

# Landscape of blood group antigens and alleles in the Indian population from whole genome sequences

Mercy Rophina<sup>1,2</sup>, Rahul C Bhoyar<sup>1</sup>, Mohamed Imran<sup>1,2</sup>, Vigneshwar Senthivel<sup>1,2</sup>, Mohit Kumar Divakar<sup>1,2</sup>, Anushree Mishra<sup>1</sup>, Bani Jolly<sup>1,2</sup>, Sridhar Sivasubbu<sup>1,2</sup>, Vinod Scaria<sup>1,2§</sup>

<sup>1</sup>CSIR Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, Delhi 110025, INDIA

<sup>2</sup>Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India.

<sup>§</sup>Address for correspondence

**Corresponding author Email:**

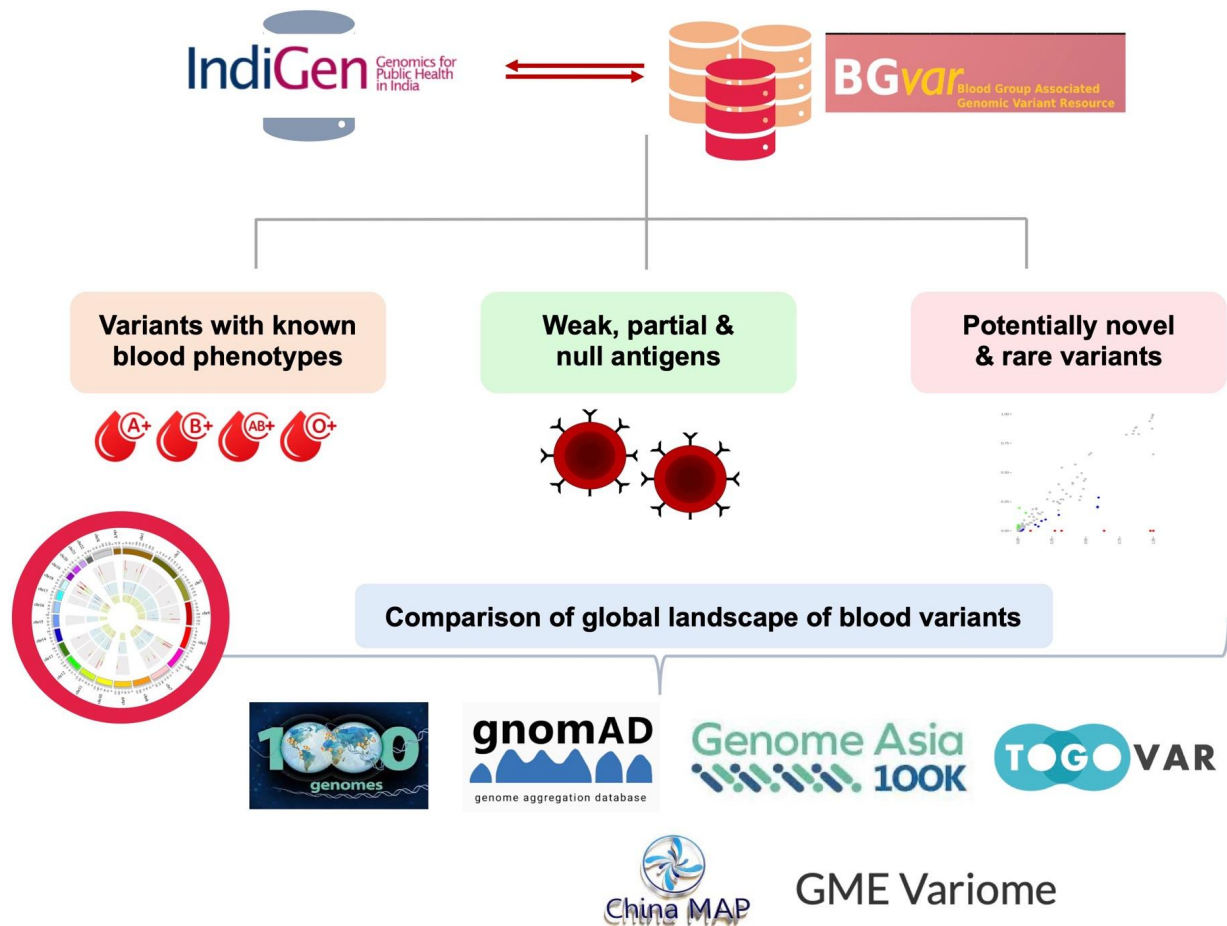
Dr. Vinod Scaria - [vinods@igib.in](mailto:vinods@igib.in) ; Tel: 91-11-29879 109

**Running Title** - Extensive blood group profile of the Indian population

## Key points

- Accurate characterization of the genomic landscape of known and rare blood group alleles and antigens in the Indian population using the whole genome sequencing data of 1029 self-declared healthy individuals
- Understanding the distinct similarities and differences in blood group genotypes and phenotypes across diverse global populations through systematic comparison of genomic datasets.

## Graphical abstract



## Abstract

Blood group antigens are genetically inherited macromolecular structures which form the underlying factor for inter individual variations in human blood. Currently there exists over 390 human blood group antigens corresponding to 44 blood group systems and 2 erythroid specific transcription factors. Distribution of these blood group antigens have been found to differ significantly among various ethnic populations. To date, there is a lack of comprehensive research that offers extensive blood group profiles for the Indian population. Whole genome sequence data (hg38) of 1029 self-declared healthy Indian individuals generated as a part of the pilot phase IndiGen programme were used for the analysis. Variants spanning the genes of 44 blood group systems and two transcription factors KLF1, GATA1 were fetched and annotated for their functional consequences. Our study reports a total of 40712 blood group related variants of which 695 were identified as non-synonymous variants in the coding region. Of the total non-synonymous variants, 105 were found to have a known blood phenotype. A total of 24 variants belonging to 12 blood groups were predicted to be deleterious by more than three computational tools. Our study was also able to identify a few rare blood phenotypes including Au(a-b+), Js(a+b+), Di(a+b-), In(a+b-) and KANNO-. This study is the first to use genomic data to understand the blood group antigen profiles of the Indian population, and it also systematically compares these profiles with those of other global populations.

## Introduction

Blood group antigens are genetically inherited macromolecular structures which form the underlying factor for inter individual variations in human blood. These antigens are primarily protein, glycolipids or glycoproteins (Dean, 2005). With constant increase in demand for blood and blood related products, characterization of human blood groups has been a routine practice in transfusion science and in blood related disorders. Blood type generally represents the complete set of red cell/platelet surface antigens carried by an individual. Any mismatch or incompatibility between these antigens during transfusion procedures or pregnancies can lead to severe immune responses which can sometimes bring life threatening complications. In addition, human blood groups are found to have profound disease associations (Mitra et al., 2014)(Tufano et al., 2013),(Hiltunen et al., 2009),(D.-S. Wang et al., 2012),(Gates et al., 2011),(Anstee, 2010),(“Complications of Blood Transfusion,” 2006)

Currently there exists about 360 human blood group antigens corresponding to 41 blood group systems and 2 erythroid specific transcription factors(Storry et al., 2019). Distribution of these blood group antigens have been found to differ significantly among various ethnic populations. Knowing the indigenous prevalence of blood groups is vital for systematic functioning of blood bank units for effectively managing alloimmunization and multi-transfusion cases. Recent years have witnessed numerous studies that have

extensively performed population scale blood phenotype frequency analysis (E.f.b.t.m. et al., 2018),(AIBilali. et al., 2017),(AISuhaibani et al., 2015), (Balci, 2010), (Yu et al., 2016),(Owaidah et al., 2020),(Bashwari et al., 2001),(Sarhan et al., 2009),(Canizalez-Román et al., 2018), (Nazli et al., 2015),(Doku et al., 2019),(Golassa et al., 2017)

India being an abode of over 4500 diverse ethnic groups with the second largest population density of 1.3 billion, demands the need to maintain indigenous population specific blood group profiles for effective regulation of transfusion practices (Mastana, 2014). There exists a handful of regional population studies in India, revealing the prevalence of major blood groups - ABO and RH (Garg et al., 2014),(Singh et al., 2016),(Chandra & Gupta, 2012),(Kaur et al., 2013),(Kaur et al., 2013; Mondal et al., 2012),(Periyavan et al., 2010),(Nanu & Thapliyal, 1997).Although ABO and RH blood groups play predominant roles in immunohematological procedures, a range of minor human blood group systems with significant impact in transfusion have been exposed in the last decade. In developing countries like India alloimmunization complications due to clinically significant minor blood groups are not completely considered. Recipient groups including young women, pregnant females and patients expected to undergo multiple transfusion procedures because of blood related disorders require routine testing for significant minor blood group antigens. Very minimal studies exist, which have looked at the landscape of clinically significant minor blood group antigens in the nation. (Lamba et al., 2013), (Thakral et al., 2007).

Owing to the diversity of human blood group antigens, obtaining a complete blood group profile using classical serological or molecular techniques becomes impractical. An alternative approach of utilizing the whole genome sequence data of an individual to retrieve the complete blood profile was demonstrated by Lane et al in 2016 (Lane et al., 2016). Large scale genomic datasets have allowed researchers to explore '*previously neglected antigens*'. One such discovery was reported by Moller et al in 2016, where a 5821 bp deletion encompassing exons 5,6, and 7 of the ABO gene was found in 20 individuals primarily of African origin in the 1000 Genomes project data (Möller et al., 2016).

Our study aims to systematically assess large scale genomic data to explore complete blood group antigen profiles of Indian indigenous population of the nation thereby paving way to understand the geographical and ethnic distribution of blood group antigens and associated phenotypes.

## **Materials and methods**

### *Reference and study datasets*

Reference dataset includes a comprehensive list of ISBT approved blood group related variants, documented in a pre-formatted template. Whole genome sequence data (hg38) of 1029 self-declared healthy Indian individuals generated as a part of the pilot phase IndiGen

programme were used for the analysis. The study dataset comprised a total of 59646267 genetic variants which includes single-nucleotide variants (SNVs), insertions and deletions. Comprehensive list of genetic variations was retrieved in the variant call format (VCF) files and were used for further analyses. In addition, a range of large global population scale datasets including the 1000 genomes project, gnomAD, GenomeAsia 100K, Greater Middle Eastern Variome, Qatar, Iranome, Singapore Sequencing Malay Project (SSMP), Singapore Sequencing Indian Project (SSIP), China Metabolic Analytics Project (ChinaMAP) and Japanese genome variations (TogoVar) were used for comparison analysis.

#### *Processing and annotation of variants*

There exist 53 unique genes associated with human blood groups. In the primary level of data processing, all the genetic variants spanning human blood group genes and erythroid specific transcription factors were filtered. **Supplementary Table 1.** provides the list of genes along with their corresponding locus genomic reference (LRG) coordinates. Filtered variants were subsequently annotated for a range of functional consequences including SIFT (Ng & Henikoff, 2003), Polyphen (Adzhubei et al., 2013) LRT, MutationTaster, Mutation Assessor (Gnad et al., 2013), FATHMM (Shihab et al., 2013), PROVEAN (Choi & Chan, 2015), CADD (Rentzsch et al., 2019), GERP (Davydov et al., 2010), PhyloP (Pollard et al., 2010) and PhastCons (Siepel et al., 2005) using Annotate Variation (ANNOVAR) (K. Wang et al., 2010)). A schematic representation of the methodology is depicted in **Figure 1.**

#### *Identification of variants with known blood phenotypes in the Indian population*

The annotated variants were systematically classified into two categories: those that are associated with a specific blood phenotype and those that are not. All the blood group systems with no blood group phenotype associated variants were conferred with the reference nomenclature as previously described {ref}.

