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

Leonie M. Hitchman[1], Allamanda Faatoese[2], Tony R. Merriman[3,4], Allison L. Miller[1], Yusmiati Liau[1,5], Oscar E.E. Graham[1], Ping Siu Kee[1], John F. Pearson[1], Tony Fakahau[6], Vicky A. Cameron[2], Martin A. Kennedy[1], Simran D.S. Maggo[1]

[1] Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand

[2] Christchurch Heart Institute, Department of Medicine, University of Otago, Christchurch, New Zealand

[3] Biochemistry Department, University of Otago, Dunedin, New Zealand.

[4] Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Alabama, US

[5] Current address: Auckland District Health Board, LabPLUS, Auckland City Hospital, Auckland, New Zealand

[6] Pacific Trust Canterbury, Montreal Street, Christchurch, New Zealand

[7] Department of Pathology, Center for Personalized Medicine, Children's Hospital Los Angeles, Los Angeles, California, USA

Keywords: *CYP2D6*, pharmacogenetics, Polynesian, Māori, Pasifika, drug response, haplotype, ancestry

Joint corresponding authors: Dr Simran Maggo, Prof Martin Kennedy

Abstract

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 Indigenous and minority populations 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 a sample of 202 Māori and Pacific people living in Aotearoa (New Zealand). We carried out a long PCR to isolate the *CYP2D6* region before performing nanopore sequencing to identify all variants and alleles in these samples. We identified eleven novel variants, 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 four novel variants were identified in two samples each. In addition, five new suballeles of *CYP2D6* were identified. One striking finding was that CYP2D6*71, an allele of unknown functional status which has been rarely observed in previous studies, occurs at a relatively high frequency (9.2%) 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).

55 The *CYP2D6* gene is highly polymorphic, with many recorded single nucleotide variants
56 (SNVs), structural variants such as small insertions/deletions, larger copy number variants
57 including whole gene deletions or duplications, as well as hybrid genes formed by
58 recombination with the closely related pseudogene (*CYP2D7)* that is in close proximity
59 (Nofziger et al., 2020). Over 150 *CYP2D6* alleles have so far been identified, each of which is
60 allocated a name using a "star" nomenclature scheme, and tracked within the PharmVar
61 database. Minor variations are often allocated a "suballele" designation (Gaedigk et al.,
62 2020;Nofziger et al., 2020).

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

71 Interethnic differences in allele distribution or frequencies are evident for *CYP2D6* (Gaedigk
72 et al., 2017;Zhou et al., 2017;Koopmans et al., 2021). For example, *CYP2D6*10* is reported
73 with a frequency of 45% in East Asian populations, compared to only 1.6% in Europeans
74 (Zhou et al., 2017). Furthermore, the PharmGKB database of *CYP2D6* variants (Whirl-Carrillo
75 et al., 2021) includes data from over 64,000 Europeans compared to less than 800
76 individuals from the Oceanian biogeographic grouping, defined as pre-colonial populations
77 of the Pacific, including Hawaii, Australia, New Zealand and Papua New Guinea (Huddart et
78 al., 2019). Many of the existing 'Oceania' studies focus on individuals from Papua New Guinea
79 and Melanesia, which will not fully represent genetic diversity of Oceanian populations.
80 Aotearoa (New Zealand) was the last major landmass to be inhabited, with Māori settlers
81 arriving about 730 years before present (BP)(Gosling and Matisoo-Smith, 2018). Māori and
82 Pacific Island populations have origins in South-East Asia, before migrations brought them
83 to remote Oceania around 3000BP (Gosling and Matisoo-Smith, 2018). There have been very
84 few studies specifically examining pharmacogenetic variability in Māori and Pacific Island

85 people (Wanwimolruk et al., 1995;Wanwimolruk et al., 1998;Lea et al., 2008). To ensure
86 equitable application of pharmacogenetic tests that detect clinically relevant alleles in people
87 of all ancestries, it is important that such pharmacogenetic variants are identified and
88 quantified by analysis of appropriate population samples.

