Allelic diversity of the pharmacogene CYP2D6 in New Zealand Māori and Pacific peoples

Leonie M. Hitchman1, Allamanda Faatoese2, Tony R. Merriman3,4, Allison L. Miller1, Yusmiati Liau1†, Oscar E. E. Graham1, Ping Siu Kee1, John F. Pearson1, Tony Fakahau5, Vicky A. Cameron2, Martin A. Kennedy1* and Simran D. S. Maggo1,6*

1Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand, 2Christchurch Heart Institute, Department of Medicine, University of Otago, Christchurch, New Zealand, 3Biochemistry Department, University of Otago, Dunedin, New Zealand, 4Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, United States, 5Paciﬁc Trust Canterbury, Montreal Street, Christchurch, New Zealand, 6Department of Pathology, Center for Personalized Medicine, Children’s Hospital Los Angeles, Los Angeles, CA, United States

The enzyme cytochrome P450 2D6 (CYP2D6) metabolises approximately 25% of commonly prescribed drugs, including analgesics, anti-hypertensives, and anti-depressants, among many others. Genetic variation in drug-metabolising genes can alter how an individual responds to prescribed drugs, including predisposing to adverse drug reactions. The majority of research on the CYP2D6 gene has been carried out in European and East Asian populations, with many Indigenous and minority populations, such as those from Oceania, greatly underrepresented. However, genetic variation is often population specific and analysis of diverse ethnic groups can reveal differences in alleles that may be of clinical significance. For this reason, we set out to examine the range and frequency of CYP2D6 variants in as a sample of 202 Māori and Pacific people living in Aotearoa (New Zealand). We carried out long PCR to isolate the CYP2D6 region before performing nanopore sequencing to identify all variants and alleles in these samples. We identified twelve variants which have previously not been reported in the PharmVar CYP2D6 database, three of which were exonic missense variations. Six of these occurred in single samples and one was found in 19 samples (9.4% of the cohort). The remaining five variants were identified in two samples each. Identified variants formed twelve new CYP2D6 suballeles and four new star alleles, now recorded in the PharmVar database. One striking finding was that CYP2D6*71, an allele of uncertain functional status which has been rarely observed in previous studies, occurs at a relatively high frequency (8.9%) within this cohort. These data will help to ensure that CYP2D6 genetic analysis for pharmacogenetic purposes can be carried out accurately and effectively in this population group.
Introduction

Pharmacogenetics is the study of genetic variants which impact on an individual’s response to drugs, with the aim of guiding prescription practices to improve healthcare quality and outcomes (Bank et al., 2018; Cacabelos et al., 2019). CYP2D6 is one of the most studied pharmacogenes. This gene encodes an enzyme (CYP2D6) which is expressed in the liver and responsible for metabolising approximately 25% of commonly prescribed drugs or prodrugs (Nofziger et al., 2020).

The CYP2D6 gene is highly polymorphic, with many recorded single nucleotide variants (SNVs), small insertions/deletions, larger copy number variants including whole gene deletions or duplications, as well as hybrid genes formed by recombination with the closely related pseudogene (CYP2D7) that is in close proximity (Nofziger et al., 2020). Over 150 CYP2D6 alleles have so far been identified, each of which is allocated a name using a “star” nomenclature scheme, and tracked within the PharmVar database. Minor variations are often allocated a “suballele” designation (Gaedigk et al., 2020; Nofziger et al., 2020).

Variants of the gene may impact CYP2D6 function, ranging from completely inactivating the enzyme through to elevating its activity, although the impact of many variants is yet to be quantified. Where the functional impact of alleles is known or can be inferred, individuals can be categorised into one of four metaboliser phenotypes—ultrarapid metaboliser (UM), normal metaboliser (NM), intermediate metaboliser (IM), and poor metaboliser (PM), each of which describes activity of the CYP2D6 enzyme (Gaedigk et al., 2008). Depending on the type of medication, individuals defined as poor or ultrarapid metabolisers are at particular risk of experiencing drug toxicity or poor drug response (Cacabelos et al., 2019).

