Mining Pathway Associations from Networks of Mutual Exclusivity Interactions

Herty Liany[1], Yu Lin[2], Anand Jeyasekharan[3], and Vaibhav Rajan[4]

[1] Department of Computer Science, School of Computing,
National University of Singapore
herty@comp.nus.edu.sg

[2] Research School of Computer Science, Australian National University, Australia
yu.lin@anu.edu.au

[3] Cancer Science Institute of Singapore, National University of Singapore
csiadj@nus.edu.sg

[4] Department of Information Systems and Analytics, School of Computing,
National University of Singapore
vaibhav.rajan@nus.edu.sg

Abstract. Study of pairwise genetic interactions such as mutual exclusivity or synthetic lethality has led to the development of targeted anticancer therapies, and mining the network of such interactions is a common approach used to obtain deeper insights into the mechanism of cancer. A number of useful graph clustering-based tools exist to mine interaction networks. These tools find subgraphs or groups of genes wherein each gene belongs to a single subgraph. However, a gene may be present in multiple groups – for instance, a gene can be involved in multiple signalling pathways. We develop a new network mining algorithm, that does not impose this constraint and can provide a novel pathway-centric view. Our approach is based on finding edge-disjoint bipartite subgraphs of highest weights in an input network of genes, where edge weights indicate the significance of the interaction and each set of nodes in every bipartite subgraph is constrained to belong to a single pathway. This problem is NP-hard and we develop an Integer Linear Program to solve this problem. We evaluate our algorithm on breast and stomach cancer data. Our algorithm mines dense between-pathway interactions that are known to play important roles in cancer and are therapeutically actionable. Our algorithm complements existing network mining tools and can be useful to study the mutational landscape of cancer and inform therapy development.

Keywords: Maximum weight bipartite subgraphs · Between-pathway interactions · Mutual exclusivity · Biological network mining.

1 Introduction

Cancer cells gain selective advantage over healthy cells and proliferate uncontrollably as a result of genetic alterations caused by endogenous and exogenous processes [23]. These alterations are observed in multiple genes that participate in important signalling pathways, and the mechanisms of these alterations differ across cancer types and subtypes [4, 39]. Investigation of combinatorial patterns of alterations, in genes and pathways, have enabled us to understand cancer better and find effective therapeutic targets.

Genetic alterations affecting pairwise interactions, such as Mutually Exclusive (ME) mutations have been studied extensively. ME can indicate functional redundancy or Synthetic Lethality (SL) [17]. SL interactions between two genes is a condition where the loss of either gene is viable but the loss of both is lethal, and has been considered a foundation for development of targeted anti-cancer therapies [7, 40]. Such interactions across pathways can result in cross-talk between pathways [30, 46] and methods have been developed to identify such between-pathway motifs [6, 24]. Many other computational approaches have also been developed to infer ME [12, 3, 32, 27, 13, 8] and SL interactions [28, 35, 45, 43, 31, 33].

A network provides a succinct global representation of all the pairwise genetic interactions observed in a set of genes. Analysis and mining of such networks can not only yield important biological insights but have also aided the development of predictive models for gene essentiality, survival and drug response [28, 31]. A common approach for exploratory mining is to find biologically interesting sub-networks, also called modules. Recent examples of methods to find modules include BeWith [15], CDPath [56] and CD-CAP [25].

BeWith [15] identifies modules with combinations of interactions both within and between modules. Their clustering approach is formulated as an Integer Linear Program (ILP) to optimize between-cluster and within-cluster scores. The scores may be based on functional relatedness, mutual exclusivity or co-occurrence. CDPath [56] uses an ILP similar to that of BeWith to identify modules that maximize coverage and mutual exclusivity within modules, and co-occurrence and functional interactions between modules. Pathways associated with each module are identified using gene-pathway association information. Further, it clusters the pathway interaction network to group pathways with functional relations. Pathways in different modules and same clusters are considered as cooperative driver pathways. Combinatorial Detection of Conserved Alteration Patterns (CD-CAP) [25] detects subnetworks of an interaction network each with an alteration pattern, within a group of genes, conserved across tumor samples. The patterns include copy number alterations, somatic mutations and aberrant gene expression. Although these methods differ in their approaches and objectives, they can all be viewed as variants or extensions of graph clustering methods.

In this paper we develop another approach to mine and analyse a given network of pairwise interactions, that complements existing approaches. While graph clustering is useful to find modules, the standard cut-based approach only finds groups of genes wherein each gene belongs to a single sub-network. They cannot find patterns of interactions across groups where a gene may be present in multiple groups, which is biologically plausible – for instance, a gene can be involved in multiple signalling pathways.

A schematic of our approach is shown in figure 1. Our approach is based on finding K edge-disjoint bipartite subgraphs of highest weights in an input network of genes, with edge weight indicating the significance of the interaction. Further, we constrain each edge set of the bipartite subgraphs to belong to a single pathway. Thus, each bipartite subgraph in our solution represents a dense between-pathway interaction. While each interaction can only be part of a single subgraph, there are no restrictions on the genes that may belong to multiple subgraphs. This problem is NP-hard and we develop an ILP-based heuristic to solve this problem.

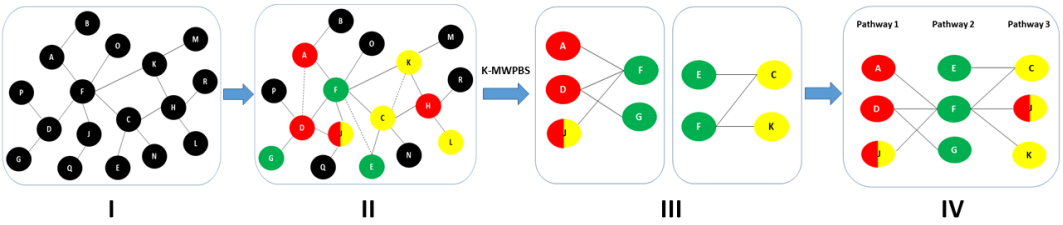

Fig. 1. (I) Input network. Nodes are genes, edges are pairwise interactions, e.g. ME, SL. Edge weights not shown. (II) Same network in I with information of input pathways, in which some of the genes participate; each color (green, red, yellow) indicates a pathway. A gene may belong to more than one pathway, e.g. nodes with 2 colors are in 2 pathways. Dashed lines indicate within-pathway edges. Genes that are not involved in any of the input pathways are in black. (III) The bipartite subgraphs obtained by our algorithm using input network and pathway information from II. Each bipartite subgraph shows the interactions between gene sets from two different pathways (left: red-green, right: green-yellow). Within-pathway edges are not considered. (IV) The bipartite subgraphs in III merged to provide a pathway-centric view. Our results in later sections are shown in this format.

We use our approach to explore ME interaction networks in breast and stomach cancers. Our algorithm uncovers relationships between signalling pathways that are known to play important roles in cancer and in targeted drug development. Such between-pathway relationships can potentially help elucidate functional mechanisms of oncogenic alterations [39]. We believe that our tool complements the existing repertoire of network mining tools that aim to gain deeper insights into the mutational landscape of cancer and inform therapy development.

2 Method

The inputs to our method are (1) an undirected graph $G = (V, E)$ with vertex set V and weighted edge set E, and (2) a list of pathways P, where a pathway P_j is represented by a list of participating genes. In our context, a vertex represents a gene and the edges E represent ME relations.

A weighted bipartite graph, (L, R, F), is a graph whose vertices can be divided into disjoint sets, denoted by left (L) and right (R) sets, and every weighted edge in the edge set F is such that it connects a vertex in L to a vertex in R. The edges, F, in a bipartite graph show us the interactions between (the left and right) sets of genes. We define the weight, $w(F)$, of a bipartite graph as the sum of the weights of all the edges in the graph. We require every gene in the left set of the mined bipartite graph to belong to one of the input pathways, that we call the *left pathway*. Similarly, every gene in the right set of the mined bipartite graph is required to belong to one of the input pathways, that we call the *right pathway*. To find those pathway pairs with the most 'dense' interactions, we mine bipartite subgraphs of G with maximum weight. Note that we do not pose any restrictions on the edges between genes within a pathway. Thus, the input network may have within-pathway ME interactions, that are not shown in the subgraphs. These requirements are modeled through the following combinatorial optimization problems.

The *Maximum Weight Pathway Constrained Bipartite Subgraph* problem (MWPBS) is to find a bipartite subgraph (L, R, F) of an input graph (V, E) that is of maximum weight and such that every gene in the left and right sets of the subgraph belongs to a pathway and the associated left and right pathways are different, i.e.,

1. $L \subset V, R \subset V, L \cap R = \phi, F = \{(u, v) | u \in L, v \in R\} \subseteq E$
2. $L \subseteq P_k, R \subseteq P_l, P_k \neq P_l$
3. $w(F)$ is maximum among all such subgraphs (satisfying conditions 1 and 2).

