Exploring Data Visualization for Malignant Pleural Mesothelioma

Jacob Watson, Computer Science
Mentor: Dr. Christopher Plaisier, Assistant Professor
School of Computing, Informatics, and Decision Systems Engineering

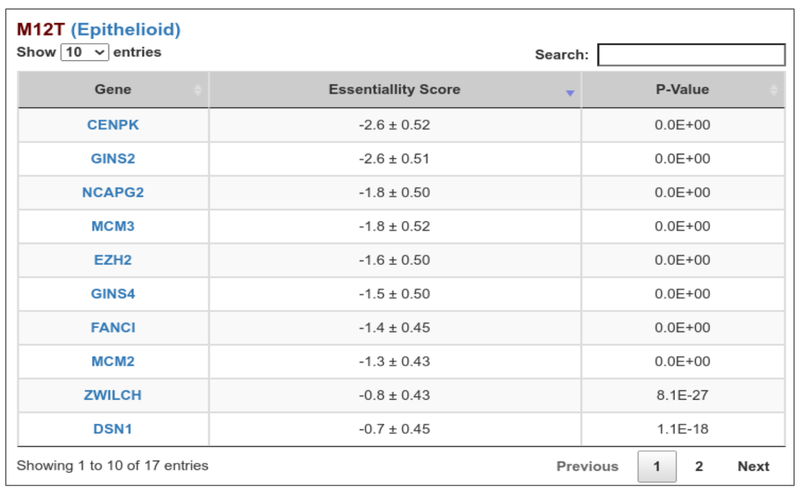
Abstract

A web application has been developed to provide data and statistics for the deadly cancer malignant pleural mesothelioma (MPM). The web application was enhanced to include additional experimental information from CRISPR-Cas9 knock-out outgrowth screens. The ultimate goal of the website is to function as a knowledge base and hypothesis generator for researchers that study MPM. This required additional visualizations and information to be displayed in efficient ways and led to changes in the database structure and web application.

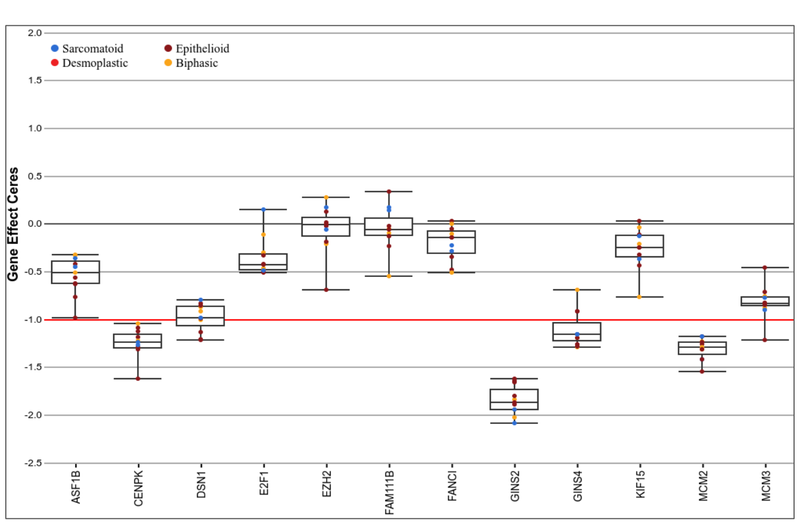
Methodology

We researched new methods of displaying data to improve the graphs for research use. We explored the usage of different visualization types, such as boxplots and tables, to present this data to researchers[1]. One graphic involved improving others' work by combining multiple graphs into one. DepMap offers researchers a look at gene essentiality scores measured between cell lines, but only lets researchers look at a single gene at a time. We improved this by comparing multiple genes at a time, letting researchers compare and contrast gene essentiality.

