Leveraging Human Phenotype Ontology (HPO) to identify phenotypes associated with chromatin modifier diseases

MATTHEW MOLERES[1], Leroy Bondhus[2], Valorie Arboleda[2,3,4]

1 BIG Summer Program, Institute for Quantitative and Computational Biosciences, UCLA

2 Department of Human Genetics David Geffen School of Medicine, UCLA

3 Department of Computational Medicine, David Geffen School of Medicine, UCLA

4 Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, UCLA, Los Angeles

Background and Motivation

There are rare diseases caused by mutations in chromatin modifiers. Chromatin modifier diseases occur when the DNA of a gene encodes for a mutated chromatin modifier protein, causing unfavorable phenotypes related to a disease. We then want to unify these chromatin modifier diseases to specific phenotypes so that an understanding of the phenotypic aberrations can be understood and related across all syndromes. This can be done by using the information in the Human Phenotype Ontology (HPO) and the Online Mendelian Inheritance of Man (OMIM).

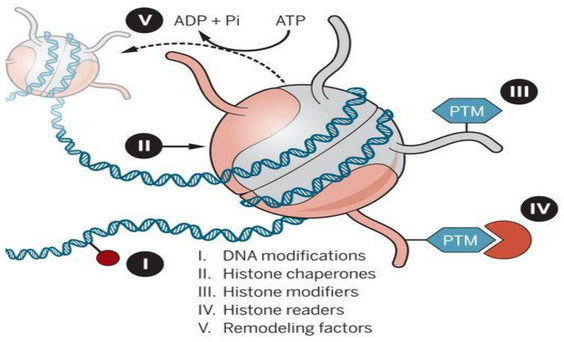

Figure 1: (*Almouzni, 2018)* , Depiction of Chromatin Modifier and all the processes that it can conduct to change the phenotype of interest.

HPO Structure

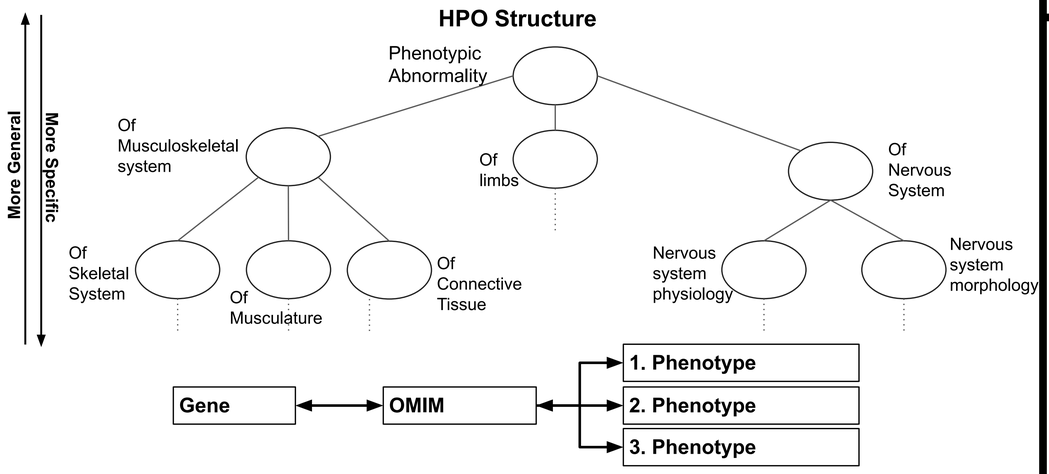

Figure 2: Overview depiction of Human Phenotype Ontology (HPO) Tree structure. Descending from broad to more specific. Online Mendelian Inheritance of Man (OMIM) overview structure relating genes to the diseases and phenotypes within OMIM.