Interethnic differences in allele distribution or frequencies are evident for CYP2D6 (Gaedigk et al., 2017; Zhou et al., 2017; Koopmans et al., 2021). For example, CYP2D6*10 is reported with a frequency of 45% in East Asian populations, compared to only 1.6% in Europeans (Zhou et al., 2017). Furthermore, the PharmGKB database of CYP2D6 variants (Whirl-Carrillo et al., 2021) includes data from over 64,000 Europeans compared to less than 800 individuals from the Oceanian biogeographic grouping, defined as pre-colonial populations of the Pacific, including Hawaii, Australia, New Zealand and Papua New Guinea (Huddart et al., 2019). Many of the existing ‘Oceania’ studies focus on individuals from Papua New Guinea and Melanesia, which will not fully represent genetic diversity of Oceanian populations. Aotearoa (New Zealand) was the last major landmass to be inhabited, with Māori settlers arriving about 730 years before present (BP) (Gosling and Matisoo-Smith, 2018). Māori and Pacific Island populations have origins in South-East Asia, before migrations brought them to remote Oceania around 3000 BP (Helsby 2016; Gosling and Matisoo-Smith, 2018). There have been very few studies specifically examining pharmacogenetic variability in Māori and Pacific Island people (Wanwimolruk et al., 1995; Wanwimolruk et al., 1998; Lea et al., 2008). To ensure equitable application of pharmacogenetic tests that detect clinically relevant alleles in people of all ancestries, it is important that such pharmacogenetic variants are identified and quantified by analysis of appropriate population samples.

The choice of technology employed for studies of interethnic diversity in CYP2D6 is critical. Many prior studies have used targeted genotyping arrays or allele-specific polymerase chain reaction (PCR) methods, targeting only known alleles of interest, often derived from analysis of Europeans (Carvalho Henriques et al., 2021). These analyses are relatively cheap and straightforward to carry out, but unknown or rare alleles will not be detected (and will be potentially reported as "*1 if not positive for another allele of interest"). A more effective approach to identify all variants and determine haplotypes is to employ long-read, single molecule nanopore DNA sequencing on a PCR product encompassing the entire CYP2D6 region (Liau et al., 2019). In this paper, we applied these methods to characterise the allelic landscape of CYP2D6 in two cohorts of New Zealand Māori and Pacific Island people.

Materials and methods

Study cohorts

A total of 202 samples were drawn from two existing studies. Thirty-five of these samples were from a study called Genetics of Gout, Diabetes, and Kidney Disease in Aotearoa New Zealand, which recruited individuals aged ≥16 years primarily from the Auckland, Waikato and Christchurch regions of Aotearoa (New Zealand) (Krishnan et al., 2018). The remaining 167 samples were from the Pasifika Heart Study (Faatoese et al., manuscript in preparation). This study involved 200 Pacific participants aged 20–64 years, selected from the patient register of a Pacific-led primary healthcare clinic (Pacific Trust Canterbury, Christchurch, NZ), a low-cost health provider largely serving Pacific residents of Christchurch. Screening clinics for the Pasifika Heart Study
were held from May 2015—June 2016 at the Pacific Trust Health
Clinic and the Nicholls Research Centre, University of Otago,
Christchurch. In both studies, ethnicity was self-reported and the
ethnicities of each participant’s four grandparents were also
documented. The majority of samples were Polynesian, self-
reporting as Samoan, Tongan, or Māori. Other included
ethnicities were Cook Island Māori, Fijian (Melanesian),
Tokelauan, Kiribati, and Niuean (Table 1).

Ethical approval for the study Genetics of Gout, Diabetes,
and Kidney Disease in Aotearoa New Zealand was given by the
NZ Multi-Region Ethics Committee (MEC/05/10/130; MEC/10/
09/092; MEC/11/04/036). The Pasifika Heart Study was
approved by the New Zealand Health and Disability Ethics
Committee (14/CEN/72/AM04). Participants in both studies
gave written, informed consent.

CYP2D6 long PCR

CYP2D6 was amplified by PCR as a 6.6 kb product, including
the whole gene as well as upstream and downstream non-coding
regions. Duplication and deletion primers were used to identify
the presence of a CYP2D6 duplication or deletion via
amplification of a secondary 3.5 kb fragment if present
(Gaedigk et al., 2007; Maggo S. D. S. et al., 2019; Liau et al.,
2019). Primer sequences are provided in Supplementary
Table S1.

