Fine-Mapping Genetic Loci showing Novel Associations with Incident Cardiovascular Disease in Type 2 Diabetes

Ryan Hernandez-Cancela[1,2], Daniel DiCorpo[3], Ching-Ti Liu[3], James Meigs[4,5,6], Josée Dupuis[3,7], Soo Heon Kwak[8]

[1]Clemson University, [2]Boston University Bioinformatics BRITE REU Program, Summer 2021, [3]Department of Biostatistics, Boston University School of Public Health, [4]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, [5]Department of Medicine, Harvard Medical School, [6]Division of General Internal Medicine, Massachusetts General Hospital, [7]Bioinformatics Graduate Program, Boston University, [8]Department of Internal Medicine, Seoul National University Hospital, South Korea

Abstract

- Individuals with type 2 diabetes (T2D) are at a greater risk of developing cardiovascular disease (CVD)
- A genome-wide association study (GWAS) of individuals of varying ancestries diagnosed with T2D was conducted, which is uncommon in the literature
  - we identified **3 novel regions** of the genome containing variants that reached **genome-wide significance** ($P < 5\times10^{-8}$): **rs147138607, rs77142250, and rs335407.**
- The goal of this project was to use statistical fine-mapping to **identify the most likely causal genetic variants of incident CVD in patients with T2D in each novel region.**
- Ancestry-specific Z-scores and ancestry-specific pairwise linkage-disequilibrium were used to run fine-mapping software MsCAVIAR.
- **For rs147138607 and rs335407** regions, the 95% **credible sets included 21 and 38 variants**, respectively.
  - These results will help prioritize likely causal variants for functional follow-up
- Future fine-mapping work will include incorporating functional annotations, as well as running ancestry-specific analyses

Background

Incident CVD

First CVD event (ex. stroke) that occurs at least a year after diagnosis of T2D

Statistical Fine-Mapping

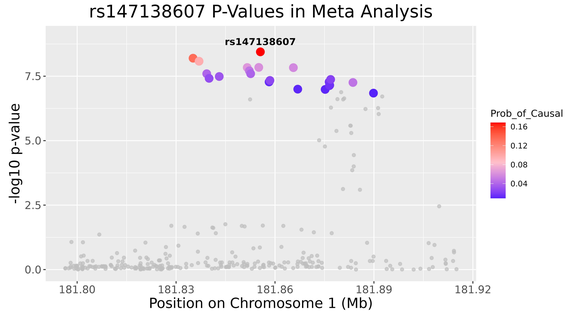
Use of summary statistics to assign probabilities of causality to variants

Pairwise Linkage Disequilibrium (LD)

Correlation between alleles of two variants

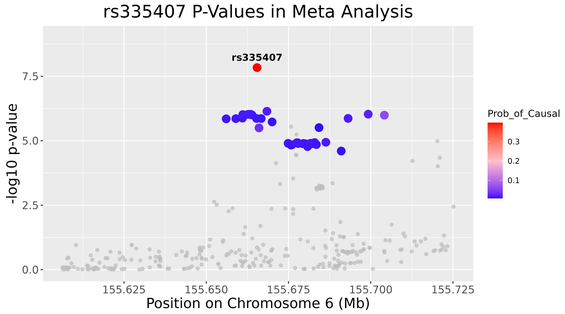
- We conducted a genome-wide association study (GWAS) of 48,138 individuals of varying ancestries (African, East Asian, European, Hispanic)
- This GWAS aimed to determine associations between single base pair genetic variants with incident CVD
- In this study, we identified 3 novel regions containing variants that reached genome-wide significance: rs147138607, rs77142250, and rs335407
  - If genome-wide significant ($P < 5\times10^{-8}$), we can say variant is associated with trait
- Statistical fine-mapping is one option for further analysis of GWAS results
  - MsCAVIAR utilizes pairwise linkage disequilibrium (LD) and z-scores from original GWAS

Results



Variant ID	*P*-value	Z-Score	Causal Probability
1:181855562:C:G (rs147138607)	3.6×10^{-9}	5.9	0.17
1:181835150:T:C	6.3×10^{-9}	5.8	0.13
1:181836968:D:I	8.3×10^{-9}	-5.8	0.10
1:181851578:T:C	1.5×10^{-8}	5.7	0.06
1:181855077:T:C	1.4×10^{-8}	5.7	0.06

Figure 1: Regional Plot of rs147138607 Region after Fine-Mapping. 294 variants were included, and colored points represent variants in the 95% credible set. Coloring based on probability of causality. Table above represents the top 5 variants in terms of causal probability.