#### *Identification of potentially novel and rare variants*

In subsequent levels of data analysis, the filtered variants were systematically classified as potentially novel/rare variants based on their SNP identification number (dbSNP ID) and minor allele frequencies (MAF). Any filtered variant lacking a dbSNP ID and literature evidence were deemed as potentially novel. Exonic novel variants were further assessed for their deleteriousness predicted by a range of computational tools. In addition, variants with MAF < 1% with absence of reported blood phenotypes were deemed as rare variants.

#### *Estimation and comparison of blood group related allele frequencies among various global populations*

To assess the distribution of blood group, neutrophil, and platelet variants in the Indian population, a comparative analysis was conducted by examining the allele frequencies of these variants across major global populations. The comparisons included data from diverse resources like 1000 Genomes project, Genome Aggregation Database (gnomAD), Genome Asia, Greater Middle Eastern Variome, Qatar Genomes and Exomes, Singapore Sequencing

Malay Project (SSMP), Singapore Sequencing Indian Project (SSIP), China Metabolic Analytics Project (ChinaMAP) and Japanese genome variations (TogoVar)

### *Statistical analysis*

Significantly distinct blood group, neutrophil and platelet antigen alleles specific for the Indian population were identified using Fisher's exact test with a P-value < 0.05. The filtered alleles/variants were subsequently checked for their clinical relevance in transfusion settings.

## **Results**

### *Human blood group associated variants and annotations*

A total of 40712 variants spanning 44 blood group genes were filtered from the preliminary level of analysis. Detailed tabulation of the variant counts in each gene is provided in **Supplementary Table 2**. Functional classification of the filtered variants revealed a total of 1197 variants spanning the exonic and splice site regions, out of which 695 were found to be non-synonymous 398 were synonymous, 1 were stop loss and 14 were stop gain variants as shown in **Supplementary Figure 1(A and B)**.

### *Overview of blood group variants in India*

A total of 29, 17 and 4 variants mapping back to ABO, RHD and RHCE genes were found to possess known blood phenotypes. In addition, 68 variants were identified spanning 27 other blood genes corresponding to 25 systems. Comprehensive summary of the list of identified blood variants is listed in **Table 1**. Out of the 44 blood group systems and 2 erythroid specific transcription factors 22 systems had no variant associated with blood group phenotype and therefore they were conferred with the reference nomenclature. Identified blood group phenotypes for all the systems are tabulated in **Table 2**.

### *Overview of blood group phenotype*

RH blood group is one of the highly polymorphic and severely immunogenic systems known till date. Although D antigen can be easily detected in most cases, variability of expression leads to weak expression of D antigens which are weakly immunogenic and are often missed during routine serological screening. Interestingly, we identified a rare partial D antigen, RHD\*DFV (RHD\*08.01, RHD\*667G) which is well known to be associated with anti-D formation in carriers (Chen & Flegel, 2005; Fichou et al., 2018; Flegel et al., 2009; Grootkerk-Tax et al., 2006; Kappler-Gratias et al., 2014; Noizat-Pirenne et al., 2002). This RH phenotype carries a single nucleotide variation 667T>G which leads to amino acid change at position 223 (p.Phe223Val). It is reported that the amino acid site 223, which is located on the seventh transmembrane domain of the RHD protein plays an integral role in the formation of RHD epitopes. Evidence shows that change of amino acid at site 223 leads to the loss of 2 RHD epitopes (1.3 and 8.1) which subsequently contributes to the formation of alloanti-D antibodies if transfused with D+ RBCs (Grootkerk-Tax et al., 2006). Another RH

weak antigen, RHD\*DFV (RHD\*18, RHD\*497C) which results in moderate reduction of D antigen densities was also identified. Carriers of this antigen were found to have no anti-D antibody formation {ref}.

#### *Analysis of MNS blood group phenotype*

The MNS blood group system is the second most complex after the RH system. MNS system consists of 49 antigens of which M/N, S/s/U are highly prevalent (Reid, 2009). It is well known that allo-anti U is one of the clinically significant antibodies associated with moderate to severe hemolytic reactions and HDFNs. {ref} A recent report from Francis LC et al., states that GYPB:c.251G>C change (p.Ser84Thr) is associated with an U-like allo antibody. This mutation which causes a change in amino acid site 84, located on the transmembrane region is found to alter the U/GPB protein (“27th Regional Congress of the ISBT, Copenhagen, Denmark, June 17-21, 2017,” 2017; Peyrard et al., 2012). MNS:4,5 (s+, U+) altered U/GPB phenotype is found in 10.11% of the Indian population. However, the clinical significance of this U-like alloantibody is still not clear.

#### *Potentially novel and rare blood group antigens*

On comparing the allele frequencies of 695 exonic, non-synonymous blood variants identified in our study with global population datasets (1000 Genomes data) we found that 197 variants were found uniquely in the Indian population. About 315 and 134 variants were found at >2x and <2x frequencies respectively when compared to overall 1000 genomes population frequencies. And, 49 variants were found to be in a similar frequency range. Further, frequency distributions were also compared with 5 major subpopulations in 1000 genomes dataset namely Africans, Americans, Europeans, East Asians and South Asians. Although, as expected, variant frequencies in the Indian population dataset shared a high level of similarity with the South Asians, we were also able to fetch variants exclusive to our study dataset. A scatter plot depicting the distribution of variant frequencies across various global populations is depicted in **Figure 2**.

Of the total variants, 143 exonic, non-synonymous variants which lacked dbSNP ID and literature evidence. These variants were found to span 37 unique blood genes and lacked any reported global population frequencies till date. Of these total variants, 24 variants belonging to 12 blood groups (LANGEREIS, GLOB, JUNIOR, CH/RG, COLTON, KELL, IN, DOMBROCK, PEL, DIEGO, BOMBAY and LEWIS) were predicted to be deleterious by more than three computational tools. Overview of the distribution of novel variants across various blood group genes is shown in **Supplementary Figure 1(C)**. Interestingly, we observed a missense variant in FUT1 gene (p.Arg221Gly) which was predicted to have deleterious effects. An alternate variant reported at the same position (p.Arg221Cys) is well known to be associated with weak H phenotype (FUT1\*01W.11) which thereby makes the identified FUT1:p.Arg221Gly a potentially novel weak H allele.

In addition, we also filtered a total of 588 exonic, non-synonymous variants with MAF<1% mapping back to 58 unique blood genes.

#### *Identification of weak/partial and null blood group antigens*

A total of 25 variants associated with weak/partial and null blood phenotypes were identified in the study as listed in **Table 3**.

Anti-Jra antibodies of the Junior blood group system are known to cause transfusion reactions or hemolytic disease of the fetus or newborn. Interestingly, we identified a weak Jr antigen (ABCG2\*01W.01 - c.421C>A; p.Gln141Lys), which is previously reported to diminish the expression of Jra antigen on the RBC membrane thereby leading to misinterpretation during serological testing {ref}.

The Kidd blood group poses several challenging transfusion scenarios. In addition to the typical acute hemolytic transfusion reactions associated with clinically relevant blood group antigens, weak Jk antigens are particularly notorious for triggering delayed hemolytic transfusion reactions (Lawicki et al., 2017; S.-S. Wang et al., 2021). Our study identifies 2 weak Jk antigens (JK\*01W.04, JK\*02W.04) and a null antigen (JK\*01N.04).

In Langereis blood group system, 3 weak (ABCB6\*01W.03, ABCB6\*01W.02, ABCB6\*01W.01) and 1 null antigen (ABCB6\*01N.13) variant was identified. Lan is one of the high prevalence antigens found in almost all global populations. Anti-Lan antibodies are well known to cause delayed hemolytic transfusion reactions in adults as well as hemolytic disease in fetuses and newborns with varied levels of clinical significance (mild-severe).

#### *Rare blood phenotypes that lacks high prevalence antigens*

High prevalence or high frequency blood antigens are those that occur in more than 99% of the population. Owing to their rarity they become difficult to identify using routine serological screenings. Patients lacking a high prevalence antigen possess a greater risk of alloimmunization while transfused. In our study dataset, 2 distinct rare blood group phenotypes lacking a high incidence of antigens were identified.

In the Diego blood group system, there exist three major high prevalence antigens - Dib, Wrb, and DISK. The rare Di(a+b-) phenotype was observed in 0.1% in the Indian populations. The Yt blood group system consists of 2 antithetical antigens - high-prevalence Yta and the significantly lower prevalence Ytb antigen (Van Buren et al., 2022). Upon analysis we found that the rare Yt(a-b+) phenotype was found in 0.29% of the population. Individuals possessing these phenotypes require antigen negative blood to avoid hemolytic transfusion reactions (HTRs). KANNO1 is a high prevalence antigen belonging to the recently categorized KANNO blood group system (Ohto et al., 2022). Lack of this high prevalence antigen has been previously reported among the East and South Asians. This rare KANNO-



phenotype was found in 0.1% of the study dataset. To this day, there have been no documented instances of hemolytic transfusion reaction or hemolytic disease of the fetus and newborn resulting from anti-KANNO1 antibodies.

### *Comparison of blood group profiles between India and global populations*

Allele frequencies of all the blood variants filtered from the study were compiled and systematically compared with global population datasets. Initially, the distribution of frequencies of the filtered variants within the subpopulations of the 1000 Genomes and gnomAD datasets was examined. As shown in **Figure 3 and 4**, although the frequencies of a handful of the variants followed a similar range between Indian and Asian datasets, differences in overall distribution was also observed. Allele frequencies observed across other population datasets are tabulated in **Table 4**. Complete list of frequencies observed in the subpopulations is tabulated in **Supplementary Tables 3a and 3b**.

Differences in the distribution of variant frequencies between India and a range of global populations namely Americans (gnomAD\_AMR), Europeans (gnomAD\_FIN, gnomAD\_NFE), Africans (gnomAD\_AFR, Gambian genomes), Middle Easterners (Qatar, GME variome, gnomAD\_MID), Asians (GenomeAsia100K), East Asians (gnomAD\_EAS, China, Japan), South East Asians (SSMP, SSIP) and South Asians (gnomAD\_SAS) was analysed by performing Fisher's exact test. Variants with p-value < 0.05 were considered significantly distinct.

One variant belonging to the Duffy blood group system *DUFFY:1:159205564G>A* responsible for encoding the Fyb antigen was found significantly different in almost all global populations. Duffy blood group gene, *ACKR1* is highly polymorphic with allele frequencies greatly varying among different populations and ethnic groups (Höher et al., 2018; Howes et al., 2011; Shimizu et al., 2000). In compliance with previous reports, our analysis revealed a high prevalence of Fyb allele frequencies in European (60%), American (54%) and Middle Eastern (66%) populations. On the contrary, East Asians exhibited a very low frequency (7.7%) of Fyb allele as previously reported. Frequency of this allele in India (32%) was found comparable to other South Asian populations (gnomAD\_SAS : 38%, 1KG\_SAS : 35%)

A total of 26 variants belonging to 13 different blood group systems were found significantly distinct in India when compared to other South Asian populations. 2 weak RHD variants were found significantly higher in Indians. As mentioned earlier, these two variants *RHD\_NM\_016124.6:c.520G>A* and *RHD\_NM\_016124.6:c.667T>G* encoding weak and partial D antigens namely RHD Type 33 and RHD\*DFV respectively are reported to form allo anti-D antibodies in carriers if transfused with D+ RBCs.

Further, on comparing the differences with other large populations based on their geographical locations including Africans, Americans, Europeans, Asians, South East Malays, East Asians, Middle East, Amish and Ashkenazi Jews we found that 37, 77, 77, 18, 26, 54, 23,

58 and 75 variants were found significantly distinct respectively. Complete list of variants that possess significantly different frequencies when compared with each global population are provided in **Table 4**. Schematic representation of the distribution of allele frequencies of significantly distinct variants is shown in **Figure 5**.

