/7A1/5A1	MYC
CH members	synergy AFTER drug treatment w.r.t
ALDH-3A2/9A1/3A1	MYC
SLC2A11	MYC
AKR1C4	MYC

6.2.9. Energy Metabolism (EM)

Under this functional category in [16], the following MYC-target genes are reported - uncoupling protein/solute carrier family 25 mitochondrial carrier (UCP/SLC25A71 and UCP3/SLC25A9).

Table 18 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 19 generated from analysis of the ranks in Table 18. The Table 18 shows rankings

 doi:10.20944/preprints202501.1733.v1

w.r.t MYC. For genes related to uncoupling protein/solute carrier family 25 mitochondrial carrier (i.e SLC25A27 (UCP4), SLC25A26, SLC25A8 (UCP2), SLC25A19 and SLC25A35); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to uncoupling protein/solute carrier family 25 mitochondrial carrier (i.e SLC25A38, SLC25A14 (UCP5), SLC25A40, SLC25A15 and SLC25A32); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 18 graphically, in Table 19.

Table 18. 2^{nd} order interaction ranking between MYC VS EM members.

RANKING MYC TARGETS IN EM					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
uncoupling protein or solute carrier family 25 ... mitochondrial carrier					
SLC25A27 (UCP4) - MYC	17	62	13	118	393
SLC25A26 - MYC	250	698	58	623	734
SLC25A8 (UCP2) - MYC	539	1596	472	1402	287
SLC25A19 - MYC	761	612	937	1136	1446
SLC25A35 - MYC	1334	472	1060	1428	1698
SLC25A38 - MYC	1577	1483	2108	2078	2260
SLC25A14 (UCP5) - MYC	1669	1754	1495	2326	2317
SLC25A40 - MYC	2151	2448	2107	2339	2506
SLC25A15 - MYC	2415	2573	2550	1165	1738
SLC25A32 - MYC	2645	2527	2551	2003	2214

Table 19. 2^{nd} order combinatorial hypotheses between MYC and EM members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
EM members	synergy BEFORE drug treatment w.r.t MYC
SLC25A-27 (UCP4)/26/8 (UCP2)/19/35	
EM members	synergy AFTER drug treatment w.r.t MYC
SLC25A-38/14 (UCP5)/40/15/32	

6.2.10. Lipid (LPD)

Under this functional category in [16], the following MYC-target genes are reported - Acyl-CoA dehydrogenase family (ACADM).

Table 20 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 21 generated from analysis of the ranks in Table 20. The Table 20 shows rankings w.r.t MYC. For genes related to Acyl-CoA dehydrogenase family (i.e ACADM and ACADSB); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to Acyl-CoA dehydrogenase family (i.e ACAD8); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination

 doi:10.20944/preprints202501.1733.v1

is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 20 graphically, in Table 21.

Table 20. 2^{nd} order interaction ranking between MYC VS LPD members.

RANKING MYC TARGETS IN LPD					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
Acyl-CoA dehydrogenase family					
ACADM - MYC	989	1298	1416	1313	304
ACADSB - MYC	110	264	117	363	416
ACAD8 - MYC	928	1943	847	1734	1744

Table 21. 2^{nd} order combinatorial hypotheses between MYC and LPD members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
LPD members	synergy BEFORE drug treatment w.r.t
ACAD-M/SB	MYC
LPD members	synergy AFTER drug treatment w.r.t
ACAD8	MYC

6.2.11. Nucleotide (NTD)

Under this functional category in [16], the following MYC-target genes are reported - phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthase (PAICS); phosphoribosyl pyrophosphate amidotransferase (PPAT); and deoxycytidine kinase (DCK).

Table 22 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 23 generated from analysis of the ranks in Table 22. The Table 22 shows rankings w.r.t MYC. For genes related to phosphoribosylaminoimidazole carboxylase and phosphoribosylaminoimidazolesuccinocarboxamide synthase (PAICS); and phosphoribosyl pyrophosphate amidotransferase (PPAT); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to deoxycytidine kinase (DCK); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 22 graphically, in Table 23.

Table 22. 2^{nd} order interaction ranking between MYC VS NTD members.

RANKING MYC TARGETS IN NTD					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
PAICS - MYC	931	186	1173	248	502
PPAT - MYC	456	519	451	449	689
DCK - MYC	2520	2258	2523	1698	1604

 doi:10.20944/preprints202501.1733.v1

Table 23. 2^{nd} order combinatorial hypotheses between MYC and NTD members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
NTD members	synergy BEFORE drug treatment w.r.t
PAICS	MYC
PPAT	MYC
NTD members	synergy AFTER drug treatment w.r.t
DCK	MYC

6.2.12. Signal Transduction (STD)

Under this functional category in [16], the following MYC-target genes are reported - cyclophilin peptidylprolyl isomerases (PPID); and phospholipases (PLA2G4A).