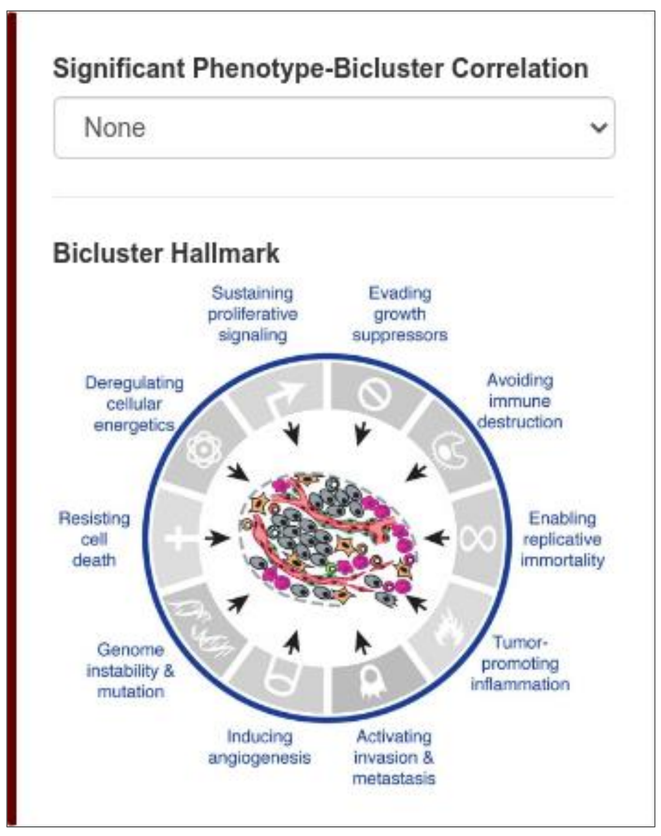
Website Results



Gene essentiality is a measure of how necessary the gene is for cell survival. This is measured from CRISPR-Cas9 knock-out screens. By interpreting the essentiality score, we can determine which genes we should target with a drug inhibitor or small RNA to stop the cancer from proliferating. Unlike DepMap, these results come from cell lines in our lab.



We have mirrored the gene essentiality data for MPM cell lines from a knowledge-base called DepMap. Researchers use this to determine the impact a gene has on cancer cells. DepMap is difficult to use, since it only shows one gene at a time. mpmSYGNAL's summary graph shows all the genes in a bicluster in one graph, allowing comparison between genes.



Searching for meaningful relationships can be difficult, so an advanced search system was developed to assist researchers. For each patient in the database, phenotypes were measured. The phenotypes values are correlated with bicluster expressions, and researchers can search for significant correlations using the advanced search. We also added the capacity to search based upon the association of the bicluster with a hallmark of cancers, such as cancer cells resisting cell death, ignoring anti-growth signals, or avoiding immune system intervention. Researchers may want to study biclusters with particular hallmarks, necessitating an advanced search.

Data Collection

To ensure the most accurate results possible, we tested our website for any flaws with the data. The website allows us to look at our dataset with ease. We found irregularities with our dataset that did not match with other Mesothelioma datasets, and from there we were able to discover problems and correct them.
Replicating results is a fundamental part of experimentation and research, and we wanted to include those results in the website. For example, we integrated the DepMap gene essentiality data alongside our own results, and use it to replicate our results whenever possible.

Conclusion

The project shows that the website is useful at summarizing data for cancer researchers. We can create visualizations easily due to how large our Mesothelioma data set is. We then present the new visualizations to researchers to help them with their research. Feedback from other members of the Plaisier lab and eventually the MPM research community will be crucial in improving the website to the point where it is an invaluable tool in studying this deadly disease.

Future Work

We plan to create a system that is easier to develop for using the lessons learned from implementing mpmSYGNAL. A large amount of data in one place makes it easy to mock out visualizations that display correlations from that data. We want to implement data from all 32 cancers covered by The Cancer Genome Atlas (TCGA). A new website will have visualizations that display correlations found in the data between cancers, which is different from the single cancer analysis mpmSYGNAL currently provides.

References

1. Olston and Mackinlay, "Visualizing Data with Bounded Uncertainty."
2. Plaisier et al., "Causal Mechanistic Regulatory Network for Glioblastoma Deciphered Using Systems Genetics Network Analysis."

FURI

Ira A. Fulton Schools of Engineering
Arizona State University