The *Top-K Maximum Weight Pathway Constrained Bipartite Subgraph* Problem (K-MWPBS) is to find K edge-disjoint maximum weight pathway constrained bipartite subgraphs of the input graph, $(L_1, R_1, F_1), \ldots, (L_K, R_K, F_K)$, i.e.,

1. $L_i \subset V, R_i \subset V, L_i \cap R_i = \phi, F_i = \{(u, v) | u \in L_i, v \in R_i\} \subseteq E, \quad \forall i \in [K]$
2. $F_i \cap F_j = \phi \quad \forall i, j \in [K]$
3. $L_i \subseteq P_k, R_j \subseteq P_l, P_k \neq P_l$
4. $w(F_1) \geq w(F_2) \geq \ldots \geq w(F_K)$ such that $\sum_{i=1}^{K} w(F_i)$ is maximum among all such K subgraphs (satisfying conditions 1–3).

Given a procedure to solve the MWPBS problem, we can iteratively solve the K-MWPBS problem in a greedy manner as shown in algorithm 1. Note that the subgraphs are edge disjoint but may share vertices.

Algorithm 1: Greedy solution to the K-MWPBS problem

input : Undirected graph $G = (V, E)$, list of pathways P and positive integer K
output: k edge disjoint pathway constrained bipartite subgraphs, $k \leq K$
initialization: i = 1;
while $E \neq \phi$ *and* $i \leq K$ **do**
  $(L_i, R_i, F_i) = \text{MWPBS}(V, E, P)$; `// MWPBS on current` E
  $E = E - F_i$; `// Remove edges in` F_i `from` E
  $i = i + 1$
end

Note that the output bipartite sub-graphs of algorithm 1 can be concatenated to form a single subnetwork of the input G, as shown in figure 1.

Hypothetically, if the pathways considered are such that all the input genes are contained in every pathway, then the MWPBS problem is equivalent to solving the Maximum Weight Bipartite Subgraph Problem, which is NP-complete [21]. Therefore the MWPBS problem is NP-complete. Note that if each gene belongs to only one pathway, then the problem can be solved in polynomial time by ranking the weights of bipartite subgraphs formed by each pair of input pathways, and selecting the one with the highest weight.

In the following we formulate an Integer Linear Program (ILP) to solve the MWPBS problem. The advantage of an ILP is that it can leverage accurate and efficient heuristics, such as LP-based Branch and Bound and Cutting Plane methods, with readily available implementations in general-purpose solvers (e.g. Gurobi [22]) to find approximate solutions.

2.1 ILP Formulation for the MWPBS Problem

We assume all edge weights are positive and non-existent edges in the graph have weight 0 and $w_{ij} = w_{ji} \quad \forall i, j$. Binary variables e_{ij} will be used to indicate the final selected edges in the bipartite subgraph. To obtain the maximum weighted subgraph we set the objective function to:

$$\text{maximize} \sum_{i,j} w_{ij} e_{ij}$$

for every pair of nodes i, j such that $i \neq j$. We now develop constraints that will ensure that the selected edges and their incident vertices form a bipartite subgraph. Define integer variable x_i for each vertex in the graph and integer variable y_{ij} for every pair of nodes i, j such that $i \neq j$. We will restrict the ranges of these variables as follows:

$$-2 \leq y_{ij} \leq 2 \quad \forall i, j, \text{ i.e., } y_{ij} \in \{-2, -1, 0, 1, 2\}$$

$$-1 \leq x_i \leq 1 \quad \forall i \text{ i.e., } x_i \in \{-1, 0, 1\}$$

Variables x_i will be used to identify three subgroups of vertices of the input graph. Denote by L and R the left and right sets respectively of the selected bipartite graph, and by O all the remaining vertices in the input graph. Variables $x_i = -1$ will indicate vertices in L, $x_i = 1$ will indicate vertices in R, while $x_i = 0$ will indicate vertices in O.

For an edge in the required bipartite graph, i.e., between vertices $x_i \in L$ and $x_j \in R$, we have $x_i - x_j = -2$ and for $x_i \in R$ and $x_j \in L$, we have $x_i - x_j = 2$. For an edge between a vertex $x_i \in L$ and a vertex $x_j \in O$, we have $x_i - x_j = -1$ and for $x_i \in O$ and $x_j \in L$, we have $x_i - x_j = 1$. Similarly, for $x_i \in O$ and $x_j \in R$, we have $x_i - x_j = -1$ and for $x_i \in R$ and $x_j \in O$, we have $x_i - x_j = 1$. Finally, for x_i, x_j in the same set (O, L or R), we have $x_i - x_j = 0$. So, we can identify the kind of edge using the following constraint:

$$x_i - x_j = y_{ij} \quad \forall i, j$$

Note that for each pair of vertices, we shall obtain two values y_{ij} and y_{ji}. Only one of them will affect the final objective function as described below.

Since we want to retain only the edges in the bipartite graph, i.e. where $y_{ij} = 2$, we will impose the constraints: if $e_{ij} = 1$ then $y_{ij} \geq 2$ and if $e_{ij} = 0$ then $y_{ij} \leq 1$. Since each edge is accounted for twice, once in each direction, excluding edges with $y_{ij} = -2$, shall not lead to missing any edges. These constraints can be imposed using an additional binary variable z_{ij} and a constant $M \geq 4$ as follows:

$$y_{ij} \leq M(1 - z_{ij}) + 1 \quad \forall i, j$$

$$e_{ij} \leq M(1 - z_{ij}) \quad \forall i, j$$

$$e_{ij} \geq 1 - M z_{ij} \quad \forall i, j$$

$$y_{ij} \geq 2 - M z_{ij} \quad \forall i, j$$

Note that when $z_{ij} = 0$ we have $e_{ij} \geq 1, y_{ij} \geq 2$ (these edges are valid in the bipartite graph) and when $z_{ij} = 1$ we have $e_{ij} \leq 0, y_{ij} \leq 1$ (invalid edges for which e_{ij} are set to 0) as required.

Pathway Constraints

We now add constraints on the vertices to enforce each vertex set L, R in the bipartite subgraph to belong to a pathway. We first define binary variables l_i to indicate whether a node belongs to L or not: $l_i = 1$ iff $x_i = -1$. Similarly, we define binary variables to identify a vertex in R: $r_i = 1$ iff $x_i = 1$. These are set using the following constraints:

$$r_i - l_i = x_i \quad \forall i$$

$$r_i + l_i \leq 1 \quad \forall i$$

Let N_p be the number of pathways considered and N_g be the number of genes. Note that $N_g = |V|$. We assume a binary matrix P of dimension $N_g \times N_p$ indicating pathway membership where the ij^{th} cell is 1 if the ith gene belongs to the jth pathway, otherwise 0. Note that the summation $\sum_i P_{it} l_i$ for the tth pathway (column in P) indicates the number of genes that are both in the tth pathway and in the left set L of the bipartite graph. To ensure that every gene in the left set must belong to a pathway, we add the following constraints:

$$\sum_i l_i - \sum_i P_{it} l_i \leq B q_t \quad \forall t \in \{1, \ldots, N_p\}$$

$$\sum_{t=1}^{N_p} q_t \leq N_p - 1$$

where q_t is a binary variable and $B \geq N_g$ is a large constant. At least one of the q_t must be set to 0 (otherwise, $\sum_t q_t > N_p - 1$) and this will make $\sum_i l_i - \sum_i P_{it} l_i \leq 0 \implies \sum_i l_i \leq \sum_i P_{it} l_i$. Since $\sum_i P_{it} l_i$ cannot be greater than $\sum_i l_i$, this will set $\sum_i l_i = \sum_i P_{it} l_i$ ensuring that all the genes in the left set are from the tth pathway.

Similarly we ensure that the genes in the right set belong to a pathway:

$$\sum_i r_i - \sum_i P_{it} s_i \leq B s_t \quad \forall t \in \{1, \ldots, N_p\}$$

$$\sum_{t=1}^{N_p} s_t \leq N_p - 1$$

where s_t is a binary variable. Finally, every gene in L belongs to a pathway (say P_l) and every gene in R belongs to a pathway (say P_r), but we want the pathways P_l, P_r to be different. This is done by the following constraint, which ensures that both q_t, s_t are not set to 0 for the same t:

$$q_t + s_t \geq 1 \quad \forall t \in \{1, \ldots, N_p\}$$

3 Experiments

3.1 Simulation Studies

We conduct simulation studies to evaluate the correctness of our algorithm. Since the MWPBS problem is NP-complete, optimal solutions cannot be found in polynomial time. We generate random networks under various settings and evaluate our ILP, both with and without pathway constraints, as well as our greedy algorithm for the K-MWPBS problem. Our results show that our ILP-based solutions can find optimal or near-optimal solutions as expected in the simulated conditions. Details are in Appendix A.

