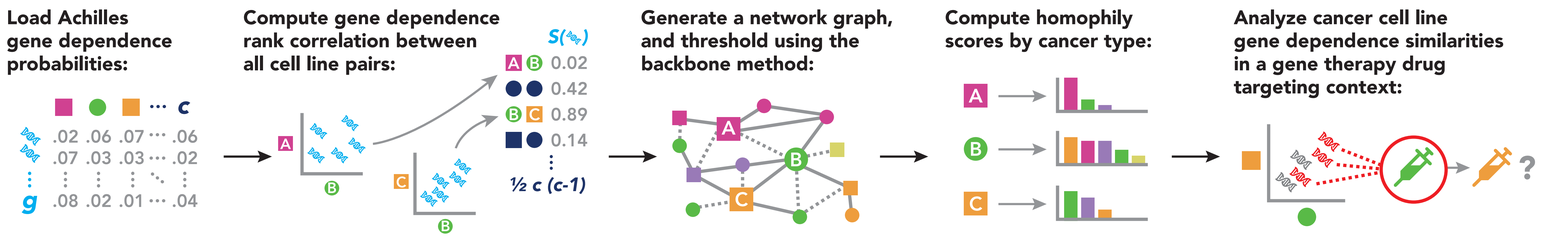

Fig 1. *An explanatory graphic describing the data processing workflow to create the gene rank correlation network of cell lines.*

Homophily and Heterogeneity in the Cancerous Cell Line Network for Gene Therapy Improvements

Jane Adams[1], Celestin Coquide[2], Diana Garcia[3], Rodrigo Migueles Ramirez[4]

[1]Vermont Complex Systems Center, University of Vermont, United States, [2]Institut UTINAM, OSU THETA, University of Bourgogne, France, [3]National Institute of Genomics Medicine, Mexico City, Mexico, [4]Quantitative Life Sciences Interfaculty Program, McGill University, Montreal, Canada.

Differences among cells across and within cancerous tumors present substantial obstacles to effective therapy treatments. Genetic profiling of cancerous cell lines reveal that they bear not only advantageous differences when compared to normal cells, but that some of these modifications also result in specific vulnerabilities. To screen these vulnerabilities, the DepMap project [1] recently analyzed the genetic dependencies of cancerous cell lines using CRISPr technology by systematically targeting a vast range of genes (18k) with sgRNAs and recording the resultant cell line survival rates.

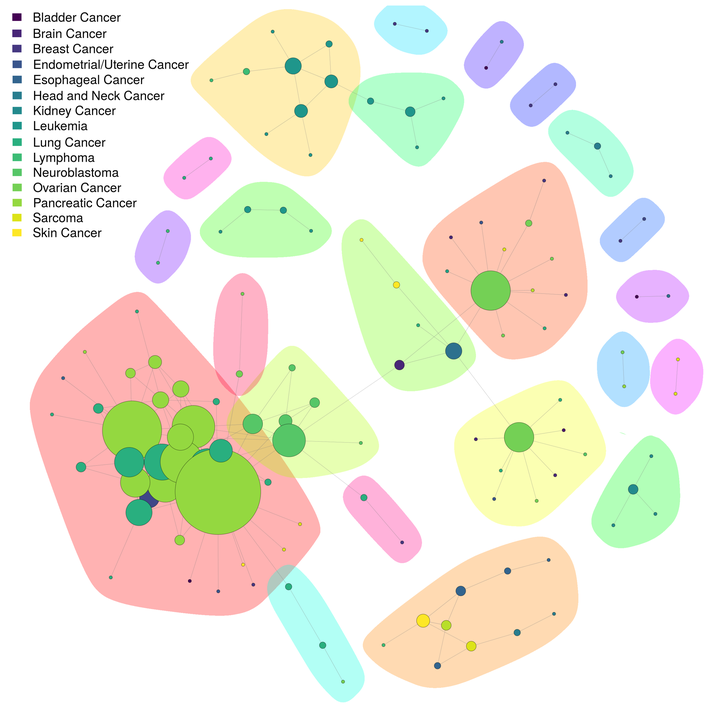

Fig 4. *The network graph, sparsified using the backbone method. There are 126 nodes (cell lines), and 186 links, weighted by rank correlation.*

Here we show that network analysis of gene dependencies for cancer cell line survival reveals interesting similarities in genetic dependency profiles among different cancer types. Using DepMap's Achilles data set, we analyze 625 different cell lines representing 28 unique cancer types, and compute similarities among cell lines by their gene dependence profiles. We use gene dependence probabilities to compute Spearman rank correlation coefficients among all pairs of cancerous cell lines, and build a cell line matrix weighted by this correlation coefficient. We then apply a backbone method [2] to the correlation matrix to preserve the cell line correlations with high structural significance, resulting in a cell line network with 19 connected components. Using the links of the network, we compute balanced homophily (β) for each cell line to compare connectivity within and across cancer types based on their anatomical origin.

