



Stevie Pederson^{1,2}, Yassine Souilmi^{1,3}, Hardip Patel², Alex Brown^{1,2}, Jimmy Breen^{1,2}

¹Black Ochre Data Labs, Telethon Kids Institute ²John Curtin School of Medical Research, Australian National University

³School of Biological Sciences, University of Adelaide

Genomic Variability

The Indigenous Australian population contains a large amount of unique genetic diversity. Given that risk factors are increasingly being shown to be polygenic and dependent on the genetic background, ignoring diversity in a large cohort may limit our ability to address this contribution.

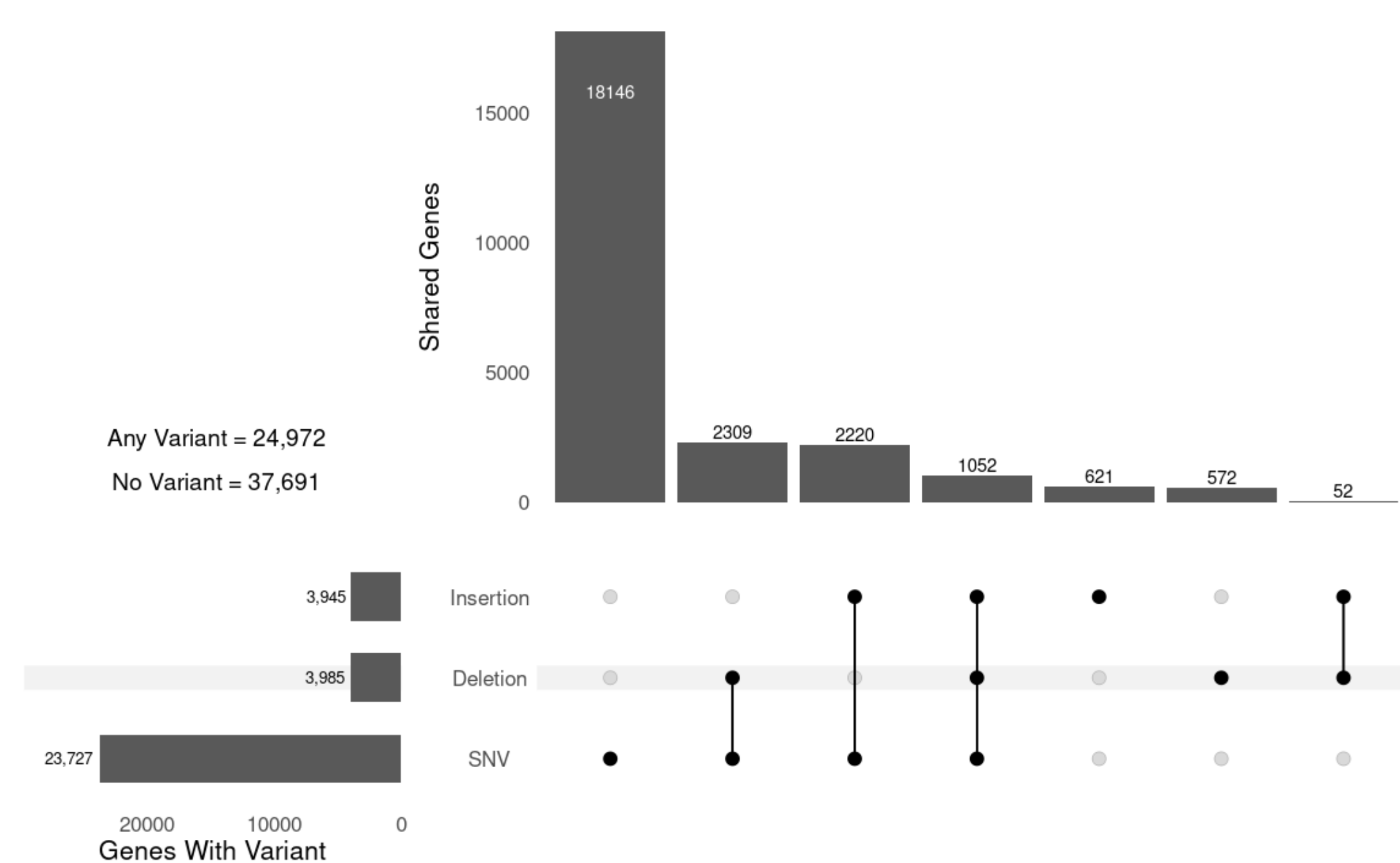


The historical relationship between researchers and the Aboriginal community has been problematic and as such, this population is poorly represented in databases such as the 1000 Genomes Project (1000GP), and the Human Genome Diversity Project (HGDP).

Variants from the 1000 Genomes Project

Taking consensus 1000GP variants (>50%) as a *proof-of-principle*, a gene-level analysis (*STARconsensus*¹) was prepared comparing the standard hg38 reference against a modified reference incorporating SNPs and InDels from the 1000GP.