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

99

100 Materials and Methods

101 Study Cohorts

102 A total of 202 samples were drawn from two existing studies. Thirty-five of these samples
103 were from a study called Genetics of Gout, Diabetes, and Kidney Disease in Aotearoa New
104 Zealand, which recruited individuals aged ≥16 years primarily from the Auckland, Waikato
105 and Christchurch regions of Aotearoa (New Zealand) (Krishnan et al., 2018). The remaining
106 167 samples were from the Pasifika Heart Study (Faatoese *et al*. manuscript in preparation).
107 This study involved 200 Pacific participants aged 20 – 64 years, selected from the patient
108 register of a Pacific-led primary healthcare clinic (Pacific Trust Canterbury, Christchurch,
109 NZ), a low-cost health provider largely serving Pacific residents of Christchurch. Screening
110 clinics for the Pasifika Heart Study were held from May 2015 – June 2016 at the Pacific Trust
111 Health Clinic and the Nicholls Research Centre, University of Otago, Christchurch. In both
112 studies, ethnicity was self-reported and the ethnicities of each participant's four
113 grandparents were also documented. The majority of samples were Polynesian, self-

114 reporting as Samoan, Tongan, or Māori. Other included ethnicities were Cook Island Māori,
115 Fijian (Melanesian), Tokelauan, Kiribati, and Niuean.

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

121 Table 1 near here

122 **CYP2D6 Long PCR**

123 *CYP2D6* was amplified by PCR as a 6.6kb product, including the whole gene as well as
124 upstream and downstream non-coding regions. Duplication and deletion primers were used
125 to identify the presence of a *CYP2D6* duplication or deletion via amplification of a secondary
126 3.5kb fragment if present (Gaedigk et al., 2007;Liau et al., 2019;Maggo et al., 2019b). Primer
127 sequences are provided in Supplementary Table 1.

128 **Barcoding of PCR amplicons**

129 Oxford Nanopore Technologies (ONT) PCR Barcoding Expansion 1-96 kit was used to add
130 specific barcodes onto the *CYP2D6* amplicons. This allowed sample pooling in later steps. For
131 a 50µL reaction, 0.5nM *CYP2D6* 6.6kb tailed amplicon was amplified in the presence of 1x
132 LongAMP buffer, 0.3mM Kapa dNTPs, 5 units of LongAMP Hot Start *Taq* DNA polymerase,
133 0.2µM barcode and ultra-pure H_2O to make up the final volume. This followed the ONT
134 protocol SQK- LSK109, (version PBAC96_9069_v109_revO_14Aug2019). The PCR began
135 with an incubation at 95°C for 3 minutes, followed by 15 cycles of 95°C for 15 seconds, 62°C
136 for 15 seconds, 65°C for 7 minutes, and finally 65°C for 7 minutes, as recommended by ONT.

137 **Magnetic Bead Purification**

138 Magnetic beads were used to purify the PCR products both prior to and following the
139 barcoding step. This removed DNA fragments shorter than 3-4kb (to avoid off-target
140 amplification) and other impurities. Magbio beads (MagBio Genomics Inc, Gaithersburg,

141 Maryland, USA) were pelleted and re-dissolved in a buffer with 10mM Tris-HCl, 1.6M NaCl,
142 1mM EDTA pH 8, 11% (w/v) PEG 8000, 0.20% (v/v) Tween-20, and ultra-pure water to
143 make a 10mL bead solution (Nagar and Schwessinger, 2018). DNA samples were purified as
144 described previously (Liau et al., 2019).

145 **Nanopore Sequencing Library Preparation**

146 DNA libraries of barcoded amplicons were prepared according to the 1D PCR barcoding (96)
147 amplicons ONT protocol, SQK-LSK109. Briefly, after pooling the barcoded amplicons, about
148 200 fmoles of the DNA pool was subjected to end-repair and ONT adapter ligation. After
149 purification, approximately 210 ng (~ 50 fmoles) of the library was introduced to the MinION
150 flowcell (R9.4.1) and sequenced for up to 48 hours on the GridION X5 nanopore sequencer
151 (ONT, UK). Some sequencing was also completed on a FLO-FLG-001 flongle (R9.4.1) loaded
152 with approximately 90 ng (~ 20 fmoles) of library.

