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

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

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

We introduce a network approach to grouping phenotype-variant cancer cell lines by gene dependency. 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 [] 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. The genetic dependency profiles are characterized by a gene effect score which measures the effect size of knocking out a gene and by the dependency probability of that gene to be essential to cell survival, given its effect score. Spearmans rank correlation coefficients were calculated on the dependency probability matrix between each pair of cell lines. The Backbone method [] 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 of the giant component. Although some same origin cell line cliques were observed (for instance leukemia cell lines rarely share any edges with non-leukemia cell lines), 93.75% of the cell lines have a tendency to be preferentially linked with cell lines of different origins than to the same origin, as measured by balanced homophily. Some of the most aggressive cancer types showed unexpected similarities to other cancer types such as such as pancreatic cancer - neuroblastoma. We group cell lines based on their susceptibility profiles and recover cell phenotypic cell lines similarities. These results suggest that this approach could allow us to identify genes that cancer types are similarly vulnerable to, thus providing hints for combinatorial therapies and novel applications for existing therapeutic strategies.

References

[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

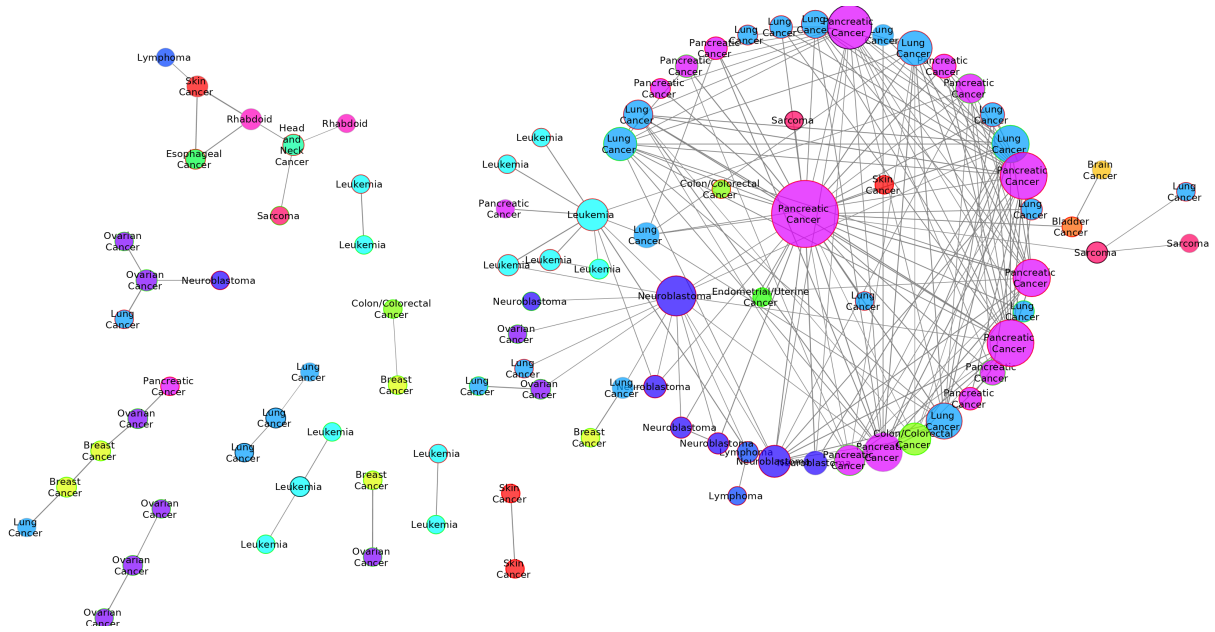

Figure 1: Network of cell-cell gene dependence similarities built from the cell line correlation matrix. Each cell line has been colored according to its cancer type. A total of 95 cell lines and 213 links.