The role of structural variation in human local adaptation

Stephanie M. Yan[1], Rachel M. Sherman[2], Sara A. Carioscia[1], Michael C. Schatz[1,2], Rajiv C. McCoy[1]

[1]Department of Biology, [2]Department of Computer Science, Johns Hopkins University, Baltimore, USA

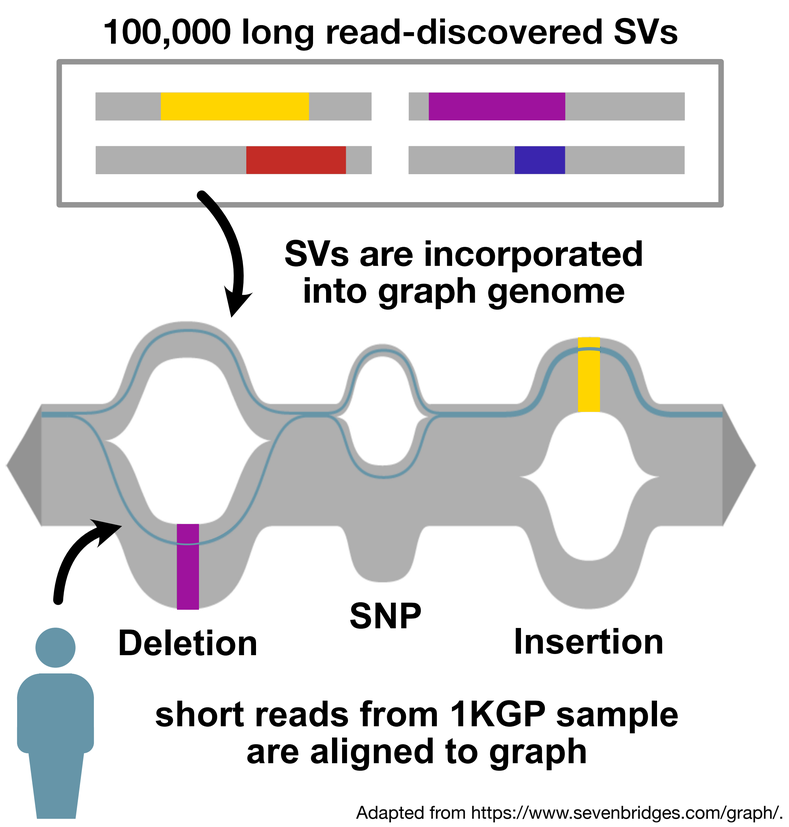
Contact: syan@jhu.edu

Background

- Structural variants (SVs)—large insertions, deletions, or inversions (>50bp)—may be prominent but hidden targets of positive selection.
- Long-read sequencing studies have expanded the catalog of SVs segregating in human populations[1], but have not yet been applied to study SVs on a large scale.

Population-wide SV genotyping

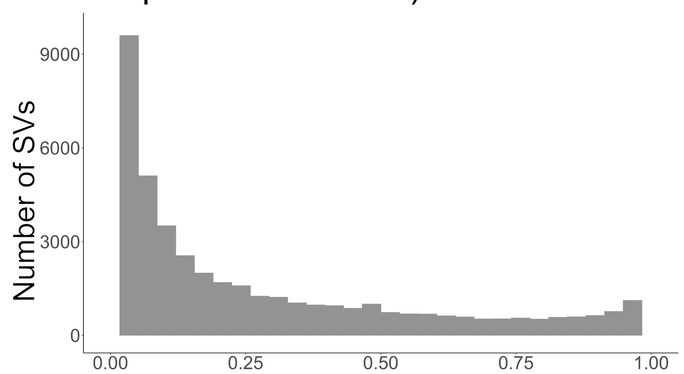
- We used Paragraph[2], a variant graph genotyper, to genotype long read-discovered SVs in short-read samples from three populations (CEU, CHB, YRI) in the 1000 Genomes Project.



Adapted from https://www.sevenbridges.com/graph/.