Table 24 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 25 generated from analysis of the ranks in Table 24. The Table 24 shows rankings w.r.t MYC. For genes related to cyclophilin peptidylprolyl isomerases (i.e PPIA, PPID, PPIL1 and PPIH); and phospholipases (i.e PLCB4); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to cyclophilin peptidylprolyl isomerases (i.e PPIL3); and phospholipases (i.e PLA2G4A, PLA2G3, PLCG1, PLCB2 and PLCH1); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 24 graphically, in Table 25.

Table 24. 2^{nd} order interaction ranking between MYC VS STD members.

RANKING MYC TARGETS IN STD					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
cyclophilin peptidylprolyl isomerases					
PPIA - MYC	224	1857	546	530	1619
PPID - MYC	853	623	657	1291	1709
PPIL1 - MYC	905	1256	967	1279	1453
PPIH - MYC	949	602	460	813	1425
PPIL3 - MYC	2429	2142	2254	2600	2603
phospholipases					
PLA2G4A - MYC	2479	2501	2656	315	1243
PLA2G3 - MYC	2583	2408	2505	1178	1310
PLCG1 - MYC	1127	1768	987	1849	1729
PLCB2 - MYC	1766	2526	1927	721	703
PLCB4 - MYC	1859	477	1747	698	1306
PLCH1 - MYC	2677	2252	2538	2461	2417

 doi:10.20944/preprints202501.1733.v1

Table 25. 2^{nd} order combinatorial hypotheses between MYC and STD members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
STD members	synergy BEFORE drug treatment w.r.t
PPI-A/D/L1/H	MYC
PLCB4	MYC
STD members	synergy AFTER drug treatment w.r.t
PLA-2G4A/2G3	MYC
PLC-G1/B2/H1	MYC

6.2.13. Nuclear Regulatory Factors (NRF)

Under this functional category in [16], the following MYC-target genes are reported - canonical high mobility group (HMGN2); sirtuins (SIRT1); and OZF (ZNF146).

Table 26 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 27 generated from analysis of the ranks in Table 26. The Table 26 shows rankings w.r.t MYC. For genes related to canonical high mobility group (i.e HMGN3, HMGB2, HMGB3 and HMGB1); and Sirtuins (i.e SIRT3); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to canonical high mobility group (i.e HMGN2 and HMGN5); and OZF (i.e ZNF146); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 26 graphically, in Table 27.

Table 26. 2^{nd} order interaction ranking between MYC VS NRF members.

RANKING MYC TARGETS IN NRF					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
canonical high mobility group					
HMGN3 - MYC	183	79	363	193	367
HMGB2 - MYC	317	819	92	150	275
HMGB3 - MYC	457	1224	774	829	624
HMGB1 - MYC	1318	1348	1093	1618	1849
HMGN2 - MYC	1604	345	1627	874	1876
HMGN5 - MYC	2532	2729	2490	338	1429
Sirtuins					
SIRT3 - MYC	333	1732	127	1290	2044
OZF					
ZNF146 - MYC	883	1908	1313	2246	2378

 doi:10.20944/preprints202501.1733.v1

Table 27. 2^{nd} order combinatorial hypotheses between MYC and NRF members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
NRF members	synergy BEFORE drug treatment w.r.t
HMG-N3/B2/B3/B1	MYC
SIRT3	MYC
NRF members	synergy AFTER drug treatment w.r.t
HMG-N2/N5	MYC
ZNF146	MYC

6.2.14. Nucleolus / RNA-Binding Protein (NRBP)

Under this functional category in [16], the following MYC-target genes are reported - dyskerin pseudouridine synthase 1 (DKC1); stem-loop histone mRNA binding protein (SLBP); nucleolin (NCL); surfeit (SURF6); nucleolar protein (NOL1); heterogeneous nuclear ribonucleoprotein (HNRPA1 and HNRNPA2B1); and RNA binding motif protein (RBM3).

Table 28 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 29 generated from analysis of the ranks in Table 28. The Table 28 shows rankings w.r.t MYC. For genes related to dyskerin pseudouridine synthase 1 (i.e DKC1); nucleolin (i.e NCL); nucleolar protein (i.e NOL8 and NOL6); heterogeneous nuclear ribonucleoprotein (i.e HNRNPA3, HNRNPA1L2, HNRNPA1 and HNRNPD); and RNA binding motif protein (i.e RBMX); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to stem-loop histone mRNA binding protein (i.e SLBP); surfeit (i.e SURF6); nucleolar protein (i.e NOL10, NOL9 and NOL11); heterogeneous nuclear ribonucleoprotein (i.e HNRNPM, HNRNPC, HNRNPA0, HNRNPH3 and HNRNPR); and RNA binding motif protein (i.e RBM19, RBM28 and RBM26); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 28 graphically, in Table 29.

 doi:10.20944/preprints202501.1733.v1

Table 28. 2^{nd} order interaction ranking between MYC VS NRBP members.