## **Discussion**

Genomic analysis of 1029 self declared healthy Indian individuals revealed a total of 695 non-synonymous blood group associated variants mapping back to exonic and splice site regions. Preliminary level of analysis uncovered a subset of these variants which were found 2x enriched and depleted in the Indian population when compared with the 1000 genomes dataset. In addition, a range of weak, partial, null and potentially novel blood variants were also identified.

### *Clinical relevance*

In India, on analysing the prevalence and incidence of the diseases that systematically require transfusion, it was found that people with bone marrow failure (73%), leukemia (60.7%), end stage renal diseases (51.6) and blood disorders like hemolytic anemia (50.2) and hemophilia (25.6) were in higher need of blood units compared to others. In addition to specific disease conditions , trauma/surgery and obstetrics and gynecology are the two major fields which have huge demand for blood components. Recent estimates state that 6.6 and 3.6 million units of blood are needed to cater the surgical and obstetrics needs of the Indian population .

In December, 2012, The Hemovigilance Program of India (HvPI) was launched with the purpose of monitoring and preventing transfusion related adverse reactions (Bisht et al., 2020). A comprehensive 5-year study conducted by HvPI states that a total of 8162 cases of transfusion adversities were reported during the period. It is also found that 44% of all the complications caused by antigen mismatches were attributed to immunological hemolysis caused by other non-ABO alloantibodies. Such findings strongly emphasize the need to devise accurate ways of performing extensive blood group profiling which would in turn help in improving population specific donor registries in blood bank networks.

## **Conclusion**

Enhanced accuracy in blood group typing greatly reduces the risk of transfusion associated complications thereby leading to increased transfusion success rates. Our study serves as the first of its kind to provide a comprehensive compilation of the genetic landscape of blood group alleles and antigens in the Indian population.

## **Data availability**

The Indian genetic variation dataset used in the study can be accessed at <https://clingen.igib.res.in/indigen/>

## Funding

This work was supported by The Council of Scientific and Industrial Research, India (Grant : MLP2001/GenomeApp).

## Author Contributions

VS conceived and designed the project. MR and VS contributed in writing the manuscript. All authors approved the final manuscript. Authors acknowledge funding from CSIR India and Ministry of Higher Education, Malaysia. The funders had no role in the preparation of the manuscript or decision to publish.

## Conflict of interest

None declared.

## Tables and Figures

**Figure 1.** Schematic representation of methodology followed in fetching human blood groups associated genetic variations in the Indian population.

**Figure 2.** Scatter plot depicting the similarities and differences in the distribution of variant frequencies between Indian and various global populations.

**Figure 3.** Heatmaps depicting the distribution of frequencies of the blood variants filtered in the study. **(A)** Frequency distribution of the 50 variants belonging to ABO and RH blood groups among various subpopulations of the 1000 Genomes dataset. **(B)** Frequency distribution of the 50 variants belonging to ABO and RH blood groups among various subpopulations of the gnomAD dataset.

**Figure 4.** Heatmaps depicting the distribution of frequencies of the blood variants filtered in the study. **(A)** Frequency distribution of the 68 variants belonging to minor blood groups among various subpopulations of the 1000 Genomes dataset. **(B)** Frequency distribution of the 68 variants belonging to minor blood groups among various subpopulations of the gnomAD dataset.

**Figure 5.** Bubble plot illustrating the schematic overview of the distribution of 56 variants across all blood group systems whose frequencies were found significantly distinct ( $p$  val < 0.05) in India in comparison to any of the other global populations.

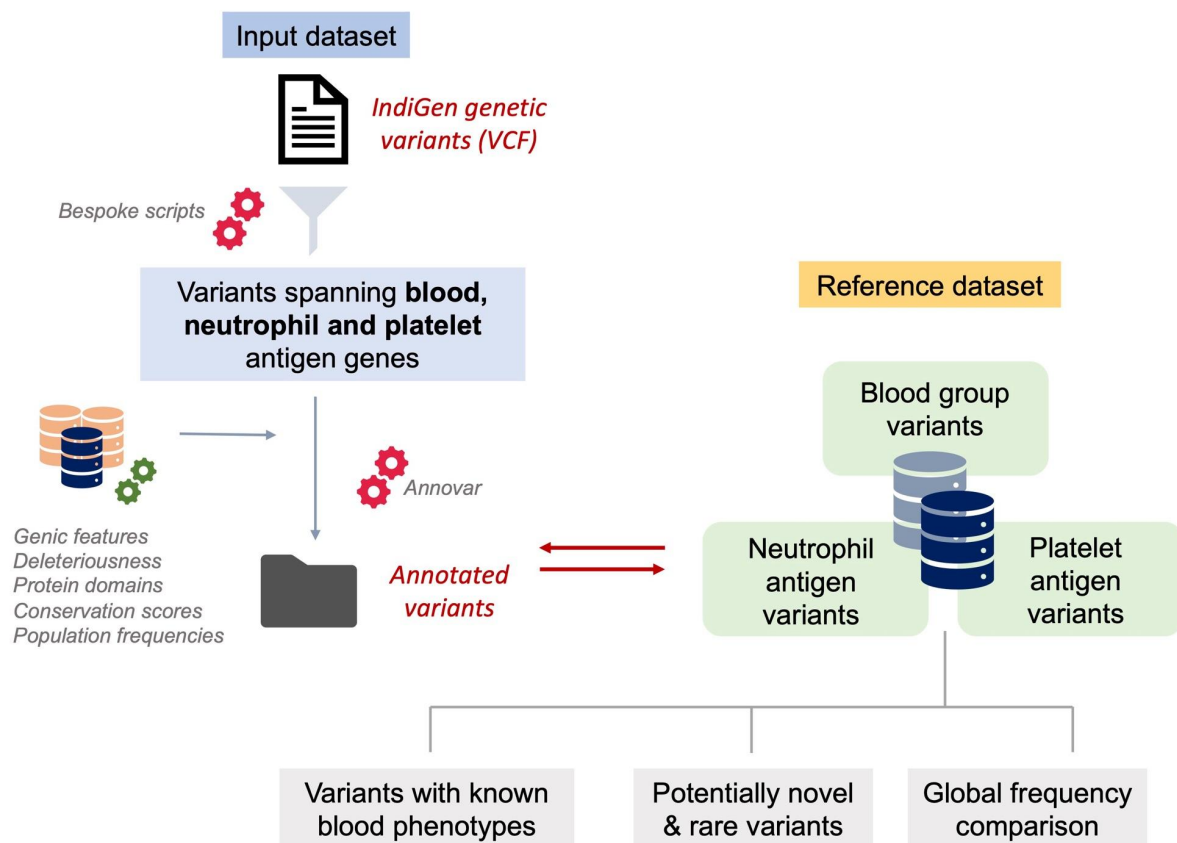
**Table 1.** Summary of blood group variants with known blood phenotypes identified from the study dataset.

**Table 2.** Tabulation of the blood phenotypes of all the 44 blood group systems identified in the Indian population.

**Table 3.** Compilation of weak, partial and null antigen frequency distribution among studies populations.

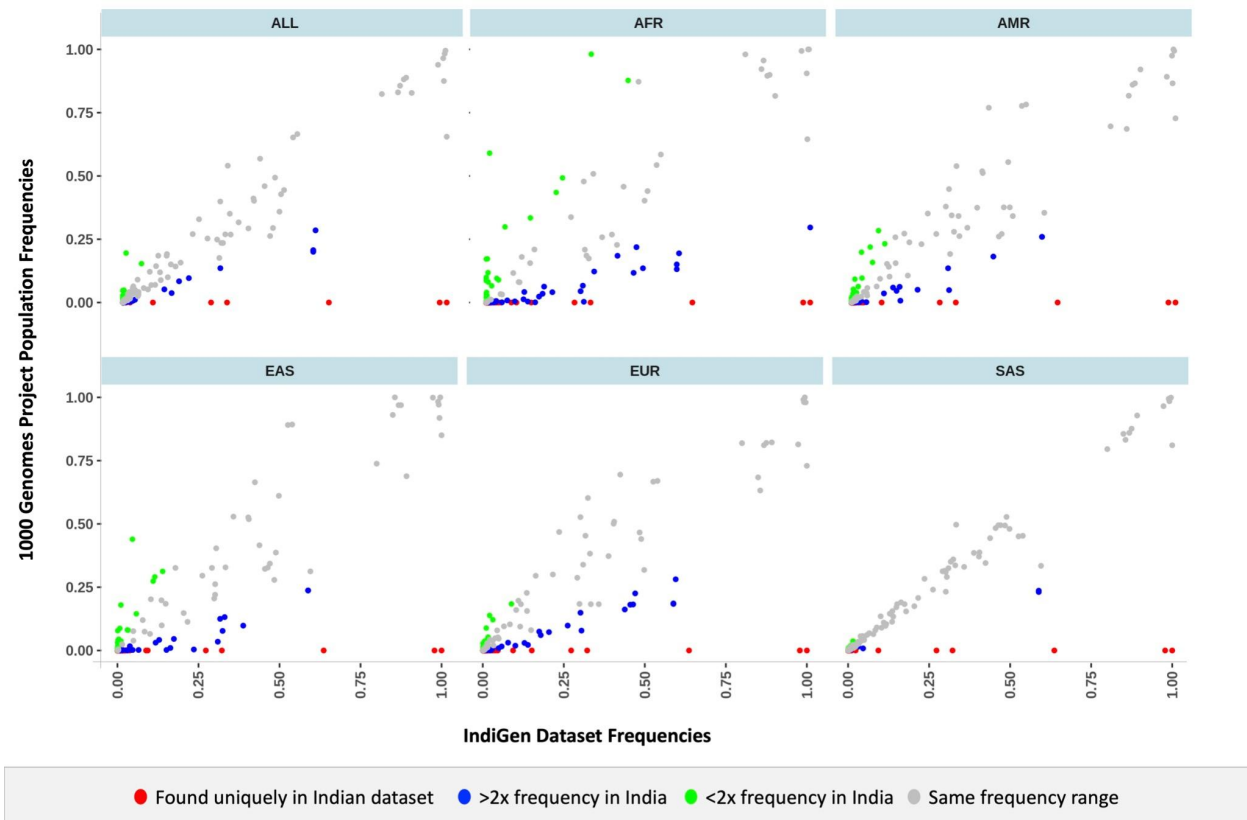
**Table 4.** Overall distribution of blood variant frequencies across various global population datasets used in the study.

**Table 5.** Tabulation of blood group variants which were found significantly distinct in the global populations along with their p-values.