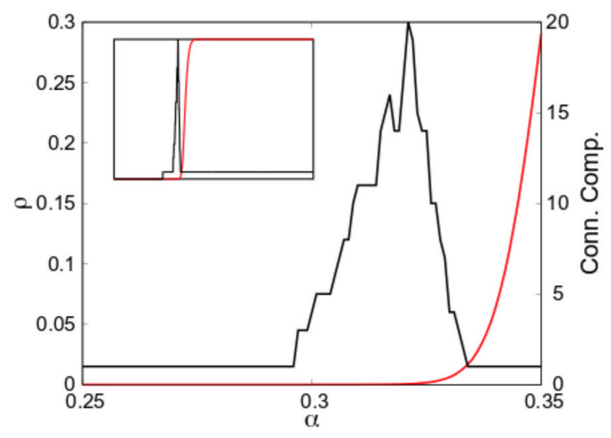

Fig 2. *A line chart showing A) in red and labeled on the left axis, the proportion of links remaining; and B) in black and labeled on the right axis, the number of connected components; as the backbone thinning parameter a increases.*

Our results show that although some cell lines share similarities preferentially with cell lines of the same origin, such as leukemia (β= 0.55), some of the most prevalent and aggressive cancer types such as lung cancer have low homophily scores (β=−0.58), suggesting unexpected and potentially useful gene dependency similarities with other cancer types. By discretizing gene dependency values, we establish gene sets classified by importance grades and identify enriched biological processes in those sets from gene ontology. With this approach, communities in our cell line association network can be related with their enriched biological processes to find shared processes between cell lines and cancer types.

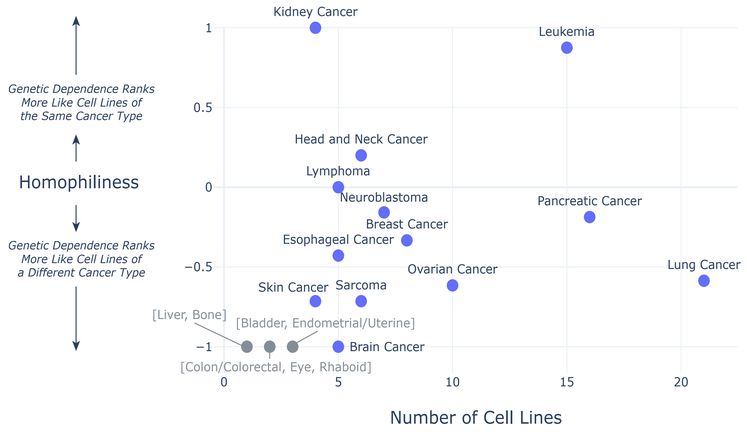

Fig 3. *A scatter plot showing cell line cancer types by the total number of cell lines in the resultant network graph, and the computed homophily score of that cancer type. Gray dots indicate cancer types with overlapping values. The five most represented cancer types in the network graph are 1) lung and 2) pancreatic cancer, 3) leukemia, 4) ovarian and 5) breast cancer. Of those, all but leukemia have a homophily score below 0, meaning that they are connected by their gene dependence rank correlation more commonly to different cancer types in the resultant network.*

From these results, we suggest that a network approach to analyzing cell line similarity could support the identification of similarities in the vulnerabilities of different cancer types, thereby expanding the therapeutic range of existing genetic cancer treatments. Given the often arduous experimental and regulatory processes of developing new gene therapies, this method offers promising potential for employing already approved gene therapy technologies in innovative ways through the targeting of cancer types with similar genetic dependency profiles.

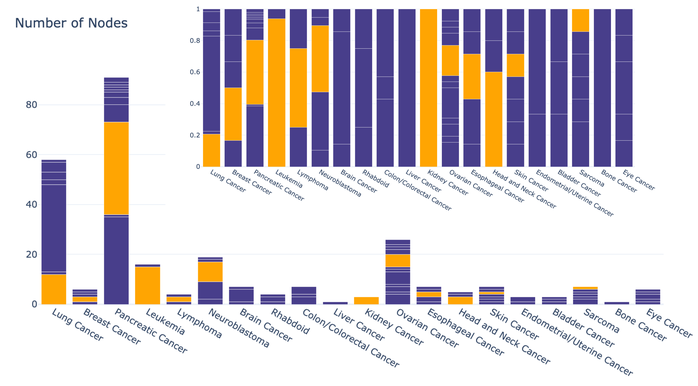

Fig 5. *A barplot showing the number of nodes in the network for each cancer type. Node counts of the same cancer type are indicated in orange, while node counts of a different cancer type are indicated in blue. Inset is a plot showing connected node type percentages.*

[1] Tsherniak, A. et al. *Defining a Cancer Dependency Map*. Cell 170, 564-576.e16 (2017). DOI: 10.1016/j.cell.2017.06.010.

[2] Serrano, M. et al. Extracting the Multiscale Backbone of Complex Weighted Networks. Proceedings of the National Academy of Sciences. Apr 2009, 106 (16) 6483-6488; DOI: 10.1073/pnas.0808904106