



The genomic landscape of blood groups in Indigenous Australians in remote communities

Sudhir Jadhao^{1,2}  | Wendy Hoy³ | Simon Lee^{1,2} | Hardip R. Patel⁴ |
Brendan J. McMorran⁵  | Robert L. Flower⁶ | Shivashankar H. Nagaraj^{1,2}

¹Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, Queensland, Australia

²Translational Research Institute, Brisbane, Queensland, Australia

³Faculty of Medicine, University of Queensland, Brisbane, Queensland, Australia

⁴National Centre for Indigenous Genomics, Australian National University, Canberra, Australian Capital Territory, Australia

⁵Department of Immunology and Infectious Disease, The John Curtin School of Medical Research, College of Health and Medicine, The Australian National University, Canberra, Australian Capital Territory, Australia

⁶Research and Development, Australian Red Cross Lifeblood Red Cell Reference Laboratory, Brisbane, Queensland, Australia

Correspondence

Shivashankar H. Nagaraj, Centre for Genomics and Personalised Health, Queensland University of Technology, Brisbane, QLD 4059, Australia.
Email: shiv.nagaraj@qut.edu.au

Funding information

Funding for this research was supported by grants from the National Health and Medical Research Council (GNT1024207) and MRFF Genomics Health Futures Mission (76757). The Centre for Chronic Disease, The University of Queensland, is supported in part by the NHMRC, Chronic Kidney Disease Centre of Research Excellence, 2016–2020 (APP1079502). The National Centre for Indigenous Genomics' genome sequencing program is supported by grants from the Australian Genomics Health Alliance, the Australian Research Data Commons (ARDC), Bio- platforms Australia (BPA), the Canberra Medical Society, the National Computational Infrastructure (NCI) through their ANU and National Merit Allocation Schemes, and the NHMRC (GNT1143734).

Abstract

Background: Red blood cell (RBC) membrane-associated blood group systems are clinically significant. Alloimmunisation is a persistent risk associated with blood transfusion owing to the antigen polymorphisms among these RBC-associated blood groups. Next-generation sequencing (NGS) offers an opportunity to characterize the blood group variant profile of a given individual. Australia comprises a large multiethnic population where most blood donors are Caucasian and blood group variants remain poorly studied among Indigenous Australians. In this study, we focused on the Tiwi Islanders, who have lived in relative isolation for thousands of years.

Methods and materials: We predicted the blood group phenotype profiles in the Tiwi (457) and 1000 Genomes Phase 3 (1KGP3-2504) cohort individuals using RBCEq (<https://www.rbceq.org/>). The predicted phenotype prevalence was compared with the previous literature report.

Results: We report, for the first time, comprehensive blood group profiles corresponding to the 35 known blood group systems among the Indigenous Tiwi islander population and identify possible novel antigen variants therein. Our results demonstrate that the genetic makeup of the Tiwi participants is distinct from that of other populations, with a low prevalence of LU (Au[a–b+])

Abbreviations: CKD, chronic kidney disease; CNVs, copy number variations; HTRs, haemolytic transfusion reactions; Indels, Insertions/deletions; NGS, Next-generation sequencing; RBC, Red blood cell; 1KGP3, 1000 genomes phase 3; SNPs, Single nucleotide polymorphisms.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Transfusion* published by Wiley Periodicals LLC on behalf of AABB.

and ABO (A2) and D+C+c+E+e- phenotype, an absence of Diego blood group variants, and a unique RHD (DIII type4) variant.

Conclusion: Our results may contribute to the development of a database of predicted phenotype donors among the Tiwi population and aid in improving transfusion safety for the ~2800 Tiwi people and the ~800,000 other Indigenous Australians throughout the nation.

KEYWORDS

alloimmunisation, blood group system, next-generation sequencing (NGS), novel antigen, rare antigen, red blood cells, single nucleotide variants (SNV), transfusion, weak antigen

1 | INTRODUCTION

Blood transfusion has historically been an important facet of hematological care for which there is no viable alternative. There is immense diversity among human blood group systems, with blood group genes displaying a range of antigenic polymorphisms, including single nucleotide polymorphisms (SNP), insertions/deletions (indels), and copy number variations (CNVs).¹ Misidentification or inappropriate matching of these variants may pose a risk of alloimmunisation that can result in serious hemolytic transfusion reactions (HTRs).^{2,3} Next-generation sequencing (NGS) is the most comprehensive approach to achieving extended red blood cell (RBC) typing, enabling the accurate characterization of the complete blood group variant profiles for a given individual. NGS approaches include whole-exome sequencing (WES), whole-genome sequencing (WGS), and targeted exome sequencing (TES).⁴ While these technologies are powerful, there is currently a lack of comprehensive software packages with the storage and processing capability necessary to handle large volumes of NGS data and to accurately predict blood group phenotypes. Recently, our group developed RBCeq, a web server-based blood group genotyping software designed to improve blood type compatibility testing, and to facilitate the accurate mass screening of blood groups in diverse populations with distinctive and complex blood group profiles.⁵

Australia is a multiethnic nation, with roughly one-third of the population having been born overseas. However, most Australian blood donors are Caucasian. It is well established that the prevalence of antigens can differ among populations,⁶ potentially hindering efforts to maintain a safe and robust blood supply in multiethnic countries.⁷ In an earlier study, we carried out extended blood group molecular typing to detect rare antigens in a large Australian Caucasian cohort.⁸ Blood group antigen profiling for Western Desert Indigenous Australians is available.⁹ However, the Tiwi Islanders are culturally distinct among these Australian Indigenous communities. They

lived in relative isolation for 7000–15,000 years and consider themselves to be distinct from mainland Aboriginal communities, with their own origin stories, language, and customs.¹⁰ Although isolated from mainland Indigenous groups for thousands of years, they had intermittent contact with visitors from Southeast Asia and Europe beginning in the seventeenth century.¹¹ The Tiwi population, along with Aboriginal people from Arnhem Land, has a higher mortality rate than the Australian mainstream.¹⁰ Tiwi people are at high risk of kidney, cardiovascular, and metabolic diseases that significantly reduce their life expectancy relative to the non-Indigenous population.^{10,12} Alpha thalassaemia has also been identified in some Aboriginal and Torres Strait Islander communities in the Northern Territory and Northwestern Australia, requiring regular transfusion for affected individuals.¹³ Studies of blood group variants among the Tiwi people have never been performed. Developing an approach to careful blood group matching is thus vital for Indigenous individuals who required blood transfusion. Here, we characterize blood samples from Tiwi people with the aim of developing a database of blood type polymorphisms among Indigenous Australians in order to facilitate accurate donor matching and thereby reduce the incidence of transfusion reactions.