Barcoding of PCR amplicons

Oxford Nanopore Technologies (ONT) PCR Barcoding
Expansion 1–96 kit was used to add specific barcodes onto
the CYP2D6 amplicons. This allowed sample pooling in later
steps. For a 50 μL reaction, 0.5 nM CYP2D6 6.6 kb tailed
amplicon was amplified in the presence of 1x LongAMP
buffer, 0.3 mM Kapa dNTPs, 5 units of LongAMP Hot Start
Taq DNA polymerase, 0.2 μM barcode and ultra-pure H2O to
make up the final volume. This followed the ONT protocol SQK-
LSK109, (version PBAC96_9069_v109_revO_14August 2019).
The PCR began with an incubation at 95°C for 3 min,
followed by 15 cycles of 95°C for 15 s, 62°C for 15 s, 65°C for
7 min, and finally 65°C for 7 min, as recommended by ONT.

Magnetic bead purification

Magnetic beads were used to purify the PCR products
both prior to and following the barcoding step. This
removed DNA fragments shorter than 3–4 kb (to avoid off-
target amplification) and other impurities. Magbio
beads (MagBio Genomics Inc., Gaithersburg, Maryland,
USA) were pelleted and re-dissolved in a buffer with
10 mM Tris-HCl, 1.6M NaCl, 1 mM EDTA pH 8, 11% (w/
v) PEG 8000, 0.20% (v/v) Tween-20, and ultra-pure water to
make a 10 ml bead solution (Nagar and Schwessinger, 2018).
DNA samples were purified as described previously (Liau
et al., 2019).

Nanopore sequencing library preparation

DNA libraries of barcoded amplicons were prepared
according to the 1D PCR barcoding (96) amplicons
ONT protocol, SQK-LSK109. Briefly, after pooling the
barcoded amplicons, about 200 fmole of the DNA
pool was subjected to end-repair and ONT adapter
ligation. After purification, approximately 210 ng (~
50 fmole) of the library was introduced to the MinION
flowcell (R9.4.1) and sequenced for up to 48 h on the

### Table 1 Ethnicity details of participants.

| Primary ethnicity | Other ethnicity 1 | Other ethnicity 2 | Other ethnicity 3 |
|-------------------|------------------|------------------|------------------|
| Samoan            | 80               | 9                | 2                | 1                |
| Tongan            | 53               | 3                | —                | —                |
| Fijian            | 35               | 6                | —                | —                |
| NZ Māori          | 23               | 1                | 1                | —                |
| Cook Island Māori | 5                | 3                | —                | —                |
| NZ European       | 3                | 1                | 1                | —                |
| Kiribati          | 1                | —                | —                | —                |
| Tokelauan         | 1                | —                | —                | —                |
| Niuean            | 1                | —                | —                | —                |
| Grand Total       | 202              | 33               | 4                | 2                |

aAs self reported by participants.
GridION X5 nanopore sequencer (ONT, UK). Some sequencing was also completed on a FLO-FLG-001 Flongle (R9.4.1) loaded with approximately 90 ng (~20 fmoles) of library.

Nanopore sequencing data analysis

A previously designed pipeline was used to analyse the data generated (Graham et al., 2020). The GridION platform conducted real-time filtering, basecalling, and demultiplexing (separating and binning each sample per barcode) using Guppy version 5.0.12 (ONT, UK). A quality threshold for sequencing was set, specifying reads between 6 and 8 kb in length and a Qscore of >9. An end-to-end pipeline for data management was designed in a conda environment using a snakemake workflow (Mölder et al., 2021). The process included utilising the output of the GridION to initially map the FASTQ files generated against a reference sequence (CYP2D6_NG008376.4) using MiniMap2 version 2.20-r1061 (Li, 2018). SAMtools (version 1.7) (Li et al., 2009) was employed to perform indexing at various stages of this pipeline. Nanopolish (version 0.13.2) (Quick et al., 2016) was used to analyse variant calls. Whatshap (version 0.17) (Martin et al., 2016) was used to phase the VCF files generated by Nanopolish. Variants were then matched to CYP2D6 star alleles using the PharmVar database (Pharmacogene Variation Consortium, 2021) and unmatched variants were identified and marked as potentially novel to be taken forward for later validation. Finally, Stargazer v1.0.8 (Lee et al., 2019), an automated tool for pharmacogenetic star allele assignment, was used to confirm the CYP2D6 star allele assigned to each sample. Star allele frequencies between studies were compared with an R9.4.1 flow cell, and the remaining 82 were sequenced with a FLO-001 Flongle, with both flow cell types run on the GridION X5.

