

Predicting Cancer Immunotherapy Resistance with Topology of Biological Networks

Aditya Karan & Kar-Tong Tan

December 2018

Abstract

Cancer immunotherapy has become a frontline therapeutic for cancer. However, resistance towards cancer immunotherapy is extremely prevalent among patients, causing them to be refractory to these treatments. Network modelling of biochemical networks governing cancer immunotherapy response can be used to better understand the mechanisms governing cancer immunotherapy resistance in cancer patients, but is hindered by the absence of the kinetic parameters underlying these networks. A recent study indicate that knowledge of network topology alone, without these kinetic parameters, is sufficient to make accurate predictions of gene perturbations. Here, we apply methods and principles highlighted by this study to develop a topology based network describing cancer immunotherapy resistance. We report the performance of topology based networks by comparison with gene knockdown experiments, and CRISPR knockout screens for gene targets that govern cancer immunotherapy resistance. Overall, the topology based approach was able to predict the direction of interactions reasonably well, however, correlations on larger networks with experimental evidences were not that strong - suggesting either a lack of information of the systems or a shortcoming of the topology based approach.

1 Introduction

Cancer immunotherapy has emerged as the long sought after "Cure for Cancer", and is an extremely promising therapeutic for millions of cancer patients. This is especially highlighted by the recent award of the Nobel Prize in Physiology or Medicine to two pioneers in the cancer immunotherapy field, James P. Allison and Tasuku Honjo, for foundational work done on checkpoint blockade inhibitors, a type of cancer immunotherapy, that has prove to be highly effective.

Checkpoint blockade inhibitors have become a standard of care for cancers like melanoma which typically carry a high number of mutations. Due to the high number of mutations these cancerous cells carry, they express many mutated proteins that causes these cells to be recognized as "foreign cells", and

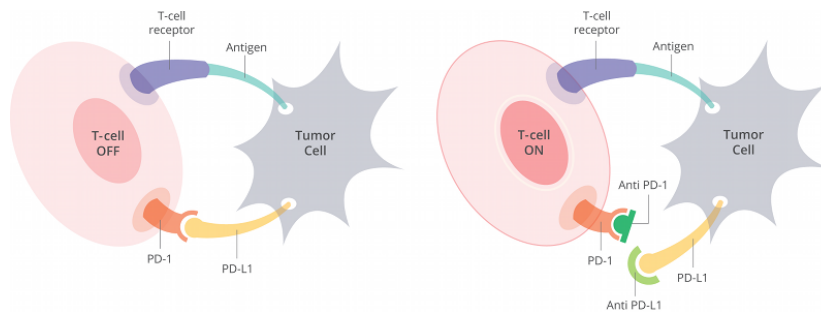


Figure 1: Mechanism of PD-L1 in checkpoint blockade inhibition in cancer immunotherapy. Figure was adapted from <https://www.smartpatients.com/targets/pd-1>

killed off by the body’s immune system. However, cancer cells typically express, PD-L1, an inhibitor of T-cells, on its cell surface which causes which allow them to escape the immune system and prevent themselves from being killed off (Figure 1). Checkpoint blockade inhibitors thus work by inhibiting the PD-L1 protein necessary for immune escape, thereby allowing cancerous cells to be recognized and eliminated by the patient’s intrinsic immune system.

While checkpoint blockade inhibition has enabled cancers of some patients to be treated and eliminated, more than 60% of patients were found to be refractory to this new standard of care in melanoma, a cancer of the skin [1]. To determine mechanisms for cancer immunotherapy resistance, a number of groups had performed high-throughput screens to identify gene targets that sensitizes cancer cells to cancer immunotherapy, leading to the discovery of a large number of targets like Ptpn2 and chromatin regulators [2, 3, 4]. However, given the large number of targets identified from these studies, such data sets are difficult to interpret directly. In addition, the relative importance of each of these genes within the biological network is difficult to establish. Thus, better knowledge of the biological networks that underlie checkpoint inhibition, and how they might govern resistance would certainly greatly improve our understanding of how resistance to checkpoint inhibition occurs.

