Detecting germline variants from whole genome sequencing data: Impact of cohort size on diagnostic yield

Robert Zeibich[1], Anna Willard[1,2,3], Marian Todaro[1,2], Sam Berkovic[6,7], Deepak Gill[5], Lisa Gordon[8], Haris Hakeem[1,2,3,4], Michael Hildebrand[6,7], Katherine Howell[6,9,11], Kavitha Kothur[5], Saul Mullen[10], Ingrid Scheffer[6,7], Adrienne Sexton[8], Lubna Shakhatreh[1,2,3,4], Penny Snell[9], Sarah Wilson[6,7], Ingrid Winship[8], Patrick Kwan[1,2,3,4], Terence O'Brien[1,2,3,4], Piero Perucca[1,2,3,4,6,7], Alison Anderson[1]

Background

The Genome Analysis Toolkit (GATK) for variant calling recommends a '**Joint-Call Cohort**' step, which uses information across multiple exomes/genomes to determine how likely each variant is a true call.

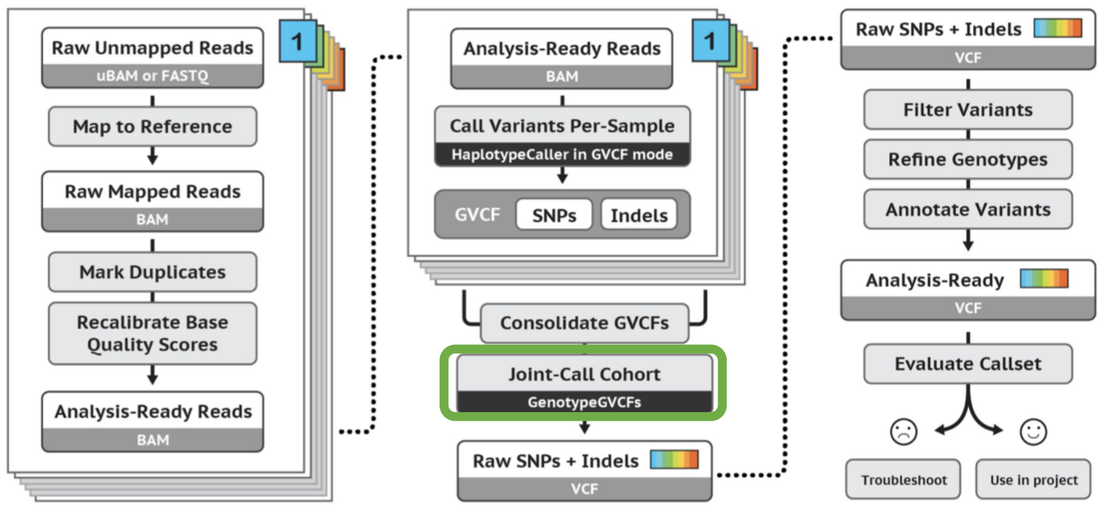

Figure 1: *Best Practices for Germline SNPs and InDels in Whole Genomes and Exomes[1]*

For exome data, it has been shown that diagnostic yield is sensitive to sample size.