153 **Nanopore Sequencing Data analysis**

154 A previously designed pipeline was used to analyse the data generated (Graham et al., 2020).
155 The GridION platform conducted real-time filtering, basecalling, and demultiplexing
156 (separating and binning each sample per barcode) using Guppy version 5.0.12 (ONT, UK). A
157 quality threshold for sequencing was set, specifying reads between 6 - 8kb in length and a
158 Qscore of >9. An end-to-end pipeline for data management was designed in a conda
159 environment using a snakemake workflow (Mölder et al., 2021). The process included
160 utilising the output of the GridION to initially map the FASTQ files generated against a
161 reference sequence (CYP2D6_NG008376.3) using MiniMap2 version 2.20-r1061 (Li, 2018).
162 SAMtools (version 1.7) (Li et al., 2009) was employed to perform indexing at various stages
163 of this pipeline. Nanopolish (version 0.13.2) (Quick et al., 2016) was used to analyse variant
164 calls. Whatshap (version 0.17) (Martin et al., 2016) was used to phase the VCF files generated
165 by Nanopolish. Variants were then matched to *CYP2D6* star alleles using the PharmVar
166 database (Pharmacogene Variation Consortium) and unmatched variants were identified
167 and marked as potentially novel to be taken forward for later validation. Finally, Stargazer
168 v1.0.8 (Lee et al., 2019), an automated tool for pharmacogenetic star allele assignment, was
169 used to confirm the *CYP2D6* star allele assigned to each sample. Star allele frequencies

170 between studies were compared with two sided Fisher's exact test and considered
171 significant at p=0.05.

172 **Sanger sequencing**

173 To validate potential novel alleles, Sanger sequencing was performed following a nested PCR
174 on the relevant *CYP2D6* gene locations, using the long amplicon as a template as described
175 (Maggo et al., 2019a; Maggo et al., 2019b). Primer sequences are provided in Supplementary
176 Table 2.

177 Results

178 From an initial 282 samples, 202 were successfully sequenced and analysed at all steps.
179 Excluded samples included those that failed to amplify the 6.6kb *CYP2D6* PCR product
180 (n=52), or samples with a low read depth (less than 50 reads; n=28). Genomic DNA samples
181 which failed to amplify were checked on a TapeStation platform (Agilent, Santa Clara, USA)
182 and were usually found to be fragmented, with this fragmentation being the likely cause of
183 long-PCR failure. 120 of the samples were sequenced with an R9.4.1 flow cell, and the
184 remaining 82 were sequenced with a FLG-001 Flongle, with both flow cell types run on the
185 GridION X5.

186 **CYP2D6 Genotyping Results**

187 Nine different *CYP2D6* star alleles were identified amongst 202 individuals, including the *5
188 gene deletion (Table 2). No *CYP2D6* gene duplications were observed. The most common
189 allele was *1 with an allele frequency of 0.48. Following this, in order of frequency, were *10,
190 *2, and *71. Four of the identified star alleles, *4, *5, *10, and *41, are known to decrease or
191 completely inactivate the function of the CYP2D6 liver enzyme, while *1, *2, and *35 have no
192 impact. *43 and *71 have unknown functional effects on CYP2D6.

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

197 Table 2 near here

198 **Predicted Metaboliser Phenotypes**

199 The metaboliser status was inferred from the genotype of each sample using the diplotype
200 activity score as previously defined (Caudle et al., 2020). In this classification, diplotypes
201 with an activity score of 0 are classified as poor metabolisers, 0.25-1 are intermediate
202 metabolisers, 1.25-2 are normal metabolisers, and scores above 2 are ultrarapid
203 metabolisers. Using this definition, the majority (74%) of individuals within the cohort were
204 classified as normal metabolisers. No individuals were identified as being ultrarapid or poor
205 metabolisers, and a small subset (8%) of our study cohort were intermediate metabolisers.
206 Almost 20% of the characterised samples had a metaboliser status that could not be assigned
207 due to the presence of one or more *71 or *43 alleles (Table 3).

