Weighted Polygenic Risk Scores

Project with Dr. Lai Jiang, Dr. Celia Greenwood

Xavier Loffree

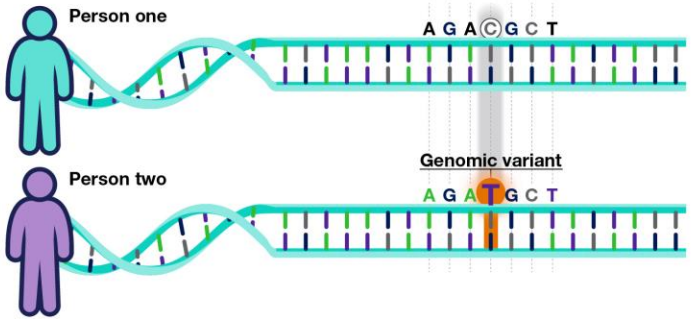
Decoding the genetic predisposition to type 2 diabetes. Medical News. (2017, March 15). https://www.medpagetoday.com/resource-centers/glucose-control/genetic-predisposition/58.

Goal

- Improve performance of polygenic risk scores (PRS) by determining a weight for each single nucleotide polymorphism (SNP) and then using those weights when calculating polygenic risk scores

Goodman, A. (2020, June 29). *Race is real, but it's not genetic*. Discover Magazine. https://www.discovermagazine.com/planet-earth/race-is-real-but-its-not-genetic.

Polygenic Diseases



- Complex causal genetic pathways
- Caused by many genes, not just one
- Polygenic diseases include type-2 diabetes, hypertension, coronary heart disease
- Monogenic diseases include sickle cell anemia, cystic fibrosis, and Huntington disease

Polygenic risk scores. Genome.gov. (n.d.). https://www.genome.gov/Health/Genomics-and-Medicine/Polygenic-risk-scores.

Polygenic Risk Scores (PRS)

- Estimate risk of an individual developing an heritable disease
- GWAS use regression to determine association of each single nucleotide polyomorphism (SNP) with particular disease
- PRS sum the strength of the association at a set of associated SNPs:

$$PRS_i = \sum_j^n \beta_j * quantity_{ij}$$

Effector Index (EI) Project

- Predict which is the most likely causal gene in a region showing GWAS associations at many SNPs and across several genes.
- Use various genomic features and annotations in a (boosted tree) model predicting gene most likely to be causal

Forgetta, V., Jiang, L., Vulpescu, N. A., Hogan, M. S., Chen, S., Morris, J. A., Grinek, S., Benner, C., Jang, D.-K., Hoang, Q., Burtt, N., Flannick, J. A., McCarthy, M. I., Fauman, E., Greenwood, C. M. T., Maurano, M. T., & Richards, J. B. (2020). An effector index to Predict CAUSAL genes AT GWAS Loci. https://doi.org/10.1101/2020.06.28.171561

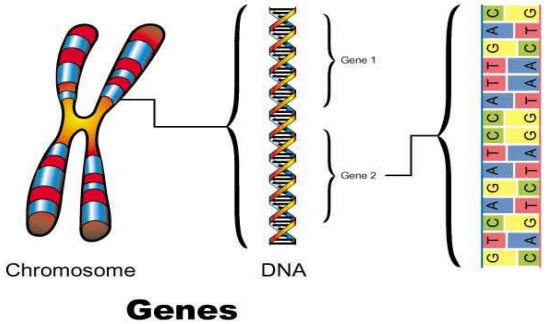
Computational biology. Lifeboat Blog RSS. (2020, January 21). https://lifeboat.com/blog/2020/01/computational-biology.