3.2 Real Data

Experiment Settings We evaluate our algorithm on two graphs containing ME relations between genes in Breast cancer (BRCA) and Stomach cancer (STCA). ME relations were inferred using DISCOVER [8], separately for each cancer. DISCOVER infers ME using mutation data and accounts for heterogeneous alteration rates across tumor samples. All non-synonymous mutations (frameshift, missense, nonsense, nonstop and splice site) were used in our analysis. Mutation data for both the cancers were obtained from TCGA [50]. A total of 15,566 genes and 977 samples in BRCA and 16,920 genes and 393 tumor samples in STCA were used. 342 gene pairs in BRCA and 132 gene pairs in STCA were detected as ME (using p-value threshold < 0.05, with FDR-corrected p-value < 0.10 for STCA and 0.25 for BRCA respectively [9]), that were used to create the respective input graphs with negative logarithm of p-values as edge weights. We test our algorithm with signalling pathways from KEGG [29] and Hallmark [23]. A total of 169 KEGG pathways and 50 Hallmark pathways were used as input (Appendix B discusses how pathways were selected). We compute *between-pathway significance scores* by combining, using Fisher's method [19], the p-values (obtained from DISCOVER) for all pairs of genes across the pathways. Table 4 in Appendix B shows the list of pathway name abbreviations used in our results.

3.3 Results

BRCA Dataset Figure 2 shows the pathways selected by our algorithm and the ME relations between the pathways. There are 14 genes and 6 pathways in our solution among which seven are well-known cancer drivers: oncogenes (TP53, PIK3CA, MAP2K4, MAP3K1) and tumor-suppressor genes (TP53, PTEN, CDH1, NF1, MAP2K4, MAP3K1) [20]. Note that TP53 gene appears in both p53 and MAPK signaling pathways.

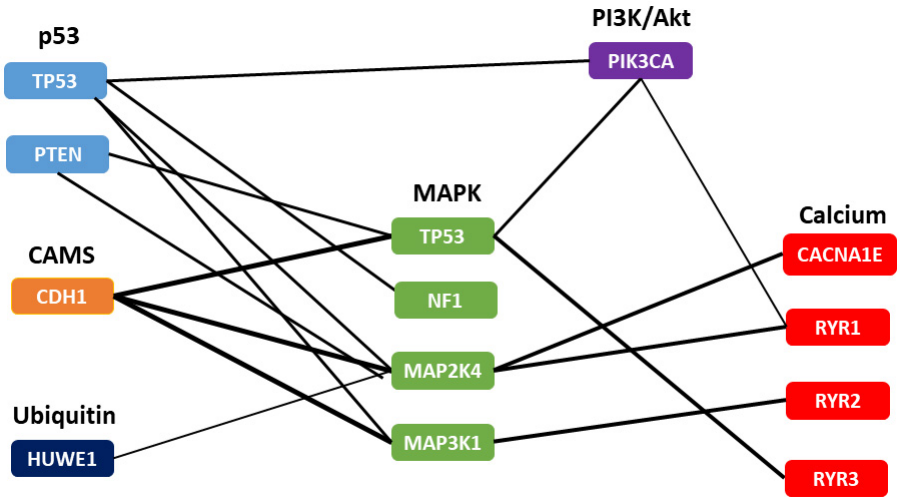

Fig. 2. Pathways and genes selected by our method from the BRCA network. Nodes denote genes, edges indicate ME relations and each pathway is identified by a color. Thickness of edges is proportional to its significance.

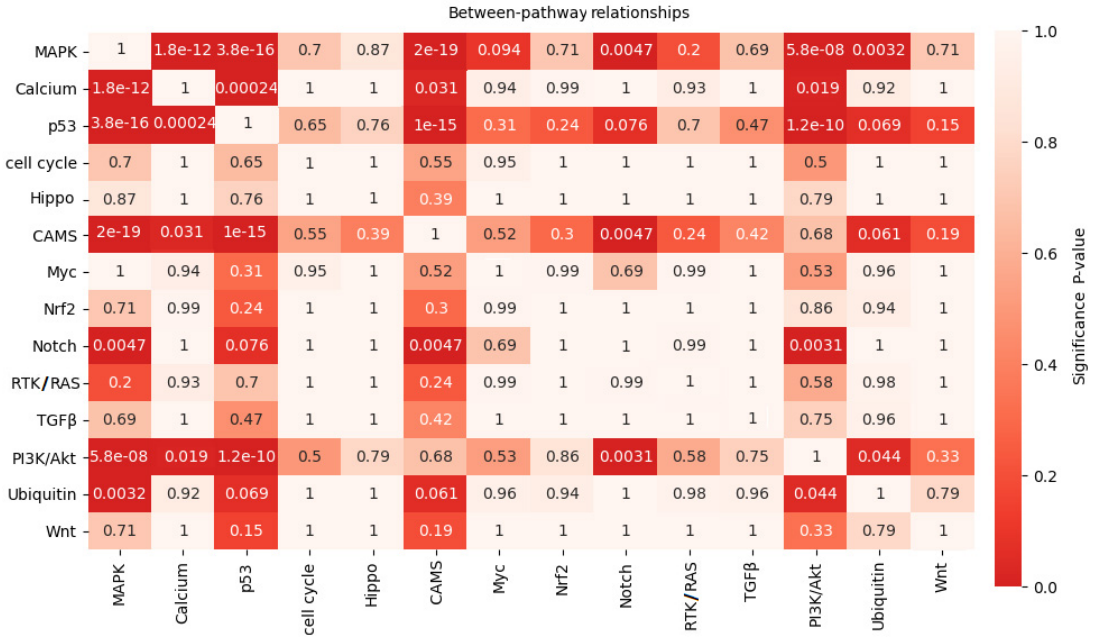

Fig. 3. Between-pathway significance scores for pathways discovered by our method from the BRCA dataset along with 10 canonical signalling pathways in [39].

The strongest ME relations between pathways was found between the CAMS and MAPK signaling pathways. It is known that CDH1 (or E-Cadherin) produces calcium-dependent cell-cell adhesion protein and is dependent on calcium ion (Ca^{2+}) to function. Loss of function of this gene is thought to contribute to cancer progression by increasing proliferation, invasion and/or metastasis [52]. The second strongest ME relations among pathways was found between the MAPK and p53 signaling pathways. MAPK signaling pathway has been reported to regulate cell proliferation through a cross-talk of the MEK-ERK and JNK-JUN signaling pathways [60]. Cancers that have lost MAP3K1 or MAP2K4 fail to activate JNK-JUN signaling pathway [55], which in turn affects p53 activation. The activation of PI3K pathway via mutations in PIK3CA can activate p53 signaling in human cells, while activation of PIK3CA results in resistance to p53-related apoptosis in PTEN deficient cells [42]. Several studies elucidate the role of calcium and MAPK signaling in relation to cancer, particularly, breast cancer, e.g. [37]. Our algorithm mines the interactions among these pathways that are known to play important roles in breast cancer.

Cross-talks between the pathway pairs uncovered by our algorithm have been found useful in the context of breast cancer drug development. For instance, it has been reported that changes of intracellular Ca^{2+} activity could trigger numerous signaling pathways, in particular MAPK signaling pathway, that in turn perturb ER homeostasis resulting in ER stress-induced cell death, an association that has been used to design the cancer drug vemurafenib [16]. Calcium channel related genes (CACNA1E, RYR1, RYR2, RYR3) have featured in recent studies as potential drug targets [34, 2, 10, 38, 5]. In particular, calcium channel blockers such as Nifedipine and Verapamil, commonly used to treat hypertension, were found to be beneficial in some combinations of drugs [47]. Drug targeting the ubiquitin-proteasome pathway for cancer therapy has become an area of intense investigation [36]. Proteasome inhibitors effect apoptosis by decreased signaling through the MAPK pathway [36, 48]. The MAPK pathway itself is targeted by MAP2K4 (and MAP3K1) inhibiting drugs like Selumetinib and Trametinib [55]. The CAMS pathway has been found to be a useful target, e.g., through FDA-approved drugs like Docetaxel, in cases when cancer cells develop resistance to chemotherapeutic drugs [1].