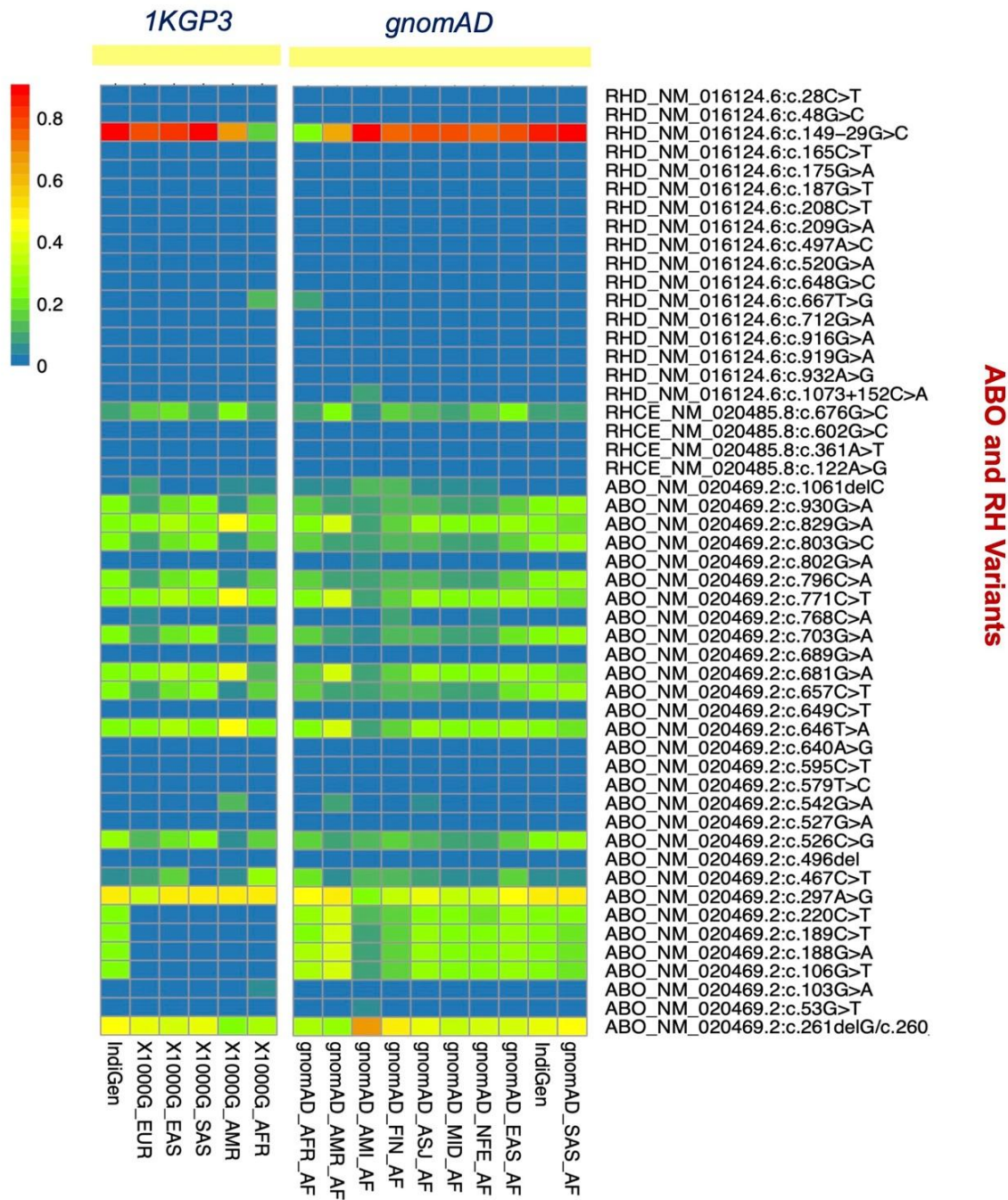


**Figure 1.** Schematic representation of methodology followed in fetching human blood groups associated genetic variations in the Indian population.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

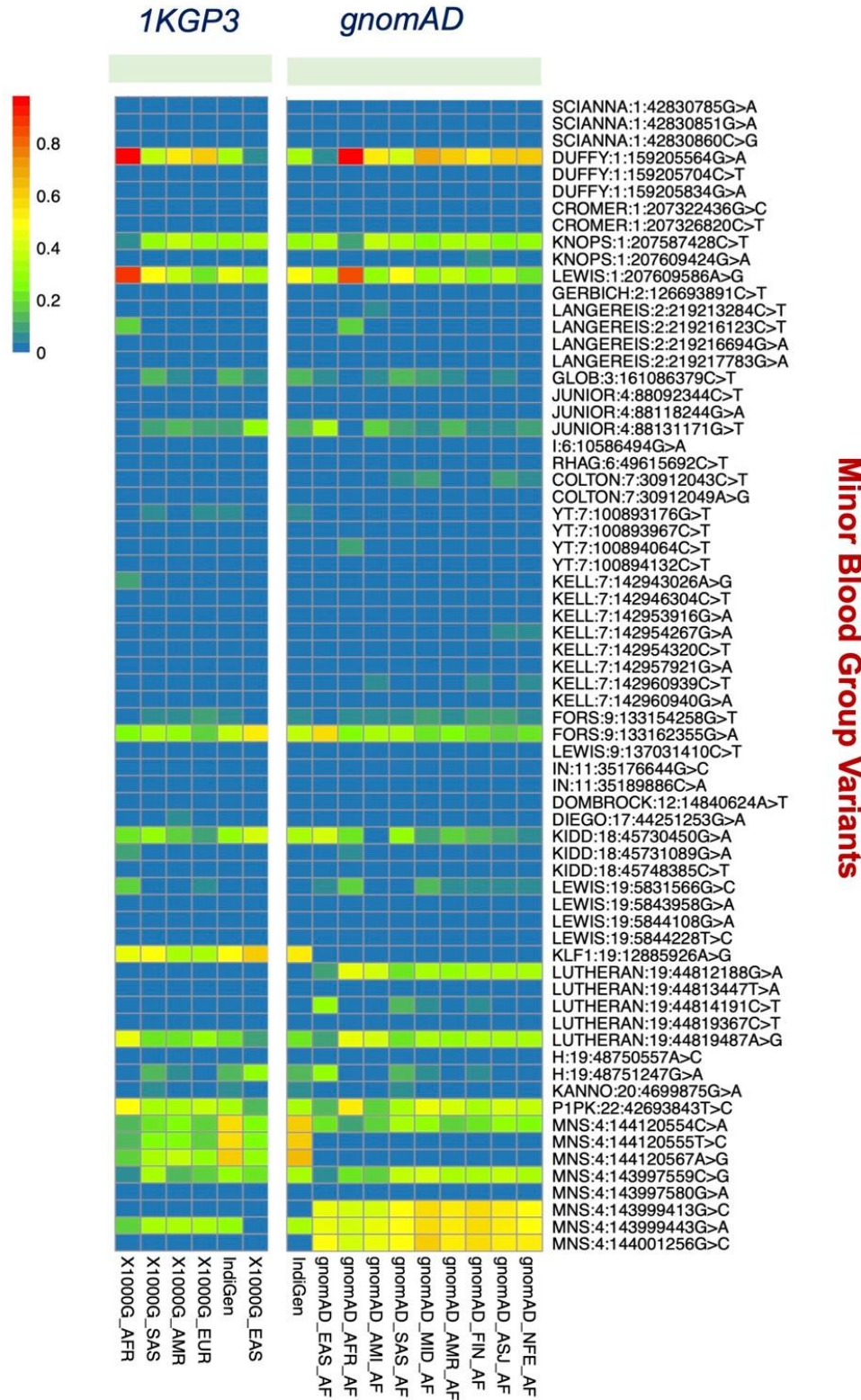


**Figure 2.** Scatter plot depicting the similarities and differences in the distribution of variant frequencies between Indian and various global populations.



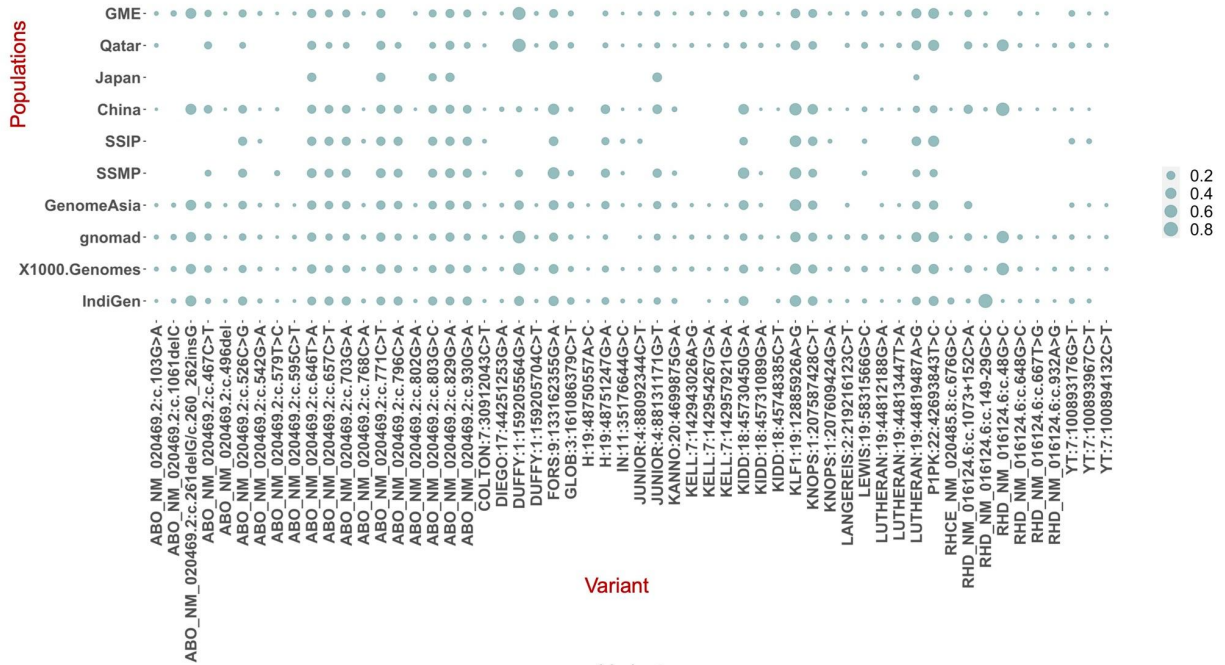
**Figure 3.** Heatmaps depicting the distribution of frequencies of the blood variants filtered in the study. **(A)** Frequency distribution of the 50 variants belonging to ABO and RH blood groups among various subpopulations of the 1000 Genomes dataset. **(B)** Frequency distribution of the 50 variants belonging to ABO and RH blood groups among various subpopulations of the gnomAD dataset.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

**Figure 4.** Heatmaps depicting the distribution of frequencies of the blood variants filtered in the study. **(A)** Frequency distribution of the 68 variants belonging to minor blood groups among various subpopulations of the 1000 Genomes dataset. **(B)** Frequency distribution of the 68 variants belonging to minor blood groups among various subpopulations of the gnomAD dataset.



**Figure 5.** Bubble plot illustrating the schematic overview of the distribution of 56 variants across all blood group systems whose frequencies were found significantly distinct ( $p$  val < 0.05) in India in comparison to any of the other global populations.