RANKING MYC TARGETS IN NRBP					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
dyskerin pseudouridine synthase 1 **DKC1 - MYC**	511	233	359	669	1601
stem-loop histone mRNA binding protein **SLBP - MYC**	2182	1734	2266	2195	31
nucleolin **NCL - MYC**	774	1120	1011	1122	1059
surfeit **SURF6 - MYC**	2073	474	1950	1796	2369
nucleolar protein					
NOL8 - MYC	1187	331	1398	1158	2072
NOL6 - MYC	1331	914	1293	947	786
NOL10 - MYC	1590	1626	1565	1952	1772
NOL9 - MYC	1896	2263	1923	2335	2240
NOL11 - MYC	2004	1456	1759	1859	1457
heterogeneous nuclear ribonucleoprotein					
HNRNPA3 - MYC	48	1299	189	1203	1373
HNRNPM - MYC	1050	2431	1043	1747	2100
HNRNPA1L2 - MYC	1072	393	846	734	244
HNRNPA1 - MYC	1083	1401	1334	1921	1461
HNRNPD - MYC	1140	1697	1419	1087	1947
HNRNPC - MYC	1236	1545	1441	2262	2315
HNRNPA0 - MYC	1838	533	1781	2712	2733
HNRNPH3 - MYC	2313	2497	2352	1226	1647
HNRNPR - MYC	2496	2177	2515	2724	2702
RNA binding motif protein					
RBMX - MYC	1460	240	1133	1681	397
RBM19 - MYC	1655	2350	1684	2687	2727
RBM28 - MYC	1898	993	1732	1599	2161
RBM26 - MYC	1918	1891	1737	2458	2314

Table 29. 2^{nd} order combinatorial hypotheses between MYC and NRBP members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
NRBP members	synergy BEFORE drug treatment w.r.t
DKC1	MYC
NCL	MYC
NOL-8/6	MYC
HNRNP-A3/A1L2/A1/D	MYC
RBMX	MYC
NRBP members	synergy AFTER drug treatment w.r.t
SLBP	MYC
SURF6	MYC
NOL-10/9/11	MYC
HNRNP-M/C/A0/H3/R	MYC
RBM-19/28/26	MYC

 doi:10.20944/preprints202501.1733.v1

6.2.15. Transcription Factors (TF)

Under this functional category in [16], the following MYC-target genes are reported - achaete-scute family bHLH transcription factor 2 (ASCL2); ETS proto-oncogene 2, transcription factor (ETS2); MYB proto-oncogene (MYBL2); E2F transcription factor (E2F1); transcription factor (TCF12); HOXL subclass homeoboxes (HOXD13); NME/NM23 nucleoside diphosphate kinase (NME1); and forkhead boxes (FOXM1).

Table 30 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 31 generated from analysis of the ranks in Table 30. The Table 30 shows rankings w.r.t MYC. For genes related to achaete-scute family bHLH transcription factor 2 (i.e ASCL2); ETS proto-oncogene 2, transcription factor (i.e ETS2); MYB proto-oncogene (i.e MYBL2, MYB and MYBL1); E2F transcription factor (i.e E2F2, E2F7, E2F1 and E2F8); transcription factor (i.e TCF19 and TCF7); HOXL subclass homeoboxes (i.e HOXB9, HOXB8, HOXB5 and HOXA9); NME/NM23 nucleoside diphosphate kinase (i.e NME1); and forkhead boxes (i.e FOXM1); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to E2F transcription factor (i.e E2F5); transcription factor (i.e TCFL5 and TCF3); HOXL subclass homeoboxes (i.e HOXB7, HOXB3, HOXB13, HOXA11 and HOXB4); NME/NM23 nucleoside diphosphate kinase (i.e NME4); and forkhead boxes (i.e FOXA2 and FOXJ1); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 30 graphically, in Table 31.

 doi:10.20944/preprints202501.1733.v1

Table 30. 2^{nd} order interaction ranking between MYC VS TF members.