208 Table 3 near here

209

210 **Novel Variants**

211 We identified eleven potentially novel variants and validated them using Sanger
212 sequencing (Table 4). Seven variants occur within an intron of the gene, and four are exonic
213 variants, three of which are non-synonymous. These have the potential to form new star
214 alleles or new suballeles.

215 Table 4 near here

216

217 Six of the eleven novel variants were singletons (detected in only one individual each), four
218 were identified in two individuals each, while a variant at position 5709 was found in
219 nineteen people. In addition, five potentially novel suballeles were also identified (data not
220 shown). These variants occur within the PharmVar CYP2D6 database but as part of different
221 star alleles to those observed in this study. The majority exist on the haplotype without any

222 core variants belonging to star alleles, so are by default *1. Two variants at position 8595
223 and 8645 are likely a new *10 suballele as all samples have the *10 core variants.

224

225 Discussion

226 Of the nine broad ethnic groups defined by Huddart *et al.,* and reported by PharmGKB,
227 Oceania is by far the least represented (PharmGKB) (Huddart et al., 2019). We have
228 successfully used nanopore sequencing to analyse the entire *CYP2D6* gene (including
229 important upstream and downstream regions) of 202 Māori and Pacific Island individuals,
230 coupled with analyses to call the star allele diplotypes of each participant. Nine star alleles
231 were identified amongst this cohort, with frequencies ranging from 0.48 for *1, to 0.002 for
232 *43. CYP2D6*71, a previously identified rare allele with unknown function, was identified at
233 a relatively high frequency within this cohort, with an allele frequency of 0.092. From each
234 individual's diplotype, we were able to categorise the majority of the cohort into metaboliser
235 phenotypes. We did not find any ultrarapid or poor metabolisers. A large percentage of
236 participants (74%) were normal metabolisers and 8% had a predicted activity score within
237 the range of intermediate metabolisers (Caudle et al., 2020). The remaining 18% were
238 unable to be assigned to a metaboliser status due to carrying one or more copies of the *43
239 or *71 allele.

240 Eleven novel variants were also discovered, four of which are exonic. Five known variants
241 were found within star alleles in which they have not been reported previously, potentially
242 forming new suballeles.

243 We did not identify any poor metabolisers in the present study, as all non-functional alleles
244 were found in diplotypes with a fully functional or partially functional allele. Non-functional
245 alleles (*4 and *5 gene deletion) were present in ~3% of the cohort, so they do exist at a low
246 frequency within the population and therefore could give rise to poor metabolisers
247 phenotypes. In this regard, it is interesting to note that a pharmacological study using the
248 probe drug debrisoquine in 101 Māori participants identified 5% as poor metabolisers
249 (Wanwimolruk et al., 1995), and a similar study in 100 Polynesian participants found none

250 were poor metabolisers (Wanwimolruk et al., 1998). Therefore, our genetic findings are
251 reasonably congruent with these earlier studies.

252 However, *CYP2D6* allele frequencies observed in our study do not match allele frequencies
253 previously reported by PharmGKB for the Oceanian region (PharmGKB). This is almost
254 certainly because the Oceanian definition encompasses populations with diverse, complex
255 migration and ancestral histories (Gosling and Matisoo-Smith, 2018). The Oceanian *CYP2D6*
256 data currently available comes from studies that in total include less than 800 individuals.
257 The largest study (Gutiérrez Rico et al., 2020) focused on 278 Ni-Vanuatu (Micronesian)
258 subjects, and found the majority of individuals to carry at least one *1 allele (allele frequency
259 of 0.804), significantly greater than our identified frequency of 0.478 ($P = 4.2 \times 10^{-26}$). In total,
260 they identified nine different star alleles amongst their cohort, the majority of which had
261 allelic frequencies of <0.01. *2 and *10 existed above this frequency (0.068 and 0.029,
262 respectively) as well as a duplication of *1 at 0.025 (Gutiérrez Rico et al., 2020). Our study
263 found no duplications and only two alleles had observed frequencies below 0.01 (*35, *43).
264 Prior reports describing high rates of *CYP2D6* duplications in Oceania have been entirely
265 based on Papua New Guinea samples (Sistonen et al., 2007;von Ahsen et al., 2010).