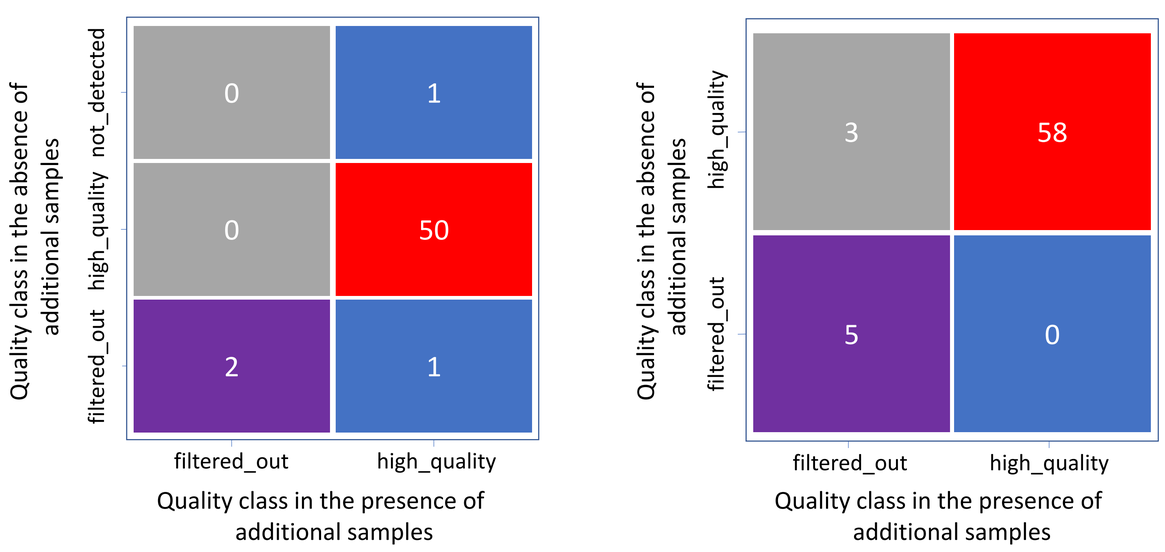

Figure 2: *Detection of rare germline pathogenic variants in cancer patients using Joint-Call Cohort[2]*

Research question

The impact of sample size on genotyping of genomic data is unclear but has implications for clinical applications, where single genomes are more likely to be processed.

Does the number of genomes used for calling impact the diagnostic yield?

Method

1. The same variant curation pipeline, used to prioritize variants for subsequent classification (based on the American College of Medical Genetics guidelines) by a multi-disciplinary team (MDT), was applied to variant call sets generated from single and joint-called cohort methods.
2. The number of variants flagged for further consideration was then compared.

Data

Whole genomes were available for 66 patients with refractory epilepsy recruited through the Genomic sequencing for Refractory Epilepsy (GREP) study.

Results

Under joint genotyping two likely pathogenic variants were filtered out. One variant was considered by the MDT to be worthy of further investigation (family segregation analysis) while the other was not relevant to the patient's phenotype.

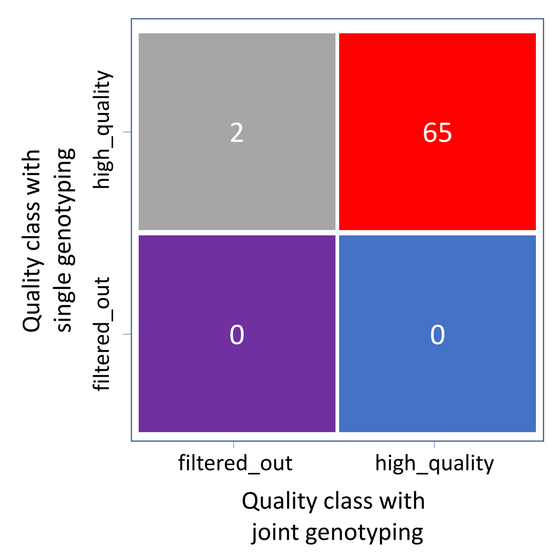

Figure 3: *Detection of rare germline pathogenic variants in GREP study patients using Joint-Call Cohort*

Conclusion

- It is possible that with genomic data, additional samples introduce noise rather than enhance performance.
- Our sample size is small, but our finding indicates that further evaluation is warranted.

Acknowledgement

We thank the patients and the clinicians and researchers for participating and making the study possible.

The study was supported by a National Health and Medical Research Council (NHMRC) funded clinical trail (APP1143934) of epilepsy.

References

1. Team, G. *Germline short variant discovery (SNPs + Indels)*. 2023 [cited 2023 16/03/2023]; Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035535932-Germline-short-variant-discovery-SNPs-Indels-#article-comments
2. Camp, S.Y., et al., *Evaluating the molecular diagnostic yield of joint genotyping-based approach for detecting rare germline pathogenic and putative loss-of-function variants.* Genet Med, 2021. **23**(5): p. 918-926.

Affiliations

1. Department of Neuroscience, Central Clinical School, Monash University; 2. Department of Neurology, The Alfred Hospital; 3. Department of Neurology, The Royal Melbourne Hospital; 4. Department of Medicine, The Royal Melbourne Hospital, The University of Melbourne; 5. Department of Neurology, The Children's Hospital at Westmead; 6. Department of Medicine, Austin Health, The University of Melbourne; 7. Comprehensive Epilepsy Program, Department of Neurology, Austin Health; 8. Department of Genetics, The Royal Melbourne Hospital; 9. Neuroscience Group, Murdoch Children's Research Institute; 10. Department of Neurology, Boxhill Hospital; 11. Department of Neurology, The Royal Children's Hospital