Our Project: Weighted PRS

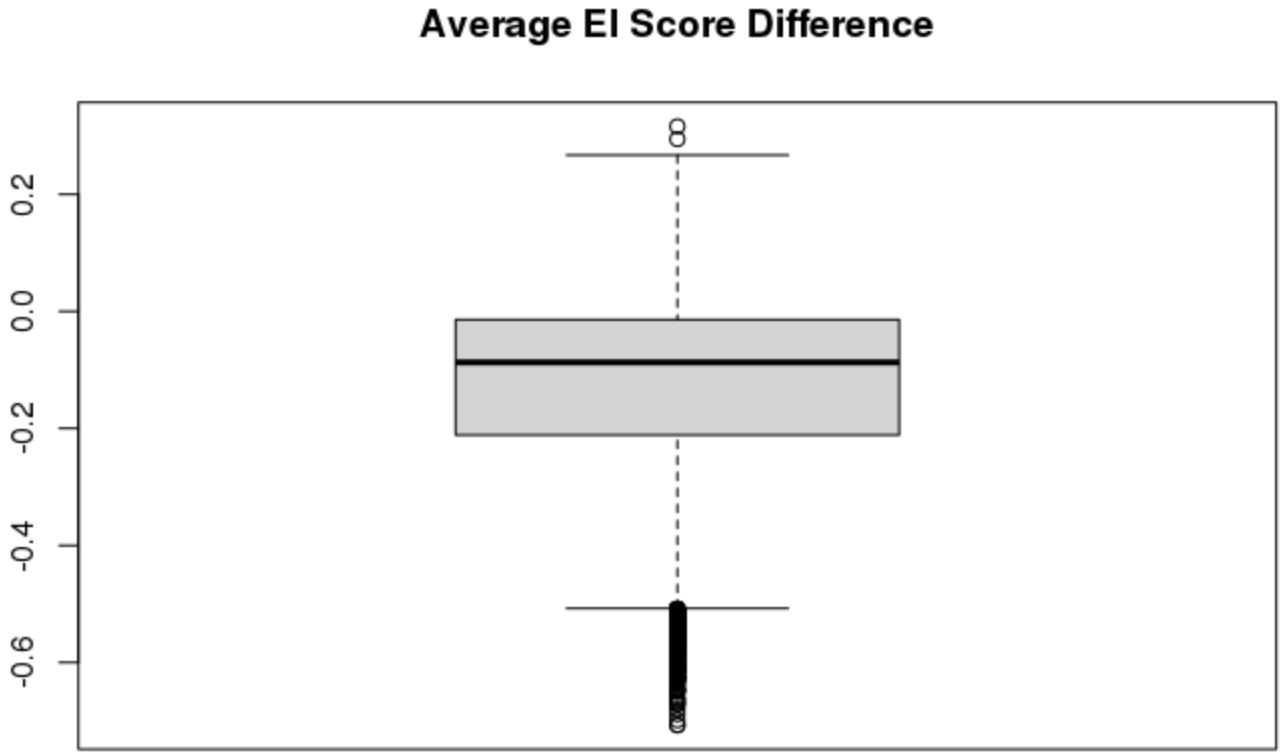
- Incorporate the causality predictions from the EI project
- Use these results to determine weights for SNPs
- SNPs with high predictions for being important in the EI algorithm are given greater weights, less important are given lower weights
- Once the weights are incorporated, this should result in a more accurate PRS

Converting Gene-Specific EI Scores to SNP-Specific EI Score

- Problem: EI predicts causal genes, but we are looking for causal SNPs
- Employed a leave-one-out algorithm to arrive at SNP-specific values from the EI algorithm
- Ran the algorithm with each SNP removed, then compared the EI values when the SNP was included in the algorithm and when it was not
- Efficiency was a challenge at first

