DATA MINING AND VISUALIZATION OF COPY NUMBER VARIANTS IN DISEASE PROCESSES

A. E. Dellinger*, M. Yu, and Y.J. Li

Center of Human Genetics, Duke University Medical Center, Duke University 595 LaSalle Street Durham, NC 27710, USA

**Email: andrew.dellinger@duke.edu*

Copy number variation (CNV) detection is a relatively new genetic tool for finding effects of variation on disease. As marker density on single nucleotide polymorphism (SNP) arrays is increasing, density of genome wide CNV data is also increasing, adding to the importance of data mining and visualization of CNV data. While many aspects of CNV data analysis present new challenges, interpretation of results and data mining for patterns connecting CNVs of interest to biologically relevant data are major tasks to achieve. The objective of this study is to create a programmatic tool to confront these challenges. This tool is a Java-based GUI. It incorporates a new statistic, the Normalized Singleton Ratio[1], to determine the optimal CNV detection method and parameters for a dataset. Data mining was used to output and display patterns in the biological effects of significant CNVs that are relevant to a disease of interest. Data from CNV and gene databases, gene ontology (GO)[2], Online Mendelian Inheritance in Man ® (OMIM®)[3], pathways, and expression studies were extracted using CNVs from the user's study and disease information input by the user. The resulting data patterns are visualized for the user to explore. The capabilities of this tool to incorporate and visualize data about a CNV's potential biological effects are a significant step in the ability to find evidence tying CNVs to disease.

Acknowledgments

Program development is supported by the US National Institute of Health (NIH) grant (1R21-EY-019086-01) to YJL

References

1. Dellinger AE, Saw SM, Goh LK, Sielsteid M, Young TL, and Li YJ. Comparative analysis of seven algorithms for identifying copy number variants from genome wide SNP arrays. Submitted *Am J Human Genet* 2009.
2. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**: 25-29.
3. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM). *Nucleic Acids Res* 2009; **37**: D793-D796.