The landscape of genomic structural variation in Indigenous Australians

MOTIVATION. Indigenous Australians harbour rich and unique genetic diversity that is largely unexplored. Indigenous peoples have been historically under-represented in genomics research worldwide, and Aboriginal and Torres Strait Islander ancestries are currently missing from leading international genomics resources, which are used to understand and diagnose disease. Inclusion of Indigenous ancestral diversity in genomics research is critical, both to advance our understanding of global genetic diversity and, importantly, to achieve equitable outcomes in genomic medicine in Australia.

TECHNOLOGY. 'Nanopore' DNA sequencing is an exciting new technology that allows us to read the most challenging, repetitive, and mysterious DNA sequences within person's genome, which cannot be studied with other existing technologies. Nanopore sequencing is the ideal tool for identifying large structural changes, like deletions or insertions of different DNA sequences within a person's genome. These large changes in DNA sequences are called structural variants (SVs, for short). SVs are currently poorly studied – due to technical limitations – but they are critically important. They account for the majority of the differences between the genomes of any two people, and many are known to be responsible for genetic diseases. Nanopore sequencing technology is now allowing us to explore and better understand SVs in health and disease.

STUDY. In this study, we worked with the NCIG, which has partnered with four Aboriginal communities in the Central Desert, Far North Queensland and two islands off the coast of the Northern Territory. NCIG has collected DNA from several individuals from these communities. We used nanopore technology to sequence DNA from all consenting individuals along with several non-Indigenous Australians. We then studied their DNA sequence libraries to characterise SVs among Indigenous Australians, for the first time. In doing so, we are beginning to establish the required data, knowledge, and analysis tools for interpreting these important DNA variations in the context of genomic medicine. The project is a unique and ongoing opportunity for NCIG to engage Indigenous communities at the very cutting-edge of genomics research.

KEY FINDINGS. We describe an abundance of SVs across the four communities, a substantial portion of which:

(i) are composed of repetitive DNA sequences that can only be resolved with nanopore technology (>60%);

(ii) have not been previously recorded by major international genomics research projects (>70%);

(iii) appear to be exclusively found in Indigenous Australians (>30%);

(iv) appear to be found only in one out of the four Indigenous communities (>90% of Indigenous SVs).

KNOWLEDGE IMPACT. These findings highlight the rich and unique genetic diversity of Indigenous Australians; Aboriginal people are highly genetically distinct from people outside Australia. This emphasises the need for the inclusion of Indigenous Australians in genomics research to create ancestry-appropriate reference data for genomic medicine, which is currently sorely lacking. Moreover, we show that different communities have highly distinct genetic architecture, reflecting their cultural and linguistic diversity. Therefore, we argue that broad and deep engagement (far beyond the four communities we have included here) will be required to properly capture Indigenous genetic diversity and, ultimately, to achieve equitable outcomes in genomic medicine.

HEALTH IMPACT. The new analysis methods and the detailed reference data that we are generating will be key resources for future use in genomic medicine for Indigenous Australians. To identify DNA variations that may cause disease in an individual or family, it is often essential to compare them to a large group of unaffected individuals of similar ancestry. Put simply, you are trying to see which DNA variations stand out from the crowd. Our study provides an immediate example of this process: we unexpectedly found that one individual carried a specific SV known to cause Machado-Joseph Disease (**Figure 5d** in our article shows how this DNA sequence stands out from other individuals in their community). This finding is expected to lead to a clinical diagnosis for the individual and their family. It also serves to demonstrate the value of our analytical approach for investigation – and one day for clinical diagnosis – of genetic disease in Indigenous Australians.