An alteration map of 10 canonical signaling pathways from 33 cancer types in TCGA has been discussed in [39] (listed in Appendix B). We take the combination of the 6 pathways uncovered by our algorithm and these 10 therapeutically actionable canonical pathways selected in their study and compute their between-pathway significance scores, shown in figure 3. We note that the highest significance scores are for those pairs of pathways that our algorithm finds in figure 2, thereby providing additional empirical evidence of the correctness of our algorithm. We observe between-pathway ME interactions in breast cancer that were not reported in [39]. For instance, PI3K has low scores with the other canonical pathways (also seen in [39]) but has high scores with MAPK and Calcium pathways. Such cross-talks can be therapeutically exploited (as discussed in [39]) and dense pathway pairs with many ME interactions, uncovered by our algorithm, may be potential drug targets to be investigated further.

STCA Dataset Figure 4 shows the pathways selected by our algorithm for the STCA dataset. There are 19 genes and 8 pathways in our solution. Seven of these are known driver genes: Oncogene (ERBB3, NOTCH1, TP53, PIK3CA) and tumor suppressor genes (MYH9, TP53, NOTCH1, CDH1, PTEN) [20]. Note that CDH1 gene appears in both CAMs and adherent junction pathways; PIK3CA gene appears in both apoptosis and focal adhesion pathways. Figure 5 shows the between-pathway significance scores for our pathways and the canonical signaling pathways [39]. We again observe that the highest significance scores are for pathway pairs obtained by our algorithm.

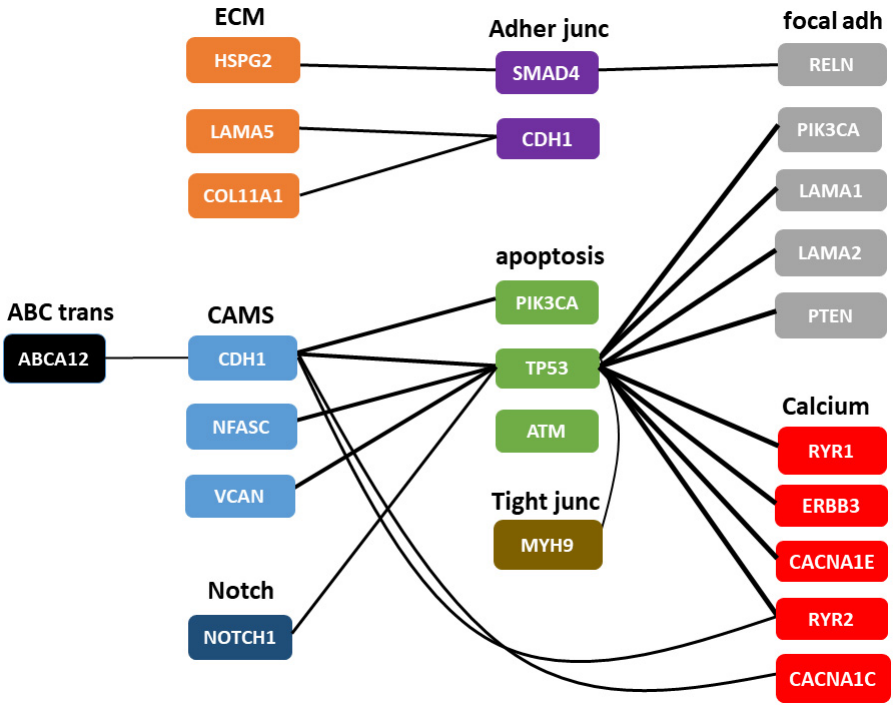

Fig. 4. Pathways and genes selected by our method from the STCA network. Nodes denote genes, edges indicate ME relations and each pathway is identified by a color. Thickness of edges is proportional to its significance.

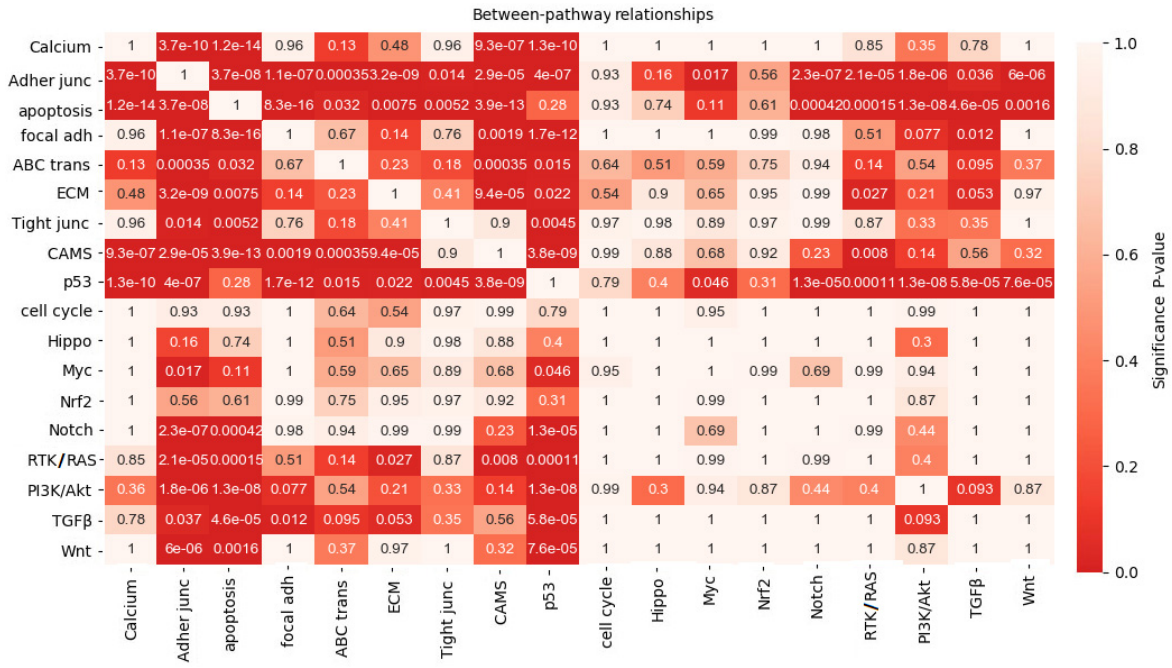

Fig. 5. Between-pathway significance scores for pathways discovered by our method from the STCA dataset along with 10 canonical signalling pathways in [39].

Cross-talks between these pathways pairs have been reported in previous literature. The study in [44] discusses the links between CAMs/Tight junction(TJ)/Focal Adhesion (FA) proteins with apoptosis pathways, cell proliferation and inflammatory responses. NOTCH1 over-expression in the Notch signaling pathway is known to protect gastric carcinoma cells and reduce apoptosis [58]. The upregulation of IP3R3-mediated Ca2+ release through ER/mitochondrial cross-talk has been observed in gastric cancer cell lines; the depletion of ER Ca2+ stores itself is sufficient to cause ER stress and to induce apoptosis pathways [14]. FA provides mechanical linkage to ECM, and as a biochemical signaling hub to concentrate and direct numerous signaling proteins at sites of integrin binding and clustering [51]. Concurrently FA, ECM and adheren junctions actively contribute to the normal development, maintenance of the gastrointestinal tract and play important roles in stomach cancer [53] and all three pathways are mined by our algorithm.

These pathways, that are significantly involved in digestive carcinogenesis, have been used in targeted drug development. The drug Luteolin was found to suppress gastric cancer progression by suppression of the Notch signaling pathway [26]. The anti-inflammatory drug Ibuprofen is also useful as an adjuvant chemotherapy drug because of its role in downregulating Notch signaling [18]. The interaction between Notch and Apoptosis pathway is also targeted by the drug Penicilazaphilone C [49]. Several drugs, such as Rapamycin, Indomethacin and Cisplatin, target the apoptosis pathway [57, 11, 59]. Therapeutic targets for gastric cancer have also been found in Calcium and TJ pathways [41, 54, 61]. These studies show that pathway pairs highlighted by our algorithm have multiple targetable alterations that have been exploited in therapy development. The high-significance between-pathway interactions in figure 5 show multiple other potentially actionable pathways.

The pathway pairs given by our algorithm, and between-pathway significance scores for these pathways, when Hallmark pathways are used (instead of KEGG) are shown in Appendix C for both STCA and BRCA datasets.

4 Conclusion

We develop a new method to mine networks of mutual exclusivity interactions to find dense between-pathway associations. We formulate this as the problem of finding K edge-disjoint maximum weight pathway constrained bipartite subgraphs of the input network (the K-MWPBS problem). We develop an ILP-based formulation to solve the K-MWPBS problem. Unlike previous methods to find modules in networks our method allows input network nodes to be part of multiple subgraphs. This is an accurate model of the biological condition of genes participating in multiple pathways. Since these pathway pairs have multiple significant mutually exclusive interactions they are potential targets for novel drug development.