CYP2D6 genotypes

Unlike short-read sequencing technologies, nanopore sequencing gives long single molecule reads that allow sample haplotypes to be discerned.

Thirteen different CYP2D6 star alleles were identified amongst 202 individuals, including the *5 gene deletion and novel star alleles (Table 2). No CYP2D6 gene duplications were observed. The most common allele was *1 with an allele frequency of 0.48. Following this, in order of frequency, were *10, *2, and *71. Four of the identified star alleles, *4, *5, *10, and *41, are known to decrease or completely inactivate the function of the CYP2D6 enzyme, while *1, *2, and *35 have no impact. *43 and *71 have uncertain functional effects on CYP2D6. *167, *168, *169, and *170 were discovered during this study and have been named by the PharmVar CYP2D6 committee, designated as novel and of unknown function.

Stargazer 1.0.8 (Lee et al., 2019) was used to call star alleles for the first 120 samples, using variants aligned to hg19. Of these 120, eleven were called incorrectly. Five of the incorrect haplotype assignments were due to the presence of a *5 allele, while the remaining six carried *71 variants.

Predicted metaboliser phenotypes

Metaboliser status was inferred from the genotype of each sample using the diplotype activity score as previously defined (Caudle et al., 2020). In this classification, diplotypes with an activity score of 0 are classified as poor metabolisers, 0.25–1 are intermediate metabolisers, 1.25–2 are normal metabolisers, and scores above 2 are ultrarapid metabolisers. Using this definition, the majority (73%) of individuals within the cohort were classified as normal metabolisers. No individuals were identified as being ultrarapid or poor metabolisers, and a small subset (8%) of our study cohort were intermediate metabolisers. Almost 20% of the characterised samples had a metaboliser status that could not be assigned due to the presence of one or more *71, *43, or novel alleles (Table 3).

Variants previously not reported in PharmVar CYP2D6 database

We identified twelve variants that have not previously been reported as PharmVar CYP2D6 alleles, however nine of these are...
recorded in dbSNP and have been assigned reference SNP numbers (rsIDs). These variants were validated in our samples using Sanger sequencing (Table 4). Seven variants occurred within an intron of the gene, one in the promoter region, and four were exonic variants, three of which are non-synonymous. These were submitted to the PharmVar committee and assigned as new \textit{CYP2D6} star alleles and suballeles. Six of the twelve variants were singletons (detected in only one individual each), five were identified in two individuals each, while a variant at position 6528 was found in nineteen individuals. In addition, four novel suballeles were also identified (data not shown). These variants occur within the PharmVar \textit{CYP2D6} database but as part of different star alleles to those observed in this study. The majority exist on the haplotype without any core variants belonging to star alleles, so are by default \#1. One known variant (position 8934) formed a new star allele (\#170.001), due to its predicted deleterious effect and prior existence as a core variant for another star allele (\#136). Two variants at position 9414 and 9464 are new \#10 suballeles as all samples have the \#10 core variants.

Of the nine broad ethnic groups defined by Huddart \textit{et al.}, and reported by PharmGKB, 2021, Oceania is by far the least represented (PharmGKB, 2021)(Huddart \textit{et al.}, 2019). We have successfully used nanopore sequencing to analyse the entire \textit{CYP2D6} gene (including important upstream and downstream regions) of 202 Maori and Pacific Island individuals, coupled with analyses to call the star allele diplotypes of each participant.

Nine previously reported star alleles were identified amongst this cohort. \textit{CYP2D6}\#71, a previously identified rare allele with an uncertain function was identified at a relatively high frequency within this cohort (8.9%). Twelve variants not previously reported in the PharmVar \textit{CYP2D6} database were also discovered, four of which are exonic. Five known variants were found within star alleles in which they have not been reported previously. From these variants, four new star alleles have been assigned (\#167-\#170), in addition to thirteen novel suballeles (Table 4).

From each individual’s diplotype, we were able to categorise the majority of the cohort into metaboliser phenotypes. We did not find any ultrarapid or poor metabolisers. A large percentage of participants (73%) were normal metabolisers and 8% had a predicted activity score within the range of intermediate metabolisers (Caudle \textit{et al.}, 2020). The remaining 19% were unable to be assigned to a metaboliser status due to carrying one or more copies of a \#43, \#71, or a novel star allele.