Chr	Position	Ref Allele	Alt Allele	Allele Count	Allele Frequency	Allele Number	Blood Group	Gene Name	Blood Phenotype
7	30912043	C	T	9	0.004	2052	COLTON	AQP1	CO:2 or Co(b+)
7	30912049	A	G	1	0	2052	COLTON	AQP1	CO:1;2;3;4
1	207322436	G	C	1	0	2048	CROMER	CD55	CROM:3 or Tc(b+)
1	207326820	C	T	4	0.002	2050	CROMER	CD55	CROM:-12 or SERF-
17	44251253	G	A	12	0.006	2050	DIEGO	SLC4A1	DI:1;-2 or Di(a+b-)



12	14840624	A	T	3	0.001	2048	DOMBROC K	ART4	DO:-8 or DOLG-
1	159205564	G	A	665	0.324	2050	DUFFY	ACKR1	FY:2 or Fy(b+)
1	159205704	C	T	12	0.006	2050	DUFFY	ACKR1	Fy(a+w)
1	159205834	G	A	2	0.001	2052	DUFFY	ACKR1	Fy(a-b-)
9	133154258	G	T	92	0.045	2052	FORS	GBGT1	FORS:-1 (FORS-)
9	133162355	G	A	718	0.358	2006	FORS	GBGT1	FORS:-1 (FORS-)
2	126693891	C	T	2	0.001	2052	GERBICH	GYPC	GE:-10 or GEPL-
3	161086379	C	T	263	0.129	2044	GLOB	B3GALN T1	GLOB:1 (P+)
19	48750557	A	C	7	0.003	2052	H	FUT1	H-
19	48751247	G	A	286	0.139	2052	H	FUT1	H+
6	10586494	G	A	5	0.002	2048	I	GCNT2	I+W
11	35176644	G	C	28	0.014	2050	IN	CD44	In(a+b-)
11	35189886	C	A	1	0	2050	IN	CD45	IN:-4 or INJA-
4	88092344	C	T	32	0.016	2046	JUNIOR	ABCG2	Jr(a+w)
4	88118244	G	A	1	0	2052	JUNIOR	ABCG2	Jr(a-)
4	88131171	G	T	236	0.115	2046	JUNIOR	ABCG2	Jr(a+w)
20	4699875	G	A	97	0.047	2054	KANNO	PRNP	KANNO-
7	142943026	A	G	1	0	2052	KELL	KEL	KEL:6,-7 or Js(a+b-)
7	142946304	C	T	1	0	2048	KELL	KEL	KEL:-26 or TOU-
7	142953916	G	A	3	0.001	2050	KELL	KEL	KEL:-22
7	142954267	G	A	2	0.001	2052	KELL	KEL	KEL:3,-4,-21 or Kp(a+b-c-)
7	142954320	C	T	1	0	2050	KELL	KEL	Kmod
7	142957921	G	A	14	0.007	2052	KELL	KEL	KEL:1,-2 or K+k-
7	142960939	C	T	4	0.002	2046	KELL	KEL	KEL:-18
7	142960940	G	A	5	0.002	2046	KELL	KEL	KEL:-18
18	45730450	G	A	625	0.305	2052	KIDD	SLC14A1	Jk(a+W)
18	45731089	G	A	1	0	2048	KIDD	SLC14A1	Jk(a+W)
18	45748385	C	T	18	0.009	2050	KIDD	SLC14A1	JK:-3 or Jk(a-b-)
19	12885926	A	G	1016	0.498	2040	KLF1	KLF1	Obsolete Normal BG phenotype
1	207587428	C	T	597	0.292	2048	KNOPS	CR1	KN:5 or Yk(a-)

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

1	207609424	G	A	10	0.005	2050	KNOPS	CR1	KN:2 or Kn(b+)
2	219213284	C	T	1	0	2050	LANGEREIS	ABCB6	Lan weak
2	219216123	C	T	8	0.004	2054	LANGEREIS	ABCB6	Lan weak
2	219216694	G	A	5	0.002	2048	LANGEREIS	ABCB6	Lan weak
2	219217783	G	A	2	0.001	2046	LANGEREIS	ABCB6	Lan-
19	44812188	G	A	2	0.001	2048	LUTHERAN	BCAM	LU:1 or Lu(a+)
19	44813447	T	A	2	0.001	2048	LUTHERAN	BCAM	LU:-8,14
19	44814191	C	T	7	0.003	2052	LUTHERAN	BCAM	LU:-6,9
19	44819367	C	T	5	0.002	2046	LUTHERAN	BCAM	LU:-26, LUBI-
19	44819487	A	G	442	0.216	2044	LUTHERAN	BCAM	LU:-18,19 or Au(a-b+)
22	42693843	T	C	674	0.331	2038	P1PK	A4GALT	P1+/-; Pk+
6	49615692	C	T	1	0	2048	RHAG	RHAG	Rhmod
1	42830785	G	A	1	0	2054	SCIANNA	ERMAP	SC:-7 or SCAN-
1	42830851	G	A	1	0	2050	SCIANNA	ERMAP	SC:2 or Sc2+
1	42830860	C	G	1	0	2050	SCIANNA	ERMAP	SC:4 or Rd+
7	100893176	G	T	96	0.047	2050	YT	ACHE (YT)	YT:2 or Yt(b+)
7	100893967	C	T	32	0.016	2048	YT	ACHE (YT)	YTEG-
7	100894064	C	T	12	0.006	2044	YT	ACHE (YT)	YT:-4 or YTLI-
7	100894132	C	T	1	0	2038	YT	ACHE (YT)	YT:-5 or YTOT
1	207609586	A	G	966	0.47	2054			
9	137031410	C	T	2	0.001	2040			
19	5831566	G	C	43	0.021	2044	LEWIS		
19	5843958	G	A	5	0.002	2028	LEWIS		
19	5844108	G	A	6	0.003	2026	LEWIS		
19	5844228	T	C	14	0.007	2040	LEWIS		
1	25390874	C	G	211	0.103	2046	RH	RHCE	RH:27 or cE RHCE*03 or RHCE*cE
1	25392026	C	G	1	0	2048	RH	RHCE	RH:3 (E+ partial, weak to neg) or RH:4 (c+ weak) RHCE*03.04

									RHCE*cE.04 RHCE*cEIV
1	25402721	T	A	4	0.002	2046	RH	RHCE	Comb Mut
1	25420665	T	C	10	0.005	2036	RH	RHCE	RH:2 (C+ partial) RH:5 (e+ partial) RH:8 (C W +) RH:-51(MAR-) RHCE*O2.08.01 RHCE*Ce.08.01 RHCE*CeCW
1	25272575	C	T	1	0.001	1916	RH	RHD	Type 61 RHD*01W.61 RHD*weak D type61
1	25272595	G	C	4	0.002	1910	RH	RHD	Normal D antigen RHD*01.01
1	25284544	G	C	1680	0.867	1938	RH	RHD	Del RHD*01EL.32 RHD*DEL32
1	25284589	C	T	1	0.001	1934	RH	RHD	Comb Mut
1	25284599	G	A	1	0.001	1934	RH	RHD	Type 151 RHD* 01W.151 RHD*weak D type 151
1	25284611	G	T	2	0.001	1934	RH	RHD	Type 152 RHD* 01W.152 RHD*weak D type 152
1	25284632	C	T	1	0.001	1930	RH	RHD	Type 122 RHD* 01W.122 RHD*weak D type 122
1	25284633	G	A	1	0.001	1930	RH	RHD	Type 85 RHD* 01W.85 RHD*weak D type 85
1	25300956	A	C	2	0.001	1906	RH	RHD	DFW RHD*18 RHD*DFW
1	25300979	G	A	5	0.003	1908	RH	RHD	Type 33 RHD*01W.33 RHD*weak D type 33
1	25301533	G	C	16	0.008	1904	RH	RHD	Type 154 RHD* 01W.154 RHD*weak D type 154
1	25301552	T	G	11	0.006	1906	RH	RHD	DFV RHD*08.01

									RHD*DFV
1	25301597	G	A	2	0.001	1908	RH	RHD	Comb Mut
1	25303436	G	A	1	0.001	1910	RH	RHD	Type 66 RHD*01W.66 RHD*weak D type 66
1	25303439	G	A	1	0.001	1910	RH	RHD	Type 8 RHD*01W.8 RHD*weak D type 8 6
1	25303452	A	G	1	0.001	1908	RH	RHD	Comb Mut
1	25306881	C	A	31	0.016	1912	RH	RHD	Comb Mut
9	133255669	CG	C	64	0.031	2040	ABO	ABO	B weak ABO*BW.18
9	133255801	C	T	507	0.248	2048	ABO	ABO	Comb Mut
9	133255902	C	T	462	0.226	2048	ABO	ABO	Comb Mut
9	133255928	C	G	503	0.245	2050	ABO	ABO	Comb Mut
9	133255929	C	T	11	0.005	2050	ABO	ABO	Comb Mut
9	133255935	G	T	502	0.245	2050	ABO	ABO	Comb Mut
9	133255960	G	A	458	0.224	2048	ABO	ABO	Comb Mut
9	133255963	G	T	35	0.017	2048	ABO	ABO	Comb Mut
9	133256028	C	T	506	0.247	2052	ABO	ABO	Comb Mut
9	133256042	C	T	1	0	2052	ABO	ABO	A x /A weak ABO*AW.30.01 or B 3 ABO*B3.02
9	133256050	C	T	463	0.226	2052	ABO	ABO	Comb Mut
9	133256074	G	A	504	0.246	2050	ABO	ABO	Comb Mut
9	133256082	G	A	1	0	2050	ABO	ABO	Comb Mut
9	133256085	A	T	462	0.225	2050	ABO	ABO	O ABO*0.06 053 (O 6)
9	133256091	T	C	1	0	2048	ABO	ABO	Comb Mut
9	133256136	G	A	3	0.001	2048	ABO	ABO	Comb Mut
9	133256152	A	G	4	0.002	2048	ABO	ABO	Comb Mut
9	133256189	C	T	51	0.025	2048	ABO	ABO	A 1 ABO*A1.02
9	133256204	C	T	8	0.004	2048	ABO	ABO	Comb Mut
9	133256205	G	C	514	0.251	2048	ABO	ABO	Comb Mut
9	133256234	GT	G	19	0.009	2048	ABO	ABO	Comb Mut
9	133256264	G	A	108	0.053	2046	ABO	ABO	Comb Mut

9	133257486	T	C	973	0.475	2048	ABO	ABO	Comb Mut
9	133258116	G	A	467	0.228	2044	ABO	ABO	Comb Mut
9	133259833	G	A	457	0.222	2054	ABO	ABO	B weak ABO*BW.26
9	133259834	C	T	458	0.223	2054	ABO	ABO	O ABO*O.01.01 O01 (O 1)
9	133261367	C	A	455	0.222	2046	ABO	ABO	Comb Mut
9	133261370	C	T	4	0.002	2046	ABO	ABO	Comb Mut
9	133262144	C	A	12	0.006	2050	ABO	ABO	Comb Mut
9	133257521	T	TC	886	0.4322	2050	ABO	ABO	Comb Mut
4	144120554	C	A	1187	0.5882	2018	MNS	GYP A	Comb Mut
4	144120555	T	C	1187	0.5882	2018	MNS	GYP A	Comb Mut
4	144120567	A	G	1206	0.5953	2026	MNS	GYP A	Comb Mut
4	143997559	C	G	610	0.2984	2044	MNS	GYP B	MNS:4,5 (s+, U+) altered U/GPB
4	143997580	G	A	3	0.0015	2046	MNS	GYP B	Comb Mut
4	143999413	G	C	1	0.0005	2044	MNS	GYP B	Comb Mut
4	143999443	G	A	634	0.3096	2048	MNS	GYP B	MNS:3 or S+
4	144001256	G	C	1	0.0005	2046	MNS	GYP B	MNS:21or Mv+

**Table 1.** Summary of blood group variants with known blood phenotypes identified from the study dataset.

Blood Group System	BG System Number	BG System Symbol	Phenotypes	Proportion (% N=1029)
ABO	001	ABO	A1	56.46 (581)
ABO	001	ABO	A1B	22.35 (230)
ABO	001	ABO	A1A2	2.24 (23)
ABO	001	ABO	B	14.38 (148)
ABO	001	ABO	O	4.57 (47)
MNS	002	MNS	M-N+	35.96 (370)
MNS	002	MNS	M+N+	43.44 (447)
MNS	002	MNS	M+N-	20.60 (212)
MNS	002	MNS	S+s-	10.79 (111)
MNS	002	MNS	S+s+	40.04 (412)
MNS	002	MNS	S-s+	49.17 (506)
MNS	002	MNS	s+U+	10.11 (104)
P1PK	003	P1PK	P1+/-; Pk+	100 (1029)
Rh	004	RH	C-c+e-E+	0.87 (9)