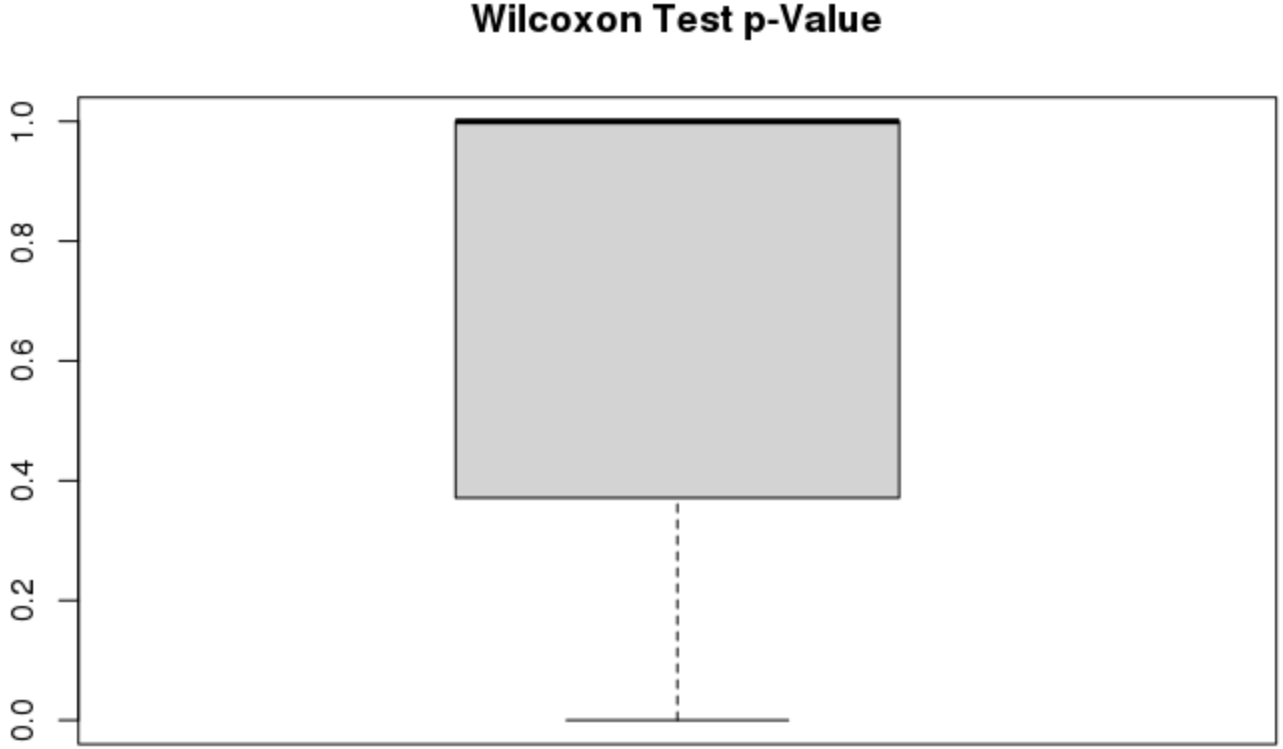


What are genes made of? MelixGX. (n.d.). https://www.melixgx.com/blogs/news/what-are-genes-made-of.


Average EI Score Difference
0.2
0.0
-0.2
-0.4
-0.6
8

Analyzing the Change

- Converted EI scores into rankings within the locus for each SNP-gene combination
- Used a Wilcoxon test to compare the EI score rankings of each SNP-gene combination within a locus before and after the new EI scores were calculated
- Wilcoxon tests give us p-values that can be used to check if the ranking change is significant


Wilcoxon Test p-Value
0.0
0.2
0.4
0.6
0.8
1.0

Next Steps

- Conduct EI-weighted PRS score for each of 12 traits associated with T2D used in the EI project
- Use UKBB data and SNPs with p-value less than 0.05 from Wilcoxon test

Decoding the genetic predisposition to type 2 diabetes. Medical News. (2017, March 15). https://www.medpagetoday.com/resource-centers/glucose-control/genetic-predisposition/58.

Thank You!

- Special thank you to Lai Jiang, Irene Zhang, Celia Greenwood, Tianyuan Lu, Kathleen Klein, Yixiao Zeng, Nicolas de Jay, Jeffrey Yu