We did not identify any poor metabolisers in the present study, as all non-functional alleles were found in diplotypes with a fully functional or partially functional allele. Non-functional alleles (\#4 and \#5 gene deletion) were present in ~3% of the cohort, so they do exist at a low frequency within the population and therefore could give rise to poor metaboliser phenotypes. In this regard, it is interesting to

| Star allele | Allele frequency (N = 202) | Number of haplotypes observed (N = 404) | Activity Score$^a$ | Impact |
|-------------|-----------------------------|----------------------------------------|-------------------|--------|
| *1          | 0.468                       | 189                                    | 1                 | Normal function |
| *2          | 0.141                       | 57                                     | 1                 | Normal function |
| *4          | 0.017                       | 7                                      | 0                 | No function |
| *5          | 0.012                       | 5                                      | 0                 | No function |
| *10         | 0.183                       | 74                                     | 0.25              | Decreased function |
| *35         | 0.007                       | 3                                      | 1                 | Normal function |
| *41         | 0.067                       | 27                                     | 0.5               | Decreased function |
| *43         | 0.002                       | 1                                      | ?                 | Uncertain function |
| *71         | 0.089                       | 36                                     | ?                 | Uncertain function |
| *167        | 0.002                       | 1                                      | ?                 | Unknown function |
| *168        | 0.002                       | 1                                      | ?                 | Unknown function |
| *169        | 0.005                       | 2                                      | ?                 | Unknown function |
| *170        | 0.002                       | 1                                      | ?                 | Unknown function |

$^a$Activity score and associated metaboliser phenotype as defined by Caudle \textit{et al.} (2020).

| Phenotype (inferred from genotype) | Number | Frequency |
|-----------------------------------|--------|-----------|
| Ultrarapid                        | 0      | 0         |
| Normal                            | 147    | 0.728     |
| Intermediate                      | 16     | 0.079     |
| Poor                              | 0      | 0         |
| Uncertain/unknown                 | 39     | 0.193     |

Discussion

Of the nine broad ethnic groups defined by Huddart \textit{et al.}, and reported by PharmGKB, 2021, Oceania is by far the least represented (PharmGKB, 2021)(Huddart \textit{et al.}, 2019). We have successfully used nanopore sequencing to analyse the entire \textit{CYP2D6} gene (including important upstream and downstream regions) of 202 Maori and Pacific Island individuals, coupled with analyses to call the star allele diplotypes of each participant.

Nine previously reported star alleles were identified amongst this cohort. \textit{CYP2D6}\#71, a previously identified rare allele with an uncertain function was identified at a relatively high frequency within this cohort (8.9%). Twelve variants not previously reported in the PharmVar \textit{CYP2D6} database were also discovered, four of which are exonic. Five known variants were found within star alleles in which they have not been reported previously. From these variants, four new star alleles have been assigned (\#167-\#170), in addition to thirteen novel suballeles (Table 4).

From each individual’s diplotype, we were able to categorise the majority of the cohort into metaboliser phenotypes. We did not find any ultrarapid or poor metabolisers. A large percentage of participants (73%) were normal metabolisers and 8% had a predicted activity score within the range of intermediate metabolisers (Caudle \textit{et al.}, 2020). The remaining 19% were unable to be assigned to a metaboliser status due to carrying one or more copies of a \#43, \#71, or a novel star allele.

We did not identify any poor metabolisers in the present study, as all non-functional alleles were found in diplotypes with a fully functional or partially functional allele. Non-functional alleles (\#4 and \#5 gene deletion) were present in ~3% of the cohort, so they do exist at a low frequency within the population and therefore could give rise to poor metaboliser phenotypes. In this regard, it is interesting to
TABLE 4 Variants not previously reported in PharmVar CYP2D6 database.