We empirically evaluate our method on two networks of mutual exclusivity interactions in breast and stomach cancer. We find that our method discovers pathways and their interactions that are known to play significant roles in cancer. These interactions have been independently discovered and studied earlier to gain deeper understanding of carcinogenesis. They have also found applications in targeted drug development. Our preliminary investigations thus illustrate the potential usefulness of our method in finding therapeutically actionable insights. The source code for our tool and experiments is available (at https://github.com/lianyh/K-MWPBS) for use by the community. Our method can be used to analyze networks of other interactions as well. We constrain our gene sets to belong to a pathway; however, other constraints can also be used (or added). Future studies can develop and analyze other such biologically relevant conditions.

References

1. Alimbetov, D., Askarova, S., Umbayev, B., Davis, T., Kipling, D.: Pharmacological Targeting Of Cell Cycle, Apoptotic and Cell adhesion Signaling Pathways Implicated In Chemoresistance Of Cancer Cells. International Journal of Molecular Sciences **19**(6), 1690 (2018)
2. Azimi, I., Roberts-Thomson, S., Monteith, G.: Calcium Influx Pathways In Breast Cancer: Opportunities For Pharmacological Intervention. British Journal Of Pharmacology **171**(4), 945–960 (2014)
3. Babur, Ö., Gönen, M., Aksoy, B.A., Schultz, N., Ciriello, G., Sander, C., Demir, E.: Systematic Identification Of Cancer Driving Signaling Pathways Based On Mutual Exclusivity Of Genomic Alterations. Genome Biology **16**(1), 45 (2015)
4. Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al.: Comprehensive Characterization Of Cancer Driver Genes And Mutations. Cell **173**(2), 371–385 (2018)
5. Bong, A.H., Monteith, G.R.: Calcium Signaling And The Therapeutic Targeting Of Cancer Cells. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1865**(11), 1786–1794 (2018)
6. Brady, A., Maxwell, K., Daniels, N., Cowen, L.J.: Fault Tolerance In Protein Interaction Networks: Stable Bipartite Subgraphs And Redundant Pathways. PloS One **4**(4) (2009)
7. Brough, R., Frankum, J.R., Costa-Cabral, S., Lord, C.J., Ashworth, A.: Searching For Synthetic Lethality In Cancer. Current Opinion in Genetics & Development **21**(1), 34–41 (2011)
8. Canisius, S., Martens, J.W., Wessels, L.F.: A Novel Independence Test For Somatic Alterations In Cancer Shows That Biology Drives Mutual Exclusivity But Chance Explains Most Co-Occurrence. Genome Biology **17**(1), 261 (2016)
9. Carlson, J.M., Heckerman, D., Shani, G.: Estimating false discovery rates for contingency tables (2009)
10. Chen, Q., Zhang, Q., Zhong, F., Guo, S., Jin, Z., Shi, W., Chen, C., He, J.: Association Between Calcium Channel Blockers And Breast Cancer: A Meta-Analysis Of Observational Studies. Pharmacoepidemiology And Drug Safety **23**(7), 711–718 (2014)
11. Chiou, S.K., Hoa, N., Hodges, A., Ge, L., Jadus, M.R.: Indomethacin Promotes Apoptosis In Gastric Cancer Cells Through Concomitant Degradation Of Survivin And Aurora B Kinase Proteins. Apoptosis **19**(9), 1378–1388 (2014)
12. Ciriello, G., Cerami, E., Sander, C., Schultz, N.: Mutual Exclusivity Analysis Identifies Oncogenic Network Modules. Genome Research **22**(2), 398–406 (2012)
13. Constantinescu, S., Szczurek, E., Mohammadi, P., Rahnenführer, J., Beerenwinkel, N.: Timex: A Waiting Time Model For Mutually Exclusive Cancer Alterations. Bioinformatics **32**(7), 968–975 (2016)
14. Cui, C., Merritt, R., Fu, L., Pan, Z.: Targeting Calcium Signaling In Cancer Therapy. Acta Pharmaceutica Sinica B **7**(1), 3–17 (2017)
15. Dao, P., Kim, Y.A., Wojtowicz, D., Madan, S., Sharan, R., Przytycka, T.M.: Bewith: A Between-Within Method To Discover Relationships Between Cancer Modules Via Integrated Analysis Of Mutual Exclusivity, Co-Occurrence And Functional Interactions. PLoS Computational Biology **13**(10), e1005695 (2017)
16. Darling, N.J., Cook, S.J.: The Role Of MAPK Signalling Pathways In The Response To Endoplasmic Reticulum Stress. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research **1843**(10), 2150–2163 (2014)
17. Deng, Y., Luo, S., Deng, C., Luo, T., Yin, W., Zhang, H., Zhang, Y., Zhang, X., Lan, Y., Ping, Y., et al.: Identifying Mutual Exclusivity Across Cancer Genomes: Computational Approaches To Discover Genetic Interaction And Reveal Tumor Vulnerability. Briefings In Bioinformatics **20**(1), 254–266 (2019)
18. Farhangian, M., Azarafrouz, F., Fallahi, H., Akrami, H.: Investigating The Effect Of Ibuprofen On DLL1 and NOTCH1 Expression In Gastric Cancer Stem Cells Derived From MKN-45 Cell Line. Middle East Journal Of Cancer **10**(4), 292–298 (2019)
19. Fisher, R.A.: 224a: Answer To Question 14 On Combining Independent Tests Of Significance. The American Statistician **2**(30) (1948)
20. Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., et al.: Cosmic: Mining Complete Cancer Genomes In The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research **39**(suppl_1), D945–D950 (2010)
21. Garey, M.R., Johnson, D.S.: Computers and Intractability; A Guide To The Theory Of NP-Completeness. W. H. Freeman & Co., USA (1990)
22. Gurobi: Gurobi Optimizer Reference Manual. http://www.gurobi.com (2015)
23. Hanahan, D., Weinberg, R.A.: Hallmarks Of Cancer: The Next Generation. Cell **144**(5), 646–674 (2011)
24. Hescott, B.J., Leiserson, M.D., Cowen, L.J., Slonim, D.K.: Evaluating Between-Pathway Models With Expression Data. Journal of Computational Biology **17**(3), 477–487 (2010)
25. Hodzic, E., Shrestha, R., Zhu, K., Cheng, K., Collins, C.C., Cenk Sahinalp, S.: Combinatorial Detection Of Conserved Alteration Patterns For Identifying Cancer Subnetworks. GigaScience **8**(4), giz024 (2019)