266 The relatively high frequency of the *71 allele we observed is one of the most striking
267 findings from this study, for several reasons. Because *71 has thus far been considered a rare
268 allele, it would not be routinely included in most *CYP2D6* genotyping panels. The functional
269 impact of *71 is unknown, so we were unable to predict metaboliser status for 18% of our
270 cohort. The first study to identify *71 was performed in a Han Chinese population (Zhou et
271 al., 2009). This group suggested the allele could be non-functional due to the p.Gly42 >Glu
272 substitution in the protein N- terminus, a region which is involved in membrane anchoring
273 of the enzyme (Zhou et al., 2009). However, this is speculative and has yet to be tested. Other
274 reports, in Micronesian populations, have described allele frequencies of 0.009 in a Ni-
275 Vanuatu sample (Gutiérrez Rico et al., 2020), significantly less than our 0.092 ($P = 3.9 \times 10^{-10}$)
276 and 0.058 in a large sample of Solomon Islanders (Charnaud et al., 2022), comparable but
277 significantly less than our 0.092 (P=0.038).

Conclusion

In conclusion, we have used a nanopore sequencing approach to identify *CYP2D6* star alleles and respective frequencies from 202 individuals of Māori or Pasifika ethnicity. We identified nine known star alleles in this cohort, including CYP2D6*71, a previously identified rare allele of unknown function, which comprised over 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 *71 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. It is to be hoped this research will contribute to a more equitable uptake of *CYP2D6* pharmacogenetics in people of Māori and Pacific ancestry.

.

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.

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

Carvalho Henriques, B., Buchner, A., Hu, X., Wang, Y., Yavorskyy, V., Wallace, K., Dong, R., Martens, K., Carr, M.S., Behroozi Asl, B., Hague, J., Sivapalan, S., Maier, W., Dernovsek, M.Z., Henigsberg, N., Hauser, J., Souery, D., Cattaneo, A., Mors, O., Rietschel, M., Pfeffer, G., Hume, S., and Aitchison, K.J. (2021). Methodology for clinical genotyping of CYP2D6 and CYP2C19. *Transl Psychiatry* 11**,** 596.

Caudle, K.E., Sangkuhl, K., Whirl-Carrillo, M., Swen, J.J., Haidar, C.E., Klein, T.E., Gammal, R.S., Relling, M.V., Scott, S.A., Hertz, D.L., Guchelaar, H.-J., and Gaedigk, A. (2020). Standardizing CYP2D6 Genotype to Phenotype Translation: Consensus Recommendations from the Clinical Pharmacogenetics Implementation Consortium and Dutch Pharmacogenetics Working Group. *Clinical and Translational Science* 13**,** 116-124.

Charnaud, S., Munro, J.E., Semenec, L., Mazhari, R., Brewster, J., Bourke, C., Ruybal-Pesantez, S., James, R., Lautu-Gumal, D., Karunajeewa, H., Mueller, I., and Bahlo, M. (2022). PacBio long-read amplicon sequencing enables scalable high-resolution population allele typing of the complex CYP2D6 locus. *Commun Biol* 5**,** 168.

Gaedigk, A., Ndjountche, L., Divakaran, K., Dianne Bradford, L., Zineh, I., Oberlander, T., Brousseau, D., Mccarver, D., Johnson, J., and Alander, S. (2007). Cytochrome P4502D6 (CYP2D6) gene locus heterogeneity: characterization of gene duplication events. *Clinical Pharmacology & Therapeutics* 81**,** 242-251.

Gaedigk, A., Sangkuhl, K., Whirl-Carrillo, M., Klein, T., and Leeder, J.S. (2017). Prediction of CYP2D6 phenotype from genotype across world populations. *Genet Med* 19**,** 69-76.