Biological networks describe the relationship between different genes and proteins. When full knowledge of the biochemical networks that underlie these network, including how their interact, and the chemical constants describing these processes are known, one can easily predict the effect of perturbing one gene on the other components of the biological networks. However kinetic parameters describing these biochemical networks are often unavailable, with only knowledge of the topology of these biological networks typically known. Interestingly, a recent study had highlighted how knowledge of the network topology alone was sufficient for attaining a 65–80% accuracy in predicting the impact of perturbation patterns [5], thus enabling the large number of topology based networks which are currently available to be utilized for the study of a large range of biological processes. In particular, the study developed three topology

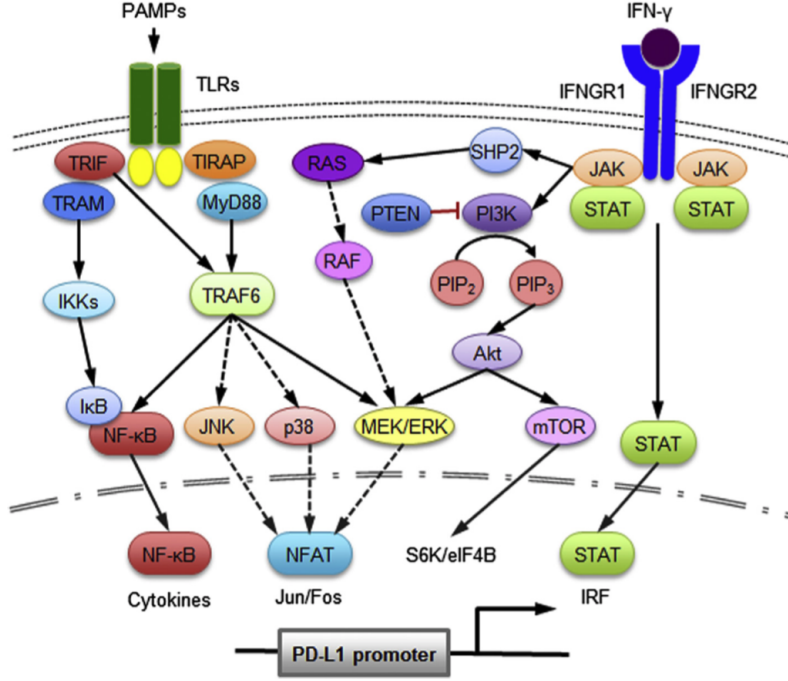


Figure 2: Biological pathway depicting genes that regulates PD-L1 expression. Figure was extracted from Ritprajak et al [6]

based approaches - Propagation, Distance, and First Neighbours (Section 3) for inferring the impact on the network on gene perturbation.

Given that detailed kinetic parameters describing biological pathways determining checkpoint blockade inhibition is lacking, we ask if the topology based approaches introduced in this previous study is sufficient for us to make predictions that are concordant with experimentation. In addition, the agreement between a less complex network model with a more detailed model would also be assessed. Overall, the results would enable us to garner a better understanding of the mechanisms underlying cancer immunotherapy resistance.

2 Datasets

2.1 PD-L1 pathway

We extracted the PD-L1 pathway which was published by Ritprajak et al [6] and also depicted in Figure 2. We then manually curated the pathway and introduced nodes and edges linking each of components in the network. Specifically, when genes form a co-complex for activation of downstream pathways, a ghost node was introduced to represent the co-complex. This co-complex was