26. Hu, L., Fan, Z.y., Wang, H.x., Zhu, Z.l., Cao, S., Wu, X.y., Li, J.f., Su, L.p., Li, C., Zhu, Z.g., et al.: Luteolin Suppresses Gastric Cancer Progression By Reversing Epithelial-Mesenchymal Transition Via Suppression Of The Notch Signaling Pathway. Journal Of Translational Medicine **15**(1), 52 (2017)
27. Hua, X., Hyland, P.L., Huang, J., Song, L., Zhu, B., Caporaso, N.E., Landi, M.T., Chatterjee, N., Shi, J.: MEGSA: A Powerful And Flexible Framework For Analyzing Mutual Exclusivity Of Tumor Mutations. The American Journal of Human Genetics **98**(3), 442–455 (2016)
28. Jerby-Arnon, L., Pfetzer, N., Waldman, Y.Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P.A., et al.: Predicting Cancer-Specific Vulnerability Via Data-Driven Detection Of Synthetic Lethality. Cell **158**(5), 1199–1209 (2014)
29. Kanehisa, M., Goto, S.: KEGG: Kyoto Encyclopedia Of Genes And Genomes. Nucleic Acids Research **28**(1), 27–30 (2000)
30. Kelley, R., Ideker, T.: Systematic Interpretation Of Genetic Interactions Using Protein Networks. Nature Biotechnology **23**(5), 561–566 (2005)
31. Lee, J.S., Das, A., Jerby-Arnon, L., Arafeh, R., Auslander, N., Davidson, M., McGarry, L., James, D., Amzallag, A., Park, S.G., et al.: Harnessing Synthetic Lethality To Predict The Response To Cancer Treatment. Nature Communications **9**(1), 1–12 (2018)
32. Leiserson, M.D., Wu, H.T., Vandin, F., RAphael, B.J.: Comet: A Statistical Approach To Identify Combinations Of Mutually Exclusive Alterations In Cancer. Genome Biology **16**(1), 160 (2015)
33. Liany, H., Jeyasekharan, A., Rajan, V.: Predicting Synthetic Lethal Interactions Using Heterogeneous Data Sources. Bioinformatics (11 2019), btz893
34. Mackrill, J.J.: Ryanodine Receptor Calcium Channels And Their Partners As Drug Targets. Biochemical Pharmacology **79**(11), 1535–1543 (2010)
35. Madhukar, N.S., Elemento, O., Pandey, G.: Prediction Of Genetic Interactions Using Machine Learning And Network Properties. Frontiers in Bioengineering and Biotechnology **3** (2015)
36. Orlowski, R.Z., Dees, E.C.: The Role Of The Ubiquitination-Proteasome Pathway In Breast Cancer: Applying Drugs That Affect The Ubiquitin-Proteasome Pathway To The Therapy Of Breast Cancer. Breast Cancer Research **5**(1), 1 (2002)
37. Paplomata, E., O'Regan, R.: The PI3K/AKT/mTOR Pathway In Breast Cancer: Targets, Trials And Biomarkers. Therapeutic Advances In Medical Oncology **6**(4), 154–166 (2014)
38. Phan, N., Wang, C., Chen, C., Sun, Z., Lai, M., Lin, Y.: Voltage-Gated Calcium Channels: Novel Targets For Cancer Therapy. Oncol Lett [Internet]. 2017; 2059–74 (2017)
39. Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al.: Oncogenic Signaling Pathways In The Cancer Genome Atlas. Cell **173**(2), 321–337 (2018)
40. Senft, D., Leiserson, M.D., Ruppin, E., Zeaev, A.R.: Precision Oncology: The Road Ahead. Trends in Molecular Medicine (2017)
41. Shabnam, B., Padmavathi, G., Banik, K., Girisa, S., Monisha, J., Sethi, G., Fan, L., Wang, L., Mao, X., Kunnumakkara, A.B.: Sorcin A Potential Molecular Target For Cancer Therapy. Translational Oncology **11**(6), 1379–1389 (2018)
42. Singh, B., Reddy, P.G., Goberdhan, A., Walsh, C., Dao, S., Ngai, I., Chou, T.C., Pornchai, O., Levine, A.J., Rao, P.H., et al.: P53 Regulates Cell Survival By Inhibiting PIK3CA In Squamous Cell Carcinomas. Genes & development **16**(8), 984–993 (2002)
43. Sinha, S., Thomas, D., Chan, S., Gao, Y., Brunen, D., Torabi, D., Reinisch, A., Hernandez, D., Chan, A., Rankin, E.B., et al.: Systematic Discovery Of Mutation-Specific Synthetic Lethals By Mining Pan-Cancer Human Primary Tumor Data. Nature Communications **8** (2017)
44. Smith, Y.E., Vellanki, S.H., Hopkins, A.M.: Dynamic Interplay Between Adhesion Surfaces in Carcinomas: Cell-cell and cell-matrix crosstalk. World Journal of Biological Chemistry **7**(1), 64 (2016)
45. Srihari, S., Singla, J., Wong, L., Ragan, M.A.: Inferring Synthetic Lethal Interactions From Mutual Exclusivity of Genetic Events In Cancer. Biology Direct **10**(1), 57 (2015)
46. Ulitsky, I., Shamir, R.: Pathway Redundancy And Protein Essentiality Revealed In The Saccharomyces Cerevisiae Interaction Networks. Molecular Systems Biology **3**(1) (2007)
47. Varghese, E., Samuel, S.M., Sadiq, Z., Kubatka, P., Liskova, A., Benacka, J., Pazinka, P., Kruzliak, P., Büsselberg, D.: Anti-Cancer Agents In Proliferation And Cell Death: The Calcium Connection. International Journal of Molecular Sciences **20**(12), 3017 (2019)
48. Voorhees, P.M., Dees, E.C., O'Neil, B., Orlowski, R.Z.: The Proteasome As A Target For Cancer Therapy. Clinical Cancer Research **9**(17), 6316–6325 (2003)
49. Wang, M., Zhao, H., Hu, J., Xu, Z., Lin, Y., Zhou, S.L.: Penicilazaphilone C, A New Azaphilone, Induces Apoptosis In Gastric Cancer By Blocking The Notch Signalling Pathway. Frontiers in Oncology **10**, 116 (2020)

50. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M., Network, C.G.A.R., et al.: The Cancer Genome Atlas Pan-Cancer Analysis Project. Nature Genetics **45**(10), 1113 (2013)
51. Wiki: Wikipedia, $https://en.wikipedia.org/wiki/Focal_adhesion$. last accessed 5th feb 2020 (2020)
52. Wong, A.S., Gumbiner, B.M.: Adhesion-Independent Mechanism For Suppression Of Tumor Cell Invasion By E-cadherin. The Journal of Cell Biology **161**(6), 1191–1203 (2003)
53. Worthley, D.L., Giraud, A.S., Wang, T.C.: The Extracellular Matrix In Digestive Cancer. Cancer Microenvironment **3**(1), 177–185 (2010)
54. Wu, Q., Yang, Z., Xia, L., Nie, Y., Wu, K., Shi, Y., Fan, D.: Methylation of miR-129-5p CpG Island Modulates Multi-Drug Resistance In Gastric Cancer By Targeting ABC Transporters. Oncotarget **5**(22), 11552 (2014)
55. Xue, Z., Vis, D.J., Bruna, A., Sustic, T., van Wageningen, S., Batra, A.S., Rueda, O.M., Bosdriesz, E., Caldas, C., Wessels, L.F., et al.: MAP3K1 And MAP2K4 Mutations Are Associated With Sensitivity To MEK Inhibitors In Multiple Cancer Models. Cell Research **28**(7), 719–729 (2018)
56. Yang, Z., Yu, G., Guo, M., Yu, J., Zhang, X., Wang, J.: CDPath: Cooperative Driver Pathways Discovery Using Integer Linear Programming And Markov Clustering. IEEE/ACM Transactions on Computational Biology And Bioinformatics (2019)
57. Yao, C., Liu, J., Shao, L.: Rapamycin Inhibits The Proliferation And Apoptosis Of Gastric Cancer Cells By Down Regulating The Expression Of Survivin. Hepato-Gastroenterology **58**(107-108), 1075–1080 (2011)
58. Yao, J., Qian, C.: Over-Activated Notch-1 Protects Gastric Carcinoma BGC-823 Cells From TNFα-Induced Apoptosis. Digestive and Liver Disease **41**(12), 867–874 (2009)
59. Yu, B., Xie, J.: Identifying Therapeutic Targets In Gastric Cancer: The Current Status And Future Direction. Acta Biochimica et Biophysica Sinica **48**(1), 90–96 (2016)
60. Zhang, W., Liu, H.T.: MAPK Signal Pathways In The Regulation Of Cell Proliferation In Mammalian Cells. Cell Research **12**(1), 9–18 (2002)
61. Zhao, X.Y., Li, L., Wang, X.B., Fu, R.J., Lv, Y.P., Jin, W., Meng, C., Chen, G.Q., Huang, L., Zhao, K.W.: Inhibition Of Snail Family Transcriptional Repressor 2 (SNAI2) Enhances Multidrug Resistance Of Hepatocellular Carcinoma Cells. PloS One **11**(10) (2016)

Appendix

Mining Pathway Associations from Networks of Mutual Exclusivity Interactions

Herty Liany, Yu Lin, Anand Jeyasekharan, Vaibhav Rajan

A Simulation Studies

We conduct simulations on synthetic networks to study the behavior of our ILP-based solution. The first simulation tests the ILP without pathway constraints. The second simulation tests the ILP with pathway constraints. Finally, the third simulation evaluates algorithm 1.

Simulation 1

We generate random bipartite graphs with parameters n, m representing the number of nodes in the Left and Right sets respectively, and with the probability p of an edge across the sets. We then add k random edges to this graph, which may make it non-bipartite. Each edge had weight 1. We generate 4 sets of such networks with $p = 0.5$ and the parameter values given in table 1.

	n	m	k	$\mu[e_S - e_G]$	$\sigma[e_S - e_G]$
I	10	20	5	1.6	0.8
II	10	20	10	3.7	2.16
III	25	25	5	0	0
IV	25	25	10	0	0

Table 1. Parameter values tested and results obtained in Simulation 1

For each set (I–IV) we generate 10 graphs. We then run the ILP without pathway constraints and evaluate the bipartite subgraph found. Let H be the input (simulated) graph to the ILP and let S be the solution returned by the ILP. Let e_G be the number of edges in simulated bipartite graph to which k random edges were added to form H. We know that H contains a bipartite subgraph with at least e_G edges. There may be more than e_G edges due to the added random edges. Let e_S be the number of edges in S.

