Mechanisms Underlying Risk SNPs for Crohn's Disease, Type I Diabetes, and Type II Diabetes

R. Bhattacharya, H. Huang, A. Sabik
Johns Hopkins University – Biomedical Engineering

Introduction

Analysis of large sets of genomes can provide information regarding variation within the human genome, which can help identify mutations that may lead to certain diseases. One such example, Single Nucleotide Polymorphisms (SNPs), are DNA sequence variations occurring commonly within a population (e.g. 1%) in which a single nucleotide – A, T, C, or G – in the genome differs between members of a biological species. Risk SNPs are SNPs which have been statistically correlated with particular diseases such as Chron's disease, Type 1 Diabetes, and Type 2 Diabetes.

Chron's disease is a type of inflammatory bowel disease that affects any part of the gastrointestinal track with symptoms including abdominal pain, diarrhea, fever, and weight loss. Type 1 Diabetes causes the immune system to destroy Beta cells, which are the producers of Insulin. Type 2 Diabetes causes the body to be unable to absorb, and eventually produce, Insulin.

Purpose

As we attempt to convert raw genomic sequence data into biological knowledge, we seek to identify the mechanisms by which risk SNPs affect disease. Through the identification of these mechanisms, we move closer towards a more personalized diagnosis and treatment of disease. Our group aims to develop an automated classification tool that is able to predict the most likely mechanism by which a SNP can cause: Type I Diabetes, Type II Diabetes, and Crohn's Disease.

Methods

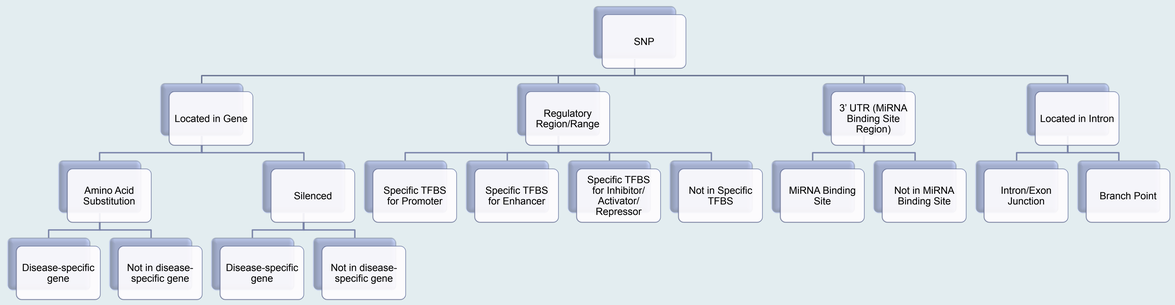
1. Reference GWAS SNPs data sets associated with the following diseases:
   - Chron's Disease
   - Type I Diabetes
   - Type II Diabetes
2. Research the following mechanisms for identifiable metrics:
   - Alteration in protein activity and stability
   - Change in gene expression level
   - Altered splicing
   - Altered microRNA binding
3. Determine genetic markers associated with listed diseases
4. Study current methods for mechanism prediction
5. Parse UCSC Table Browser to obtain intersections between SNPs and relevant marker tracks
6. Build Python parsers to cross-reference intersection data with GWAS SNPs data sets
7. Build decision tree classifier based on research of marker prioritization
8. Use classifier to identify associated mechanisms for given data set SNPs

Results

Below is a matrix of percentages that are the ratios of SNPs with a particular classified mechanism to all SNPs with any classified mechanism.

	Chron's Disease	Type 1 Diabetes	Type 2 Diabetes
Protein Activity	.285	.230	.305
Gene Expression	.595	.650	.536
Altered Splicing	.119	.120	.159
miRNA Binding	.001	0	0

Decision Tree



Discussion

Using previous GWAS studies in conjunction with research on the forces that affect phenotypes, it is possible to implicate the mechanism by which SNPs influence the expression of certain diseases. We got very few hits for miRNA binding, which was partially expected because of the relatively small number of miRNA binding sites. On the other hand, gene expression appears to be the dominant mechanism for all three diseases, which may be due to the fact that there are more ways to affect regulation than the transcription of a disease-specific protein. Additionally, there may be error within the bias of assigning a reward value of "1" per depth of the classification.

Future Direction

1. Use linkage disequilibrium data to identify other SNPs potentially correlated with disease and classify mechanism
2. Identify if SNPs in the protein activity mechanism category cause animo acid substitutions utilizing databases such as DbSNP and Sift
3. Utilize data from HAPMAP SQL databases
4. Extend methods to other diseases
5. Experiment with reward assignment
6. Cross reference results with existing tools

References

[1] "Chron's Disease". National Digestive Diseases Information Clearinghouse (NDDIC). July 10, 2013. Retrieved 12 June 2014.
[2] Barreiro LB; Laval G; Quach H; Patin E; Quintana-Murci L. (2008). "Natural selection has driven population differentiation in modern humans". *Nature Genetics* **40** (3): 340–345.
[3] Freedman, Matthew L., et al. "Principles for the post-GWAS functional characterization of cancer risk loci." Nature genetics 43.6 (2011): 513-518.
[4] https://genome.ucsc.edu/
[5] Kumar, Vinay; Fausto, Nelson; Abbas, Abul K.; Cotran, Ramzi S. ; Robbins, Stanley L. (2005). Robbins and Cotran Pathologic Basis of Disease (7th ed.). Philadelphia, Pa.: Saunders. pp. 1194–1195.
[6] Dean L, McEntyre J. The Genetic Landscape of Diabetes [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2004. Chapter 3, Genetic Factors in Type 2 Diabetes. 2004 Jul 7. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1665/