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

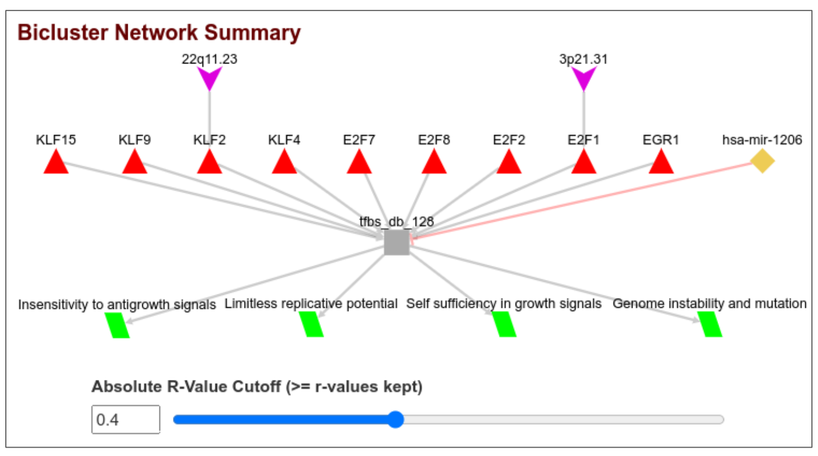
Abstract

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 hypothesis knowledge base and hypothesis generator for researchers that study MPM. This required additional graphics and information to be displayed in efficient ways and led to changes in the database structure and web application.

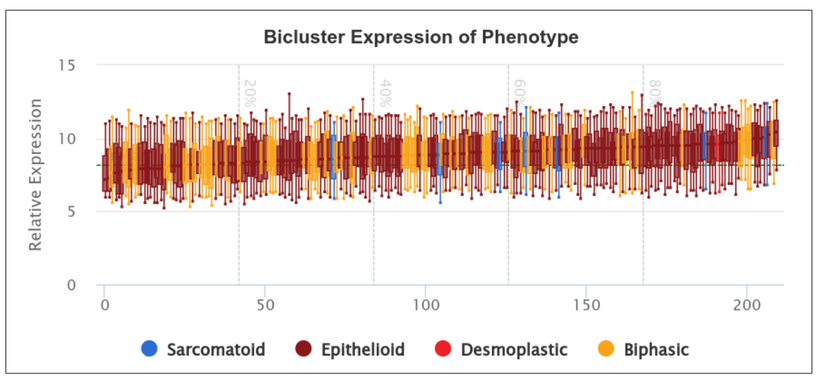
Methodology

In order to improve the visuals on the website, we researched new methods to display data to researchers. This included exploring the usage of different graphs, as well as presenting important correlation statistics to users[1]. We were able to make a number of improvements through this research. One improvement is the r-value cutoff filter, which limits what regulators are shown to the user. This allows researchers to easily search regulators that have a significant correlation with the bicluster they regulate. We used open-source libraries to accomplish this.

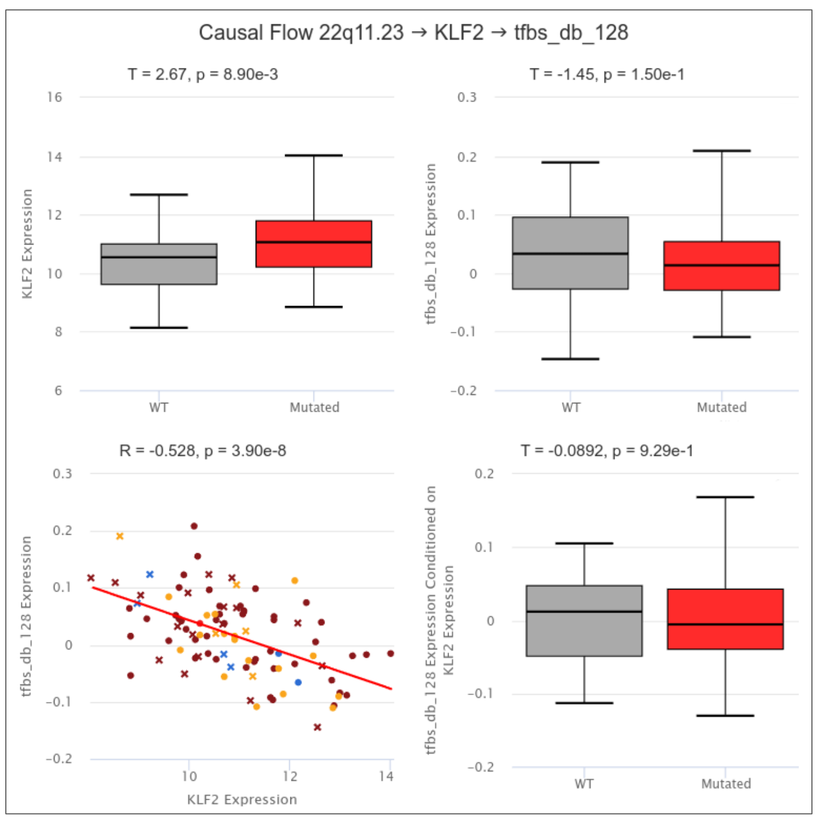
Website Results



The bicluster network summary gives researchers a graphical way of visualizing the relationships within a bicluster. The pink chevrons represent mutations, the red triangles and yellow diamonds represent regualtors, the gray box represents the bicluster, and the green parallelograms represent the hallmarks of the bicluster. A path through a mutation, regulator, and bicluster is a causal flow.



Displays the various gene expressions per-patient for the currently selected bicluster. Each bicluster is composed of multiple genes, and each patient has expression values measured for each gene in a bicluster. The user can select phenotypes like patient sex, or others if they are statistically significant. The figure shows gene expression with tumor subtype selected as the phenotype.



Causal flows are relationships between mutations, regulators, and biclusters, in that order. Researchers need to identify causal flows to know the direction of this relationship flow. Using this relationship, the website plots graphs showing researchers gene expression values. The top graphs show the gene expression of the regulator and bicluster in the causal flow respectively. The correlation between the two is then plotted as a scatter plot on the bottom left, with the residual shown on the bottom right. As you can tell, there is a strong correlation between the causal flow's regulator and bicluster. The view based off of Figure 5 from Plaisier et al.

DB Improvement

In order to meet our goals, the existing database needed to be changed. We needed to fit more data into it, so new tables were created. Speed was a concern since some pages took a long time to load for users. By optimizing the database structure and MySQL queries, we were able to improve the speed in some cases by a factor of 100. Queries were written with performance in mind, and the structure was changed to include indices for improving JOIN performance. Performing queries efficiently improves website responsivity, which users will appreciate when browsing the website.

Conclusion

The project shows that websites dedicated to the presentation of cancer research data are useful. The platform allows us to rapidly prototype unique graphs for researchers, as well as tinker with existing graphs to display data better. Researchers will be able to sift through data faster with the website than they would have been able to without it.

Future Work

Future work includes adding additional data from The Cancer Genome Atlas (TCGA) to our MPM dataset, which will improve the amount of data we show researchers about miRNA regulators. We also plan on expanding the website to cover 32 additional cancers characterized by TCGA. Along with this, we will implement new graphs and views to the website that can be shared across the 32 cancers. This requires generalizing the website to accept a common data format for the cancers.

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