RANKING MYC TARGETS IN TF					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
achaete-scute family bHLH transcription factor 2					
ASCL2 - MYC	153	356	280	1	476
ETS proto-oncogene 2, transcription factor					
ETS2 - MYC	1118	1412	1415	1020	1614
MYB proto-oncogene					
MYBL2 - MYC	59	26	6	121	545
MYB - MYC	261	638	529	261	1105
MYBL1 - MYC	484	177	249	462	973
E2F transcription factor					
E2F2 - MYC	94	188	76	630	1196
E2F7 - MYC	232	479	795	400	842
E2F1 - MYC	316	1461	294	645	1002
E2F8 - MYC	444	344	241	214	588
E2F5 - MYC	2050	2084	1600	2584	2473
transcription factor					
TCF19 - MYC	1051	1363	1585	2159	2257
TCF7 - MYC	1139	1516	1406	1007	1181
TCFL5 - MYC	2141	1205	2245	1993	1511
TCF3 - MYC	2562	2269	2463	2601	2605
HOXL subclass homeoboxes					
HOXB7 - MYC	284	2116	456	1816	2122
HOXB9 - MYC	295	306	321	533	338
HOXB8 - MYC	918	462	767	472	447
HOXB5 - MYC	937	1220	645	1665	791
HOXA9 - MYC	1387	1054	1326	1677	1316
HOXB3 - MYC	1554	2454	1128	2736	2726
HOXB13 - MYC	2093	1812	2002	2680	2564
HOXA11 - MYC	2224	1386	2069	2021	2281
HOXB4 - MYC	2727	2671	2728	1504	946
NME/NM23 nucleoside diphosphate kinase					
NME1 - MYC	911	351	896	385	672
NME4 - MYC	2287	2386	2181	2233	2110
forkhead boxes					
FOXM1 - MYC	127	65	262	68	499
FOXA2 - MYC	1064	1644	864	2164	1813
FOXJ1 - MYC	2533	2736	2610	324	953

 doi:10.20944/preprints202501.1733.v1

Table 31. 2^{nd} order combinatorial hypotheses between MYC and TF members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
TF members	synergy BEFORE drug treatment w.r.t
ASCL2	MYC
ETS2	MYC
MYB, MYBL-1/2	MYC
HNRNP-A3/A1L2/A1/D	MYC
E2F-2/7/1/8	MYC
TCF-19/7	MYC
HOX-B9/B8/B5/A9	MYC
NME1	MYC
FOXM1	MYC
TF members	synergy AFTER drug treatment w.r.t
E2F5	MYC
TCF-L5/3	MYC
HOX-B7/B3/B13/A11/B4	MYC
NME4	MYC
FOX-A2/J1	MYC

6.2.16. DNA Maintenance / Repair (DMR)

Under this functional category in [16], the following MYC-target genes are reported - apurinic/apyrimidinic endodeoxyribonuclease (APEX1); telomerase reverse transcriptase (TERT); prothymosin alpha (PTMA); DNA polymerases (POLB and POLD2); H2A histones (H2AZ); minichromosome maintenance complex component (MCM7); BRCA1/BRCA2-containing complex (BRCA2); and DNA topoisomerase (TOP1).

Table 32 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 33 generated from analysis of the ranks in Table 32. The Table 32 shows rankings w.r.t MYC. For genes related to apurinic/apyrimidinic endodeoxyribonuclease (i.e APEX1); DNA polymerases (i.e POLQ, POLE2, POLA1, POLD1, POLG2, POLE3, POLA2 and POLD2); H2A histones (i.e H2AFV, H2AFZ and H2AFX); minichromosome maintenance complex component (i.e MCM4, MCM8, MCM3, MCM10, MCM2, MCM5, MCM6 and MCM7); BRCA1/BRCA2-containing complex (i.e BRCA1 and BRCA2); and DNA topoisomerase (i.e TOP2A and TOP1MT); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to telomerase reverse transcriptase (i.e TERT); prothymosin alpha transcription factor (i.e PTMA); DNA polymerases (i.e POLB); and DNA topoisomerase (i.e TOP2B); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 32 graphically, in Table 33.

 doi:10.20944/preprints202501.1733.v1

Table 32. 2^{nd} order interaction ranking between MYC VS DMR members.

RANKING MYC TARGETS IN DMR					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
apurinic/apyrimidinic endodeoxyribonuclease					
APEX1 - MYC	514	433	918	478	1401
telomerase reverse transcriptase					
TERT - MYC	2724	2730	2740	217	525
prothymosin alpha transcription factor					
PTMA - MYC	2300	1665	2261	2017	2453
DNA polymerases					
POLQ - MYC	235	227	317	3	554
POLE2 - MYC	238	528	83	440	1207
POLA1 - MYC	404	386	431	463	324
POLD1 - MYC	489	281	470	654	855
POLG2 - MYC	776	621	466	1157	1517
POLE3 - MYC	1094	1310	1407	1419	1084
POLA2 - MYC	1259	830	1233	1591	188
POLB - MYC	1380	1998	1469	1984	2092
POLD2 - MYC	1463	1435	1202	1075	1822
H2A histones					
H2AFV - MYC	962	1060	1262	1423	139
H2AFZ - MYC	1093	1008	1016	1527	1520
H2AFX - MYC	2134	859	2096	1468	246
minichromosome maintenance complex component					
MCM4 - MYC	80	1194	259	271	570
MCM8 - MYC	114	531	29	438	1089
MCM3 - MYC	248	736	413	335	975
MCM10 - MYC	313	575	126	79	284
MCM2 - MYC	315	560	133	122	1369
MCM5 - MYC	337	229	462	495	724
MCM6 - MYC	500	843	507	170	889
MCM7 - MYC	605	505	1062	614	872
BRCA1/BRCA2-containing complex					
BRCA1 - MYC	383	967	661	88	1094
BRCA2 - MYC	339	29	596	71	513
DNA topoisomerase					
TOP2A - MYC	131	6	341	44	718
TOP1MT - MYC	783	641	739	756	1154
TOP2B - MYC	1772	2097	1548	2632	2681