Gaedigk, A., Simon, S.D., Pearce, R.E., Bradford, L.D., Kennedy, M.J., and Leeder, J.S. (2008). The CYP2D6 activity score: translating genotype information into a qualitative measure of phenotype. *Clin Pharmacol Ther* 83**,** 234-242.

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.

Gosling, A.L., and Matisoo-Smith, E.A. (2018). The evolutionary history and human settlement of Australia and the Pacific. *Curr Opin Genet Dev* 53**,** 53-59.

Graham, O.E.E., Pitcher, T.L., Liau, Y., Miller, A.L., Dalrymple-Alford, J.C., Anderson, T.J., and Kennedy, M.A. (2020). Nanopore sequencing of the glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort. *Parkinsonism Relat Disord* 70**,** 36-41.

Gutiérrez Rico, E.M., Kikuchi, A., Saito, T., Kumondai, M., Hishinuma, E., Kaneko, A., Chan, C.W., Gitaka, J., Nakayoshi, T., Oda, A., Saito, S., Hirasawa, N., and Hiratsuka, M. (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 Metabolism and Pharmacokinetics* 35**,** 89-101.

Huddart, R., Fohner, A.E., Whirl-Carrillo, M., Wojcik, G.L., Gignoux, C.R., Popejoy, A.B., Bustamante, C.D., Altman, R.B., and Klein, T.E. (2019). Standardized Biogeographic

Grouping System for Annotating Populations in Pharmacogenetic Research. *Clin Pharmacol Ther* 105**,** 1256-1262.

Koopmans, A.B., Braakman, 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 CYP2C19. *Transl Psychiatry* 11**,** 141.

Krishnan, M., Major, T.J., Topless, R.K., Dewes, O., Yu, L., Thompson, J.M.D., Mccowan, L., De Zoysa, J., Stamp, L.K., Dalbeth, N., Harre Hindmarsh, J., Rapana, N., Deka, R., Eng, W.W.H., Weeks, D.E., Minster, R.L., Mcgarvey, S.T., Viali, S., Naseri, T., Sefuiva Reupena, M., Wilcox, P., Grattan, D., Shepherd, P.R., Shelling, A.N., Murphy, R., and Merriman, T.R. (2018). Discordant association of the CREBRF 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.

Lea, 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. *The New Zealand medical journal* 121**,** 33-37.

Lee, S.-B., Wheeler, M.M., Thummel, K.E., and Nickerson, D.A. (2019). Calling Star Alleles With Stargazer in 28 Pharmacogenes With Whole Genome Sequences. *Clinical Pharmacology & Therapeutics* 106**,** 1328-1337.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34**,** 3094-3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25**,** 2078-2079.

Liau, Y., Maggo, S., Miller, A.L., Pearson, J.F., Kennedy, M.A., and Cree, S.L. (2019). Nanopore sequencing of the pharmacogene CYP2D6 allows simultaneous haplotyping and detection of duplications. *Pharmacogenomics* 20**,** 1033-1047.

Maggo, S., Kennedy, M.A., Barczyk, Z.A., Miller, A.L., Rucklidge, J.J., Mulder, R.T., and Foulds, J.A. (2019a). Common CYP2D6, CYP2C9, and CYP2C19 Gene Variants, Health Anxiety, and Neuroticism Are Not Associated With Self-Reported Antidepressant Side Effects. *Frontiers in Genetics* 10.

Maggo, S.D.S., Sycamore, K.L.V., Miller, A.L., and Kennedy, M.A. (2019b). The Three Ps: Psychiatry, Pharmacy, and Pharmacogenomics, a Brief Report From New Zealand. *Front Psychiatry* 10**,** 690.

Martin, M., Patterson, M., Garg, S., Fischer, S., Pisanti, N., Klau, G., Schöenhuth, A., and Marschall, T. (2016). "WhatsHap: fast and accurate read-based phasing. bioRxiv, 085050".).

Mölder, F., Jablonski, K., Letcher, B., Hall, M., Tomkins-Tinch, C., Sochat, V., Forster, J., Lee, S., Twardziok, S., Kanitz, A., Wilm, A., Holtgrewe, M., Rahmann, S., Nahnsen, S., and Köster, J. (2021). Sustainable data analysis with Snakemake [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research* 10.

