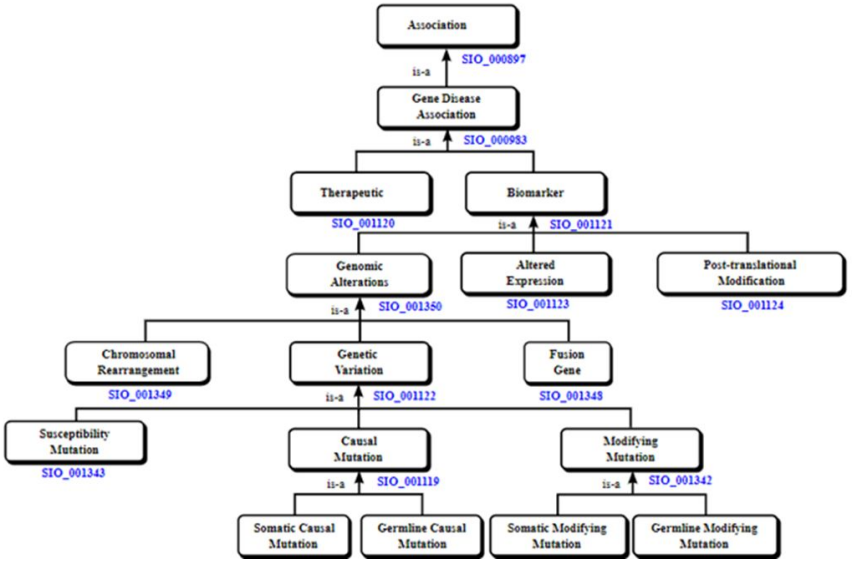
Exploration and Analysis of Gene-Disease Associations

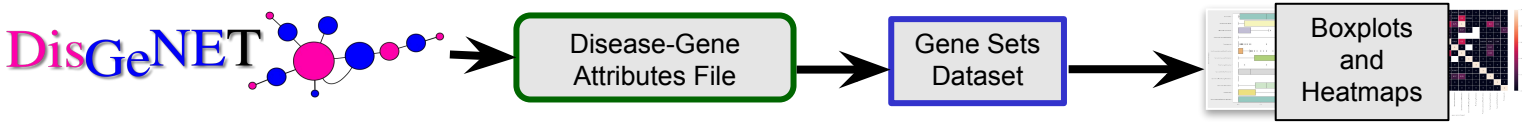
Arun Agarwal, Anna Yannakopoulos, Arjun Krishnan

Introduction

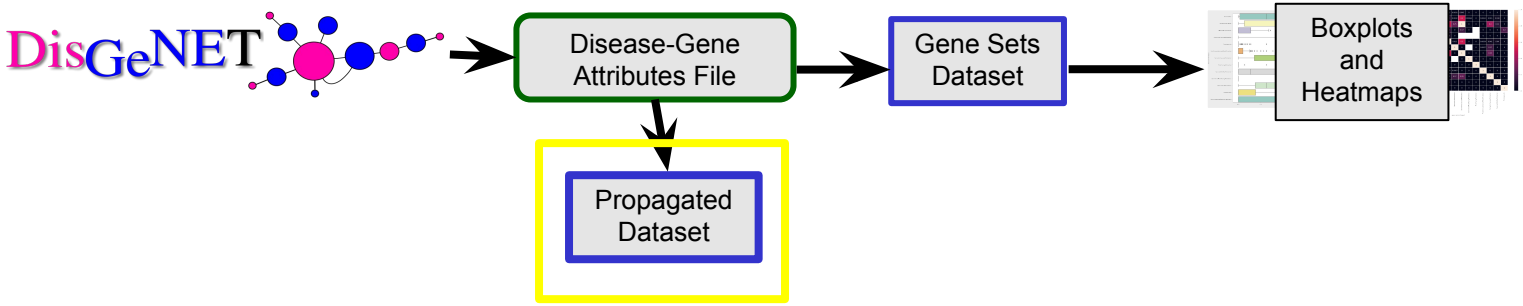
- Many diseases arise due to disruptions of multiple genes
- To work towards cures, the human genome has been studied through various experimental methods that provide different understandings of the disease's genetic cause
- **Hypothesis**: The genomic basis of complex diseases can be understood on a holistic level through the exploration and combination of genetic data from distinct sources, experimental methods, and association types
- DisGeNET: database used with human gene-disease associations