 doi:10.20944/preprints202501.1733.v1

Table 33. 2^{nd} order combinatorial hypotheses between MYC and DMR members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
DMR members	synergy BEFORE drug treatment w.r.t
APEX1	MYC
POL-Q/E2/A1/D1/G2/E3/A2/D2	MYC
H2A-FV/FZ/FX	MYC
MCM4/8/3/10/2/5/6/7	MYC
BRCA-1/2	MYC
TOP-2A/1MT	MYC
DMR members	synergy AFTER drug treatment w.r.t
TERT	MYC
PTMA	MYC
POLB	MYC
TOP2B	MYC

Additionally, since members of DNA polymerases are MYC targets, I also analysed rankings of combinations of members of DNA-dependent RNA polymerase (POLR), with MYC.

Table 34 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 35 generated from analysis of the ranks in Table 34. The Table 34 shows rankings w.r.t MYC. For genes related to DNA-dependent RNA polymerase (i.e POLR1D, POLR3K, POLR1E, POLR1B, POLR2G, POLR2K, POLR3E, POLR3A and POLR1C); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to DNA-dependent RNA polymerase (i.e POLR2D, POLR2F and POLR1A); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 34 graphically, in Table 35.

Table 34. 2^{nd} order interaction ranking between MYC VS DMR members.

RANKING MYC WITH POLR MEMBERS					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
DNA-dependent RNA polymerase					
POLR1D - MYC	364	1383	553	556	1489
POLR3K - MYC	370	165	619	504	1567
POLR1E - MYC	545	561	749	1038	437
POLR1B - MYC	932	795	711	688	656
POLR2G - MYC	1037	1075	1236	1474	1180
POLR2K - MYC	1270	1678	935	2368	102
POLR3E - MYC	1286	1369	1225	2590	2511
POLR2D - MYC	1325	1641	1630	2211	2533
POLR3A - MYC	1692	1024	1107	1738	1036
POLR1C - MYC	1758	971	1461	1238	971
POLR2F - MYC	1842	2233	1711	2689	2711
POLR1A - MYC	1885	1145	1815	1101	1665

 doi:10.20944/preprints202501.1733.v1

Table 35. 2^{nd} order combinatorial hypotheses between MYC and POLR members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
POLR members	synergy BEFORE drug treatment w.r.t MYC
POLR-1D/3K/1E/1B/2G/2K/3E/3A/1C	
POLR members	synergy AFTER drug treatment w.r.t MYC
POLR-2D/2F/1A	

6.2.17. Other (OT)

Under this functional category in [16], the following MYC-target genes are reported - epoxide hydrolase (EPHX1); inositol monophosphatase (IMPA2); pyruvate dehydrogenase (PDHA1); microsomal glutathione S-transferase (MGST1); and solute carrier family 19 (SLC19A1).

Table 36 shows rankings of these combinations. Followed by this is the unexplored combinatorial hypotheses in Table 37 generated from analysis of the ranks in Table 36. The Table 36 shows rankings w.r.t MYC. For genes related to epoxide hydrolase (i.e EPHX2); inositol monophosphatase (i.e IMPA2); pyruvate dehydrogenase (i.e PDHA1); microsomal glutathione S-transferase (i.e MGST1); and solute carrier family 19 (i.e SLC19A1 and SLC19A3); all show majority low rankings (below half way mark) across the SA methods. That is, the low rankings indicate that a combination is weakly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in up regulated manner in CRC BEFORE the drug treatment.

On the other hand, genes related to microsomal glutathione S-transferase (i.e MGST2); all show majority high rankings (above half way mark). That is, the high rankings indicate that a combination is strongly down regulated after ETC-1922159 treatment. Thus, they might be working synergistically in down regulated manner in CRC AFTER the drug treatment.

One can also interpret the results of the Table 36 graphically, in Table 37.

Table 36. 2^{nd} order interaction ranking between MYC VS OT members.