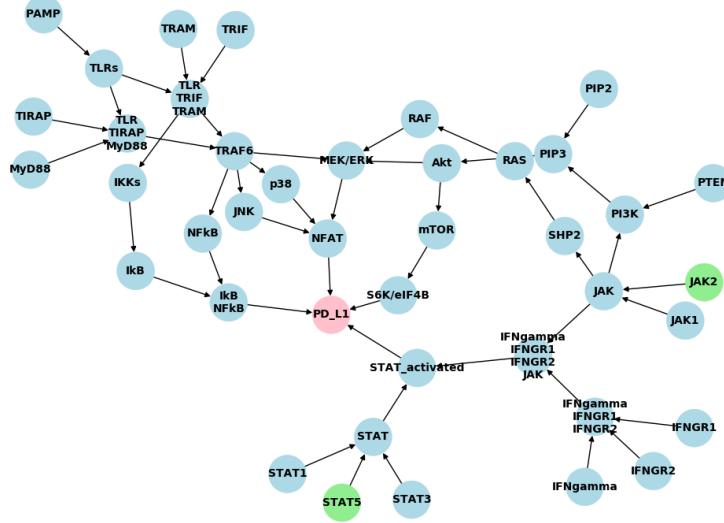


Figure 3: PD-L1 network generated from the PD-L1 biological pathway. The genes that were perturbed in this study are highlighted in green and the PD-L1 gene is highlighted in pink.

then linked to downstream components. In addition, we also noted cases in which certain genes indicated in this schematic was represented by multiple degenerate components. In this case, we again created a ghost node to represent the degenerate component, with constituent components pointing towards this generic component (e.g. STAT1 and STAT5 to the generic component STAT5). Overall, this allowed us to generate the network as observed in Figure 3.

2.2 Gene expression data

Two previously published gene expression datasets were utilized to determine the impact of gene perturbations on PD-L1 expression, and possible impact on immunotherapy response. Specifically, JAK2 knockdown (GSE54645), and STAT5 knockdown (GSE68993) datasets were utilized for this study. Gene expression datasets were downloaded in R using the GEOquery library in Bioconductor (cite). The probeset information for each dataset was similarly downloaded using GEOquery. Gene names were then annotated using the probeset information. Where gene names information were not available, gene symbols were annotated by comparing the chromosome, start, end, and strand information in the probeset to the UCSC known gene annotations.

2.3 Genes that confer resistance to immunotherapy from CRISPR knockout screens

The list of genes that confers sensitivity to cancer immunotherapy, as identified from CRISPR knockout screens, were extracted from three different publications [2, 3, 4]. These genes were then manually curated and compared against the list of genes in our biological network and pathway.

3 Numerical Methods

3.1 Validating the influence networks

We first evaluated Santolini's [5] approach against an existing BioModel (BIOMD0000000313) from the BioModels Database [7, 8] an IL-13 pathway involving PD-L1. Using the provided code, we converted this BioModel into a Matlab function which represents a system of ODEs. The database also provided a steady state.

To verify this numerical steady state - we tried both numerically solving the steady state via non-linear equation solving, as well as running an ODE solver (45, 23) and comparing the results. A direct, non-linear solver proved numerically unstable while the ODE solver seeded at the provided steady state matched closely with the database provided value.

For the influence networks (labeled DYNAMO or dynamical-agnostic models) we implemented 3 classes of models found in the original study (Figure 4). For each of these networks we compute a specific Sensitivity matrix - a square matrix, where the (i, j) component represents the influence of component j on quantity i . This sensitivity matrix can then be used to make predictions of the direction of a perturbation (the sign of the (i, j) entry) as well as the relative strength (the absolute value of the entry) [5]

The classes of models are:

- Propagation networks. From an adjacency matrix, A , we construct degree matrices. D_1 and D_2 . D_1 is a diagonal matrix where the i th component is the sum of all the in-degrees for node i and for D_2 the i th component is the sum of the out-degrees of node i . We then compute the normalized weight $W = D_1^{-\frac{1}{2}} A D_2^{-\frac{1}{2}}$ where $D_1^{-\frac{1}{2}}$ is simply $\frac{1}{\sqrt{d_{i_j}}}$ (since D_1 is just a diagonal matrix). To compute the sensitivity matrix, we calculate $S = (1 - \alpha)(I - \alpha W)^{-1}$ with $\alpha = 0.9$
- Distance matrices: Define matrix D where D_{ij} as the number of edges needed to go from i to j . The Sensitivity matrix is then the inverse of the distance - more specifically $S_{ij} = 1/(1 + D_{ij})$
- First neighbor: A simple adjacency network as the sensitivity matrix.