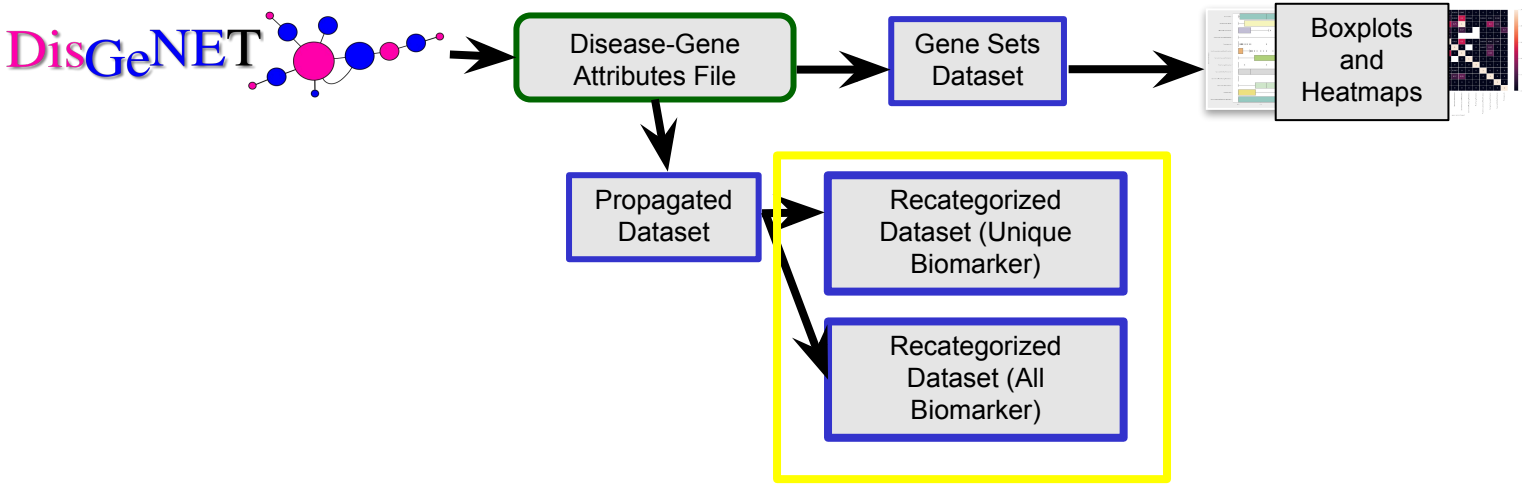
Association Type Exploration



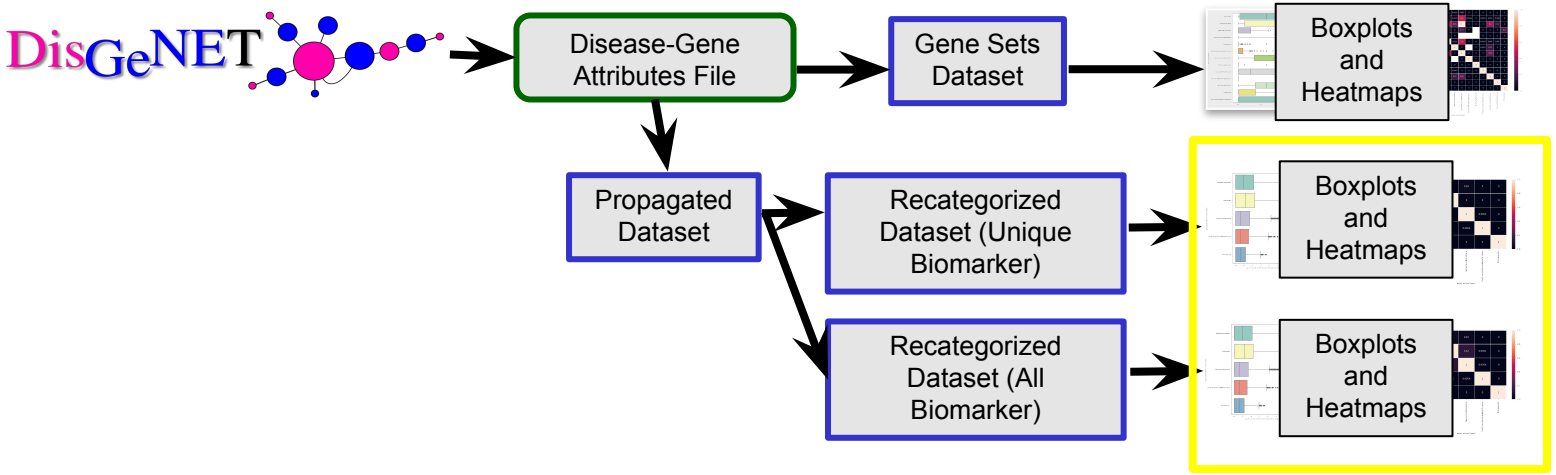
Association Type Exploration