2 | METHODS

2.1 | Study population and datasets

Blood samples were collected between 2013 and 2014 as detailed in a previous study investigating the prevalence of chronic kidney disease (CKD) in the Tiwi population.¹² Whole-genome sequencing of 189 samples was performed in three batches by three different research groups. Additionally, 268 Tiwi samples were genotyped using Infinium Global Screening Array v3.0. Publicly available variant dataset from the 1000 Genomes Phase 3 (1KGP3) project¹⁴ was downloaded and used for comparative analyses. Allele

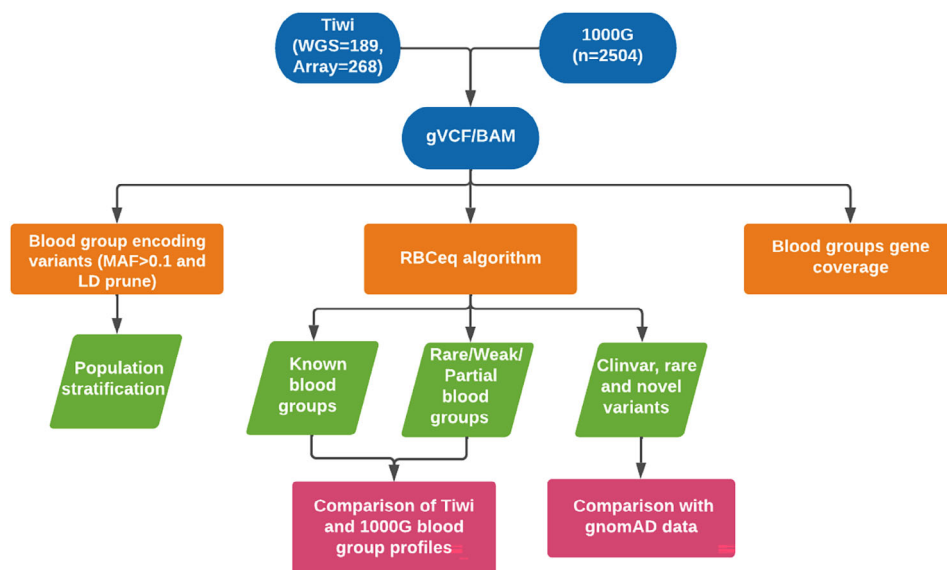


FIGURE 1 An overview of the workflow to comprehensively characterize population-specific blood group variants and phenotypes [Color figure can be viewed at wileyonlinelibrary.com]

frequencies for gnomAD¹⁵ populations were obtained from ANNOVAR¹⁶ (gnomAD genome collection [v2.1.1]).

2.2 | Variant calling and annotation

Variant calling was performed according to GATK best practices v4.0.2.0¹⁷ and Sentieon DNaseq v.201808.08¹⁸ using the GRCh37 human reference genome. The resulting variant calls were filtered based on variant quality score recalibration (VQSR). Autosomal variants for 39 samples aligned to hg38 in the form of a processed variant call format (VCF) files were obtained from the National Centre for Indigenous Genomics. To achieve comprehensive blood group allele annotation, all variant calls were converted to the GRCh37 assembly for autosomal variants.

2.3 | Prediction of blood group alleles and their phenotype using RBCeq

We predicted the blood group antigen profiles for the Tiwi (457) and 1KGP3 (2504) cohort individuals using RBCeq (<https://www.rbceq.org/>).⁵ The genomic variant call format (gvcf) of each sample was given as input with the blood group gene coverage calculated using the BAM-Trimmed tool within RBCeq. RBCeq considers SNV, indels from 45 genes and calculates CNV to profile 36 blood groups and two transcription factors (Table S1). RBCeq first, scans each blood group for reference and alternative allele in the input vcf file. The zygosity of detected reference-alternate allele guides the sub allele selection. The selected alleles are paired against each other and are assigned with a score. The score helps to select allele pair which considers most of the known

present variants from the input vcf file. The detailed method is described in Jadhao et al., 2022.⁵

The RBCeq output included the prediction of International Society of Blood Transfusion (ISBT) known blood group antigen alleles and phenotype. RBCeq also classified the filtered high-quality deleterious variants which are not mapped to known ISBT alleles. First, it classified the variant as clinically significant by querying against ClinVar,¹⁹ then rare variants which have gnomAD $MAF \leq 0.05$ and potential novel variants which are without any public database annotation (Figure 1). The Rh copy number variation (D zygosity and C/c) prediction accuracy by RBCeq was confirmed using the RH blood group gene coverage plot. All samples RH genes coverage were considered using BedTools²⁰ (bin size: 1), which were further normalized for every 300 bases pairs and plotted using R libraries.