We observe that the ILP correctly returns bipartite graphs in all the 40 solutions. Table 1 shows the mean (μ) and standard deviation (σ) of $e_S - e_G$ over 10 runs for each set I–IV. In no simulation did S have lesser number of edges than e_G. This suggests that the ILP found the maximum weighted bipartite subgraphs in each case for these relatively small networks.

Simulation 2

We generate random bipartite graphs with parameters n, m representing the number of nodes in the Left and Right sets respectively, and with the probability p of an edge across the sets. Each of the n nodes in the Left set is assigned color red and each of the m nodes in the Right set is assigned color yellow, where a color denotes a distinct pathway. We add j random new nodes to the graph of colors red, yellow and black in equal proportions. We select r random nodes with replacement and randomly assign a new color (red or yellow or black) to the node, to have nodes with multiple colors. We then add k random edges to this graph, which may make it non-bipartite. Each edge had weight 1. We generate 4 sets of such networks with $p = 0.5$ and the parameter values given in table 2.

	n	m	j	k	r	$\mu[e_S - e_G]$	$\sigma[e_S - e_G]$
I	10	20	6	5	15	1.7	1.05
II	10	20	12	10	20	3.8	1.5
III	25	25	6	5	30	0	0
IV	25	25	12	10	35	0	0

Table 2. Parameter values tested and results obtained in Simulation 2

For each set (I–IV) we generate 10 graphs. We then run the ILP for the MWPBS problem and evaluate the bipartite subgraph found. The input list of pathways has 3 pathways, one for each color (red, yellow, black).

Let H be the input (simulated) graph to the ILP and let S be the solution returned by the ILP. Let e_G be the number of edges in simulated bipartite graph to which k random edges were added to form H. We know that H contains a bipartite subgraph with at least e_G edges. There may be more than e_G edges due to the added random edges. Let e_S be the number of edges in S.

We observe that the ILP correctly returns bipartite graphs in all the 40 solutions over simulation sets I–IV, and all the nodes in the left/right node sets are of one color, with the two colors being different. Table 1 shows the mean (μ) and standard deviation (σ) of $e_S - e_G$ over 10 runs for each set I–IV. In no simulation did S have lesser number of edges than e_G. This suggests that the ILP found the maximum weighted bipartite subgraphs in each case for these relatively small networks.

Simulation 3

We generate 6 sets of nodes where each set contains 10 nodes and has a distinct color. We generate 6 bipartite graphs with number of edges $100, 50, 25, 15, 10, 5$ by randomly selecting pairs of node sets. The same set may be selected more than once but the same pair of node sets is not selected more than once. These graphs are merged into one graph. We add j random new nodes to the graph where one of the six colors is randomly assigned to each added node. We select $r = 30$ random nodes with replacement and randomly assign a new color (from the six colors) to the node, to have nodes with multiple colors. We then add k random edges to this graph. Each edge has weight 1. We generate 4 sets of such networks with the parameter values given in table 3. For each set (I–IV)

	j	k	$\mu[E_S - E_G]$	$\sigma[E_S - E_G]$
I	20	10	20	0
II	40	20	39.5	0.7
III	30	10	30	0
IV	60	20	19.4	4.25

Table 3. Parameter values tested and results obtained in Simulation 3

we generate 10 graphs. We then run our algorithm 1 for the K-MWPBS problem, with $K = 6$ and evaluate the bipartite subgraphs found. The input list of pathways has 6 pathways, one for each color used.

Let H be the input graph to the algorithm and let $S_1 - -S_6$ be the 6 bipartite graphs returned by the ILP. Let E_G be the total number of edges in all the bipartite graphs merged to which k random edges were added to form H. We know that H contains 6 bipartite subgraphs with at least E_G edges. There may be more than E_G edges due to the added random edges. Let E_S be the total number of edges in $S_1 - -S_6$.

We observe that our algorithm correctly returns bipartite subgraphs in all the 40 solutions over simulation sets I–IV, and all the nodes in the left/right node sets, in each subgraph, are of one color, with the two colors being different. The number of bipartite subgraphs may be less than six as larger bipartite subgraphs may be formed due to random addition of colors and edges in the simulation. Table 3 shows the mean (μ) and standard deviation (σ) of $E_S - E_G$ over 10 runs for each set I–IV. In no simulation was E_S less than E_G. This suggests that our algorithm found six maximum weighted bipartite subgraphs in each case for these relatively small networks.

B Experiment Settings

To focus on "signaling pathways", we excluded the following 17 pathways in KEGG: huntingtons_disease, pathways_in_cancer, colorectal_cancer, pancreatic_cancer, endometrial_cancer, prostrate_cancer, thyroid_cancer, melanoma, bladder_cancer,
chronic_myeloid_leukemia, acute_myloid_leukemia, small_cell_lung_cancer,
non_small_cell_lung_cancer, hypertrophic_cardiomyopathy_HCM, arrhythmogenic_right_
ventricular_ cardiomyopathy_arvc, dilated_cardiomyopathy, viral_myocarditis. From Hallmark we do not exclude any pathways but utilize all existing 50 pathways.

Table 4 shows the list of pathway name abbreviations used in our figures.

Table 4. Pathway name abbreviations used in figures.

Pathway abbreviations	Full Pathway names
MAPK	MAPK Signaling Pathway
Calcium	Calcium Signaling Pathway
p53	p53 Signaling Pathway
cell cycle	Cell Cycle Pathway
Hippo	Hippo Signaling Pathway
CAMS	Cell adhesion molecules Pathway
Myc	Myc Pathways
Nrf2	NRF2 Pathway
RTK/RAS	RTK-RAS Signaling Pathway
TGFβ	TGF Beta Signaling Pathway
PI3K/Akt	PI3K-Akt Pathway
Ubiquitin	Ubiquitin Mediated Proteolysis Pathway
Wnt	Wnt Signaling Pathway
Adher junc	Adherens Junction Pathway
apoptosis	Apoptosis Pathway
focal adh	Focal Adhesion Pathway
ECM	ECM-receptor interaction Pathway
ABC trans	ABC Transporter Pathway
Tight junc	Tight Junction Pathway
Notch	Notch Signaling Pathway
H p53	Hallmark p53 Pathway
H Apical junc	Hallmark Apical Junction Pathway
H Compl	Hallmark Complement Pathway
H EMT	Hallmark Epithelial-mesenchymal Transition Pathway
H IL2 STAT5	Hallmark IL2 STAT5 Signaling Pathway
H Myogenesis	Hallmark Myogenesis Pathway
H KRAS s UP	Hallmark KRAS Signaling UP Pathway
H KRAS s DN	Hallmark KRAS Signaling DN Pathway
H UV R DN	Hallmark UV Response DN Pathway
H UV R UP	Hallmark UV Response UP Pathway

The 10 canonical signaling pathways discussed in [39] are:

1. Cell Cycle Pathway
2. Hippo Signaling Pathway
3. Myc Pathway
4. NRF2 Pathway
5. RTK-RAS Signaling Pathway
6. TGF Beta Signaling Pathway
7. PI3K-Akt Pathway
8. Notch Signaling Pathway
9. p53 Signaling Pathway
10. Wnt Signaling Pathway

C Analysis with Hallmark pathways

The pathway pairs given by our algorithm when Hallmark pathways are used are shown in figure 6 for BRCA and in figure 8 for STCA. Between-pathway significance scores for these pathways and the canonical signaling pathways are shown in figure 7 for BRCA and figure 9 for STCA.

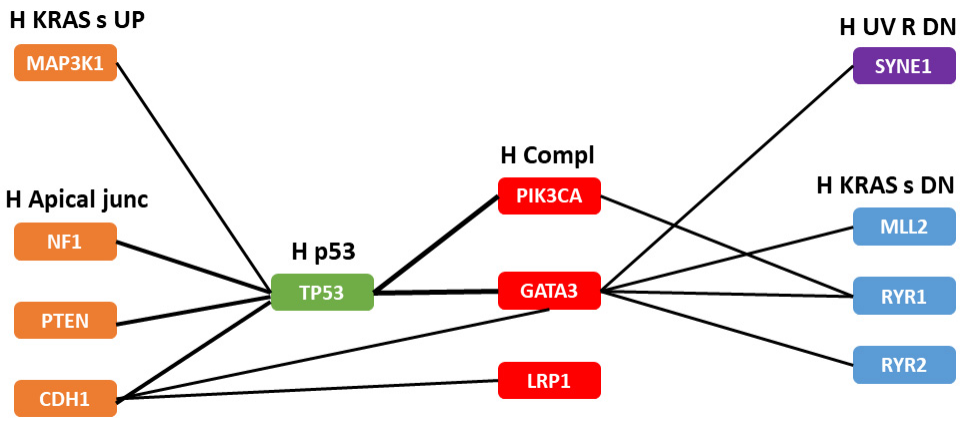

Fig. 6. Pathways and genes selected by our method from the BRCA network. Nodes denote genes, edges indicate ME relations and each pathway is identified by a color. Thickness of edges is proportional to its significance.