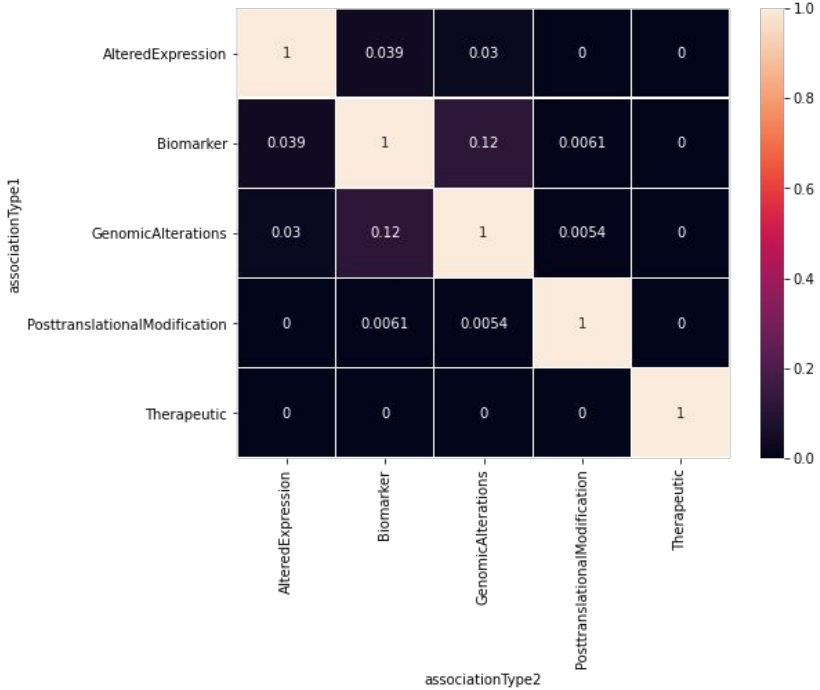
Association Type Exploration



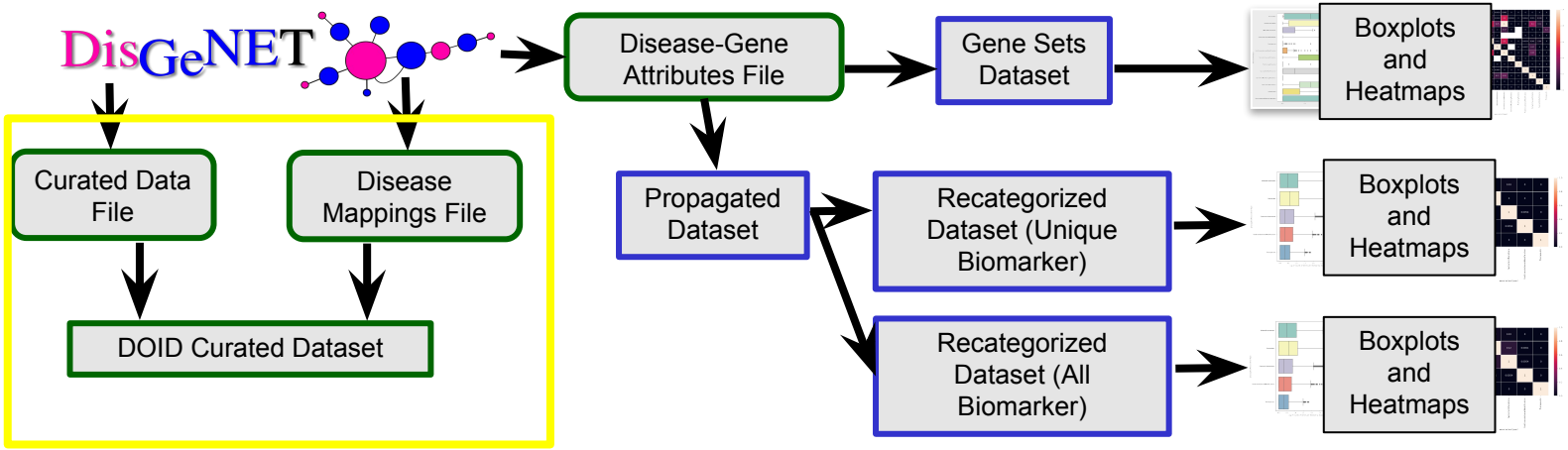