| Variant position (CYP2D6 Reference)* | SNP rsID | Type of variant | Number observed in cohort | Allelic designation | Functional prediction |
|--------------------------------------|----------|----------------|--------------------------|---------------------|-----------------------|
| 4818 (Promoter)                      | rs377204775 | G>A            | 2<sup>a</sup>            | *4.033              | —                     |
| 6071 (Exon 2)                        | rs/na     | Missense A>G (Gln > Arg) | 1                         | *167.001             | —                     |
| 6528 (Intron 2)                      | rs369508051 | G>A            | 19<sup>b</sup>           | *41.006              | —                     |
| 6947 (Exon 7)                        | rs75518310 | Missense G>A (Glu > Lys) | 1                         | *168.001             | —                     |
| 7058 (Intron 4)                      | rs64994275 | A>C            | 1                         | *4.032               | —                     |
| 7061 (Intron 4)                      | rs77808161 | G>C            | 1                         | *1.050               | —                     |
| 7238 (Intron 4)                      | rs/na     | G>C            | 2                         | *1.051               | —                     |
| 8376 (Intron 7)                      | rs376909251 | G>A            | 1                         | *1.052               | —                     |
| 8406 (Intron 7)                      | rs/na     | Insertion AGCACC | 2                         | *1.053               | —                     |
| 8849 (Exon 8)                        | rs1602566413 | Missense G>C (Val > Leu) | 2<sup>a</sup>              | *169.001             | —                     |
| 8887 (Exon 8)                        | rs1037093492 | Synonymous C>T (Arg > Arg) | 2<sup>a</sup> | *2.031               | No predicted effect |
| 9021 (Intron 8)                      | rs769045995 | A>G            | 1                         | *41.007              | —                     |

*Reference sequence NG_008376.4 (1 = sequence start).
<sup>a</sup>Variant also identified by Charnaud et al. (2022).
Italic values are Star allele designations.

Note that a pharmacological study using the probe drug debrisoquine in 101 Māori participants identified 5% as poor metabolisers (Wanwimolruk et al., 1995), and a similar study in 100 Polynesian participants found no poor metabolisers (Wanwimolruk et al., 1998). Therefore, our genetic findings are reasonably congruent with these earlier studies.

However, CYP2D6 allele frequencies observed in our study do not match allele frequencies previously reported by PharmGKB, 2021 for the Oceanian region (PharmGKB, 2021). This is almost certainly because the Oceanian definition encompasses populations with diverse, complex migration and ancestral histories (Gosling and Matiss-Smith, 2018). The Oceanian CYP2D6 data currently available comes from studies that in total include less than 800 individuals. The largest study (Gutiérrez Rico et al., 2020) focused on 278 Ni-Vanuatu (Micronesian) subjects, and found the majority of individuals to carry at least one *1 allele (allele frequency of 0.804), significantly greater than our identified frequency of 0.468 ($p = 3.8 \times 10^{-15}$). In total, they identified nine different star alleles amongst their cohort, the majority of which had allelic frequencies of <0.01. *2 and *10 existed above this frequency (0.068 and 0.029, respectively) as well as a duplication of *1 at 0.025 (Gutiérrez Rico et al., 2020). Our study found no duplications and only two non-novel alleles had observed frequencies below 0.01 (*35, *43). Prior reports describing high rates of CYP2D6 duplications in Oceania have been entirely based on Papua New Guinea samples (Sistonen et al., 2007; von Ahsen et al., 2010).

The relatively high frequency of the *71 allele we observed is one of the most striking findings from this study, for several reasons. Because *71 has thus far been considered a rare allele, it would not be routinely included in most CYP2D6 genotyping panels. The functional impact of *71 is uncertain, so we were unable to predict the metaboliser status for 19% of our cohort. The first study to identify *71 was performed in a Han Chinese population (Zhou et al., 2009). This group suggested the allele could be non-functional due to the p.Gly42 > Glu substitution in the protein N-terminus, a region which is involved in membrane anchoring of the enzyme (Zhou et al., 2009). However, this is speculative and has yet to be tested. Other reports, in Micronesian populations, have described allele frequencies of 0.009 in a Ni-Vanuatu sample (Gutiérrez Rico et al., 2020), significantly less than our 0.089 ($p = 2.0 \times 10^{-7}$) and 0.058 in a large sample of Solomon Islanders (Charnaud et al., 2022), comparable but significantly less than our 0.089 ($p = 0.043$).

