# A Network Approach to Cancer: Clustering Cell Lines by Gene Dependency Rank Distances

*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*

Cancer heterogeneity is one of the main obstacles that prevent efficient treatment: each tumor is usually composed of widely divergent cells. Conversely, different types of cancer can present unexpected genetic similarities.

Recently, the DepMap project[1] evaluated the gene dependency for survival in cancer cell lines using CRISPr technology to target a wide number of genes (18k) with sgRNAs. In this project, we use the Achilles dataset from the DepMap project to find similarities between 625 different cancer cell lines based on their genetic dependency profiles.

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

The genetic dependency profiles are characterized by a gene effect score and a gene dependency probability.

Spearman's rank correlation coefficients were calculated on the dependency probability matrix between each pair of cell lines. A backbone method[2] was applied to the correlation matrix resulting in a network with 12 connected components.

Half of the components were composed of heterogeneous tumour origins, accounting for 84.21%of the cell lines. One prominent component containing 60 nodes from 12 different types of cell lines features pancreatic cancer as the central hub. Although some same-origin cell lines form cliques, like leukemia cell lines, 93.75% of the cell lines have a tendency to be preferentially linked with cell lines of different origins, as measured by balanced homophily.

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

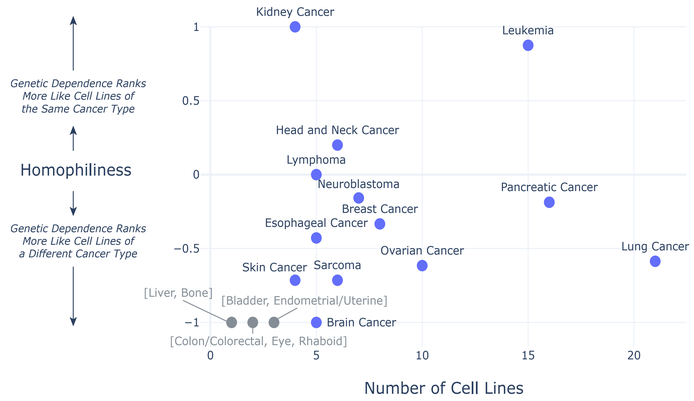


**Fig 3.** *A scatter plot showing cell line cancer types by the total number of cell lines in the resultant network graph, and the computed homophily score of that cancer type. The five most represented cancer types in the network graph are 1) lung and 2) pancreatic cancer, 3) leukemia, 4) ovarian and 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.*

These results suggest that this approach could allow us to identify genes to which cancer types are similarly vulnerable, thus providing hints for combinatorial therapies and novel applications for existing therapeutic strategies.

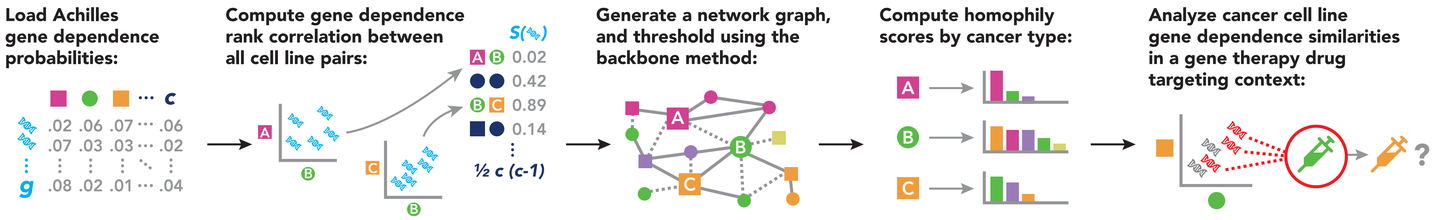


**Fig 4.** *An explanatory graphic describing the data processing workflow to create the gene rank correlation network of cell lines.*

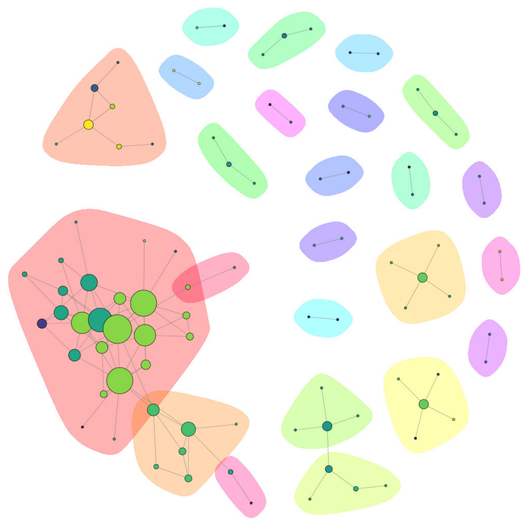

**Fig 1.** *The network graph, sparsified using the backbone method. There are 126 nodes (cell lines), and 186 links, weighted by rank correlation.*

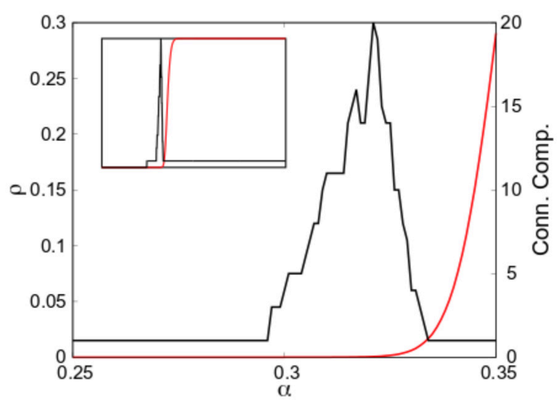


**Fig 2.** *A line chart showing A) in red and labeled on the left axis, the proportion of links remaining; and B) in black and labeled on the right axis, the number of connected components; as the backbone thinning parameter α increases.*

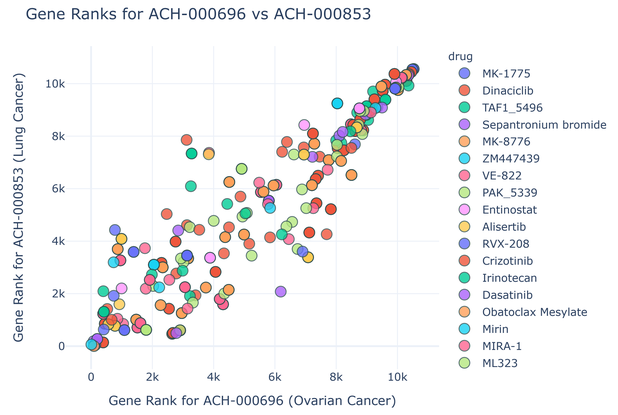


**Fig 5.** *A scatter plot showing a subset of gene dependence ranks for two cell lines (one ovarian cancer cell line and one lung cancer cell line). This pair of cell lines is connected in the network graph for its high gene dependence rank correlation. Genes upon which both cell lines are dependent for survival are in the upper right corner. The subset of gene dependence ranks shown represent 386 genes for which there exist gene therapy drugs. The legend shown is a list of gene therapy drugs, in order of the highest gene dependence rank for cell line ACH-000696 of that drug's gene targets (here, MK-1775, which targets genes WEE1 and PLK1).*