Association Type Exploration



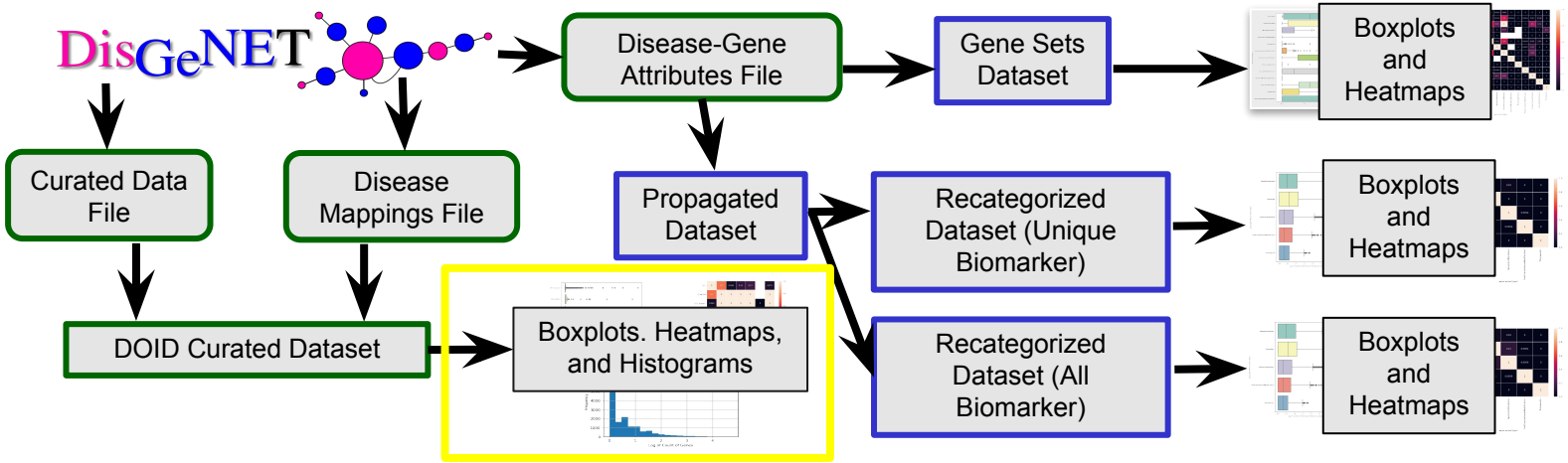
Heatmap of Median Correlation Between Gene Sets for Chosen Association Types