Conclusion

In conclusion, we have used a nanopore sequencing approach to identify CYP2D6 star alleles and respective frequencies from 202 individuals of Māori and Pacific Island ethnicity. We identified nine known star alleles in this cohort, including CYP2D6*71, a previously identified rare allele of uncertain function, which comprised ~9% of the alleles detected. The frequency of this allele prevented us from inferring metabolic
function for nearly 20% of this cohort, and clearly, understanding the functional impact of *CYP2D6* using clinical phenotyping approaches or *in vitro* analyses is a priority area for future research. Should this allele have an altered function, it would be important to ensure it is included in pharmacogenetic testing panels, particularly if to be used in the Oceania region.

Our analysis did not reveal any *CYP2D6* duplications in this cohort, and overall our findings highlight the significant limitations of the Oceania biogeographic grouping for pharmacogenetic research, given the wide allele frequency differences we observe relative to studies on other Oceanian ancestral groups.

Finally, this study has further demonstrated the utility of nanopore sequencing for highly variable genes like *CYP2D6*, with long-read sequencing providing the advantage of analysing the entire gene, including intronic regions, at high throughput and relatively low cost. In addition, this method allows direct observation of haplotypes, without the need for allele-specific long PCR or other approaches which are often applied to resolve haplotypes detected in short read sequencing data. It is to be hoped this research will contribute to a more equitable uptake of *CYP2D6* pharmacogenetics in people of Māori and Pacific ancestry.

### Data availability statement

The datasets presented in this article are not readily available because ethical approval for the underlying studies prohibit public sharing of data. Requests to access the datasets should be directed to Dr Kennedy (martin.kennedy@otago.ac.nz) or Dr Faatoese (allamanda.faatoese@otago.ac.nz).

### Ethics statement

The studies involving human participants were reviewed and approved by NZ Multi-Region Ethics Committee (MEC/05/10/130; MEC/10/09/092; MEC/11/04/036) (Genetics of Gout, Diabetes and Kidney Disease in Aotearoa New Zealand cohort). New Zealand Health and Disability Ethics Committee (14/CEN/72/AM04) (Pasifika Heart Study cohort). The patients/participants provided their written informed consent to participate in this study.

### References

Bank, P. C. D., Swen, J. J., and Guchelaar, H. J. (2018). Implementation of pharmacogenomics in everyday clinical settings. *Adv. Pharmacol.* 83, 219–246. doi:10.1016/bs.apha.2018.04.003

Cacabelos, R., Cacabelos, N., and Carril, J. C. (2019). The role of pharmacogenomics in adverse drug reactions. *Expert Rev. Clin. Pharmacol.* 12, 407–442. doi:10.1080/17512433.2019.1597706

Carvalho Henriques, B., Buchner, A., Hu, X., Wang, Y., Yavorskyy, V., Wallace, K., et al. (2021). Methodology for clinical genotyping of *CYP2D6* and *CYP2C19*. *Transl. Psychiatry* 11, 596. doi:10.1038/s41398-021-01717-9

Caudle, K. E., Sangkuhl, K., Whirl-Carrillo, M., Swen, J. J., Häidar, C. E., Klein, T. E., et al. (2020). Standardizing *CYP2D6* genotype to phenotype translation: Consensus recommendations from the clinical pharmacogenetics implementation consortium and Dutch pharmacogenetics working group. *Clin. Transl. Sci.* 13, 116–124. doi:10.1111/cts.12692

Charnaud, S., Munro, J. E., Semencic, L., Mazhari, R., Brewster, J., Bourke, C., et al. (2022). PacBio long-read amplicon sequencing enables scalable high-resolution population allele typing of the complex *CYP2D6* locus. *Commun. Biol.* 5, 168. doi:10.1038/s42003-022-03102-8

Hitchman et al. 10.3389/fgene.2022.1016416

### Author contributions

LH carried out the lab work and contributed greatly to the manuscript. AF and TM provided the DNA samples and provided advice on the manuscript. AM supported and advised on laboratory work. YL set up workflow and helped troubleshoot issues. OG established bioinformatic workflow. PK performed Sanger sequencing. JP helped with bioinformatics and statistical analyses. TF contributed to establishing the Pasifika Heart Study. VC advised on manuscript and contributed to establishing and maintaining Pasifika Heart Study. MK and SM supervised LH during study and contributed significantly to manuscript.

### Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

### Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

### Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1016416/full#supplementary-material

**SUPPLEMENTARY TABLE S1**

Primer Sequences for *CYP2D6* amplification.

