Finding rare variants of large effect by identity-by-descent mapping

Suyash S. Shringarpure*, David Hinds, Vladimir Vacic, Steve Pitts, Robert Gentleman, Adam Auton, the 23andMe Research Team

23andMe, Inc., Mountain View, CA – research.23andme.com *sshringarpure@23andme.com

Introduction

GWAS studies are not well-powered to discover rare variant associations due to limited imputation accuracy for rare variants.

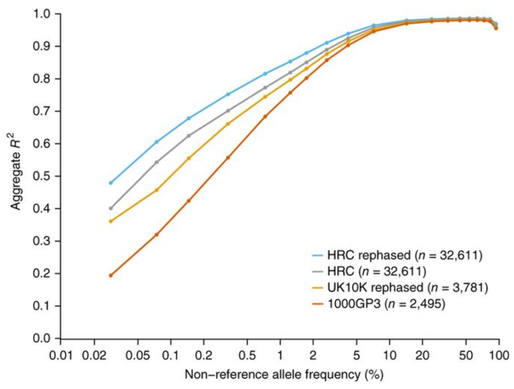


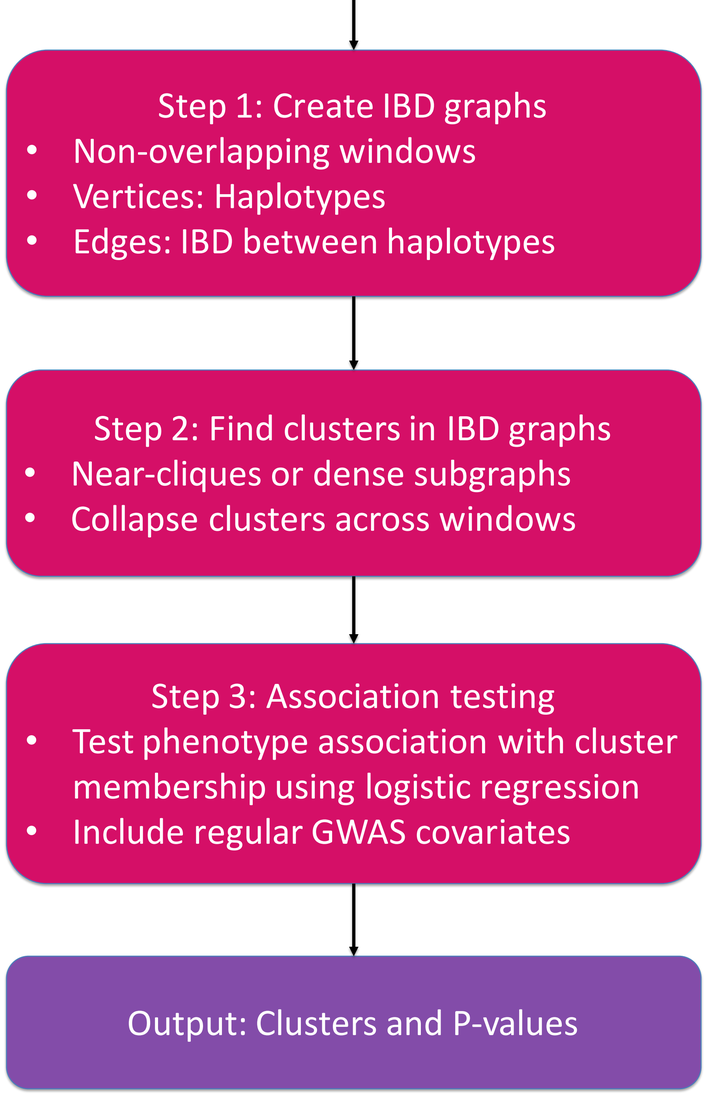
Figure 1. Imputation accuracy with state-of-the-art imputation panels. Imputation accuracy decreases with minor allele frequency and is low for rare variants.

We demonstrate the utility of identity-by-descent (IBD) mapping methods to discover rare associations using simulations and application on data from 23andMe research participants.

Through simulations, we show that IBD mapping can discover associations rarer than 0.5% in frequency, which cannot be detected using GWAS on imputed data.

We applied our method on consented research participants from the 23andMe cohort and analyzed 200 phenotypes. We find rare associations in regions with known common associations, as well as completely novel rare associations.