The propagation networks can come in the form of a signed directed networked (aware of whether the species up or down regulates) a unsigned directed

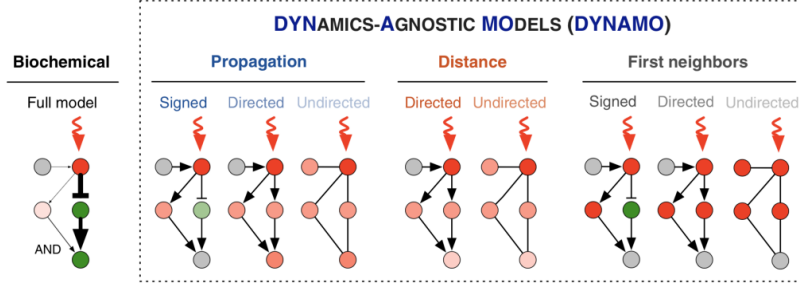


Figure 4: Illustration of biochemical network models, and the propagation, distance, and neighbour topology based network models

network (knowledge of the direction of regulation up-regulates or down regulates) or a unsigned undirected networks (knowledge that A and B interact but no direction)

We re-implemented this method in python (see supporting code) to solve the various sensitivity matrices for the different models.

In addition - we decided to try two other approaches for comparisons sake

- Eigenvectors of the adjacency matrix.
- Katz-centrality scaled. Here we computed the Katz centrality for each node, and then applied this weighting factor to the adjacency matrix. The Katz centrality of a node can be define as $C(x_i) = \alpha \sum A_{j,i} C(x_j) + \beta$ where β is some bias term. This can be written as a matrix eigenvector problem, where we have $C = \alpha A^T C + \beta \mathbf{1}$. Solving this out we have $C = \beta(I - \alpha A^T)^{-1} \mathbf{1}$ Because of the inverse involved - most solvers (including the one we use) instead use a power iteration to calculate this value. Once we have the centrality values of each nodes, we scale the adjacency by the centrality. In effect this creates a first neighbor model, where the value of a neighbor of node i is scaled by the centrality score.

In both these cases we computed the sensitivity matrix using the same formulas as the propagation network ($S = (1 - \alpha)(I - \alpha W)^{-1}$ where W is either the eigenvectors or the Katz weighed neighbors. We believe that these models may also perform well because they encode some information about the "global" influence (ex: page-rank like approach) may prove to be an interesting comparison. Katz centrality in particular has been shown to be a good predictor in neuron firing rates and perhaps similar dynamics may be found. [9]

To evaluate these proposed models - we employ two different statistics to test both the accuracy of the "strength" of influence as well as the direction. To test the direction, we simply looked at the signs of the different similarity matrices and check to see how many components are matching. To test the strength, we employ the Spearmans rank correlation coefficient on the absolute value of the various sensitivity matrices. By doing so, we can observe whether

the relative ordering of predicted strength of an interaction is preserved across the biochemical model and the newly proposed ones. We consider the absolute value of the coefficients in order to isolate the analysis of the strength of the correlations from the direction.

3.2 Evaluating on less understood network

In order to see whether this analysis could work in a more complex system - a published path - but one without dynamical information known, we performed further analysis using the PD-L1 biological network extracted from a recent publication [6]. Figure 3 shows the relationship between the various components. Compared to the known BioModel pathway - it is significantly more complex and sparse.

Using JAK2 knockout data, we computed the various influence models as described in the previous sections. To define the directional changes - we looked the change in mRNA expression from experimental to control for various genes. if the change in expression was more than 10% we consider it to be up-regulated. Any change in expression down 10% or more was considered down-regulated. Values outside of that range were labeled as unchanged. To evaluate the accuracy of the strength prediction - we employed Spearman to see if the rank ordering of the strengths predicted by the influence models were similar the experimental data. Similar analyses were also performed using a STAT5 knock-down dataset.