Approach

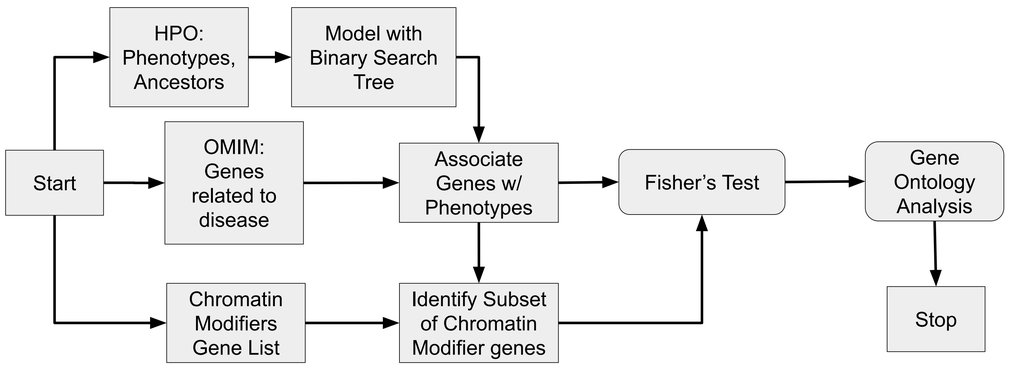

Figure 3: With the Human Phenotype Ontology (HPO) and Online Mendelian Inheritance of Man (OMIM), we used to comprise a binary search tree. In this search tree, each node was comprised of an identification number, the name of the phenotype, the ancestral number to the node, genes associated with the phenotype, the chromatin modifier genes present in the genes of the phenotype, if any, and lastly a Fisher's Test of that node. From the test, it could be seen which phenotypes were highly enriched and significant with chromatin modifier diseases and see which phenotypes were most significantly associated with chromatin modifier diseases. Then a Gene Ontology (GO) analysis was done for any phenotype in the tree to see how its genes relate to biological processes.

Fisher's Test

Genes associated

Chromatin Modifier		True	False
	True	**A** chromatin modifiers associated	**B** chromatin modifiers not associated
	False	**C** non chromatin modifiers associated	**D** Non chromatin modifiers not associated

$$OR = \frac{(A*D)}{(B*C)}$$

Results

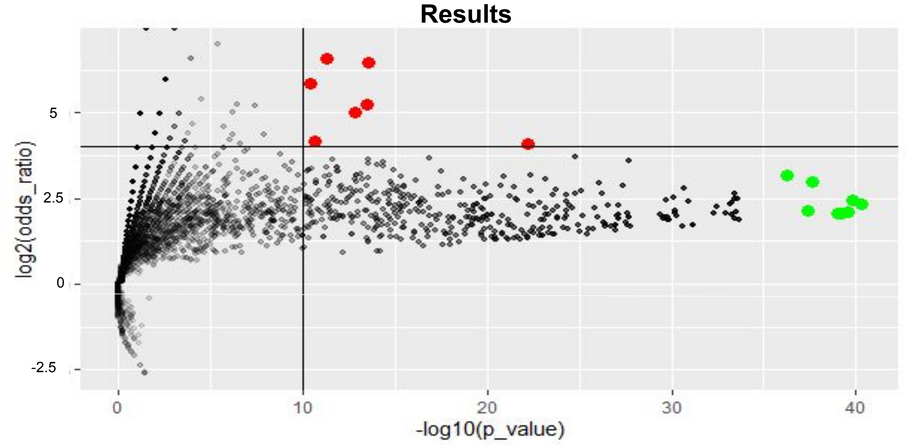

Figure 4: Scatter Plot showing how enriched and significant a phenotype is. Red dots indicate phenotypes that are highly enriched and still significant. While green dots depict phenotypes that are still enriched but the most significant. Dots above 0 signify enrichment, at 0 signify 1 to 1 ratio, and below 0 show depletion.