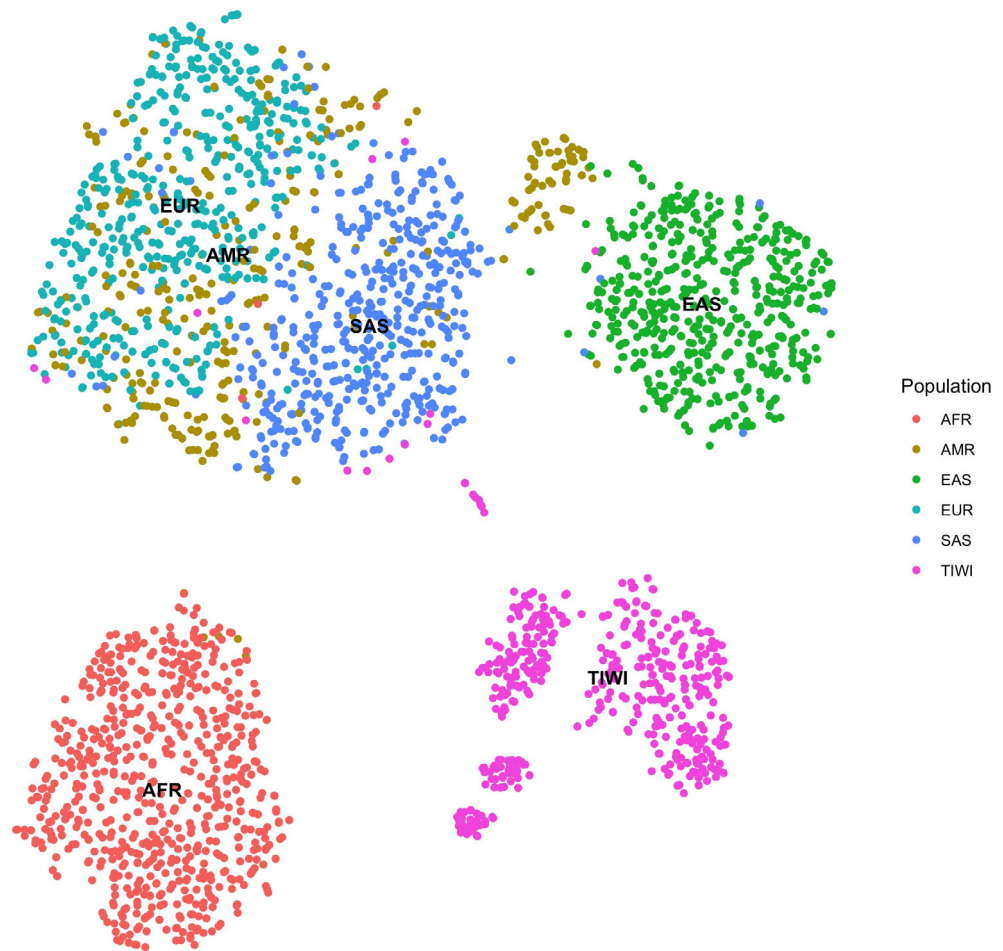
To do the population stratification between the Tiwi dataset and other global populations, we carried out principal component analysis (PCA) of 695 variants (LD pruned and $MAF \geq 0.1$) from 42 autosomal blood groups encoding gene sequences of TIWI and 1KGP3 dataset using PLINK v1.9 (Figure 1). The PCA results were plotted using ggplot2.²¹ The blood group encoding gene variants annotation/categorization was done using ANNOVAR.¹⁶ The Circos plot was created to show variants frequency and the gene annotation using R libraries like circlize,²² tidyverse,²³ dplyr,²⁴ ComplexHeatmap,²⁵ and stringr.²³

3 | RESULTS

3.1 | The blood group gene variant landscape in the Tiwi population

Results from a PCA plot analysis revealed that the Tiwi participants exhibited a unique blood group genomic

FIGURE 2 PCA plots showing 695 (MAF ≥ 0.1 and LD prune) markers in 454 individuals from the Tiwi and 1KGP3 datasets; the X-axis denotes the value of PC1, while Y-axis denotes the value of PC2, with each dot in the figure representing one individual. The first two principal components showed here account for $\sim 80\%$ of the observed variance in the combined dataset. AMR, American; AFR, African; EAS, east Asian; EUR, European; LD, linkage disequilibrium; MAF, minor allele frequency; PCA, principal component analysis; SAS, south Asian [Color figure can be viewed at wileyonlinelibrary.com]



identity that was not closely correlated with that of the European population, who account for most blood donors in Australia. While clear differentiation was observed among the Tiwi, African, and East Asian participants, all three groups also distinctly differed from the rest of the global populations (Figure 2).

3.2 | Distribution of weak and partial antigens

We observed that the distribution of ABO antigens in Tiwi was distinct from that in other populations. Four sub-phenotypes were predicted in ABO blood groups among this population: O (81.18%), A1 (18.38%), A2 (0.22%), and B (0.22%). The distribution of the O phenotype was very high compared with the other populations (Caucasian: 44%, African: 49%, and Asian: 43%). A1 and A2 phenotypes account for 99% of all groups A individuals.²⁶ The A2 subgroup is considered a weak A subgroup that is most identified via the presence of an unexpected anti-A or anti-A1. This is the first study to report the prevalence of A2 in an Indigenous population, revealing it to be very rare (0.22%) among this Tiwi

population (Table 1). The rare Jk(a-b-) was missing in all six populations, but Jk(a+^wb-) predicted from JK*01 W.01 (c.130G > A) had a much higher frequency (16.19%) in Tiwi as compared to others 4.39% (African), 4.74% (American), 1.19% (European), and 8.38% (South Asian) except in East Asians (16.47%) (Table 1). Jk(a+^wb-) antigen causes weak or partial expression of Jk(a+b-) and can be mistyped as Jk(a-).

Genetic variants that lead to weakened or partial antigen expression can be difficult to type.²⁸ D category IV is one of the most clinically relevant and phylogenetically diverse partial D.²⁹ We detected a DIII type 4 frequency of 1.75% ($n = 8$) at a homozygous level among the Tiwi population (Table 1). Phenotyping a patient with partial D as D- is safe for their management and enables the prevention of anti-D alloimmunisation.³⁰ However, these patients are often misidentified as being D+ and they are at risk for alloanti-D development. The mother with anti-D and fetus with partial D are at risk for hemolytic disease of the fetus and newborn.³ Antibodies to Lutheran antigens are rarely clinically significant; most of its antigens have high prevalence, but Au(a-b+) phenotype is polymorphic and rare.³¹ A multiethnic population-level frequency for the Au(a-b+) phenotype

TABLE 1 Comparison of blood group profiles with previously reported prevalence in other populations