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

Rh	004	RH	C-c+e+E+	18.46 (190)
Rh	004	RH	C+c+e+E+	0.19 (2)
Rh	004	RH	C-c+e+Ew+	0.1 (1)
Rh	004	RH	C-c+E-e+	80.37 (827)
Rh	004	RH	Del RHD*01EL.32	73.95 (761)
Rh	004	RH	Type 33 RHD*01W.33	0.1 (1)
Rh	004	RH	Type 154 RHD*01W.154	0.19 (2)
Rh	004	RH	DFV RHD*08.01	0.1 (1)
Rh	004	RH	D+	25.66 (264)
Lutheran	005	LU	Lu(a-b+)	99.81 (1027)
Lutheran	005	LU	Lu(a+b+)	0.19 (2)
Lutheran	005	LU	Au(a+b+)	33.43 (344)
Lutheran	005	LU	Au(a-b+)	4.76 (49)
Kell	006	KEL	k+K-	98.64 (1015)
Kell	006	KEL	k+K+	1.36 (14)
Lewis	007	LE	Le(a-b+)	100 (1029)
Duffy	008	FY	Fy(a+b-)	46.55 (479)
Duffy	008	FY	Fy(a+b+)	42.27 (435)
Duffy	008	FY	Fy(a-b+)	11.18 (115)
Kidd	009	JK	Jk(a+b-)	90.96 (936)
Kidd	009	JK	Jk(aw+b-)	9.04 (93)
Diego	010	DI	Di(a-b+)	98.93 (1018)
Diego	010	DI	Di(a+b+)	0.97 (10)
Diego	010	DI	Di(a+b-)	0.1 (1)
Yt	011	YT	Yt(a+b-)	90.96 (936)
Yt	011	YT	Yt(a+b+)	8.75 (90)
Yt	011	YT	Yt(a-b+)	0.29 (3)
Yt	011	YT	YTEG-	0.1 (1)
Xg	012	XG	Xga	100 (1029)
Scianna	013	SC	Sc1+	100 (1029)
Dombrock	014	DO	Do(a-b+)	100 (1029)
Colton	015	CO	Co(a+b-)	99.13 (1020)
Colton	015	CO	Co(a+b+)	0.87 (9)
Landsteiner- Wiener	016	LW	Lw(a+b-)	100 (1029)
Chido/Rodgers	017	CH/RG	Ch+Rg-	100 (1029)
H	018	H	H+	100 (1029)
Kx	019	XK	Kx+	100 (1029)











chr2	2192	3	C	T	LAN	B6	Lan weak	ABCB	0.0	0.04	0.04	0.01	0.00	0.00	0.00	0.00	0.02	0.00
	1612				GER			ABC	6*01									
					EIS			W.02	0	89	40	20	00	00	02	00	01	86
chr2	2192	94	G	A	LAN	B6	Lan weak	ABCB	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	166				GER			ABC	6*01									
					EIS			W.01	0	38	94	26	00	00	01	00	06	41
chr2	2192	3	G	A	LAN	B6	Lan-	ABCB	0.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1778				GER			ABC	6*01N									
					EIS			.13	0	06	16	03	00	00	00	00	00	15

**Table 3.** Compilation of weak, partial and null antigen frequency distribution among studies populations.

ABO-RH Variants										
Variant	IndiG en	gnom ad	Genome Asia	SSMP	SSIP	China	Japan	Qatar	GME	Gambia
RHCE_NM_02048 5.8:c.122A>G	0.00 500	0.00 002	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 0	0.0000 0	0.000000
RHD_NM_016124. 6:c.28C>T	0.001 00	0.00 048	0.0000 0	0.0000 0	0.0000 0	0.00080 0	0.0000 0	0.00000 0.00171	0.00171	0.000000
RHD_NM_016124. 6:c.48G>C	0.002 00	0.60 035	0.0000 0	0.88542	0.0000 0	0.75194 0	0.0000 0	0.54940 0	0.0000 0	0.180000
RHD_NM_016124. 6:c.149-29G>C	0.867 00	0.00 066	0.0000 0	0.0000 0	0.0000 0	0.00857 0	0.0000 0	0.01330 3	0.0075	0.000000
RHD_NM_016124. 6:c.165C>T	0.001 00	0.00 010	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 4	0.0005	0.000000
RHD_NM_016124. 6:c.175G>A	0.001 00	0.00 001	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 0	0.0000	0.000000
RHD_NM_016124. 6:c.187G>T	0.001 00	0.00 009	0.0000 0	0.0000 0	0.0000 0	0.00019 0	0.0000 0	0.00000 0	0.0000	0.000000
RHD_NM_016124. 6:c.208C>T	0.001 00	0.00 025	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 0	0.0000	0.000000
RHD_NM_016124. 6:c.209G>A	0.001 00	0.00 000	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 0	0.0000	0.000000
RHD_NM_016124. 6:c.497A>C	0.001 00	0.00 335	0.0003 0	0.0000 0	0.0000 0	0.00208 0	0.0000 0	0.00000 0	0.0000	0.000000
RHD_NM_016124. 6:c.520G>A	0.00 300	0.00 025	0.0000 0	0.0000 0	0.0000 0	0.00000 0	0.0000 0	0.00000 7	0.0010	0.000000

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

RHD_NM_016124. 6:c.648G>C	0.00 800	0.033 63	0.0000 0	0.0000 0	0.0000 0	0.00166	0.0000 0	0.01710	0.0235 5	0.130000
RHD_NM_016124. 6:c.667T>G	0.00 600	0.00 017	0.0000 0	0.0000 0	0.0000 0	0.00047	0.0000 0	0.00130	0.0010 7	0.000000
RHD_NM_016124. 6:c.712G>A	0.001 00	0.001 12	0.0000 0	0.0000 0	0.0000 0	0.00062	0.0024 6	0.00000	0.0005 4	0.010000
RHD_NM_016124. 6:c.916G>A	0.001 00	0.00 000	0.0000 0	0.0000 0	0.0000 0	0.00000	0.0000 0	0.00000	0.0000 0	0.000000
RHD_NM_016124. 6:c.919G>A	0.001 00	0.00 009	0.0000 0	0.0000 0	0.0000 0	0.00000	0.0000 0	0.00000	0.0000 0	0.000000
RHD_NM_016124. 6:c.932A>G	0.001 00	0.012 43	0.0000 0	0.0000 0	0.0000 0	0.00066	0.0000 0	0.01850	0.0000 0	0.040000
RHD_NM_016124. 6:c.1073+152C>A	0.016 00	0.149 44	0.17300	0.12376	0.0000 0	0.25845	0.0000 0	0.17160	0.12991	0.080000
RHCE_NM_02048 5.8:c.676G>C	0.103 00	0.00 037	0.0000 0	0.0000 0	0.0000 0	0.00005	0.0000 0	0.00000	0.0050 4	0.020000
RHCE_NM_02048 5.8:c.602G>C	0.00 000	0.001 12	0.01100	0.0000 0	0.0000 0	0.00699	0.0000 0	0.00060	0.0005 0	0.000000
RHCE_NM_02048 5.8:c.361A>T	0.002 00	0.00 931	0.0000 0	0.0000 0	0.0000 0	0.00014	0.0000 0	0.00410	0.0085 0	0.000000
ABO_NM_020469. 2:c.1061delC	0.031 00	0.062 80	0.0200 0	0.0000 0	0.0000 0	0.00000	0.0000 0	0.00000	0.0621 8	0.050000
ABO_NM_020469. 2:c.930G>A	0.248 00	0.115 10	0.19500	0.23958	0.2500 0	0.20948	0.0000 0	0.08520	0.15222	0.170000
ABO_NM_020469. 2:c.829G>A	0.226 00	0.250 26	0.26100	0.30208	0.2638 9	0.23968	0.2630 0	0.23860	0.2401 8	0.260000
ABO_NM_020469. 2:c.803G>C	0.245 00	0.114 87	0.19700	0.23438	0.2500 0	0.20991	0.17700	0.08520	0.1532 3	0.170000
ABO_NM_020469. 2:c.802G>A	0.00 500	0.015 05	0.00577	0.0000 0	0.0000 0	0.00014	0.0000 0	0.00000	0.01159	0.000000
ABO_NM_020469. 2:c.796C>A	0.245 00	0.114 89	0.19600	0.23438	0.2500 0	0.20977	0.0000 0	0.08520	0.0000 0	0.170000
ABO_NM_020469. 2:c.771C>T	0.224 00	0.250 44	0.26200	0.30729	0.2638 9	0.23933	0.2630 0	0.23860	0.2416 8	0.260000
ABO_NM_020469. 2:c.768C>A	0.017 00	0.027 99	0.0060 6	0.0000 0	0.0138 9	0.00038	0.0000 0	0.00000	0.0055 5	0.000000

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

ABO_NM_020469. 2:c.703G>A	0.247 00	0.115 15	0.19700	0.22917	0.23611	0.20981	0.0000 0	0.08520	0.15691	0.170000
ABO_NM_020469. 2:c.689G>A	0.00 000	0.00 305	0.0008 7	0.0000 0	0.0000 0	0.00009	0.0000 0	0.00000	0.0040 4	0.000000
ABO_NM_020469. 2:c.681G>A	0.226 00	0.223 02	0.25600	0.30729	0.2638 9	0.23947	0.0000 0	0.21590	0.23131	0.170000
ABO_NM_020469. 2:c.657C>T	0.246 00	0.115 38	0.19700	0.23438	0.2500 0	0.20991	0.0000 0	0.08520	0.1587 7	0.170000
ABO_NM_020469. 2:c.649C>T	0.00 000	0.001 12	0.0000 0	0.0000 0	0.0000 0	0.00005	0.0000 0	0.00000	0.0005 0	0.000000
ABO_NM_020469. 2:c.646T>A	0.225 00	0.250 89	0.26100	0.30729	0.2638 9	0.23956	0.26100	0.23860	0.2389 1	0.260000
ABO_NM_020469. 2:c.640A>G	0.00 000	0.00 001	0.0000 0	0.0000 0	0.0000 0	0.00000	0.0000 0	0.00000	0.0000 0	0.000000
ABO_NM_020469. 2:c.595C>T	0.001 00	0.014 25	0.0005 8	0.0000 0	0.0000 0	0.00000	0.0000 0	0.00000	0.0045 4	0.000000
ABO_NM_020469. 2:c.579T>C	0.002 00	0.00 034	0.01100	0.05208	0.0000 0	0.00482	0.0000 0	0.00000	0.0010 1	0.000000
ABO_NM_020469. 2:c.542G>A	0.025 00	0.023 38	0.01200	0.0000 0	0.0138 9	0.00033	0.0000 0	0.00000	0.02117	0.000000
ABO_NM_020469. 2:c.527G>A	0.00 400	0.00 060	0.00202	0.0000 0	0.0000 0	0.00000	0.0000 0	0.01700	0.0060 5	0.000000
ABO_NM_020469. 2:c.526C>G	0.251 00	0.130 50	0.20300	0.23438	0.2500 0	0.20991	0.0000 0	0.08520	0.1628 0	0.170000
ABO_NM_020469. 2:c.496del	0.00 900	0.001 90	0.00462	0.0000 0	0.0000 0	0.00014	0.0000 0	0.00000	0.0050 4	0.000000
ABO_NM_020469. 2:c.467C>T	0.053 00	0.115 98	0.0940 0	0.08854	0.0000 0	0.20415	0.0000 0	0.15910	0.11139	0.260000
ABO_NM_020469. 2:c.297A>G	0.475 00	0.397 86	0.47200	0.54688	0.5138 9	0.45221	0.4540 0	0.35230	0.4058 4	0.500000
ABO_NM_020469. 2:c.261delG/c.260_262insG	0.432 20	0.353 90	0.3950 0	0.0000 0	0.0000 0	0.43365	0.0000 0	0.00000	0.4139 0	0.310000
ABO_NM_020469. 2:c.220C>T	0.228 00	0.239 76	0.0000 0	0.0000 0	0.0000 0	0.24197	0.0000 0	0.00000	0.0000 0	0.280000
ABO_NM_020469. 2:c.189C>T	0.222 00	0.249 20	0.0000 0	0.0000 0	0.0000 0	0.24244	0.0000 0	0.00000	0.0000 0	0.240000