RANKING MYC TARGETS IN OT					
MYC - X	HSIC			SOBOL	
	laplace	linear	rbf	2002	jansen
epoxide hydrolase					
EPHX2 - MYC	1155	1510	1255	1668	1440
inositol monophosphatase					
IMPA2 - MYC	445	198	743	1120	1187
pyruvate dehydrogenase					
PDHA1 - MYC	1242	915	892	862	1843
microsomal glutathione					
S-transferase					
MGST1 - MYC	1002	1832	1018	2626	13
MGST2 - MYC	2289	1634	2334	2606	2703
solute carrier family 19					
SLC19A1 - MYC	111	24	475	360	228
SLC19A3 - MYC	62	80	59	200	451

 doi:10.20944/preprints202501.1733.v1

Table 37. 2^{nd} order combinatorial hypotheses between MYC and OT members.

UNEXPLORED COMBINATORIAL HYPOTHESES	
OT members	synergy BEFORE drug treatment w.r.t
EPHX2	MYC
IMPA2	MYC
PDHA1	MYC
MGST1	MYC
SLC19A-1/3	MYC
OT members	synergy AFTER drug treatment w.r.t
MGST2	MYC

7. Conclusion

Findings of this work point to conserved machine learning rankings of gene combinations across different sensitivity methods. These findings point to the existence of biological synergy among the genes, for experimentally tested combinations as well as those that have to be explored/tested. Results for MYC related combinations in CRC cells treated with ETC-1922159, are presented. A theoretically sound and a practical framework has been developed to prioritize higher order combinations of regulated genes after the administration of ETC-1922159 PORCN-WNT inhibitor in cancer cells. The prioritization uses advanced sensitivity methods that exploit nonlinear relations in reproducing kernel hilbert spaces via kernel trick and support vector ranking method to rank and reveal various combinations of identified and unidentified factors that are affected after the drug treatment. This gives medical specialists/oncologists/biologists a way to navigate in a guided manner in a vast combinatorial search forest, thus cutting down cost in time/investment/energy as well as avoid cherry picking unknown biological hypotheses. Biologists/oncologists would not have to struggle to search for gene combinations of interest, which they might want to test in wet lab.

Author Contributions: SS designed, developed and implemented the insilico experimental setup, wrote the code, generated and analysed the results and wrote the manuscript.

Data Availability Statement: Data used in this research work has been released online publicaly, in a publication in [2]. This data was made available in the form of supplementary table. Related to this data, on NCBI Gene Expression Omnibus (GEO) Series GSE69687 https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE69687, click on **Download RNA-seq counts** button, it opens a page that contains **Human gene annotation table** (at the bottom of the page). The file **Human.GRCh38.p13.annot.tsv.gz** contains the range of genes with ENSEMBLGENEID all starting with ENSG. This ENSG identifier is used to index the recording of the regulated genes, in the data made available in the supplementary table in [2]. The data itself is available as supplementary material in the journal, however, the indexing of the genes used in the supplementary material is available on NCBI NIH database.

Acknowledgments: Special thanks to Mrs. Rita Sinha and late Mr. Prabhat Sinha for supporting the author financially, without which this work could not have been made possible.

Conflicts of Interest: There are no conflicts to declare.

Code Availability: Code of the search engine used to generate the rankings has been made available on CERN based Zenodo at https://zenodo.org/records/14636112.