Blood group	Predicted phenotype	RBCeq prediction (%)						Previously published (%) ²⁷			
		Tiwi <i>n</i> = 457	AFR <i>n</i> = 661	AMR <i>n</i> = 347	EUR <i>n</i> = 503	EAS <i>n</i> = 504	SAS <i>n</i> = 489	Caucasians	Africa	Asian	rare
ABO	A2	0.22 (<i>n</i> = 1)	11.50 (<i>n</i> = 76)	5.64 (<i>n</i> = 19)	13.92 (<i>n</i> = 70)	1.19 (<i>n</i> = 6)	5.32 (<i>n</i> = 26)	10	8	27	
ABO	B	0.22 (<i>n</i> = 1)	23.90 (<i>n</i> = 158)	5.93 (<i>n</i> = 20)	42.86 (<i>n</i> = 216)	10.54 (<i>n</i> = 53)	47.44 (<i>n</i> = 233)	9	20	27	
RHD	DIII type4	1.75 (<i>n</i> = 8)	2.87 (Four homozygous and 15 heterozygous)	-	-	-	-	-	-	-	-
LU	LU:18-, 19 or Au(a-b+)	0.22 (<i>n</i> = 1)	19.21 (<i>n</i> = 127)	7.12 (<i>n</i> = 24)	10.34 (<i>n</i> = 52)	0.99 (<i>n</i> = 5)	4.91 (<i>n</i> = 24)	-	-	-	-
FY	Fy(a-b+)	0.88 (<i>n</i> = 3)	-	24.04 (<i>n</i> = 81)	0.4 (<i>n</i> = 2)	33 (<i>n</i> = 166)	12.68 (<i>n</i> = 62)	34	22	0.3-3	
JK	Jk(a-b+)	5.69 (<i>n</i> = 26)	5.14 (<i>n</i> = 34)	30.27 (<i>n</i> = 102)	25.79 (<i>n</i> = 130)	25.45 (<i>n</i> = 128)	14.31 (<i>n</i> = 70)	23	8	27	
JK	Jk(a+ ^w b-)	16.19 (<i>n</i> = 76)	4.39 (<i>n</i> = 29)	4.74 (<i>n</i> = 16)	1.19 (<i>n</i> = 6)	16.47 (<i>n</i> = 83)	8.38 (41)	-	-	-	-
DO	Do(a+b-)	50.32 (<i>n</i> = 230)	6.51 (<i>n</i> = 43)	14.54 (<i>n</i> = 49)	1.19 (<i>n</i> = 6)	13.92 (<i>n</i> = 70)	12.88 (<i>n</i> = 63)	18	11	0.5-2	
DO	Do(a-b+)	10.5 (<i>n</i> = 48)	53.86 (<i>n</i> = 356)	40.35 (<i>n</i> = 136)	81.55 (<i>n</i> = 411)	39.36 (<i>n</i> = 198)	35.99 (<i>n</i> = 176)	33	45	77-87	

Abbreviations: AMR, American; AFR, African; EAS, East Asian; EUR, European; SAS, South Asian.

has not been reported previously. The Tiwi distribution of Au(a-b+) phenotype was 0.22% (*n* = 1), which was very low relative to the frequency in other populations (African, American, East Asian, European, South Asian) (Table 1).

3.3 | Predicted blood group profiles contrasting with previous reports in other population

The protein sequence of the A and B glycosyltransferases (GTA and GTB) are encoded by the ABO gene, which differs by four amino acid residues, and a mutation in these sequences resulting in protein truncation is what leads to the O phenotype.³² B antigen is common in most of the population²⁷ and we also observed the same through the 1KG population dataset analysis. However, the prevalence of B antigen was very low among the Tiwi population such that only one individual harboring the B antigen was detected (Table 1).

Low immunogenic antigens from the Dombrock blood group systems can cause mild-to-moderate and acute-to-delayed HTRs.³³ Dombrock antigens are difficult to detect,³⁴ and typing is not routinely performed in the context of routine pretransfusion practice. Additionally, their antigenic frequency represents a significant ethnic marker.³⁴ Both Do(a+) and Do(b+) antigens are common among most global populations.²⁷ However, we detected at least two times higher Do(a+b-) prevalence (50.32%) among Tiwi as compared with the 1KGP3 population and previous reports.²⁷ Conversely, the Do(a-b+) predicted phenotype prevalence among the Tiwi population was less than half of that reported for other global populations and detected in the 1KGP3 population in the current study (Table 1).

It is well established that the Duffy (Fy) antigens can cause maternal immunization³⁵ and subsequent rare HDFN. The Fy(a-b-) null phenotype is more common among African populations. An anti-Fya African sickle cell anemia patient may require multiple blood transfusions.³⁶ The Fy(a-b+) phenotype is observed in about 34% of Caucasians, 22% of African Americans, and 3% of the Asian population.²⁷ Here, we detected a low prevalence of Fy(a-b+) among the Tiwi population (0.88%) which was comparable to that in the East Asian population (0.4%) (Table 1).

The clinically significant anti-Jka and anti-Jkb are weak antigens that may trigger HTRs and delayed transfusion reactions.³⁷ Anti-Jka and anti-Jkb are difficult to detect in routine serology tests and notorious for their evanescence.³⁸ We predicted a three times lower prevalence (5.69%) of the Jk(a-b+) phenotype among the Tiwi

TABLE 2 RHD blood group phenotype frequency prevalence in Tiwi and 1KGP3 dataset

Blood group	Gene	Genotype	Predicted phenotype	RBCseq prediction (%)						Previously published (%) ²⁷		
				Tiwi n = 151	AFR n = 661	AMR n = 347	EUR n = 503	EAS n = 504	SAS n = 489		Western Desert Indigenous Australians	Caucasians Africans Asian
RHD	DD	D+	100 (n = 151)	96.52 (n = 628)	93.95 (n = 326)	83.90 (n = 422)	99.80 (n = 503)	94.48 (n = 462)	97.22 (n = 72)	85.00	92.00	99.00
RHD	D-	D-	0 (n = 0)	3.48 (n = 23)	6.05 (n = 21)	16.10 (n = 81)	0.20 (n = 1)	5.52 (n = 27)	2.77 (n = 2)	15.00	8.00	1.00

Abbreviations: AMR, American; AFR, African; EAS, East Asian; EUR, European; SAS, South Asian.

population compared with that predicted in the 1KGP3 population, and a 1.3 times lower prevalence than reported in previous studies²⁷ (Table 1).