For the PD-L1 test, we then compared the list of genes that were predicted to cause PD-L1 downregulation, and thus sensitivity to cancer immunotherapy to a number of recent studies. These studies had utilized whole genome CRISPR knockout screens to identify genes that cause cancerous cells to be more sensitive to these therapeutics. Indeed, if PD-L1 is predicted to be down-regulated in our network, then it should confer cancerous cells sensitivity towards checkpoint blockade cancer immunotherapy as is observed in these experiments.

4 Results

4.1 Biomodel

Figure 5 shows the result of the Biomodel comparison. Looking at the correlations in Figure 5a we see that we see that the propagation network, both respecting direction with one respective sign and the other one not - perform reasonably well (0.9 and above). A directed distance method also performed fairly well. Interestingly enough, an approach based on the sensitivity of the eigenvectors performed poorly but an approach based on Katz sensitivity performs extremely well.

Figure 5b shows the percent agreement in sign across the various model. Here we see that the signed propagation network and the distance metric performs reasonably well. The fact that the unsigned metrics also perform relatively

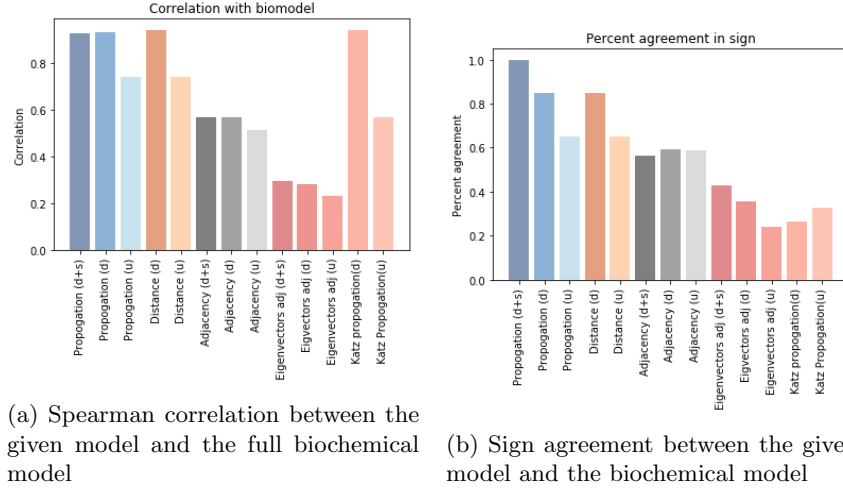


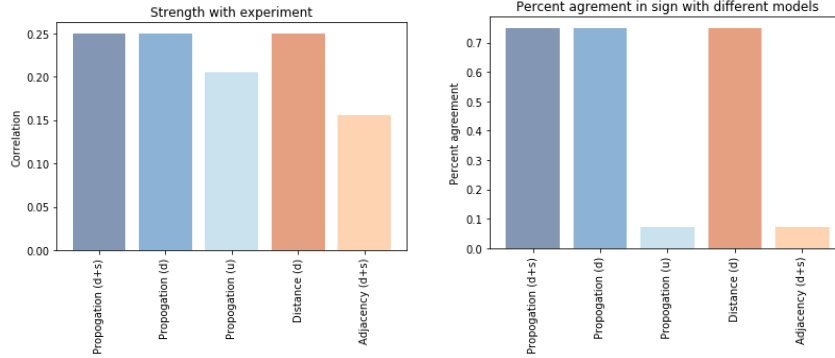
Figure 5: Results of comparing biochemical model with the the various influence networks (propagation networks -blue, distance - orange, First neighbors - grey, eigenvector based - red)

high suggests that the incremental improvement in having a signed understanding is relatively marginal (around 0.15 correlation pickup) - suggesting that this network was fairly one directional to start with. Interestingly the Katz based approach performed fairly poorly in predicting the sign while doing well at predicted strength.

4.2 Jak2 experimental verification

For the JAK2 knockdown comparison (which we limited to just the 3 classes of models originally proposed), the correlations are, however, significantly worse. In Figure 6a we see that the highest correlations are 0.25 with the Propagation (d + s) network. The sign agreement performs slightly better in figure 6b where the top performing models (signed propagation, unsigned directed and distance) predicted roughly 70% of the directional.