Nagar, R., and Schwessinger, B. (2018). DNA size selection (> 3–4 kb) and purification of DNA using an improved homemade SPRI beads solution. v. 1. *Protocols. io* 10.

Nofziger, C., Turner, A.J., Sangkuhl, K., Whirl-Carrillo, M., Agundez, J.a.G., Black, J.L., Dunnenberger, H.M., Ruano, G., Kennedy, M.A., Phillips, M.S., Hachad, H., Klein, T.E., and Gaedigk, A. (2020). PharmVar GeneFocus: CYP2D6. *Clin Pharmacol Ther* 107**,** 154-170.

Pharmacogene Variation Consortium *CYP2D6* [Online]. Available: https://www.pharmvar.org/gene/CYP2D6 [Accessed 6 October 2021].

Pharmgkb *CYP2D6 Frequency Table* [Online]. Available: https://api.pharmgkb.org/v1/download/file/attachment/CYP2D6_frequency_table.xlsx [Accessed 30th Sept 2021].

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., and Mikhail, A. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530**,** 228-232.

Sistonen, J., Sajantila, A., Lao, O., Corander, J., Barbujani, G., and Fuselli, S. (2007). CYP2D6 worldwide genetic variation shows high frequency of altered activity variants and no continental structure. *Pharmacogenet Genomics* 17**,** 93-101.

Steen, V., Andreassen, O., Daly, A., Tefre, T., Borresen, A., Idl, E.J., and Gulbrandsen, A. (1995). Detection of the poor metabolizer-associated CYP2D6(D) gene deletion allele by long-PCR technology. *Pharmacogenetics* 5**,** 214-223.

Von Ahsen, N., Tzvetkov, M., Karunajeewa, H.A., Gomorrai, S., Ura, A., Brockmöller, J., Davis, T.M.E., Mueller, I., Ilett, K.F., and Oellerich, M. (2010). CYP2D6 and CYP2C19 in Papua New Guinea: High frequency of previously uncharacterized CYP2D6 alleles and heterozygote excess. *International journal of molecular epidemiology and genetics* 1**,** 310-319.

Wanwimolruk, S., Bhawan, S., Coville, P.F., and Chalcroft, S.C. (1998). Genetic polymorphism of debrisoquine (CYP2D6) and proguanil (CYP2C19) in South Pacific Polynesian populations. *Eur J Clin Pharmacol* 54**,** 431-435.

Wanwimolruk, S., Pratt, E.L., Denton, J.R., Chalcroft, S.C., Barron, P.A., and Broughton, J.R. (1995). Evidence for the polymorphic oxidation of debrisoquine and proguanil in a New Zealand Maori population. *Pharmacogenetics* 5**,** 193-198.

Whirl-Carrillo, M., Huddart, R., Gong, L., Sangkuhl, K., Thorn, C.F., Whaley, R., and Klein, T.E. (2021). An Evidence-Based Framework for Evaluating Pharmacogenomics Knowledge for Personalized Medicine. *Clin Pharmacol Ther* 110**,** 563-572.

Wright, G.E.B., Niehaus, D.J.H., Drögemöller, B.I., Koen, L., Gaedigk, A., and Warnich, L. (2010). Elucidation of CYP2D6 Genetic Diversity in a Unique African Population: Implications for the Future Application of Pharmacogenetics in the Xhosa Population. *Annals of Human Genetics* 74**,** 340-350.

Zhou, Q., Yu, X.M., Lin, H.B., Wang, L., Yun, Q.Z., Hu, S.N., and Wang, D.M. (2009). Genetic polymorphism, linkage disequilibrium, haplotype structure and novel allele analysis of CYP2C19 and CYP2D6 in Han Chinese. *Pharmacogenomics J* 9**,** 380-394.

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.