Methods



Results

IBD mapping simulations

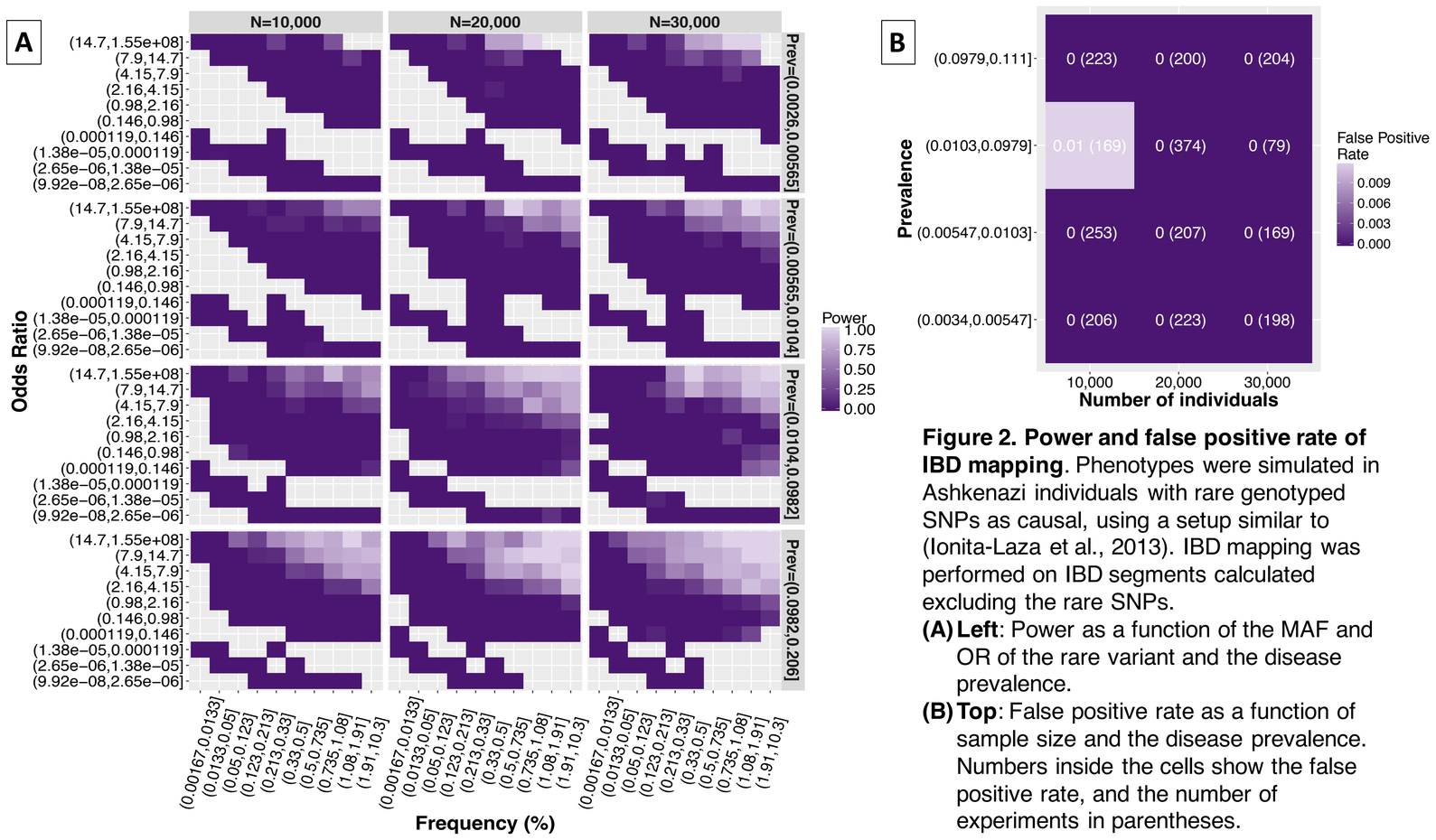


Figure 2. Power and false positive rate of IBD mapping. Phenotypes were simulated in Ashkenazi individuals with rare genotyped SNPs as causal, using a setup similar to (Ionita-Laza et al., 2013). IBD mapping was performed on IBD segments calculated excluding the rare SNPs.

(A) Left: Power as a function of the MAF and OR of the rare variant and the disease prevalence.

(B) Top: False positive rate as a function of sample size and the disease prevalence. Numbers inside the cells show the false positive rate, and the number of experiments in parentheses.