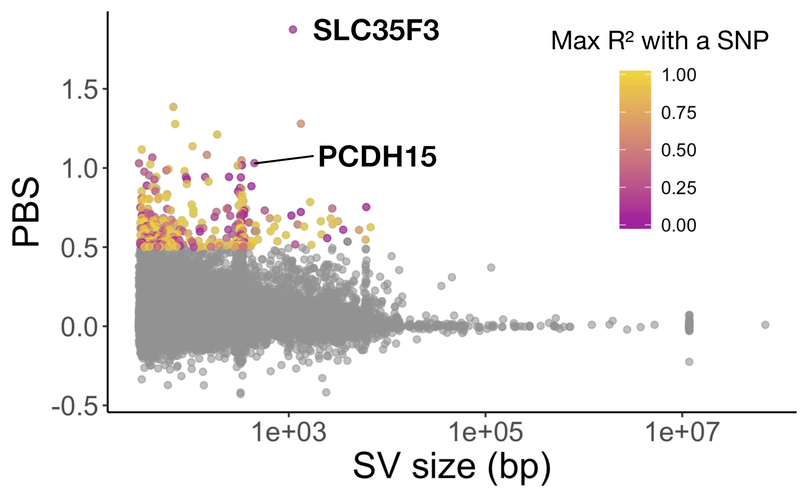
Most SVs are in weak LD with known SNPs

- We required all SVs to be genotyped in at least 98% of samples. We removed variants with excess heterozygote calls (Hardy-Weinberg equilibrium p-value < 0.0001).

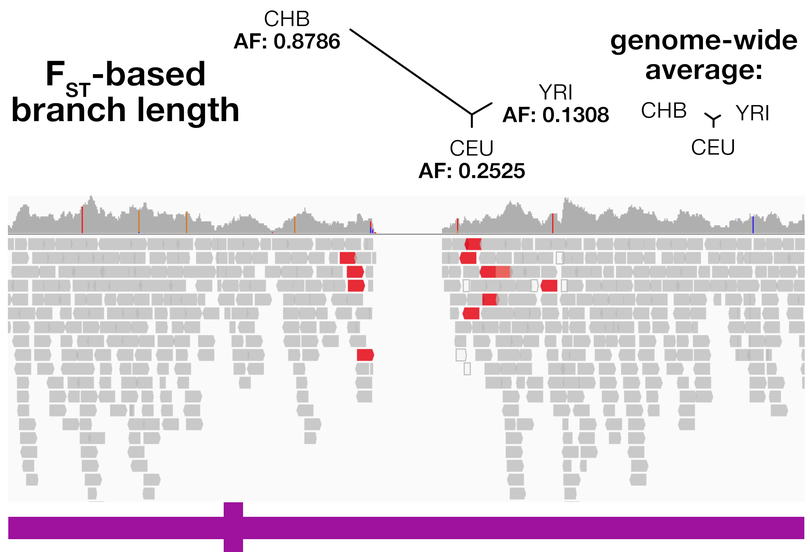


Identifying SVs with signatures of positive selection

- We used the population branch statistic (PBS) to identify 330 SVs with extreme allele frequency differentiation between populations (PBS > 0.5).



A 447bp deletion in PCDH15



- Codes for protocadherin-15, a protein involved in retinal and cochlear function.[3]

Conclusions

- We applied variant graph genotyping to genotype structural variants on a population-wide scale. Several SVs show signatures of positive selection.
- Many SVs are not strongly linked to known SNPs, indicating that they could constitute previously unknown adaptive loci.

Acknowledgments

Thank you to members of the McCoy lab for helpful feedback. This work is supported by NIH grant R35GM133747 to R.C.M.

References

1. Audano *et al.* (2019). Characterizing the Major Structural Variant Alleles of the Human Genome. *Cell* **176**, 663-675.e19.
2. Chen *et al.* (2019). Paragraph: a graph-based structural variant genotyper for short-read sequence data. *Genome Biol.* **20**, 291.
3. https://ghr.nlm.nih.gov/gene/PCDH15.