 doi:10.20944/preprints202501.1733.v1

References

1. Sinha, S. Machine learning ranking of plausible (un)explored synergistic gene combinations using sensitivity indices of time series measurements of Wnt signaling pathway. *Integrative Biology* **2024**.
2. Madan, B.; Ke, Z.; Harmston, N.; Ho, S.Y.; Frois, A.; Alam, J.; Jeyaraj, D.A.; Pendharkar, V.; Ghosh, K.; Virshup, I.H.; et al. Wnt addiction of genetically defined cancers reversed by PORCN inhibition. *Oncogene* **2016**, *35*, 2197.
3. Sinha, S. Sensitivity analysis based ranking reveals unknown biological hypotheses for down regulated genes in time buffer during administration of PORCN-WNT inhibitor ETC-1922159 in CRC. *bioRxiv* **2017**, p. 180927.
4. Chen, B.; Dodge, M.E.; Tang, W.; Lu, J.; Ma, Z.; Fan, C.W.; Wei, S.; Hao, W.; Kilgore, J.; Williams, N.S.; et al. Small molecule–mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nature chemical biology* **2009**, *5*, 100–107.
5. Wang, X.; Moon, J.; Dodge, M.E.; Pan, X.; Zhang, L.; Hanson, J.M.; Tuladhar, R.; Ma, Z.; Shi, H.; Williams, N.S.; et al. The development of highly potent inhibitors for porcupine. *Journal of medicinal chemistry* **2013**, *56*, 2700–2704.
6. Proffitt, K.D.; Madan, B.; Ke, Z.; Pendharkar, V.; Ding, L.; Lee, M.A.; Hannoush, R.N.; Virshup, D.M. Pharmacological inhibition of the Wnt acyltransferase PORCN prevents growth of WNT-driven mammary cancer. *Cancer research* **2013**, *73*, 502–507.
7. Liu, J.; Pan, S.; Hsieh, M.H.; Ng, N.; Sun, F.; Wang, T.; Kasibhatla, S.; Schuller, A.G.; Li, A.G.; Cheng, D.; et al. Targeting Wnt-driven cancer through the inhibition of Porcupine by LGK974. *Proceedings of the National Academy of Sciences* **2013**, *110*, 20224–20229.
8. Duraiswamy, A.J.; Lee, M.A.; Madan, B.; Ang, S.H.; Tan, E.S.W.; Cheong, W.W.V.; Ke, Z.; Pendharkar, V.; Ding, L.J.; Chew, Y.S.; et al. Discovery and optimization of a porcupine inhibitor. *Journal of medicinal chemistry* **2015**, *58*, 5889–5899.
9. Nair, S.K.; Burley, S.K. X-ray structures of Myc-Max and Mad-Max recognizing DNA: molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **2003**, *112*, 193–205.
10. Dang, C.V. c-Myc target genes involved in cell growth, apoptosis, and metabolism. *Molecular and cellular biology* **1999**, *19*, 1–11.
11. Hoffman, B.; Liebermann, D. Apoptotic signaling by c-MYC. *Oncogene* **2008**, *27*, 6462–6473.
12. Miller, D.M.; Thomas, S.D.; Islam, A.; Muench, D.; Sedoris, K. c-Myc and cancer metabolism, 2012.
13. Knoepfler, P.S.; Zhang, X.y.; Cheng, P.F.; Gafken, P.R.; McMahon, S.B.; Eisenman, R.N. Myc influences global chromatin structure. *The EMBO journal* **2006**, *25*, 2723–2734.
14. Sikora, K.; Chan, S.; Evan, G.; Gabra, H.; Markham, N.; Stewart, J.; Watson, J. c-myc oncogene expression in colorectal cancer. *Cancer* **1987**, *59*, 1289–1295.
15. Smith, D.; Myint, T.; Goh, H. Over-expression of the c-myc proto-oncogene in colorectal carcinoma. *British journal of cancer* **1993**, *68*, 407.
16. Fernandez, P.C.; Frank, S.R.; Wang, L.; Schroeder, M.; Liu, S.; Greene, J.; Cocito, A.; Amati, B. Genomic targets of the human c-Myc protein. *Genes & development* **2003**, *17*, 1115–1129.
17. Sinha, S. Hilbert-Schmidt and Sobol sensitivity indices for static and time series Wnt signaling measurements in colorectal cancer-part A. *BMC systems biology* **2017**, *11*, 120.
18. Joachims, T. Training linear SVMs in linear time. In Proceedings of the Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006, pp. 217–226.
19. Sobol', I.M. On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie* **1990**, *2*, 112–118.
20. Iooss, B.; Lemaître, P. A review on global sensitivity analysis methods. *arXiv preprint arXiv:1404.2405* **2014**.
21. Morris, M.D. Factorial sampling plans for preliminary computational experiments. *Technometrics* **1991**, *33*, 161–174.
22. Moon, H.; Dean, A.M.; Santner, T.J. Two-stage sensitivity-based group screening in computer experiments. *Technometrics* **2012**, *54*, 376–387.
23. Dean, A.; Lewis, S. *Screening: methods for experimentation in industry, drug discovery, and genetics*; Springer Science & Business Media, 2006.
24. Andres, T.H.; Hajas, W.C. Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model **1993**.