3.4 | Frequency of RHD/RHCE blood group phenotypes

We also equated the RHD/RHCE blood group predicted phenotype frequencies of Tiwi with those from 1KGP3 dataset and previous reports.²⁷ In the comparison of the prevalence of the D with respect to homozygous and null expression, our results revealed that 99.80% of East Asian, 96.52% of African, 93.95% of American, and 94.48% of South Asian participants were homozygous for the *RHD*-gene (Table 2). However, we found a very high (100%) prevalence of homozygous *RHD*-gene (Table 2; Figure S1) in 151 WGS Tiwi samples. The previous indigenous study also reported only two D-negative individuals (Table 2).⁹

The highly polymorphic C antigen expression is driven by the number of exon 2- *RHCE* reads which are mapped to exon 2 of *RHD*^{39–41} (Figure S2). The prevalence of the D+C–c+E+e+ phenotype was very low in Tiwi and East Asian populations as compared to other populations (Table 3). The prevalence of the D+C+c+E–e+ phenotype was two times higher (10.28) as compared to other populations other than American and African populations (Table 3). The rare D+C+c+E+e– phenotype was predicted in two individuals of the Tiwi population however it was absent in the 1KGP3 dataset.

3.5 | The MNS phenotype of the Tiwi people

We observed at least a 1.5 times higher prevalence of the M–N+S–s+ among the Tiwi participants as compared with the 1KGP3 population (Table S2). Conversely, the prevalence of M+N–S–s+ phenotype was predicted to be two times lower (2.41%) in the Tiwi participants as compared with other populations except in European²⁷ (Table S2).

3.6 | Investigation of novel blood group variants with potential clinical relevance

In a secondary analysis, we characterized variants with the potential to affect the antigen specificity but with no previous blood group allele associations within the Tiwi population dataset (Figure 3). We detected 13 variants that had clinical associations¹⁹ (Table S3), 5 of which had

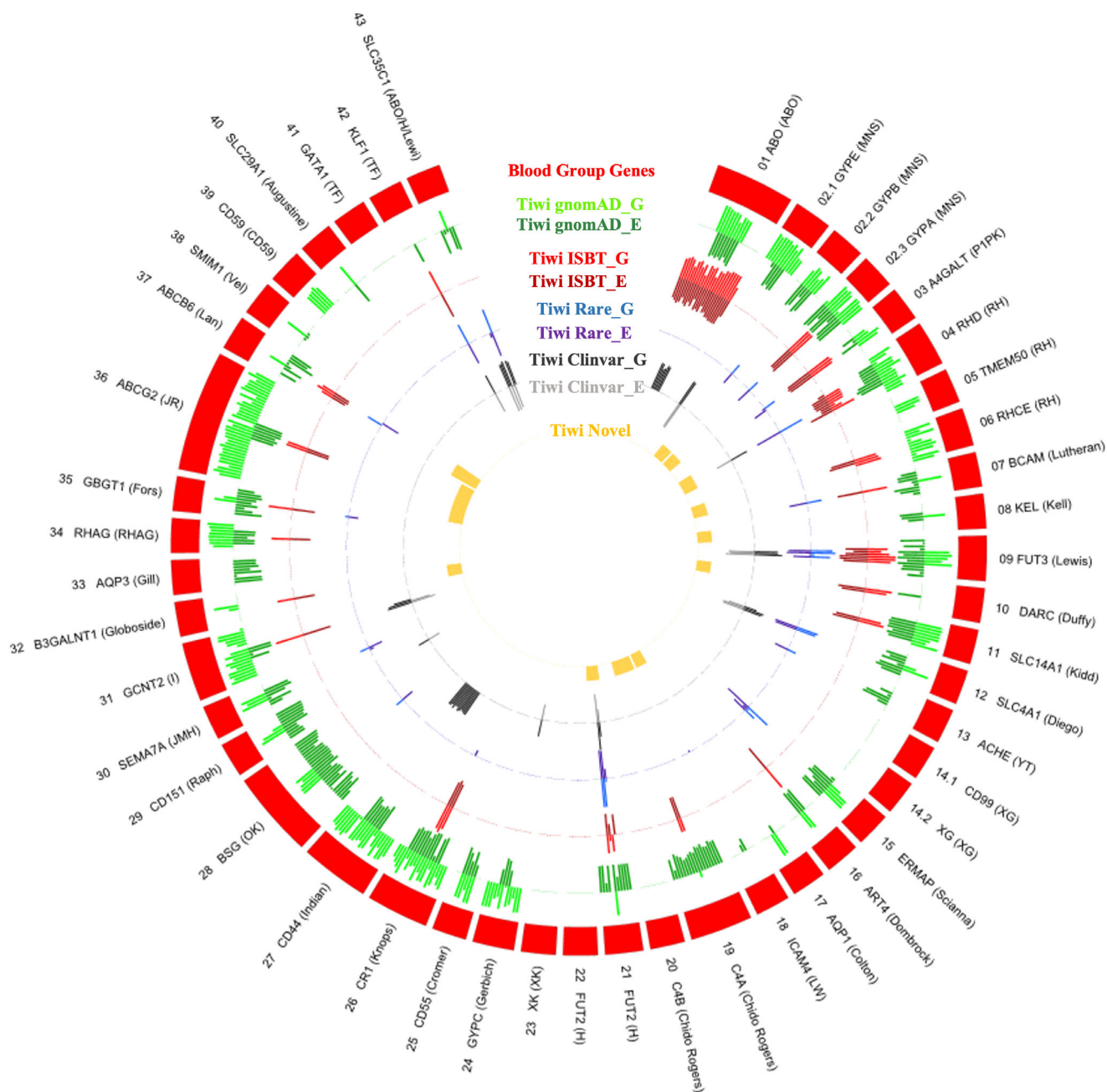


FIGURE 3 Tiwi blood group gene encoding variants distribution as known (ISBT), GnomAD mapped, rare (≤ 0.05 in gnomAD), ClinVar annotated and potential novel variants. The RBC antigen encoding genes are shown by the outer ring (red); the number of variations found is represented by the box length; G denotes gnomAD genome frequency, and E denotes gnomAD exome frequency. The outside green (light/dark) circle represents the Tiwi variants mapped to gnomAD data's across distinct blood group genes. The red (light/dark) circle indicates variants with ISBT annotation and their frequency in gnomAD dataset. The blue (light/dark) circle indicates the distribution of Tiwi variants with no ISBT annotation and is rare in gnomAD dataset. The dark gray circle indicates the number of non-ISBT variants annotated to the ClinVar database with gnomAD frequency. The yellow circle shows the distribution of the number of potential novel variants [Color figure can be viewed at wileyonlinelibrary.com]

two times higher frequency in Tiwi. And 22 variants had Human Gene Mutation Database (HGMD) annotation⁴³ (Table S4), five were associated with blood group antigen structure rs3088190 (Dombrock), rs5036 (Diego), rs8138197 (P1PK), rs5751348 (P1PK), CD44*c.255C > G (Indian). The rs201915844 (FUT2) was also detected

which impairs the H antigen activity. Other 17 were associated with enhancer activity, porphyria, non-secretory phenotype, and sedimentation (Table S4). There were 28 rare variants with frequencies of ≤ 0.05 among the gnomAD dataset (Table S5). Most importantly, 13 potential novel variants were identified that were predicted to