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

ABO_NM_020469. 2:c.188G>A	0.223 00	0.269 79	0.0000 0	0.0000 0	0.0000 0	0.24244	0.0000 0	0.00000 0	0.0000 0	0.320000
ABO_NM_020469. 2:c.106G>T	0.222 00	0.269 82	0.0000 0	0.0000 0	0.0000 0	0.24249	0.0000 0	0.00000 0	0.0000 0	0.320000
ABO_NM_020469. 2:c.103G>A	0.002 00	0.021 03	0.00462	0.0000 0	0.0000 0	0.00005	0.0000 0	0.01140	0.0136 0	0.040000
ABO_NM_020469. 2:c.53G>T	0.00 600	0.015 28	0.0063 5	0.0000 0	0.0000 0	0.00014	0.0000 0	0.00570	0.0000 0	0.000000
<b>Minor BG Variants</b>										
<b>Variant</b>	<b>IndiG en</b>	<b>gnom ad</b>	<b>Genome Asia</b>	<b>SSMP</b>	<b>SSIP</b>	<b>China</b>	<b>Japan</b>	<b>Qatar</b>	<b>GME</b>	<b>Gambia</b>
SCIANNA:1:42830 785G>A	0.00 00	0.00 43	0.0014	0.0052	0.4444	0.0000	0.0000	0.0000	0.0051	0
SCIANNA:1:42830 851G>A	0.00 00	0.001 8	0.0014	0.0000	0.0000	0.0000	0.0000	0.0000	0.0020	0.0000
SCIANNA:1:42830 860C>G	0.00 00	0.00 04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0005	0.0000
DUFFY:1:15920556 4G>A	0.324 0	0.641 9	0.2750	0.1510	0.0000	0.0580	0.0000	0.7328	0.6863	1
DUFFY:1:15920570 4C>T	0.00 60	0.00 99	0.0017	0.0000	0.0000	0.0008	0.0000	0.0090	0.0156	0.0000
DUFFY:1:15920583 4G>A	0.001 0	0.00 00	0.0029	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CROMER:1:207322 436G>C	0.00 00	0.00 05	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
CROMER:1:207326 820C>T	0.002 0	0.00 01	0.0012	0.0052	0.0139	0.0000	0.0000	0.0000	0.0000	0.0000
KNOPS:1:2075874 28C>T	0.292 0	0.230 1	0.2620	0.2135	0.2778	0.3813	0.0000	0.1989	0.2221	0.02
KNOPS:1:2076094 24G>A	0.00 50	0.021 2	0.0000	0.0000	0.0139	0.0003	0.0000	0.0000	0.0081	0.0000
LEWIS:1:2076095 86A>G	0.470 0	0.40 02	0.4830	0.4375	0.5556	0.2342	0.0000	0.3750	0.3243	0.92
GERBICH:2:126693 891C>T	0.001 0	0.00 00	0.0003	0.0000	0.1389	0.0000	0.0000	0.0000	0.0000	0.0000
LANGEREIS:2:2192 13284C>T	0.00 00	0.00 42	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000



KELL:7:142954267 G>A	0.001 0	0.00 68	0.0014	0.0000	0.0000	0.0000	0.0000	0.0023	0.0035	0
KELL:7:142954320 C>T	0.00 00	0.00 00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
KELL:7:142957921 G>A	0.007 0	0.026 1	0.0049	0.0000	0.0000	0.0000	0.0000	0.0478	0.0448	0.0000
KELL:7:14296093 9C>T	0.002 0	0.00 01	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
KELL:7:14296094 0G>A	0.002 0	0.00 01	0.0026	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FORS:9:133154258 G>T	0.045 0	0.059 8	0.0000	0.0052	0.0000	0.0007	0.0000	0.1328	0.0918	0.03
FORS:9:133162355 G>A	0.358 0	0.228 8	0.3110	0.5469	0.2917	0.4618	0.0000	0.2193	0.2226	0.27
LEWIS:9:13703141 0C>T	0.001 0	0.00 04	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
IN:11:35176644G> C	0.014 0	0.00 00	0.0038	0.0052	0.0278	0.0001	0.0000	0.0011	0.0005	0.0000
IN:11:35189886C> A	0.00 00	0.00 00	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DOMBROCK:12:14 840624A>T	0.001 0	0.00 00	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
DIEGO:17:4425125 3G>A	0.00 60	0.00 35	0.0130	0.0000	0.0000	0.0255	0.0000	0.0000	0.0035	0.0000
KIDD:18:4573045 0G>A	0.305 0	0.137 6	0.3160	0.4844	0.1944	0.3938	0.0000	0.0796	0.1219	0.23
KIDD:18:45731089 G>A	0.00 00	0.022 0	0.0092	0.0052	0.0000	0.0000	0.0000	0.0299	0.0292	0.11
KIDD:18:45748385 C>T	0.00 90	0.00 02	0.0000	0.0000	0.0000	0.0002	0.0000	0.0057	0.0010	0.0000
LEWIS:19:5831566 G>C	0.021 0	0.073 7	0.0000	0.0313	0.0278	0.0386	0.0000	0.0852	0.0660	0.21
LEWIS:19:584395 8G>A	0.002 0	0.001 6	0.0000	0.0000	0.0000	0.0029	0.0000	0.0000	0.0032	0
LEWIS:19:5844108 G>A	0.00 30	0.00 68	0.0000	0.0000	0.0000	0.0007	0.0000	0.0057	0.0000	0.01



It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

LEWIS:19:584422 8T>C	0.007 0	0.019 3	0.0000	0.0000	0.0000	0.0000	0.0000	0.0057	0.0059	0.04
KLF1:19:12885926 A>G	0.498 0	0.364 4	0.5220	0.5365	0.5000	0.6059	0.0000	0.2835	0.2963	0.44
LUTHERAN:19:448 12188G>A	0.001 0	0.026 5	0.0055	0.0000	0.0000	0.0002	0.0000	0.0057	0.0121	0.03
LUTHERAN:19:448 13447T>A	0.001 0	0.00 94	0.0026	0.0000	0.0000	0.0001	0.0000	0.0350	0.0277	0.0000
LUTHERAN:19:448 14191C>T	0.00 30	0.001 2	0.0032	0.0000	0.0000	0.0000	0.0000	0.0022	0.0015	0.0000
LUTHERAN:19:448 19367C>T	0.002 0	0.00 08	0.0026	0.0000	0.0278	0.0005	0.0000	0.0000	0.0035	0.0000
LUTHERAN:19:448 19487A>G	0.216 0	0.321 1	0.1900	0.1510	0.2917	0.1049	0.0530	0.2964	0.3303	0.41
H:19:48750557A> C	0.00 30	0.00 01	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
H:19:48751247G>A	0.139 0	0.031 7	0.1820	0.2448	0.1667	0.2853	0.0000	0.0462	0.0418	0.0000
KANNO:20:46998 75G>A	0.047 0	0.00 31	0.0450	0.0260	0.0000	0.0393	0.0000	0.0044	0.0025	0.0000
P1PK:22:4269384 3T>C	0.331 0	0.385 2	0.2640	0.1719	0.4583	0.1504	0.0000	0.4485	0.4181	0.48
MNS:4:144120554 C>A	0.588 2	0.451 1	0.0000	0.0000	0.0000	0.2774	0.0000	0.2185	0.0000	0.2100
MNS:4:144120555 T>C	0.588 2	0.452 1	0.0000	0.0000	0.0000	0.0000	0.5000	0.3005	0.0457	0.2100
MNS:4:144120567 A>G	0.595 3	0.455 3	0.0000	0.2188	0.0000	0.4225	0.0000	0.4751	0.3819	0.2600
MNS:4:143997559 C>G	0.298 4	0.182 1	0.2500	0.3021	0.0000	0.1951	0.2150	0.2500	0.2704	0.0900
MNS:4:143997580 G>A	0.001 5	0.001 1	0.0012	0.0000	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000
MNS:4:143999413 G>C	0.00 05	0.00 01	0.0032	0.0104	0.0000	0.0012	0.0000	0.0000	0.0000	0.0000
MNS:4:143999443 G>A	0.309 6	0.274 7	0.2060	0.0677	0.3472	0.0482	0.0570	0.3807	0.3338	0.1400

MNS:4:144001256	0.00	0.00								
G>C	05	34	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

**Table 4.** Overall distribution of blood variant frequencies across various global population datasets used in the study.

Fishers Exact test of ABO and RH variants									
Variant	KG_fishe rs	gnomad_ fishers	GA_fishe rs	SSMP_fi shers	SSIP_fis hers	China_fis hers	Japan_fis hers	Qatar_fis hers	GME_fis hers
RHCE_NM_020485.8: c.122A>G	1.00000	0.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.2 8C>T	1.00000	0.60611	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.35422
RHD_NM_016124.6:c.4 8G>C	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000	1.00000
RHD_NM_016124.6:c.1 49-29G>C	0.00000	0.00000	1.00000	1.00000	1.00000	0.00000	1.00000	0.00000	0.00000
RHD_NM_016124.6:c.1 65C>T	1.00000	0.19245	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.1 75G>A	1.00000	0.02812	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.1 87G>T	0.62270	0.01632	1.00000	1.00000	1.00000	0.08409	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.2 08C>T	1.00000	0.38392	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.2 09G>A	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.4 97A>C	0.62057	0.10325	0.29480	1.00000	1.00000	0.58660	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.5 20G>A	0.79235	0.00020	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.45294
RHD_NM_016124.6:c.6 48G>C	0.00000	0.00000	1.00000	1.00000	1.00000	0.00000	1.00000	0.01993	0.00023
RHD_NM_016124.6:c.6 67T>G	0.00001	0.00000	1.00000	1.00000	1.00000	0.00000	1.00000	0.04661	0.02236
RHD_NM_016124.6:c.7 12G>A	1.00000	1.00000	1.00000	1.00000	1.00000	0.35683	0.31025	1.00000	1.00000

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .