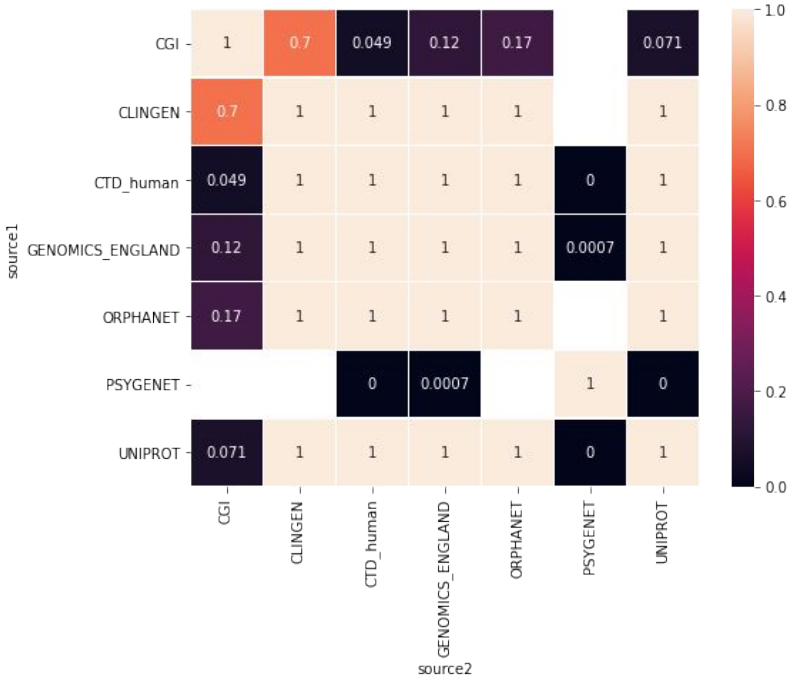
Sources Exploration



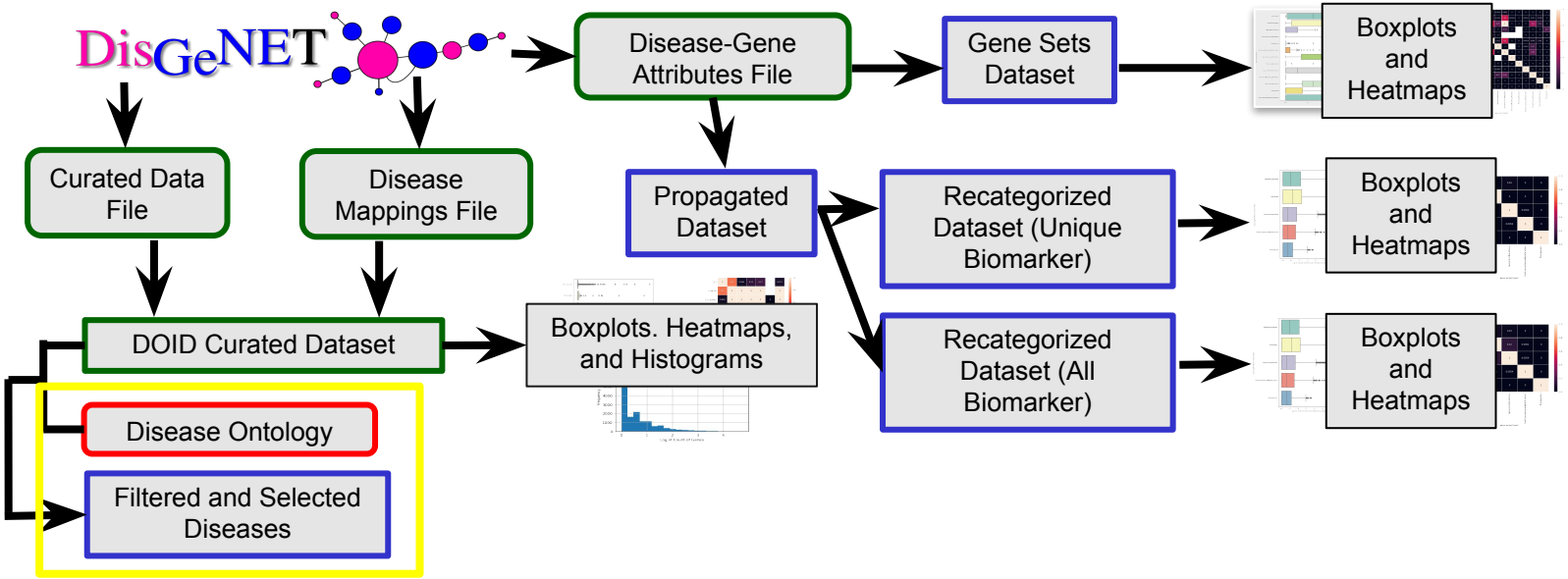
Sources Exploration