TABLE 3 RHCE blood group phenotype frequency prevalence in Tiwi and 1KGP3 dataset

Phenotype	RBCeq prediction (%)						Previously published (%) ^{27,42}
	TIWI (457)	AFR (n = 661)	AMR (n = 347)	EUR (n = 503)	SAS (n = 489)	EAS (n = 504)	Caucasians
D+C-c+E+e+	0.44 (n = 2)	8.02 (n = 53)	5.48 (n = 19)	12.33 (n = 62)	4.7 (n = 23)	0.6 (n = 3)	10
D+C+c+ E-e	10.28 (n = 47)	13.77 (n = 91)	8.65 (n = 30)	4.57 (n = 23)	5.32 (n = 26)	5.16 (n = 26)	34.9
D+C+c+E+e-	0.44 (n = 2)	-	-	-	-	-	<0.1

Abbreviations: AMR, American; AFR, African; EAS, East Asian; EUR, European; SAS, South Asian.

TABLE 4 Blood group antigen prevalence comparisons for the Tiwi and other previously reported populations

Blood group	Tiwi (%)	Western Desert (%) ⁹	Indigenous literature (%) ⁵⁰⁻⁵⁵	Caucasian (%) ²⁷
Yk(a-)	10.07%	38%	Missing	8%
Fy(a-)	0.88	1.39	0%	34%
Di(a+)	0%	0%	0%	<0.1%
KYO+	Missing	<0.1	Missing	0%
K+k-	0%	0%	0%	0.2

be deleterious by any of the SIFT,⁴⁴ Polyphen2,⁴⁵ MutationTaster2,⁴⁶ LRT,⁴⁷ FATHMM,⁴⁸ or PROVEAN⁴⁹ in silico methods (Figure 3, Table S6).

3.7 | Comparison of Tiwi blood group antigens with other indigenous populations

We compared the blood group antigen profiles of the Tiwi participants with those of other Indigenous Australian populations reported previously,⁵⁰⁻⁵⁵ including Western Desert Indigenous Australians⁹ for whom a comprehensive blood group antigen profile prediction was performed using WES (Table 4).⁹ Overall, 10% of Tiwi samples in this dataset lacked the high-frequency Yk(a+) antigen belonging to the KN blood group system as compared to 38% of the Western Desert population. Yk(a+) is common in most populations other than the African population. The Fy(a-b+) negative phenotype common among Caucasians, was predicted to be rare in the Tiwi population as well as in other Indigenous Australian populations. No Diego blood group antigen was detected among the Tiwi dataset except reference antigen; the same finding was observed in the western desert report. Interestingly, the rare KYO+ phenotype predicted in the Western Desert population, was absent among the Tiwi people. KYO+ was previously only been reported in two Japanese individuals.²⁷ K+k- was completely missing in Tiwi but present in Caucasians (0.2%)⁵⁶ (Table 4). The Mc+ antigen from the MNS blood group system was present in Tiwi but at the heterozygous level which was also observed in the western desert.

4 | DISCUSSION

Blood group antigens are poorly characterized for Indigenous Australians. The most recent serological blood typing for Indigenous populations was performed in the 1960s.⁵¹ Here, we compared the blood group genetic profile of the Indigenous Tiwi participants with those of African, American, Asian, and European populations from the 1KGP3 dataset and other previously published reports. Our population stratification analysis revealed the genetic makeup of the Tiwi participants to be distinct from that of other populations, as was that of the African participants, consistent with prior reports.⁵⁷ We analyzed the distribution of the low-prevalence phenotype among these six populations. The ABO phenotype A2 has been present primarily among European and African populations. These ABO subgroups exhibit distinct frequencies among different populations, thus requiring data from and experience with individual populations to ensure safe blood transfusion practices.³²

We identified the polymorphic and rare Au(a-b+) phenotype, belonging to the Lutheran blood group system, in all six populations with different frequencies among these groups. Although the Lutheran blood group system comprises 20 clinically well-defined antigens, no specific data is available regarding the clinical significance of anti-Au(a-b+). Our results predicted the highest prevalence of the weak Jk(a+^wb-) phenotype, among the East Asian and Tiwi participants.

The most clinically significant blood group variants involved in HDFN are the RHD antigens, and other RH antigens including the c, E, e and C antigens.⁵⁸ The anti-D, and

anti-c have >50% risk of developing mild to severe HDFN if the fetus inherits the target paternal red cell antigen.⁵⁸ Our copy number variation analysis for the homozygosity and null percentages of the RHD gene in all six populations indicated that the majority of the Tiwi (100%), East Asian (90.28%), African (96.52%), American (99.35%) and South Asian (99.48%) participants were homozygous for RHD gene (D+ phenotype). Only one Tiwi individual was hemizygous for the RHD gene. We also detected DIII type4 in eight Tiwi individuals at the homozygous level. DIII type4 is often mistyped as D+ and may lead to the risk of alloanti-D production.^{39,59} Additionally, maternal anti-D with a partial D-type leads to the risk of HDFN in a D-positive fetus. We also report the prevalence of the rare D+C+c+E+e- phenotype in the Tiwi population, which was completely absent in 1KGP3 populations.

In the secondary analysis, blood group gene variants that are not previously assigned to known blood group antigens and are predicted to be deleterious using in silico methods were scanned for clinical significance, rarity, and potential novelty with global population frequency. We uncovered 13 variants that had pathogenic clinical association mostly involved in leukocyte adhesion deficiency, hemolytic anemia, and secretor/nonsecretor polymorphism. We also detected 29 rare variants with minimum allele frequencies of ≤ 0.05 in any of the five populations (African, American, East Asian, South Asian, and European) and 11 of them had two times higher frequency in Tiwi. Most interestingly, we identified 13 potential novel variants using in silico prediction techniques, of which three were common, and eight were rare in the Tiwi dataset.