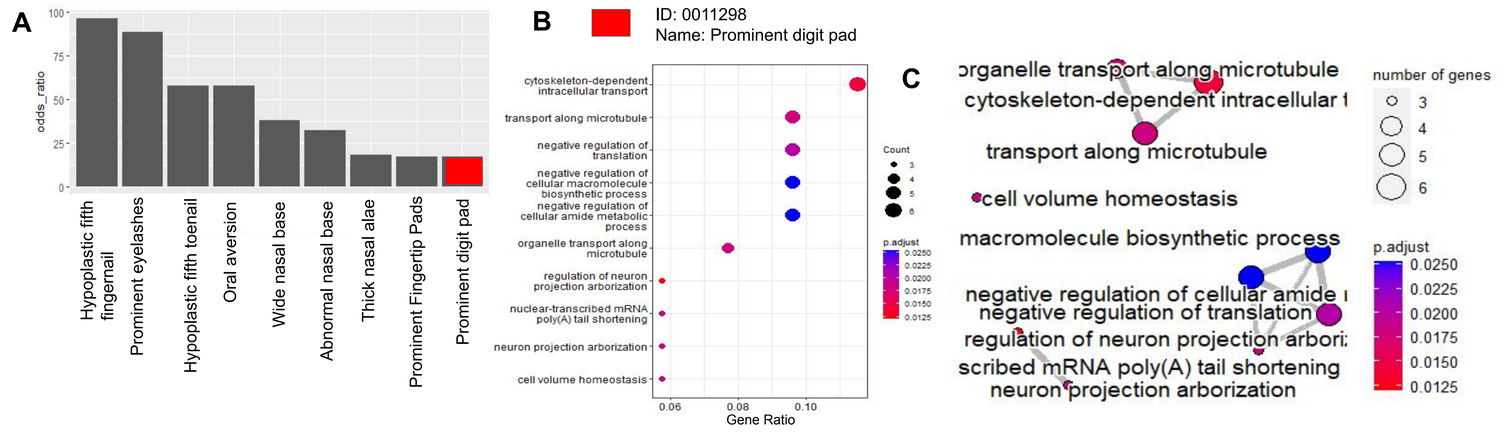

Figure 5: A) Shows which phenotypes are the most highly enriched and still significant. B) Depiction of phenotype and its gene relation to various biological processes, along with various gene ratios. C) Shows connectedness of genes to processes. Different clusters symbolizes how unsimilar genes are to the process.

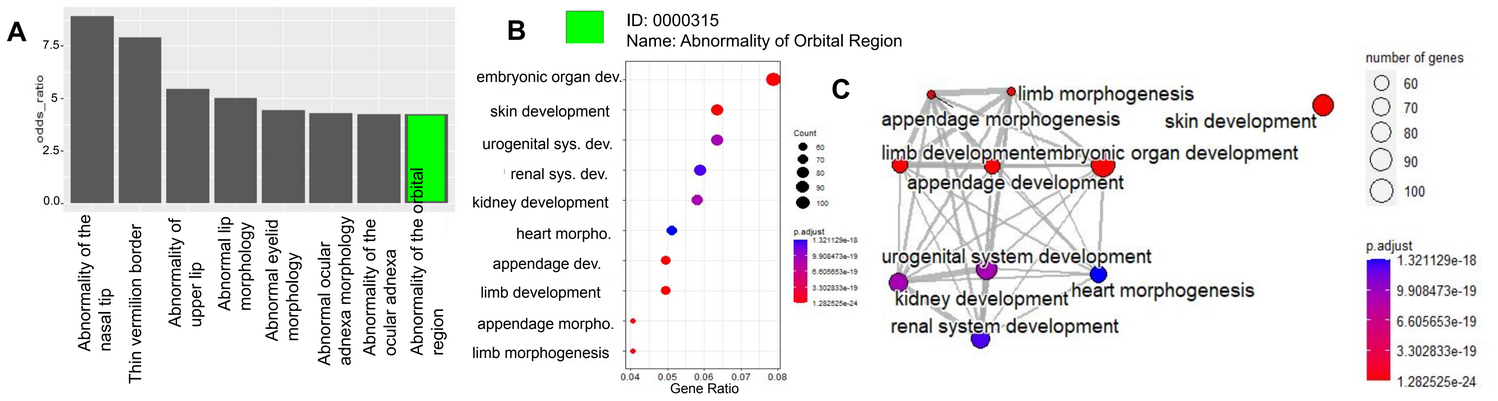

Figure 6: A) Shows which phenotypes are the most significant from Figure 3. Y-axis is the odds ratio, x-axis different phenotypes. B) Depiction of phenotype and its gene relation to various biological processes C) association of genes between processes, thickness of lines indicates number of genes, connectedness = similarity.

Conclusion

We have associated the chromatin modifier diseases with specific phenotypes that are enriched in chromatin modifier diseases. From this we can tell which biological processes are related to the genes of the phenotype we want to look at. We then can observe how closely the genes of each phenotype is related to one another and their given gene ratio. Allowing us to get the enrichment of various phenotypes within chromatin modifier diseases

Future Directions

- For future research These chromatin modifier diseases and phenotypes can be looked at over a developmental timeline to see where the genes in the phenotypes have come from and their role in biological processes over time.
- Create a hierarchical analysis for each of the phenotypes to bridge similarities between related phenotypes

Acknowledgements

I would like to acknowledge all the members of the Arboleda Lab for direction and useful feedback and the B.I.G summer program! **Alexander Hoffmann**

Dr. Valerie Arboleda (PI), Leroy Bondhus (Mentor), Meghna Singh, Maria Palafox, Isabella Lin, Aileen Nava, Angela Wei, Sarah Spendlove, Samantha Jensen, Maneesha Thaker, Roshni Varma, Yenifer Hernandez,Samantha Chung, Michelle Nguyen**. Supported by the B.I.G Summer Research Program**

Yadav, T., Quivy, J. P., & Almouzni, G. (2018). Chromatin plasticity: A versatile landscape that underlies cell fate and identity. *Science*, *361*(6409), 1332-1336.