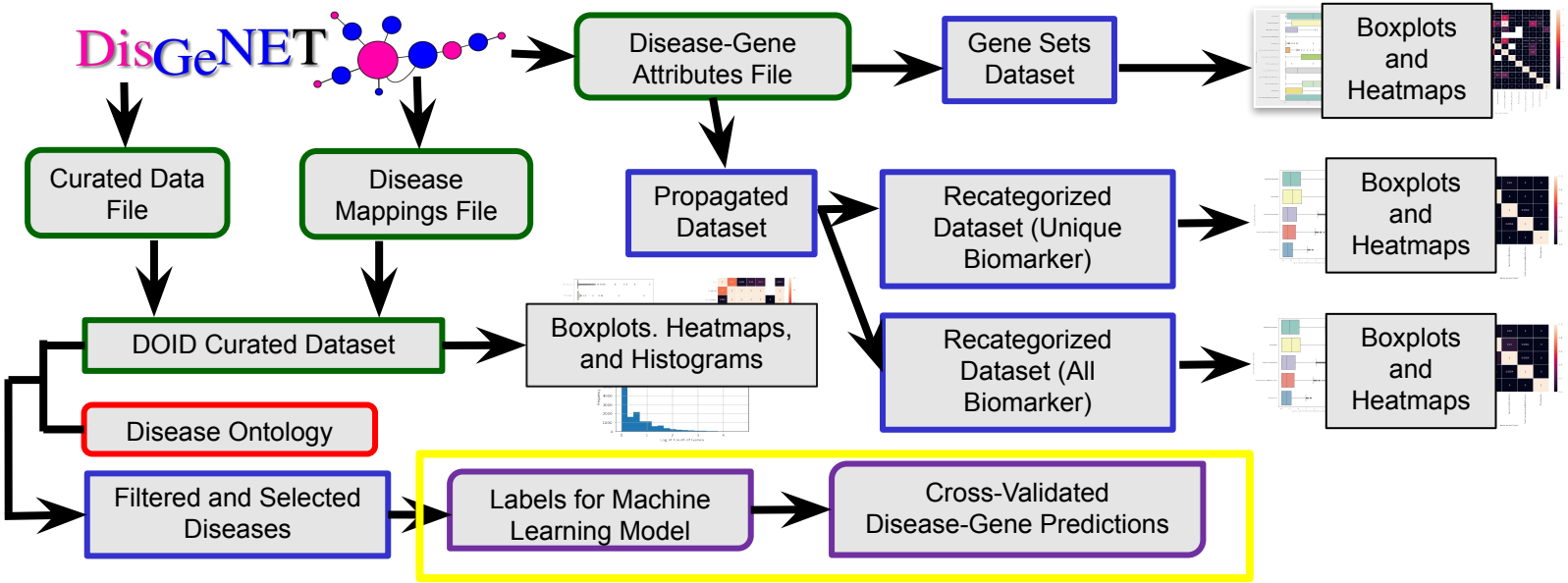
Heatmap of Median Correlation Between Gene Sets for All Sources