Region	SNV	Insertion	Deletion	Total
No Classified Region	1,281,298	110,072	113,176	1,504,546
Promoter	151,537	13,775	14,616	179,928
Upstream Promoter	100,273	9,848	9,845	119,966
Exon	68,628	4,682	4,814	78,124
UTR	24,213	2,197	2,261	28,671
Splice Junction	1,628	236	265	2,129
Stop Codon	36	1	5	42
Start Codon	16	3	2	21



Genes with variants overlapping an exon

The PROPHECY Study

The PROPHECY study (**P**reventing **R**enal, **O**phthalmic and **H**eart **E**vents in **C**ommunit**Y**) consists of ~ 1400 indigenous participants drawn from regional, remote and urban locations within South Australia. Amongst community, the study is colloquially known as the *Aboriginal Diabetes Study*.

The PROPHECY Study is a multi-omics study including genomic variants, DNA methylation, bulk RNA-Seq, proteomics, lipidomics, metabolomics and multiple other layers, all derived from blood samples taken from the same participants.

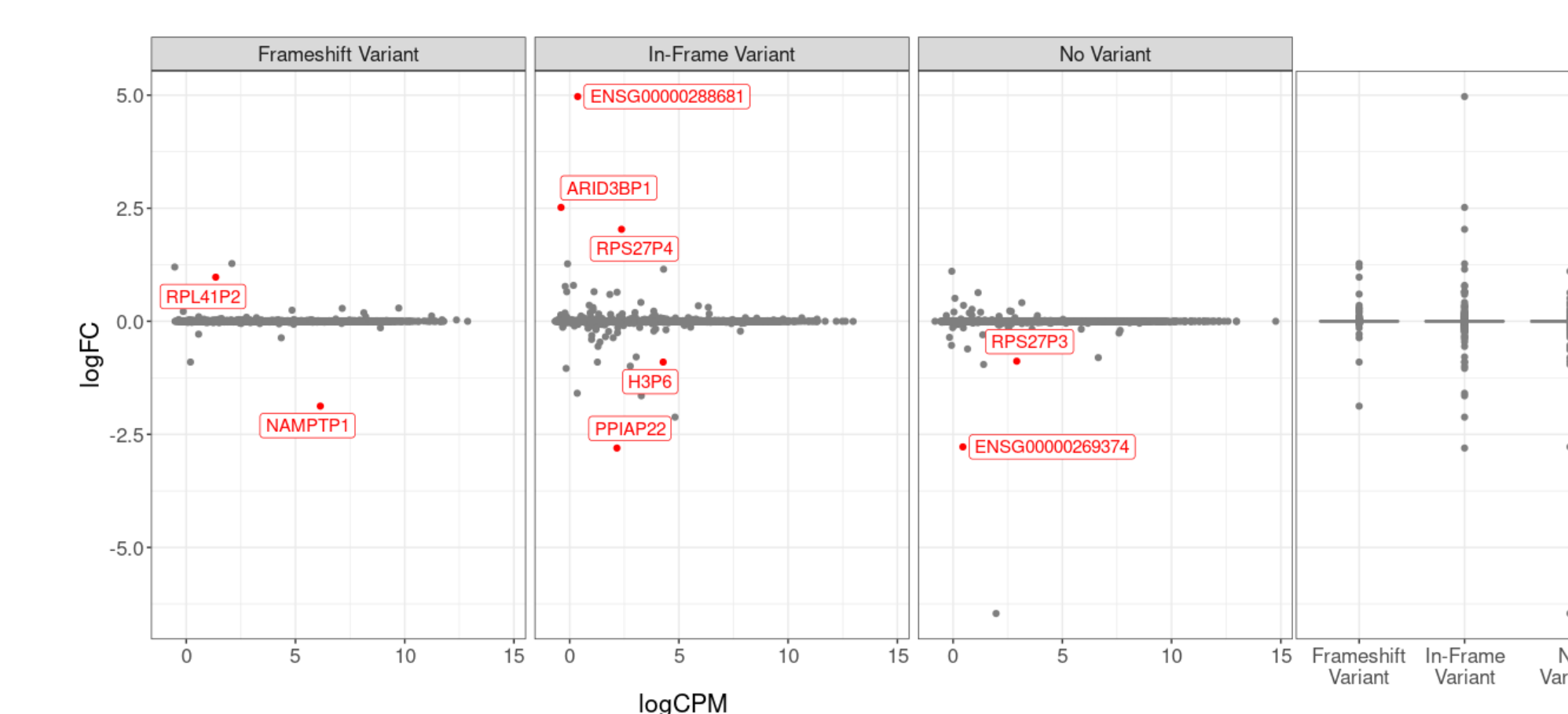
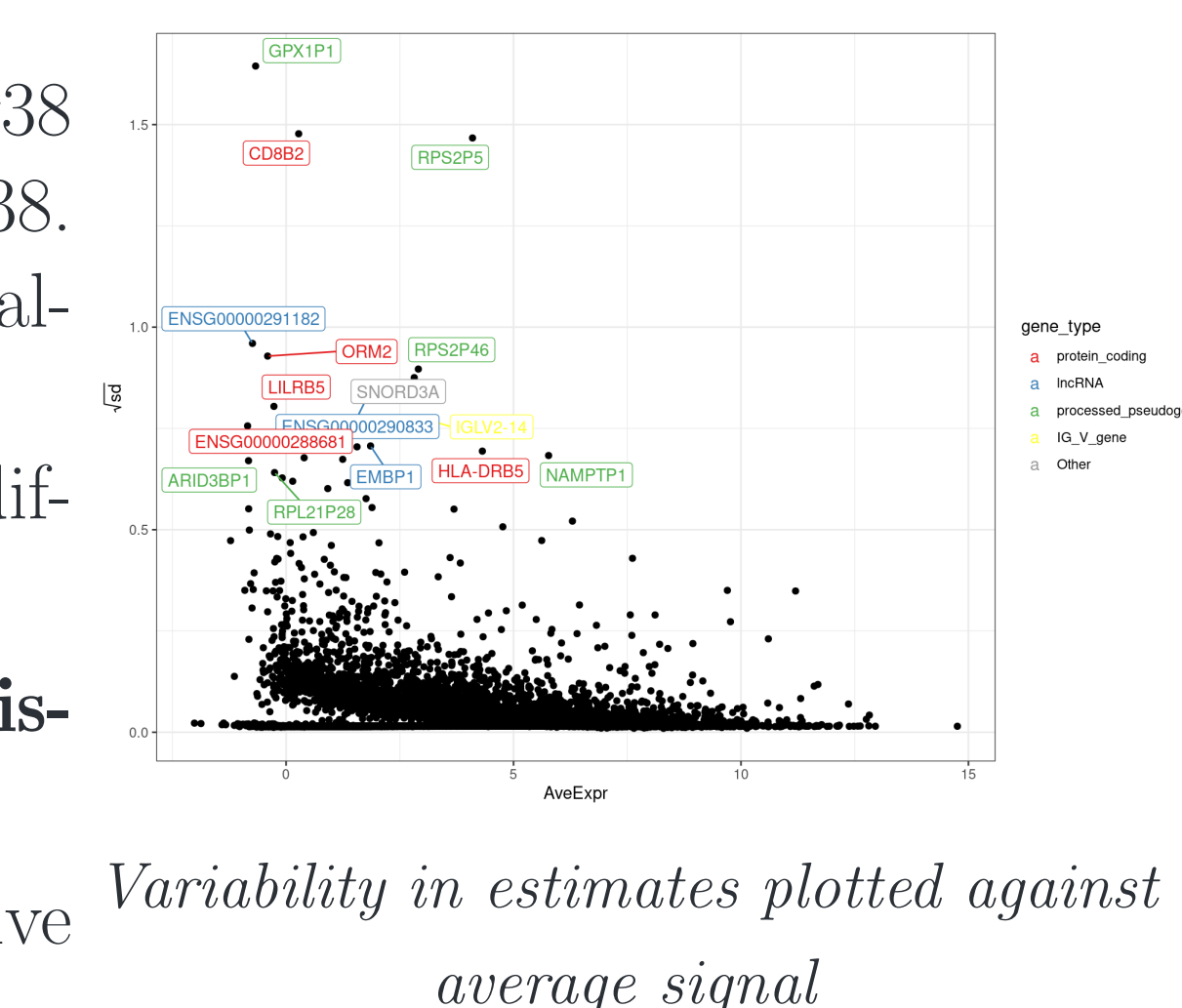
Haploid Reference Strategies