IBD mapping in the Ashkenazi population

Example 1: Morning person

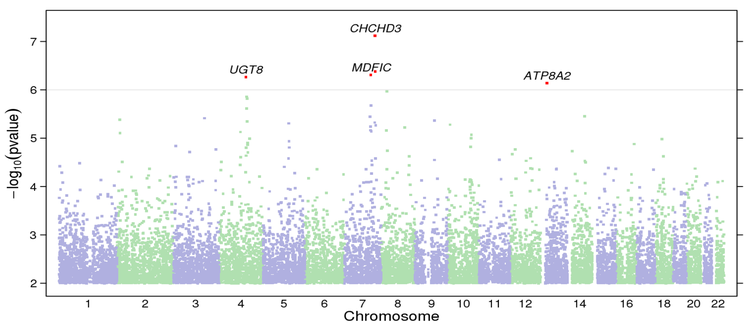


Figure 3. Morning person IBD mapping analysis. Manhattan plot for IBD association analysis with 11,625 cases and 11,059 controls in Ashkenazi individuals. Each point is an IBD cluster.

Chr	Position (Mb)	Cluster size	P-value	Cluster Freq	OR (95%CI)	Gene context
7	132.5	238	7.62E-08	0.01049	0.48 (0.36-0.63)	EXOC4
7	114.5	27	4.91E-07	0.00119	12.0 (3.14-46.1)	FOXP2
4	115.5	62	5.45E-07	0.00273	4.40 (2.28-8.48)	UGT8
13	26.5	57	7.26E-07	0.00251	4.57 (2.30-9.07)	ATP8A2

Table 1. Morning person IBD association hits. Common variants overlapping the first 3 hits are reported in the recent meta-analysis for chronotype (Jones et al., 2018).

Example 2: Gout

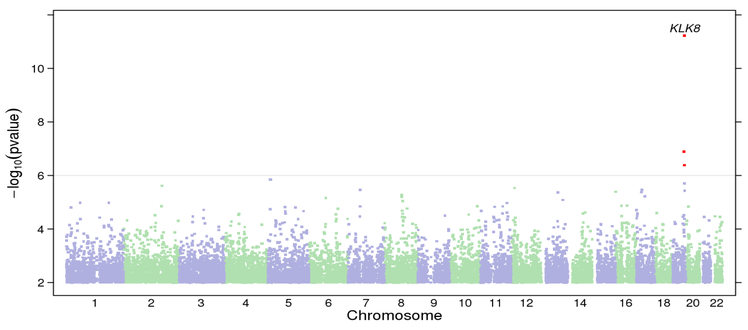


Figure 4. Gout IBD mapping analysis. Manhattan plot for IBD association analysis with 1,471 cases and 24,717 controls in Ashkenazi individuals. Each point is an IBD cluster.

Chr	Position (Mb)	Cluster size	P-value	Cluster Freq	OR (95% CI)	Gene context
19	51.5	145	5.97E-12	0.0055	5.73 (3.70-8.89)	ALDH16A1

Table 2. Gout IBD association hit. One sequenced individual in the cluster contains a low-frequency missense variant (c.1580C>G) in ALDH16A1, which has been previously reported as associated with gout in Icelanders (Sulem et al., 2011).

Conclusion

We show using simulations that IBD mapping has power to detect rare variant associations of large effect. These associations occur at frequencies where imputation does not work well. We show IBD mapping associations in the morningness phenotype and gout as examples of large-effect associations found in an Ashkenazi population.

Acknowledgments

We thank 23andMe customers who consented to participate in research for enabling this study. We also thank employees of 23andMe who contributed to the development of the infrastructure that made this research possible.

References

1. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., & Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. The American Journal of Human Genetics, 92(6), 841-853.
2. Sulem, P., Gudbjartsson, D. F., Walters, G. B., Helgadottir, H. T., Helgason, A., Gudjonsson, S. A., ... & Magnusson, G. (2011). Identification of low-frequency variants associated with gout and serum uric acid levels. Nature genetics, 43(11), 1127.
3. Jones, S. E., Lane, J. M., Wood, A. R., van Hees, V. T., Tyrrell, J., Beaumont, R. N., ... & Tuke, M. A. (2018). Genome-wide association analyses of chronotype in 697,828 individuals provides new insights into circadian rhythms in humans and links to disease. bioRxiv, 303941.