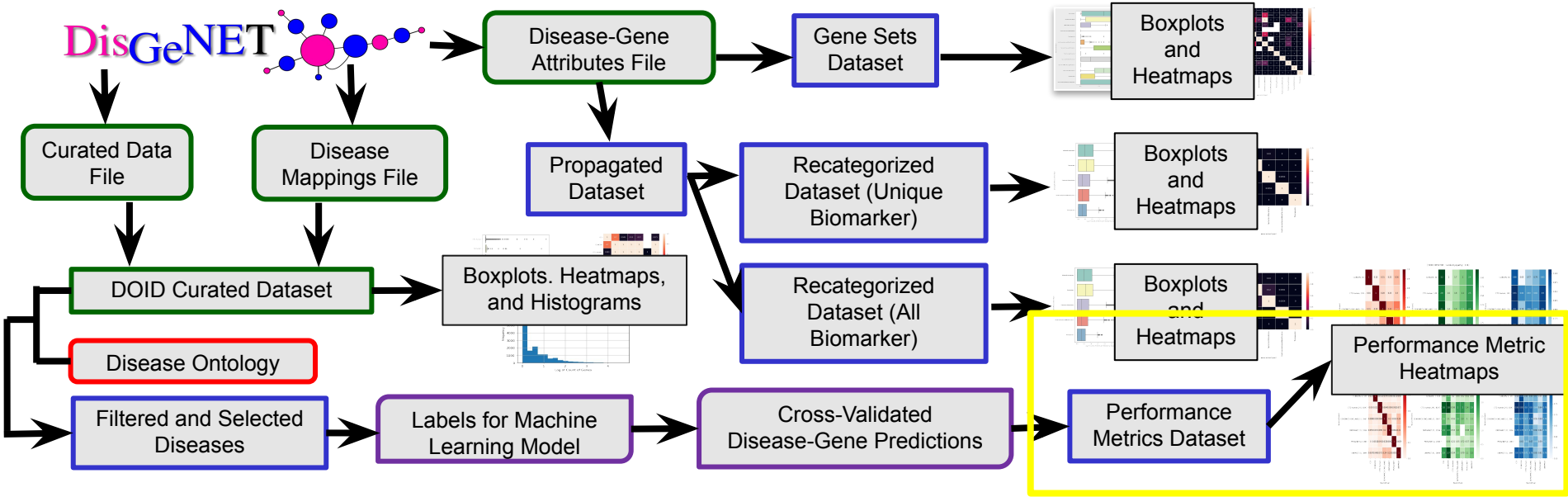
Sources Exploration



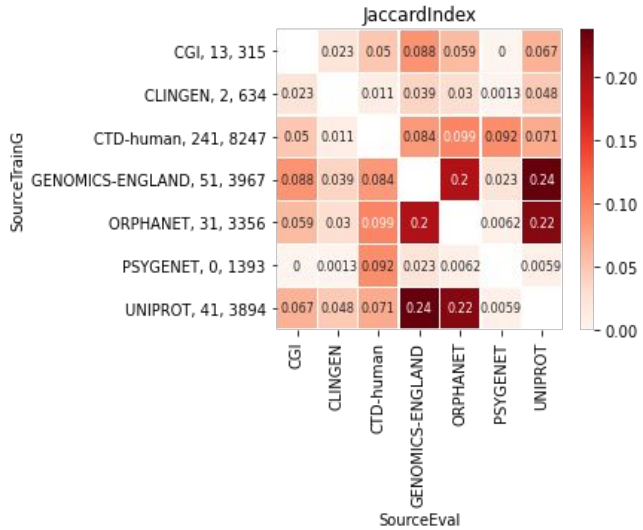
Sources Exploration

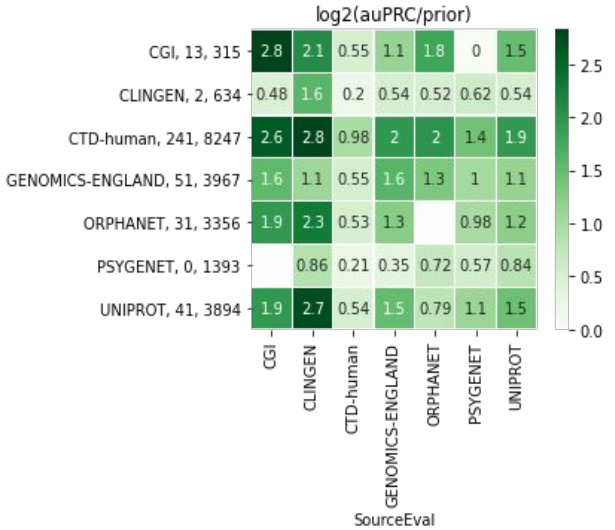


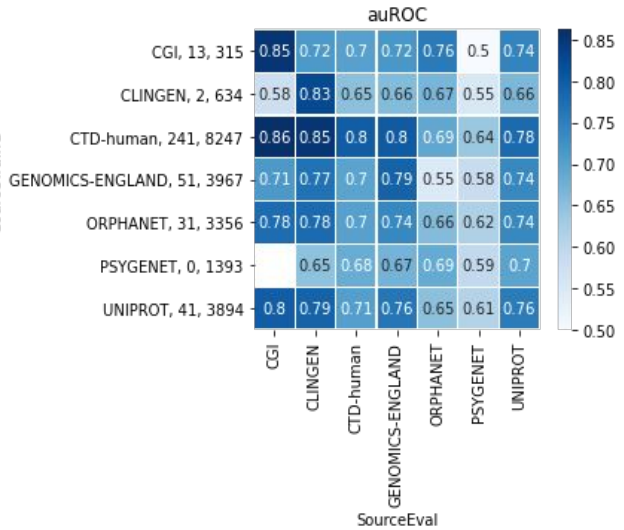
Sources Exploration



Heatmaps of Median Correlation Score Between Gene Sets, log2(auPRC/prior) Score, and auROC Score for All Sources







Conclusion

We tried to answer:

1. Do different experimental methods provide the same view of the underlying biology, or do they present different aspects of it?
2. Can models trained on one experimental method predict the genes associated to another?

Next Steps

1. Create an ensemble model
2. Make my work available for further disease study