The common genes identified in BRCA data by our method in both KEGG and Hallmark pathways are TP53, PIK3CA, MAP3K1, PTEN, CDH1, NF1, RYR1, and RYR2. Some genes such as GATA3, SYNE1 only exist in Hallmark whereas CACNA1E, MAP2K4 genes only exist in KEGG.

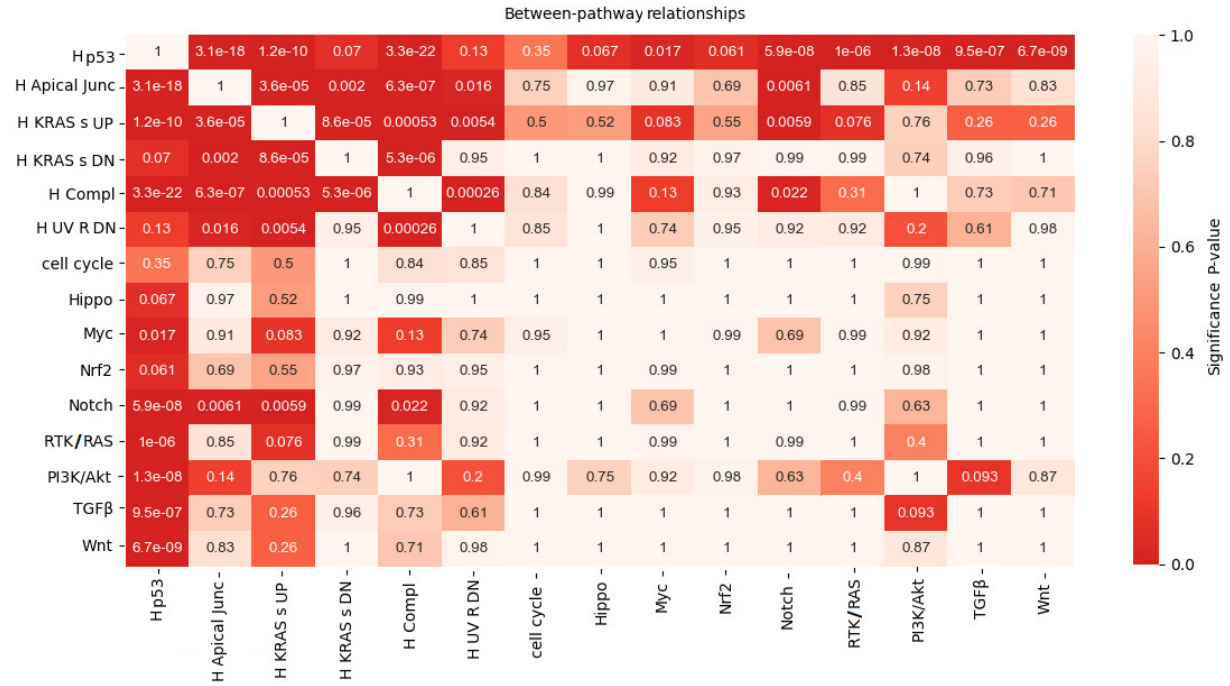

Fig. 7. Between-pathway significance scores for pathways discovered by our method from the BRCA dataset along with 10 canonical signalling pathways in [39].

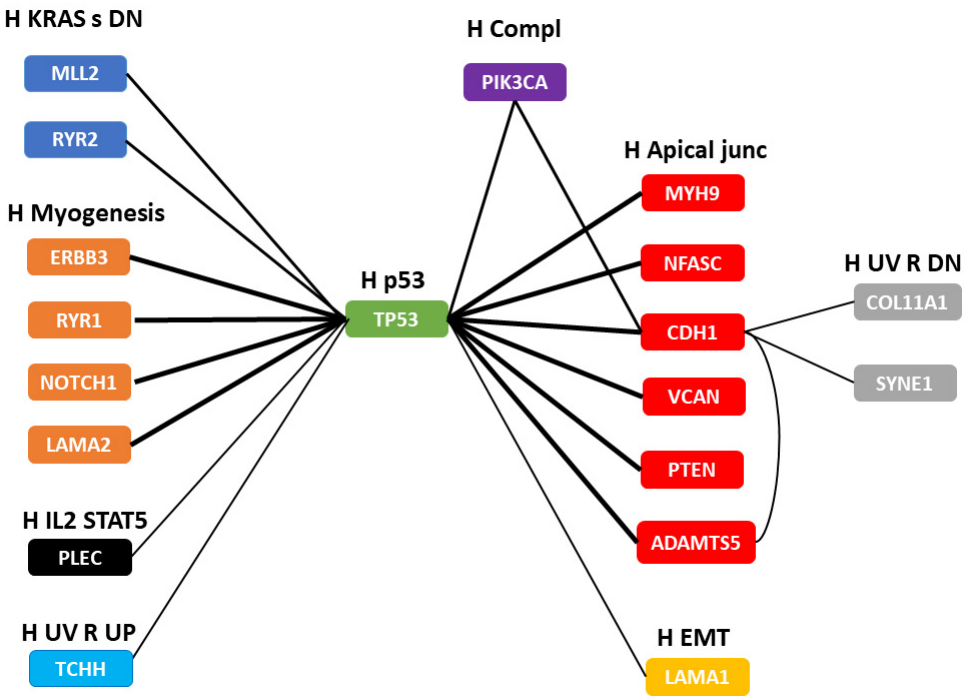

Fig. 8. Pathways and genes selected by our method from the STCA network. Nodes denote genes, edges indicate ME relations and each pathway is identified by a color. Thickness of edges is proportional to its significance.

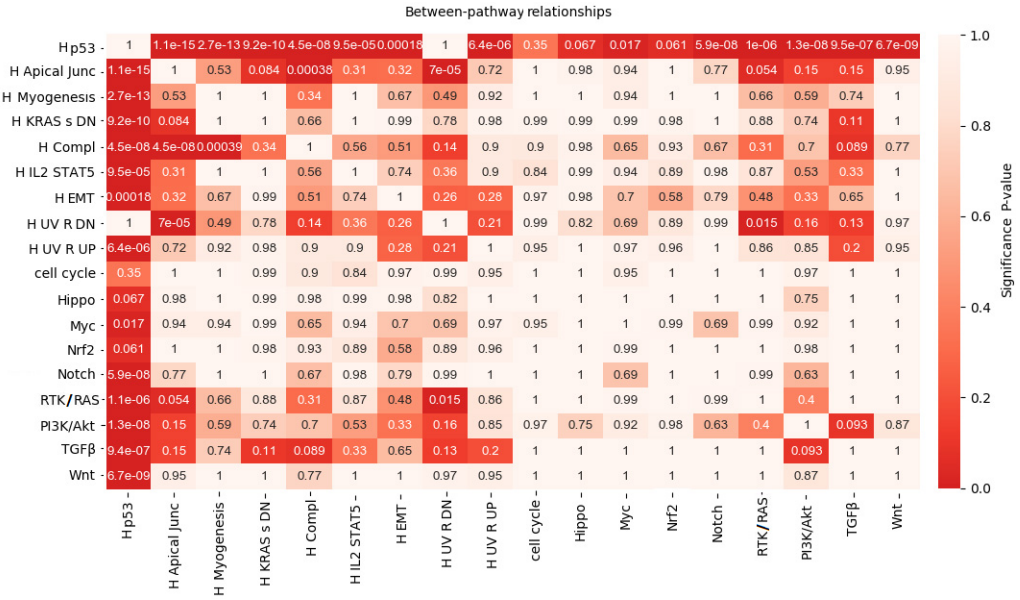

Fig. 9. Between-pathway significance scores for pathways discovered by our method from the STCA dataset along with 10 canonical signalling pathways in [39].

In the STCA dataset, common genes identified by our method in both KEGG and Hallmark are TP53, PIK3CA, CDH1, NOTCH1, PTEN, MYH9, ERBB3, NFASC, COL11A1, LAMA2, and VCAN. Some other identified genes,SYNE1, ADAMTS5, MLL2, PLEC only exist in Hallmark and CACNA1E, CACNA1C, ABCA12, SMAD4 genes only exist in KEGG.