Although one of the first comprehensive blood group studies of a Tiwi-Indigenous Australian population, the current study has certain limitations. The RH copy number variation analysis was restricted to 151 samples due to the availability of only WGS data, and it did not include an investigation of hybrid RH and MNS systems. The predicted clinically significant and novel variant was done solely by applying the in silico tools method from RBCEq, further investigation is needed for their antigenic properties. Blood group calls could not be made for the XG, GATA1 and XK systems because the variants were called for only autosomal chromosomes. Additionally, serology testing and exon-level detection were not performed to validate these RBCEq output data. Despite the limitations, our results have the potential to help identify donors from Indigenous Australian populations and to aid in blood group matching for patients with rare phenotypes. Our study also suggests that, in comparison to other ethnicities, East Asian populations have the next closest blood group profile to Tiwi. In addition to the associated anthropological interest, the clarification of well-known blood group polymorphisms in diverse populations represents a key step towards ensuring the safety of blood transfusion. This work will

provide critical data necessary for healthcare facilities to implement regional and national blood transfusion programs aimed at securing transfusion safety for the Tiwi population and other Indigenous Australian communities.

ACKNOWLEDGMENTS

The authors would like to acknowledge the following people: Barry Ullungurra for his help as the key contact person with the Tiwi Islanders; Bev Mcleod and Ceri Flowers for their project management and sample and data collection; Maria Scarlett for her considerable advice and guidance on the ethics of this project; and Beverley Hayhurst for the original sample collection and most notably the study participants and the Tiwi Land Council for their time and ongoing support for this project; Australian governments fund Australian Red Cross Lifeblood to provide blood, blood products and services to the Australian community. Open access publishing facilitated by Queensland University of Technology, as part of the Wiley - Queensland University of Technology agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ETHICS STATEMENT

The study entitled “*To confirm the genetic regions found to cause renal disease in the Tiwi Island Aboriginal Population*” received the full support of the Tiwi Island Land Council and was approved by the human research ethics committees of The Northern Territory Department of Health (2012–1767), The Australian National University (2014–663), The University of Queensland (2012001146), The University of Tasmania (H0012832), and Queensland University of Technology (1800000895). The blood group study was approved by Tiwi Island Land Council on 22nd June 2017 through personal communication to Dr Nagaraj.

ORCID

Sudhir Jadhao  <https://orcid.org/0000-0002-1497-3044>

Brendan J. McMorran  <https://orcid.org/0000-0003-1845-8872>

REFERENCES

1. Daniels G. The molecular genetics of blood group polymorphism. *Transpl Immunol.* 2005;14:143–53.
2. Strobel E. Hemolytic transfusion reactions. *Transfus Med Hemother.* 2008;35:346–53.
3. Basu S, Kaur R, Kaur G. Hemolytic disease of the fetus and newborn: current trends and perspectives. *Asian J Transfus Sci.* 2011;5:3–7.
4. Lane WJ, Westhoff CM, Uy JM, Aguad M, Smeland-Wagman R, Kaufman RM, et al. Comprehensive red blood cell

