LINKING DISEASE- ASSOCIATED GENETIC VARIANTS TO GENES USING SEQUENCE AND 3D CONTACTS

Chinaza E. Nnawulezi[1, *], Arjun Krishnan[2,], Jianrong Wang[2,]

[1] Department of Biological Science, Hampton University, Hampton, VA
[2] Department of CSME, Michigan State University, East Lansing, MI
*Student Email: Chinaza97@gmail.com

Introduction

- Currently, it is known that humans share over 90% of genes
- How humans can be genetically virtually replicates, yet still be vessels to a variety of diseases and other phenotypes?
- Single nucleotide polymorphisms(SNPs) are the genetic variants that are the leading factor in the concerning question
- Genome Wide Association Studies(GWAS) have been shown to be extremely useful in associating SNPS with different phenotypes
- It is now possible to associate these SNPs to genes and ultimately biological pathways
- This study focuses on relating SNPs of many different phenotypes to genes and biological pathways that are in both close and long ranges from a specified SNP

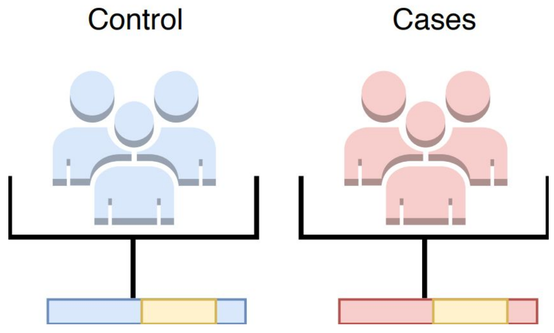


Figure 1. Diagram of GWAS methodology

Short Range Linkage Method

- GWAS studies from the UK Biobank were analyzed with PASCAL
- These gene scores are ranked based respective to their p-value
- Pathway enrichment analysis is done using the Gene Ontology to relate the correlations and bring biological significance.

Long Range Linkage Method

- By utilizing Chia-PET technology, chromatin interactions across a genome can be analyzed
- Upon analyzing these interactions, pathway enrichment can be performed to find the associated biological pathways
- Genes and SNPs from the short range analysis can be implemented here to find greater gene linkages.

Method Comparison

The short range linkage method, finds genes in the linear structure of DNA, however, it is known that DNA has a complex and folded structure. It should be stated that variants in DNA that are in LD are very likely to be linked to one another. Solely looking at one the three-dimensional structure of DNA, however, might not detect short range, but influential genes.

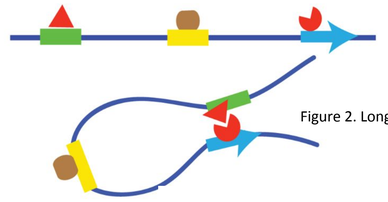


Figure 2. Long range genome interactions

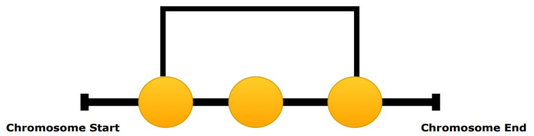
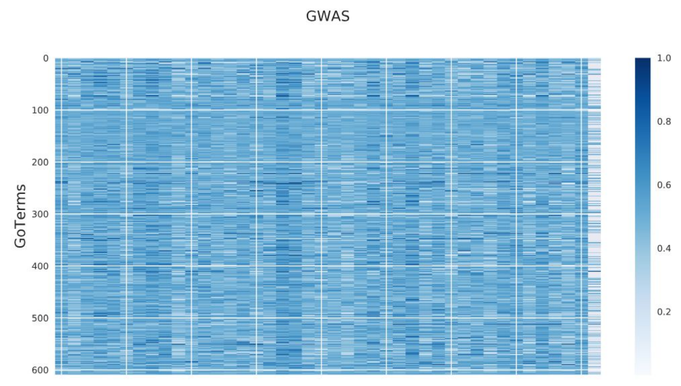


Figure 3. Short range genome interactions

Short Range Linkage Results



Graph 1. Heat map of GO Terms and GWAS studies

Long Range Linkage Results

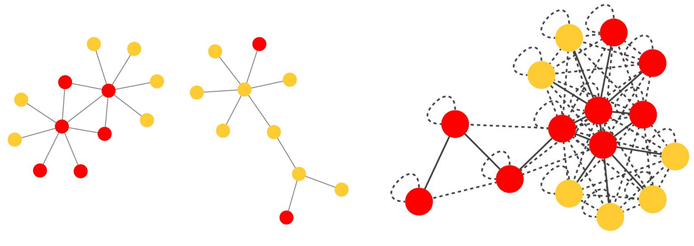
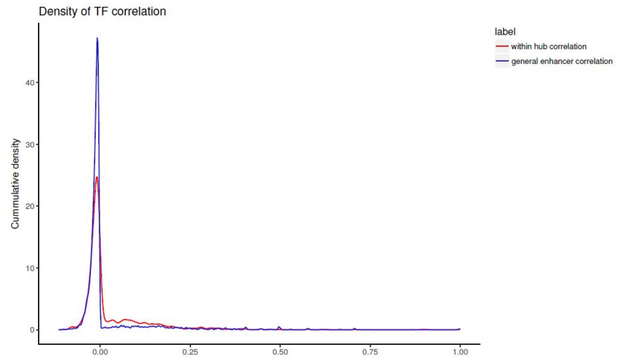


Figure 4. (left) Long range promoter and enhancer interactions
(right) Second order promoter and enhancer interactions



Graph 2. TF Density Correlation

Conclusions & Future Directions

- Using both long and short range interactions, a very detailed and informative gene network can be produced
- Gene scores as well as SNPs that are found in linkage disequilibrium can be incorporated into the larger model to note gene and pathway interactions that were once unknown
- With this information, developments can be made to better treatment technology, medications, as well as understanding regulatory processes of many diseases and other phenotype
- Future directions include performing noise reduction and clustering analysis and incorporation of more GWAS study outputs into the long range network

Acknowledgments and References

We acknowledge support from the MSU ACRES and BEACON programs, which is supported by the National Science Foundation through grant ACI -1560168.

We acknowledge support from Binbin Huang and Hao Wang of the Wang Lab, as well as support from Kayla Johnson, Chris Manusco, and Anna Yannakopoulos of the Krishnan Lab.

1. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z., & Bergmann, S. (2016). Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PLoS Computational Biology, 12(1), e1004714. http://doi.org/10.1371/journal.pcbi.1004714
2. Matharu, N., & Ahituv, N. (2015). Minor Loops in Major Folds: Enhancer–Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genetics*, *11*(12), e1005640. http://doi.org/10.1371/journal.pgen.1005640