We hypothesize the poor performance of the correlation may come due the differing natures of the networks - the BIOMODEL's networks and most models in that repository tend to be smaller (fewer nodes) and more connected (more edges) - which may imply that the localized network is relatively well understood. There's a partial selection bias as well - the BIOMODEL's networks had have to been extremely well studied to come up with rate constants in the first place, while in comparison - a pure topology network may not have the same level of detail.



(a) Rank correlation result relative to RNA expression. RNA expression data was compiled for various genes in relation to a knockout of JAK2. The sensitivity matrices were then computed for the various method, and the rank correlations were computed to understand whether relative strengths were being captured

(b) Percent agreement in sign relative to RNA expression. In the JAK2 knockout, an increase of above 10% corresponded to a 1 for comparison purposes, below 10% resulted in a -1 and everything else 0. This was compared to the corresponding output in the sensitivities matrices of various models

Figure 6: Results of Jak2 knockdown experimental verification

4.3 Stat5 experimental verification

Stat5 experimental verification was also performed using a Stat5 knockdown dataset (GSE68993). However, there was a relatively poor overlap between genes present in this dataset, with genes in our manually curated biological pathway. The results of the analysis were thus not very informative, and not presented here.

4.4 PD-L1 experimental validation

To determine if topology based networks utilized in our study is able to make prediction of genes that confer resistance to cancer immunotherapy, we computed the sensitivity matrix of the various models and ask if the results are concordant with experimentally defined list of genes that confer resistance to cancer immunotherapy. We first artificially down-regulated the expression of each node, and then assessed the expression of PD-L1 following gene perturbation (Figure 7). Notably, our network indicate that PD-L1 expression remains unchanged when most of the genes are perturbed. These results were also observed across different types of models (Propagation, Distance, neighbours). Only in the 'Distance - unsigned model' was the PD-L1 expression predicted to decrease in all cases.

Interestingly, our model predicted that down-regulation of mTOR, PIP2,

Akt, and NFAT to have a strong impact on PD-L1 downregulation (Figure 7). However, only the AKT genes (AKT1, AKT2, AKT3) were identified in these experimental studies (Figure 8). Conversely, many more weak candidates identified in our network were identified in these three studies. Nonetheless, we also note poor concordance between these 3 experimental studies. Thus, the poor concordance of our topology based methodology with these studies may be due to a lack of sensitivity and/or specificity of these original experimental studies.

4.5 Limitations

While the evaluation on the biomodel seemed to perform reasonably well, the performance on the PD-L1 model against expression data did not perform as well. We postulate several possible reasons:

- Systems difference - while the original models were tested primarily on bacterial systems, mammalian systems display a much more complex regulatory network - which may necessitate knowing more dynamical information that a simple topology approach may have.
- Less information about the system. Systems that already kinetic rate constants by their nature would be far better understood than a system that doesn't. There may be parts in the PD-L1 pathway that we still do not fully grasp.
- mRNA expression data utilized in this study is a proxy for protein expression. However, proteins levels which truly govern the biochemical networks may not be properly represented by mRNA expression. Utilization of protein expression data, like those collected using mass spectrometry is more likely to yield more meaningful results.

5 Conclusions and future works

While the topology based dynamic approach sounded appealing, when evaluating against these larger, more complex networks than the ones original studies, it becomes less clear. When evaluating against expression data, directional outcomes could be well predicted but not strength. When evaluating predicted impact on PD-L1, the various models predicted many more interactions than what was observed. It may be difficult to say whether it was a short coming in those studies or our models, but based on initial evidence, it is hard to conclude if these topology based dynamical networks could be used. Further studies on mTOR, PIP2, and NFAT relationship with PD-L1 would help verify or refute our model's predictions. In addition, mRNA as a proxy may not fully capture the protein protein

One bright spot that would be exploring further would be evaluating further the Katz's based models - in particular their incorporation of signed network