**SUPPLEMENTARY TABLE S2**

Primer Sequences for novel variant confirmation.
Gaedigk, A., Ndjountche, L., Divakaran, K., Dianne Bradford, L., Zineh, I., Oberlander, T., et al. (2007). Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: Characterization of gene duplication events. *Clin. Pharmacol. Ther.* 81, 242–251. doi:10.1016/j.cpt.2006.08.1910

Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., and Lee, J. S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genet. Med.* 19, 69–76. doi:10.1038/gim.2016.80

Gaedigk, A., Simon, S. D., Pearce, R. E., Bradford, L. D., Kennedy, M. J., and Lee, J. S. (2008). The CYP2D6 activity score: Translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther.* 83, 234–242. doi:10.1038/sj.clpt.6100406

Gaedigk, A., Whirl-Carrillo, M., Pratt, V. M., Miller, N. A., and Klein, T. E. (2020). PharmVar and the landscape of pharmacogenetic resources. *Clin. Pharmacol. Ther.* 107, 43–46. doi:10.1002/cpt.1654

Gosling, A. L., and Matison-Smith, E. A. (2018). The evolutionary history and human settlement of Australia and the Pacific. *Curr. Opin. Genet. Dev.* 53, 53–59. doi:10.1016/j.gde.2018.06.015

Graham, O. E. E., Pitcher, T. L., Liao, Y., Miller, A. L., Dalrymple-Alford, J. C., Anderson, T. J., et al. (2020). Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson’s disease cohort. *Park. Relat. Disord.* 70, 36–41. doi:10.1016/j.parkreldis.2019.11.022

Gutierrez Rico, E. M., Kikuchi, A., Satto, T., Komondai, M., Hishinuma, E., Kaneko, A., et al. (2020). CYP2D6 genotyping analysis and functional characterization of novel allelic variants in a Ni-Vanuatu and Kenyan population by assessing dextromethorphan O-demethylation activity. *Drug Metab. Pharmacokinet.* 35, 89–101. doi:10.1016/j.dmpk.2019.07.003

Helby, N. A. (2016). CYP2C19 and CYP2D6 genotypes in Pacific peoples. *Br. J. Clin. Pharmacol.* 82, 1303–1307. doi:10.1111/bcp.13045

Huddart, R., Fechner, A. E., Whirl-Carrillo, M., Wojcik, G. L., Gignoux, C. R., Popejoy, A. B., et al. (2019). Standardized biogeographic grouping system for annotating populations in pharmacogenetic research. *Clin. Pharmacol. Ther.* 105, 1256–1262. doi:10.1002/cpt.1322

Koopmans, A. B., Braackman, M. H., Vinkers, D. J., Hoek, H. W., and Van Harten, P. N. (2021). Meta-analysis of probability estimates of worldwide variation of CYP2D6 and CYP2D19. *Transl. Psychiatry* 11, 141. doi:10.1038/s41398-020-01291-9

Krishnan, M., Major, T. J., Toplesk, R. K., Dewes, O. Y., Liu, T., Thompson, J. M. D., et al. (2018). Discordant association of the CREBF rs373863828 A allele with increased BMI and protection from type 2 diabetes in Maori and Pacific (Polynesian) people living in Aotearoa/New Zealand. *Diabetologia* 61, 1603–1613. doi:10.1007/s00125-018-4623-2

Lee, R., Roberts, R., Green, M., Kennedy, M., and Chambers, G. (2008). Allele frequency differences of cytochrome P450 polymorphisms in a sample of New Zealand Maori. *N. Z. Med. J.* 121, 33–37.

Lee, S.-B., Wheeler, M. M., Thummel, K. E., and Nickerson, D. A. (2019). Calling genetic polymorphism, linkage disequilibrium, haplotype structure and novel allele variants of debrisoquine (CYP2D6) and proguanil (CYP2C19) in South Pacific Polynesian populations. *Eur. J. Clin. Pharmacol.* 54, 431–435. doi:10.1007/s00228-0054-088

Zhou, Y., Ingelman-Sundberg, M., and Lauschke, V. M. (2017). Worldwide distribution of cytochrome P450 alleles: A meta-analysis of population-scale sequencing projects. *Clin. Pharmacol. Ther.* 102, 688–700. doi:10.1002/cpt.690