Tables:

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		1
NZ Māori	23	1	1	
Cook Island Māori	5	3		
NZ European	3	11	1	
Kiribati	1			
Tokelauan	1			
Niuean	1			
Grand Total	**202**	**33**	**4**	**2**

Table 2 Star allele frequencies and associated information

Star allele	Allele frequency (n=202)	Number of haplotypes observed (n=404)	Activity score[1]	Impact
*1	0.478	193	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	?	Unknown function
*71	0.092	37	?	Unknown function

[1] Activity score and associated metaboliser phenotype as defined by Caudle et al (2020).

Table 3 Inferred metaboliser frequencies within cohort

Inferred Phenotype (from genotype)	Number	Frequency
Ultrarapid	0	0
Normal	150	0.743
Intermediate	16	0.079
Poor	0	0
Unknown	36	0.178

Table 4 Novel Variants

Variant position (CYP2D6 reference)[1]	SNP rsID	Type of variant	Number of times observed	Allelic designation
5252 (Exon 2)	rs/na	Missense A>G (Gln > Arg)	1	*new
5709 (Intron 2)	rs369508051	G>A	19[2]	*41.new
6128 (Exon 4)	rs755518310	Missense G>A (Gly > Glu)	1	*71.new
6239 (Intron 4)	rs564994275	A>C	1	*4.new
6242 (Intron 4)	rs778008161	G>C	1	*1.new
6419 (Intron 4)	rs/na	G>C	2	*1.new
7557 (Intron 7)	rs376909251	G>A	1	*1.new
7574 (Intron 7)	rs/na	Indel T>TCAGCAC	2	*1.new
8030 (Exon 8)	rs1602566413	Missense G>C (Val > Leu)	2[2]	*new
8068 (Exon 8)	rs1037093492	Synonymous C>T (Leu > Leu)	2[2]	*2.new
8202 (Intron 8)	rs769045995	A>G	1	*41.new

[1]Reference sequence AY545216 (1=sequence start) NG_008376.3

[2]Variant also identified by Charnaud et al. (2022)

Supplementary Table 1 Primer Sequences for CYP2D6 amplification

Primer Name	Primer Sequence (5'-3')	Reference
6.6kb F	ATGGCAGCTGCCATACAATCCACCTG	(Gaedigk et al., 2007)
6.6kb R	CGACTGAGCCCTGGGAGGTAGGTAG	(Gaedigk et al., 2007)
Tailed F	TTTCTGTTGGTGCTGATATTGC	ONT
Tailed R	ACTTGCCTGTCGCTCTATCTTC	ONT
CYP-13 (Del)	ACCGGGCACCTGTACTCCTCA	(Steen et al., 1995)
CYP-24 (Del)	GCATGAGCTAAGGCACCCAGAC	(Steen et al., 1995)
FragB F (Dup)	CCATGGAAGCCCAGGACTGAGC	(Gaedigk et al., 2007)
FragB R (Dup)	CGGCAGTGGTCAGCTAATGAC	(Gaedigk et al., 2007)

Supplementary Table 2. Primer Sequences for novel variant confirmation

Primer Name	Primer Sequence (5'-3')	References
Exon 2 Forward	TCCTCCTTCCACCTGCTCAC	(Wright et al., 2010)
Exon 2 Reverse	CTTTGCCCCACCTCGTCTCT	(Wright et al., 2010)
Exon 3 4 (old) Forward	AGCTGGAATCCGGTGTCGAA	(Wright et al., 2010)
Exon 3 4 (old) Reverse	AGCCATCTCCAGGTAGACCCAG	(Wright et al., 2010)
Exon 3 4 (new) Forward	ATAGGGTTGGAGTGGGTGGT	This publication
Exon 3 4 (new) Reverse	AAATCCTGCTCTTCCGAGGC	This publication
Exon 7 Forward	CCAACATAGGAGGCAAGAAG	(Wright et al., 2010)
Exon 7 Reverse	ACTGGACTCTAGGATGCTGG	(Wright et al., 2010)
Exon 8 Forward	GTCTAGTGGGGAGACAAACCAG	This publication
Exon 8 Reverse	TGCCCTGAGGAGGATGATC	This publication