Script for Video

"A Network Approach to Cancer: Clustering Cell Lines by Gene Dependency Rank" by Jane Adams, Celestin Coquide, Diana García and Rodrigo Migueles Ramirez.

Cancer cells bear modifications that confer them with phenotypes that compose the hallmarks of cancer. However, these modifications also result in vulnerabilities. In this work, we use networks to compare cancerous cell lines in terms of their vulnerabilities. In particular, we are interested in similarities among cell lines of different cancer types, as a potential avenue of exploration for repurposing existing gene therapies for additional oncological applications. We use the Achilles DepMap gene dependence data set to analyze 625 different cancer cell lines and their gene dependence survival probabilities. The DepMap data was generated using CRISPr technology to target approximately 18 thousand unique genes by sgRNA, in each cell line. Gene dependence probabilities are assigned, based on the cell survival when each gene is removed. We use these gene dependence probabilities to assign a Spearman rank to each gene with a dependence probability above 0.5, the recommended threshold provided by DepMap project authors. Therefore, every cancer cell line in our set of 625 has a gene rank list. We then calculate the Spearman rank correlation between each pair of cancer cell lines, to build a network weighted by gene dependence probability rank correlation. In order to identify the unusual connections in this network, we sparsify the matrix using the Serrano et. al. backbone method. Unlike universal thresholding sparsification methods, this method is locally aware; it preserves edges based on the average edge weights associated with each node, and keeps edges which are most unusual within the local network context. The tuning parameter alpha, on the x axis, represents the 'unusualness' threshold of edge weights. The red line represents the proportion of edges included in the graph at each alpha level, while the black line shows the component count. We threshold the network at an alpha value which maximizes the component count, which here is 12, to produce a cell line network with communities for further exploration. The resultant network graph displays 126 cancer cell lines and 183 edges. Interestingly, cell lines of certain cancer types appear to be more commonly linked to cell lines of a different cancer type, while others seem predominantly homophilic. We show here for each cancer type the number of nodes in the network graph and the homophily score of each cancer type. This is not to imply a causal relationship between these two variables, as the number of nodes is dependent upon the number of cell lines for that cancer type in the original cell line data set, but to provide a map to visualize metadata about cancer types. Of note, lung cancer and pancreatic cancer cell lines, both cancer types for which we have significant amounts of data,

are more frequently connected to cell lines of different cancer types, as evidenced by a homophily score below 0, indicating that gene dependence probability rankings among lung cancer and pancreatic cancer cell lines most resemble the gene dependence rankings of other cancer types. Conversely Leukemia, another well-represented cell line, has a homophily score close to 1, indicating that gene dependence probability rankings among leukemia cell lines most resemble the rankings of other leukemia cell lines. These similarities provide an interesting exploratory analysis of gene dependence rankings when evaluating drug treatments in a gene therapy context. Finding gene dependence probability rank correlations between cancer cell lines of different types may help to identify similarities in vulnerability, thus providing hints for combinatorial therapies and novel applications for existing therapeutic strategies.