 doi:10.20944/preprints202501.1733.v1

25. Bettonvil, B.; Kleijnen, J.P. Searching for important factors in simulation models with many factors: Sequential bifurcation. *European Journal of Operational Research* **1997**, *96*, 180–194.
26. Cotter, S.C. A screening design for factorial experiments with interactions. *Biometrika* **1979**, *66*, 317–320.
27. Christensen, R. *Linear models for multivariate, time series, and spatial data*; Springer Science & Business Media, 1991.
28. Saltelli, A.; Chan, K.; Scott, E. Sensitivity analysis Wiley series in probability and statistics. *Willey, New York* **2000**.
29. Helton, J.C.; Davis, F.J. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety* **2003**, *81*, 23–69.
30. McKay, M.D.; Beckman, R.J.; Conover, W.J. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **1979**, *21*, 239–245.
31. Homma, T.; Saltelli, A. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety* **1996**, *52*, 1–17.
32. Sobol, I.M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and computers in simulation* **2001**, *55*, 271–280.
33. Saltelli, A. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications* **2002**, *145*, 280–297.
34. Saltelli, A.; Ratto, M.; Tarantola, S.; Campolongo, F. Sensitivity analysis for chemical models. *Chemical reviews* **2005**, *105*, 2811–2828.
35. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global sensitivity analysis: the primer*; John Wiley & Sons, 2008.
36. Cukier, R.; Fortuin, C.; Shuler, K.E.; Petschek, A.; Schaibly, J. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory. *The Journal of Chemical Physics* **1973**, *59*, 3873–3878.
37. Saltelli, A.; Tarantola, S.; Chan, K.S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics* **1999**, *41*, 39–56.
38. Tarantola, S.; Gatelli, D.; Mara, T.A. Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety* **2006**, *91*, 717–727.
39. Saltelli, A.; Annoni, P.; Azzini, I.; Campolongo, F.; Ratto, M.; Tarantola, S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications* **2010**, *181*, 259–270.
40. Janon, A.; Klein, T.; Lagnoux, A.; Nodet, M.; Prieur, C. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics* **2014**, *18*, 342–364.
41. Owen, A.B. Better estimation of small Sobol'sensitivity indices. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **2013**, *23*, 11.
42. Tissot, J.Y.; Prieur, C. Bias correction for the estimation of sensitivity indices based on random balance designs. *Reliability Engineering & System Safety* **2012**, *107*, 205–213.
43. Da Veiga, S.; Gamboa, F. Efficient estimation of sensitivity indices. *Journal of Nonparametric Statistics* **2013**, *25*, 573–595.
44. Archer, G.; Saltelli, A.; Sobol, I. Sensitivity measures, ANOVA-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation* **1997**, *58*, 99–120.
45. Tarantola, S.; Gatelli, D.; Kucherenko, S.; Mauntz, W.; et al. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety* **2007**, *92*, 957–960.
46. Saltelli, A.; Annoni, P. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software* **2010**, *25*, 1508–1517.
47. Jansen, M.J. Analysis of variance designs for model output. *Computer Physics Communications* **1999**, *117*, 35–43.
48. Storlie, C.B.; Helton, J.C. Multiple predictor smoothing methods for sensitivity analysis: Description of techniques. *Reliability Engineering & System Safety* **2008**, *93*, 28–54.
49. Da Veiga, S.; Wahl, F.; Gamboa, F. Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics* **2009**, *51*, 452–463.
50. Li, G.; Rosenthal, C.; Rabitz, H. High dimensional model representations. *The Journal of Physical Chemistry A* **2001**, *105*, 7765–7777.

 doi:10.20944/preprints202501.1733.v1

51. Hajikolaei, K.H.; Wang, G.G. High Dimensional Model Representation With Principal Component Analysis. *Journal of Mechanical Design* **2014**, *136*, 011003.
52. Borgonovo, E. A new uncertainty importance measure. *Reliability Engineering & System Safety* **2007**, *92*, 771–784.
53. Sobol, I.; Kucherenko, S. Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation* **2009**, *79*, 3009–3017.
54. Fort, J.C.; Klein, T.; Rachdi, N. New sensitivity analysis subordinated to a contrast. *arXiv preprint arXiv:1305.2329* **2013**.
55. Da Veiga, S. Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation* **2015**, *85*, 1283–1305.
56. Székely, G.J.; Rizzo, M.L.; Bakirov, N.K.; et al. Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **2007**, *35*, 2769–2794.
57. Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. Measuring statistical dependence with Hilbert-Schmidt norms. In Proceedings of the Algorithmic learning theory. Springer, 2005, pp. 63–77.
58. Csiszár, I.; et al. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.* **1967**, *2*, 299–318.
59. Aizerman, A.; Braverman, E.M.; Rozoner, L. Theoretical foundations of the potential function method in pattern recognition learning. *Automation and remote control* **1964**, *25*, 821–837.
60. Pujol, G.; Iooss, B.; Janon, A.; Boumhaout, K.; Da Veiga, S.; Fruth, J.; Gilquin, L.; Guillaume, J.; Le Gratiet, L.; Lemaitre, P.; et al. Sensitivity: global sensitivity analysis of model outputs. *R package version* **2017**, *1*.
61. Tusher, V.G.; Tibshirani, R.; Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proceedings of the National Academy of Sciences* **2001**, *98*, 5116–5121.
62. Choe, S.E.; Boutros, M.; Michelson, A.M.; Church, G.M.; Halfon, M.S. Preferred analysis methods for Affymetrix GeneChips revealed by a wholly defined control dataset. *Genome biology* **2005**, *6*, R16.
63. Witten, D.; Tibshirani, R. A comparison of fold-change and the t-statistic for microarray data analysis. *Analysis* **2007**, *1776*, 58–85.