- and platelet antigen prediction from whole genome sequencing: proof of principle. *Transfusion*. 2016;56:743–54.
5. Jadhao S, Davison CL, Roulis EV, Schoeman EM, Divate M, Haring M, et al. RBCeq: a robust and scalable algorithm for accurate genetic blood typing. *EBioMedicine*. 2022;76:103759.
 6. Delaney M, Harris S, Haile A, Johnsen J, Teramura G, Nelson K. Red blood cell antigen genotype analysis for 9087 Asian, Asian American, and native American blood donors. *Transfusion*. 2015;55:2369–75.
 7. Volken T, Crawford RJ, Amar S, Mosimann E, Tschaggelar A, Taleghani BM. Blood group distribution in Switzerland - a historical comparison. *Transfus Med Hemother*. 2017;44:210–6.
 8. Jadhao S, Davison C, Roulis EV, Lee S, Lacaze P, Riaz M, et al. Using whole genome sequencing to characterize clinically significant blood groups among healthy older Australians. medRxiv. 2021. <https://doi.org/10.1101/2021.04.18.21255241>
 9. Schoeman EM, Roulis EV, Perry MA, Flower RL, Hyland CA. Comprehensive blood group antigen profile predictions for Western Desert indigenous Australians from whole exome sequence data. *Transfusion*. 2019;59:768–78.
 10. Hoy WE, Mott SA, McLeod BJ. Transformation of mortality in a remote Australian aboriginal community: a retrospective observational study. *BMJ Open*. 2017;7:e016094.
 11. Morris J. The Tiwi: from isolation to cultural change. A history of encounters between an island people and outside forces. Darwin, N.T.: Northern Territory University Press (NTU Press); 2001.
 12. Thomson RJ, McMorran B, Hoy W, Jose M, Whittock L, Thornton T, et al. New genetic loci associated with chronic kidney disease in an indigenous Australian population. *Front Genet*. 2019;10:330.
 13. Gaff C, Newstead J, Metcalfe S. Genetics in Family Medicine: The Australian Handbook for General Practitioners. Ed. Commonwealth of Australia Biotechnology Australia; 2007.
 14. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
 15. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581:434–43.
 16. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.
 17. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8.
 18. Kendig KI, Baheti S, Bockol MA, Drucker TM, Hart SN, Heldenbrand JR, et al. Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy. *Front Genet*. 2019;10:736.
 19. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44:D862–8.
 20. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
 21. Hadley W, Danielle N, Thomas L. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
 22. Gu Z, Gu L, Eils R, Schlesner M, Brors B. Circlize implements and enhances circular visualization in R. *Bioinformatics*. 2014;30:2811–2.
 23. Wickham H, Averick M, Bryan J, Chang W, LD MG, François R, et al. Welcome to the tidyverse. *J Open Source Software*. 2019;4(43):1686. <https://doi.org/10.21105/joss.01686>
 24. Hadley Wickham, R.F., Lionel Henry and Kirill Müller. (2021) dplyr: A Grammar of Data Manipulation.
 25. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847–9.
 26. Jeon H, Calhoun B, Pothiwala M, Herschel M, Baron BW. Significant ABO hemolytic disease of the newborn in a group B infant with a group A2 mother. *Immunohematology*. 2000;16:105–8.
 27. Reid ME, Lomas F, Olsson M. The blood group antigen facts book. Vol 3. London: Elsevier; 2012.
 28. McBean RS, Hyland CA, Flower RL. Approaches to determination of a full profile of blood group genotypes: single nucleotide variant mapping and massively parallel sequencing. *Comput Struct Biotechnol J*. 2014;11:147–51.
 29. von Zabern I, Wagner FF, Moulds JM, Moulds JJ, Flegel WA. D category IV: a group of clinically relevant and phylogenetically diverse partial D. *Transfusion*. 2013;53:2960–73.
 30. Dean L. Blood Groups and Red Cell Antigens, Chapter 7. Bethesda (MD): National Center for Biotechnology Information (US); 2005.
 31. Needs, M. (2018) The Lutheran blood group system.
 32. Ying YL, Hong XZ, Xu XG, Chen S, He J, Zhu FM, et al. Molecular basis of ABO variants including identification of 16 novel ABO subgroup alleles in Chinese Han population. *Transfus Med Hemother*. 2020;47:160–6.
 33. Strupp A, Cash K, Uehlinger J. Difficulties in identifying antibodies in the Dombrock blood group system in multiply alloimmunized patients. *Transfusion*. 1998;38:1022–5.
 34. Langer IBV, Visentainer JEL, Zacarias JMV, Grilo KTM, Hatschbach PR, Zimmermann RS, et al. Genotyping of Dombrock and Lutheran blood group systems in blood donors from the southwestern region of the state of Parana, southern Brazil. *Hematol Transfus Cell Ther*. 2019;41:25–30.
 35. Duffy PE. Maternal immunization and malaria in pregnancy. *Vaccine*. 2003;21:3358–61.
 36. Afenyi-Annan A, Kail M, Combs MR, Orringer EP, Ashley-Koch A, Telen MJ. Lack of Duffy antigen expression is associated with organ damage in patients with sickle cell disease. *Transfusion*. 2008;48:917–24.
 37. Villa MA, Moulds M, Coluccio EB, Pizzi MN, Paccapelo C, Revelli N, et al. An acute haemolytic transfusion reaction due to anti-Jk. *Blood Transfus*. 2007;5:102–6.
 38. Yates J, Howell P, Overfield J, Voak D, Downie DM, Austin EB. IgG anti-Jka/Jkb antibodies are unlikely to fix complement. *Transfus Med*. 1998;8:133–40.
 39. Schoeman EM, Lopez GH, McGowan EC, Millard GM, O'Brien H, Roulis EV, et al. Evaluation of targeted exome sequencing for 28 protein-based blood group systems, including the homologous gene systems, for blood group genotyping. *Transfusion*. 2017;57:1078–88.
 40. Chou ST, Flanagan JM, Vege S, Luban NLC, Brown RC, Ware RE, et al. Whole-exome sequencing for RH genotyping

- and alloimmunization risk in children with sickle cell anemia. *Blood Adv.* 2017;1:1414–22.
41. Wagner FF, Flegel WA. Review: the molecular basis of the Rh blood group phenotypes. *Immunohematology.* 2004;20:23–36.
 42. Mbalibulha Y, Muwanguzi E, Mugenyi G. Rhesus blood group haplotype frequencies among blood donors in southwestern Uganda. *J Blood Med.* 2018;9:91–4.
 43. Stenson PD, Mort M, Ball EV, Chapman M, Evans K, Azevedo L, et al. The human gene mutation database (HGMD ([R])): optimizing its use in a clinical diagnostic or research setting. *Hum Genet.* 2020;139:1197–207.
 44. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc.* 2016;11:1–9.
 45. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods.* 2010;7:248–9.
 46. Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutation Taster2: mutation prediction for the deep-sequencing age. *Nat Methods.* 2014;11:361–2.
 47. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. *Genome Res.* 2009;19:1553–61.
 48. Rogers MF, Shihab HA, Mort M, Cooper DN, Gaunt TR, Campbell C. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics.* 2018;34:511–3.
 49. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS ONE.* 2012;7:e46688.
 50. Erber WN, Buck AM, Threlfall TJ. The haematology of indigenous Australians. *Hematology.* 2004;9:339–50.
 51. Scott ID, Scott MK. Blood groups of some Australian aborigines of the Western Desert. *Nature.* 1966;212:545.
 52. Kirk GHVaRL. A naturally-occurring anti-E which distinguishes a variant of the E antigen in Australian aborigines. *Vox Sang.* 1962;7:22–32.
 53. Simmons RT, Semple NM, Cleland JB, Casley-Smith JR. A blood group genetical survey in Australian aborigines at Haast's bluff, Central Australia. *Am J Phys Anthropol.* 1957;15: 547–53.
 54. Sanger R, Walsh RJ, Kay MP. Blood types of natives of Australia and New Guinea. *Am J Phys Anthropol.* 1951;9: 71–8.
 55. Boyd, J.B.B.A.W.C. *Blood Groups in the Australian Aborigines.*
 56. Dean L. *Blood groups and red cell antigens.* Bethesda (MD): National Center for Biotechnology Information (US); 2005.
 57. Bodmer W. Genetic characterization of human populations: from ABO to a genetic map of the British people. *Genetics.* 2015;199:267–79.
 58. Johnsen JM. Using red blood cell genomics in transfusion medicine. *Hematology Am Soc Hematol Educ Program.* 2015;2015: 168–76.
 59. Lane WJ, Westhoff CM, Gleadall NS, Aguiar M, Smeland-Wagman R, Vege S, et al. Automated typing of red blood cell and platelet antigens: a whole-genome sequencing study. *Lancet Haematol.* 2018;5:e241–51.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Jadhao S, Hoy W, Lee S, Patel HR, McMorran BJ, Flower RL, et al. The genomic landscape of blood groups in Indigenous Australians in remote communities. *Transfusion.* 2022;62:1110–20. <https://doi.org/10.1111/trf.16873>