Variant ID	*P*-value	Z-Score	Causal Probability
6:155665441:T:C (rs335407)	1.5×10^{-8}	5.7	0.40
6:155704129:A:G	1.0×10^{-6}	-4.9	0.06
6:155666036:A:G	3.2×10^{-6}	-4.7	0.04
6:155699223:A:C	9.3×10^{-7}	-4.9	0.02
6:155668440:A:G	7.2×10^{-7}	-5.0	0.02

Figure 2: Regional Plot of rs335407 Region after Fine-Mapping. 319 variants were included, and colored points represent variants in the 95% credible set. Coloring based on probability of causality. Table above represents the top 5 variants in terms of causal probability.

Conclusion

- **rs147138607** and **rs335407** had **highest probabilities of causality** in their respective regions, so they're the most-likely causal variants

Most-likely causal in Chromosome 1 Region: **rs147138607**

Probability of Causality: **17%**

Most-likely causal in Chromosome 6 Region: **rs335407**

Probability of Causality: **40%**

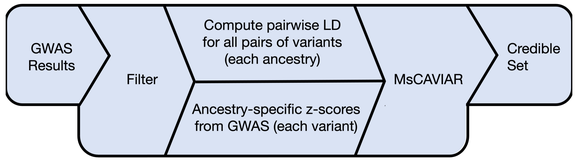
- Region containing **rs147138607** – more *broad* distribution of probabilities
  - Related to *higher* LD between variants
- Region containing **rs335407** – more *focused* distribution of probabilities
  - Related to *lower* LD between variants

- Both regions had **relatively large 95% credible sets**
  - Region containing **rs147138607** had **21 variants**
  - Region containing **rs335407** had **38 variants**
  - Smaller credible sets of 2-3 variants are favorable due to high cost of laboratory follow-up

- Large **credible set size** likely **due to limitations of analysis**, including
  - Exclusion of variants not present in all ancestries (in GWAS results)
  - Exclusion of variants with more than 2 alleles
  - Relatively small sample size

- **Future work** will attempt to address limitations by
  - Conducting **ancestry-specific analyses** – less variants excluded
  - Incorporating **functional annotations** – further reduce credible set

Methods



Filter: variant must be within 60 kilobases from lead (most associated) variant and must have Minor Allele Count (MAC) > 40 in ALL ancestries

References

1. Rawshani A, Rawshani A, Franzén S, Sattar N, Eliasson B, Svensson AM, Zethelius B, Miftaraj M, McGuire DK, Rosengren A, Gudbjörnsdottir S (2018). "Risk Factors, Mortality, and Cardiovascular Outcomes in Patients with Type 2 Diabetes". N Engl J Med, 379, 633-644.
2. Gu JK, Charles LE, Fekedulegn D, Allison P, Ma CC, Violanti JM, Andrew ME (2021). "Temporal trends in prevalence of cardiovascular disease (CVD) and CVD risk factors among U.S. older workers: NHIS 2004–2018". Annals of Epidemiology, 55, 78-82.
3. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD, Duggan DJ, Spencer KL, Dumitrescu L, Eaton CB, Thomas F, Young A, Carty C, Heiss G, Le Marchand L, Crawford DC, Hindorff LA, Kooperberg CL, PAGE Consortium (2013). "Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study". PLoS Biol, 11(9).
4. The 1000 Genomes Project Consortium (2015). "A global reference for human genetic variation". Nature, 526, 68-74.
5. LaPierre et. al. (2020). "Identifying Causal Variants by Fine Mapping Across Multiple Studies". bioRxiv.

Acknowledgements

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program, and partially supported by NIH grants R01 HL151855 and UM1 DK078616.