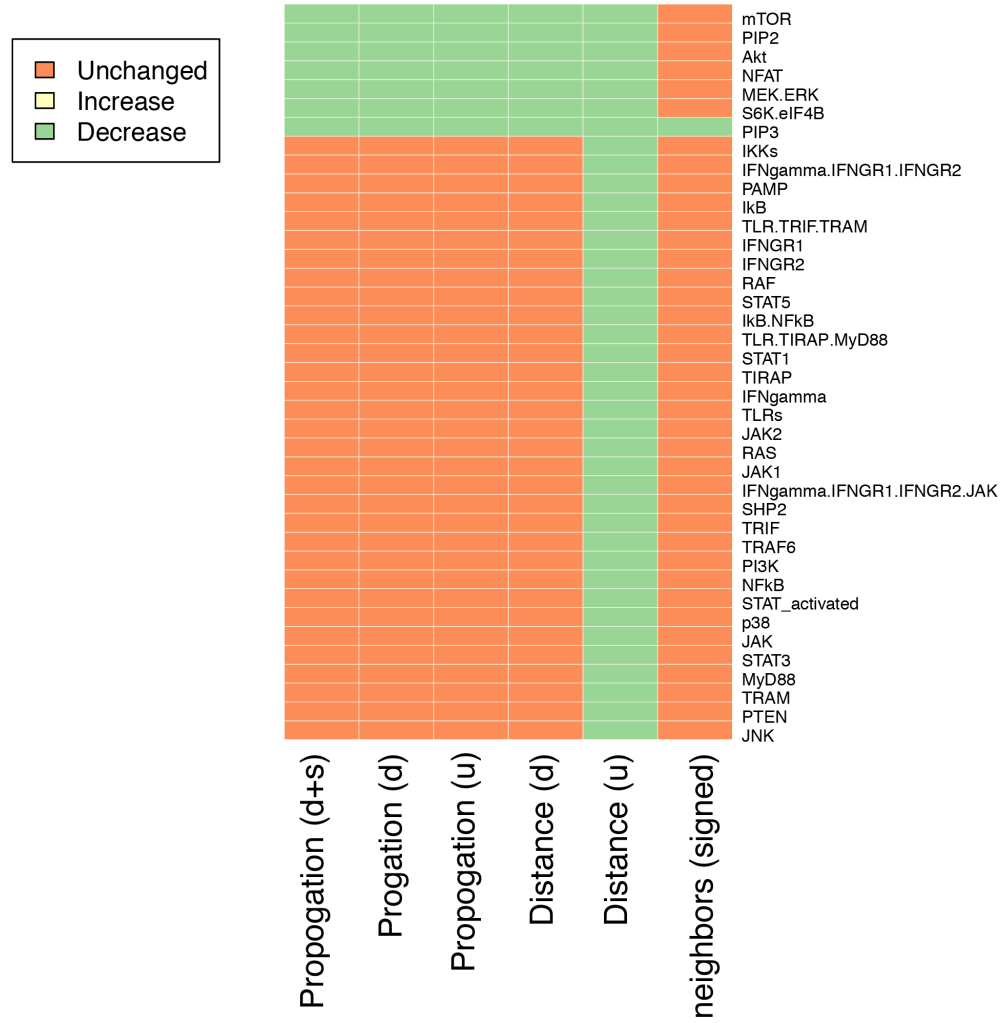


Figure 7: Predicted impact of down regulation of each gene on PD-L1 gene expression, using different topology based network approaches.

Gene Symbol	Topology Model predictions (out of 6)	Manguso et al	Pan et al	Patel et al
AKT1,AKT2,AKT3	5	Yes	Yes	Yes
IFNGR1	1	-	Yes	-
IFNGR2	1	-	Yes	Yes
IKBKB	1	-	Yes	
JAK1	1	-	Yes	Yes
JAK2	1	-	Yes	Yes
MAPK8	1	-	-	Yes
MAPK14	1	Yes	-	-
PIK3CA,PIK3CB,PIK3CD,PIK3CG	1	Yes	Yes	-
PTEN	1	-		Yes
BRAF,RAF1	1	-	Yes	Yes
KRAS,NRAS	1	Yes	-	-
PTPN11	1	-	Yes	-
STAT1	1	-	Yes	Yes
STAT3	1	-	Yes	-
TRAF6	1	Yes	-	-

Figure 8: Comparison of genes that are predicted to confer sensitivity from topology networks, with genes that are identified from three different studies. Genes that were found found to confer sensitivity to cancer immunotherapy were labelled with a "Yes"

in some fashion may help in improving the sign prediction accuracy. Further eigenvector based approach’s may field similarly interesting results.