Develop a figure showing

STAR Consensus Results

6 Pilot Samples were aligned to hg38 along with the 1000GP-modified hg38. Standard Differential Expression Analysis performed.

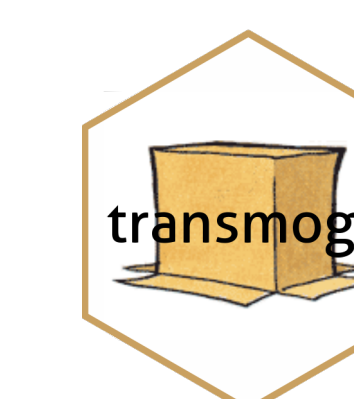
- **1 in 500 reads** were aligned to different locations
- Direction of change **was inconsistent** between samples
- Some key genes were highly sensitive



Differential Gene Expression results showing genes which passed an FDR of 0.05 for consideration as differentially expressed

Developing a Variant-Modified Reference Transcriptome

The R package **transmogR** has now been developed for using variants to modify a standard reference transcriptome, including decoy transcripts.² This approach is coordinate-agnostic \implies mapping back to regulatory features and GWAS results remains uncomplicated



References

- [1] Benjamin Kaminow et al. “Pan-human consensus genome significantly improves the accuracy of RNA-seq analyses”. en. In: *Genome Res.* 32.4 (Apr. 2022), pp. 738–749.
- [2] Avi Srivastava et al. “Alignment and mapping methodology influence transcript abundance estimation”. en. In: *Genome Biol.* 21.1 (Sept. 2020), p. 239.