6 Appendix

The github can be found at <https://github.com/ktan8/AM205-Project>

References

- [1] A. Ribas, O. Hamid, A. Daud, F. S. Hodi, J. D. Wolchok, R. Kefford, A. M. Joshua, A. Patnaik, W.-J. Hwu, J. S. Weber, *et al.*, “Association of pembrolizumab with tumor response and survival among patients with advanced melanoma,” *Jama*, vol. 315, no. 15, pp. 1600–1609, 2016.
- [2] S. J. Patel, N. E. Sanjana, R. J. Kishton, A. Eidizadeh, S. K. Vodnala, M. Cam, J. J. Gartner, L. Jia, S. M. Steinberg, T. N. Yamamoto, *et al.*, “Identification of essential genes for cancer immunotherapy,” *Nature*, vol. 548, no. 7669, p. 537, 2017.
- [3] R. T. Manguso, H. W. Pope, M. D. Zimmer, F. D. Brown, K. B. Yates, B. C. Miller, N. B. Collins, K. Bi, M. W. LaFleur, V. R. Juneja, *et al.*, “In vivo crispr screening identifies ptpn2 as a cancer immunotherapy target,” *Nature*, vol. 547, no. 7664, p. 413, 2017.
- [4] D. Pan, A. Kobayashi, P. Jiang, L. F. de Andrade, R. E. Tay, A. M. Luoma, D. Tsoucas, X. Qiu, K. Lim, P. Rao, *et al.*, “A major chromatin regulator determines resistance of tumor cells to t cell-mediated killing,” *Science*, vol. 359, no. 6377, pp. 770–775, 2018.
- [5] M. Santolini and A.-L. Barabási, “Predicting perturbation patterns from the topology of biological networks,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 27, pp. E6375–E6383, 2018.
- [6] P. Ritprajak and M. Azuma, “Intrinsic and extrinsic control of expression of the immunoregulatory molecule pd-l1 in epithelial cells and squamous cell carcinoma,” *Oral oncology*, vol. 51, no. 3, pp. 221–228, 2015.
- [7] F. Büchel, N. Rodriguez, N. Swainston, C. Wrzodek, T. Czauderna, R. Keller, F. Mittag, M. Schubert, M. Glont, M. Golebiewski, M. van Iersel, S. Keating, M. Rall, M. Wybrow, H. Hermjakob, M. Hucka, D. B. Kell, W. Müller, P. Mendes, A. Zell, C. Chaouiya, J. Saez-Rodriguez, F. Schreiber, C. Laibe, A. Dräger, and N. Le Novère, “Path2Models: large-scale generation of computational models from biochemical pathway maps,” *BMC Systems Biology*, vol. 7, p. 116, Nov 2013.
- [8] V. Raia, M. Schilling, M. Böhm, B. Hahn, A. Kowarsch, A. Raue, C. Sticht, S. Bohl, M. Saile, P. Möller, N. Gretz, J. Timmer, F. Theis, W.-D.

- Lehmann, P. Lichter, and U. Klingmüller, “Dynamic mathematical modeling of il13-induced signaling in hodgkin and primary mediastinal b-cell lymphoma allows prediction of therapeutic targets,” *Cancer research*, vol. 71, p. 693—704, February 2011.
- [9] J. M. Fletcher and T. Wennekers, “From structure to activity: Using centrality measures to predict neuronal activity,” *International Journal of Neural Systems*, vol. 28, no. 02, p. 1750013, 2018. PMID: 28076982.