RHD_NM_016124.6:c.9 16G>A	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.9 19G>A	1.00000	0.17113	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
RHD_NM_016124.6:c.9 32A>G	0.00000	0.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.00186	1.00000
RHD_NM_016124.6:c.1 073+152C>A	0.00000	0.00000	0.00000	0.00000	1.00000	0.00000	1.00000	0.00000	0.00000
RHCE_NM_020485.8: c.676G>C	0.00000	0.00000	1.00000	1.00000	1.00000	0.00000	1.00000	1.00000	0.00000
RHCE_NM_020485.8: c.602G>C	0.08286	0.73269	0.00000	1.00000	1.00000	0.00003	1.00000	1.00000	1.00000
RHCE_NM_020485.8: c.361A>T	1.00000	0.00009	1.00000	1.00000	1.00000	0.00170	1.00000	0.24345	0.00497
ABO_NM_020469.2:c. 1061delC	0.00233	0.00000	0.01079	1.00000	1.00000	1.00000	1.00000	1.00000	0.71828
ABO_NM_020469.2:c. 930G>A	0.00000	0.00000	0.00029	0.93223	1.00000	0.00156	1.00000	0.00001	0.00000
ABO_NM_020469.2:c. 829G>A	0.00314	0.04718	0.02387	0.07449	0.58196	0.26588	0.00488	0.78484	0.00000
ABO_NM_020469.2:c. 803G>C	0.00000	0.00000	0.00068	0.86433	0.89348	0.00322	0.00000	0.00001	0.00000
ABO_NM_020469.2:c. 802G>A	0.03775	0.00009	1.00000	1.00000	1.00000	0.00000	1.00000	1.00000	0.03809
ABO_NM_020469.2:c. 796C>A	0.00000	0.00000	0.00068	0.86431	0.89332	0.00350	1.00000	0.00002	1.00000
ABO_NM_020469.2:c. 771C>T	0.00191	0.02879	0.01558	0.05070	0.49357	0.21251	0.00318	0.71544	0.00000
ABO_NM_020469.2:c. 768C>A	0.32755	0.00270	0.00014	1.00000	1.00000	0.00000	1.00000	1.00000	0.00057
ABO_NM_020469.2:c. 703G>A	0.00000	0.00000	0.00048	0.73244	1.00000	0.00227	1.00000	0.00001	0.00000
ABO_NM_020469.2:c. 689G>A	0.08094	0.02475	1.00000	1.00000	1.00000	0.24239	1.00000	1.00000	0.01951
ABO_NM_020469.2:c. 681G>A	0.35078	0.81506	0.05085	0.05189	0.58201	0.27768	1.00000	0.85456	0.00000

ABO_NM_020469.2:c.657C>T	0.00000	0.00000	0.00061	0.86436	0.89364	0.00295	1.00000	0.00001	0.00000
ABO_NM_020469.2:c.649C>T	1.00000	0.73276	1.00000	1.00000	1.00000	0.16880	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.646T>A	0.00314	0.03962	0.02203	0.05169	0.58184	0.26590	0.00667	0.71685	0.00000
ABO_NM_020469.2:c.640A>G	1.00000	0.03937	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.595C>T	0.00039	0.00000	0.36677	1.00000	1.00000	1.00000	1.00000	1.00000	0.08745
ABO_NM_020469.2:c.579T>C	0.74003	0.00664	0.00008	0.00000	1.00000	0.08259	1.00000	1.00000	0.68752
ABO_NM_020469.2:c.542G>A	0.60176	0.65874	0.00072	1.00000	1.00000	0.00000	1.00000	1.00000	0.46401
ABO_NM_020469.2:c.527G>A	0.11225	0.00006	0.28339	1.00000	1.00000	1.00000	1.00000	0.05161	0.37604
ABO_NM_020469.2:c.526C>G	0.00000	0.00000	0.00088	0.73466	1.00000	0.00074	1.00000	0.00001	0.00000
ABO_NM_020469.2:c.496del	0.00222	0.00000	0.05182	1.00000	1.00000	0.00000	1.00000	1.00000	0.13638
ABO_NM_020469.2:c.467C>T	0.00000	0.00000	0.00000	0.07236	1.00000	0.00000	1.00000	0.00001	0.00005
ABO_NM_020469.2:c.297A>G	0.31644	0.00001	0.88281	0.27031	0.75448	0.23283	0.28510	0.05096	0.00000
ABO_NM_020469.2:c.261delG/c.260_262insG	0.00000	0.00000	0.07550	1.00000	1.00000	0.94968	1.00000	1.00000	0.00000
ABO_NM_020469.2:c.220C>T	1.00000	0.35986	1.00000	1.00000	1.00000	0.29089	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.189C>T	1.00000	0.02871	1.00000	1.00000	1.00000	0.11309	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.188G>A	1.00000	0.00021	1.00000	1.00000	1.00000	0.12562	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.106G>T	1.00000	0.00018	1.00000	1.00000	1.00000	0.11241	1.00000	1.00000	1.00000
ABO_NM_020469.2:c.103G>A	0.00000	0.00000	0.16270	1.00000	1.00000	0.00028	1.00000	0.07687	0.00002

ABO_NM_020469.2:c. 53G>T	0.05618	0.00016	1.00000	1.00000	1.00000	0.00000	1.00000	1.00000	1.00000
<b>Fishers Exact test of other minor blood variants</b>									
SCIANNA:1:42830785 G>A	0.19381	0.00301	0.42237	0.16437	0.00000	0.16909	1.00000	1.00000	0.00530
SCIANNA:1:42830851 G>A	1.00000	0.19457	0.42261	1.00000	1.00000	0.16880	1.00000	1.00000	0.21143
SCIANNA:1:42830860 C>G	1.00000	0.58667	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
DUFFY:1:159205564G >A	0.00000	0.00000	0.00577	0.00007	1.00000	0.00000	1.00000	0.00000	0.01645
DUFFY:1:159205704C >T	0.57553	0.07037	0.01324	1.00000	1.00000	0.00000	1.00000	0.27577	0.00322
DUFFY:1:159205834G >A	0.58497	0.00172	0.23050	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
CROMER:1:20732243 6G>C	1.00000	0.62315	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
CROMER:1:20732682 0C>T	0.11423	0.00010	0.47953	0.36202	0.16036	1.00000	1.00000	1.00000	1.00000
KNOPS:1:207587428C >T	0.00504	0.00000	0.07694	0.08416	1.00000	0.00000	1.00000	0.04722	0.00000
KNOPS:1:207609424 G>A	0.00200	0.00000	1.00000	1.00000	0.31885	0.00000	1.00000	1.00000	0.24073
LEWIS:1:207609586A >G	0.30132	0.00005	0.62421	0.63682	0.41084	0.00000	1.00000	0.13212	0.00000
GERBICH:2:12669389 1C>T	0.20449	0.00357	0.55962	1.00000	0.00000	0.02206	1.00000	1.00000	1.00000
LANGEREIS:2:2192132 84C>T	0.45157	0.00450	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.61915
LANGEREIS:2:2192161 23C>T	0.00000	0.00000	0.00080	1.00000	1.00000	0.00000	1.00000	0.00000	0.07065
LANGEREIS:2:2192166 94G>A	0.50079	0.00027	1.00000	1.00000	1.00000	0.00024	1.00000	0.23026	0.41582
LANGEREIS:2:2192177 83G>A	0.63155	0.77604	0.55944	1.00000	1.00000	1.00000	1.00000	1.00000	0.68291
GLOB:3:161086379C> T	0.00000	0.00000	0.00001	0.02625	1.00000	0.00000	1.00000	0.00000	0.00000

JUNIOR:4:88092344C >T	0.00000	0.00000	0.00006	1.00000	0.32764	0.00000	1.00000	0.00653	0.06056
JUNIOR:4:88118244G> A	0.49694	0.34872	1.00000	1.00000	0.00001	0.52296	1.00000	1.00000	0.61897
JUNIOR:4:88131171G> T	0.68653	0.00223	0.00000	0.00001	1.00000	0.00000	0.00000	0.00000	0.00000
I:6:10586494G>A	1.00000	0.29584	0.15631	1.00000	1.00000	0.00010	1.00000	0.40867	1.00000
RHAG:6:49615692C>T	1.00000	0.24505	1.00000	1.00000	1.00000	0.04141	1.00000	1.00000	1.00000
COLTON:7:30912043C >T	0.00584	0.00000	0.35270	0.59222	0.29448	0.00000	1.00000	1.00000	0.16226
COLTON:7:30912049A >G	1.00000	0.23500	1.00000	1.00000	1.00000	1.00000	1.00000	0.60952	0.21144
YT:7:100893176G>T	0.00001	0.00124	0.00017	1.00000	0.39557	0.00000	1.00000	0.00390	0.00000
YT:7:100893967C>T	0.00000	0.00000	0.00008	1.00000	0.11989	0.00000	1.00000	0.01190	0.00045
YT:7:100894064C>T	0.08015	0.00000	0.70972	1.00000	1.00000	0.00000	1.00000	0.34751	0.35926
YT:7:100894132C>T	0.00052	0.00000	0.06408	1.00000	1.00000	1.00000	1.00000	0.00000	0.00085
KELL:7:142943026A> G	0.00000	0.00000	0.00528	1.00000	1.00000	1.00000	1.00000	0.00191	0.00000
KELL:7:142946304C> T	0.49637	0.08938	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
KELL:7:142953916G> A	1.00000	0.00026	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
KELL:7:142954267G> A	0.01050	0.00019	0.71825	1.00000	1.00000	1.00000	1.00000	0.42037	0.10405
KELL:7:142954320C> T	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
KELL:7:142957921G> A	0.04207	0.00000	0.35720	1.00000	1.00000	1.00000	1.00000	0.00000	0.01318
KELL:7:142960939C> T	0.06300	0.00005	0.20309	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
KELL:7:142960940G> A	0.13285	0.00001	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
FORS:9:133154258G> T	0.74837	0.00615	1.00000	0.00587	1.00000	0.00000	1.00000	0.00000	0.18670
FORS:9:133162355G> A	0.02017	0.00000	0.01262	0.00123	0.47285	0.00000	1.00000	0.00000	0.00000

LEWIS:9:137031410C>T	0.32992	0.17898	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
IN:11:35176644G>C	0.00000	1.00000	0.00006	0.50785	0.27602	0.00000	1.00000	0.00000	0.00000
IN:11:35189886C>A	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
DOMBROCK:12:14840624A>T	0.07665	0.00005	0.36677	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
DIEGO:17:44251253G>A	0.00784	0.08529	0.00920	1.00000	1.00000	0.00000	1.00000	1.00000	0.35982
KIDD:18:45730450G>A	0.00001	0.00000	0.56741	0.00079	0.15232	0.00000	1.00000	0.00000	0.00000
KIDD:18:45731089G>A	0.00000	0.00000	0.00001	0.16479	1.00000	0.16866	1.00000	0.00000	0.00000
KIDD:18:45748385C>T	0.00000	0.00000	1.00000	1.00000	1.00000	0.00000	1.00000	1.00000	0.00041
LEWIS:19:5831566G>C	0.00000	0.00000	1.00000	0.31151	0.66575	0.00004	1.00000	0.00005	0.00229
LEWIS:19:5843958G>A	1.00000	0.25977	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.76645
LEWIS:19:5844108G>A	1.00000	0.03814	1.00000	1.00000	1.00000	0.00590	1.00000	0.44315	1.00000
LEWIS:19:5844228T>C	1.00000	0.00001	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	0.84144
KLF1:19:12885926A>G	0.01382	0.00000	0.37441	0.56126	1.00000	0.00000	1.00000	0.00000	0.00000
LUTHERAN:19:44812188G>A	0.00000	0.00000	0.01065	1.00000	1.00000	0.09191	1.00000	0.22008	0.00000
LUTHERAN:19:44813447T>A	0.17852	0.00000	0.23041	1.00000	1.00000	0.06501	1.00000	0.00000	0.00000
LUTHERAN:19:44814191C>T	0.15555	0.01635	1.00000	1.00000	1.00000	0.00000	1.00000	0.55548	0.34388
LUTHERAN:19:44819367C>T	1.00000	0.03390	1.00000	1.00000	0.02243	0.00754	1.00000	1.00000	0.57541
LUTHERAN:19:44819487A>G	0.00021	0.00000	0.05655	0.09500	0.26962	0.00000	0.00000	0.00001	0.00000
H:19:48750557A>C	0.00883	0.00000	0.11347	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
H:19:48751247G>A	0.00003	0.00000	0.00056	0.00223	0.61064	0.00000	1.00000	0.00000	0.00000

KANNO:20:4699875G >A	0.00000	0.00000	0.68937	0.27044	1.00000	0.09996	1.00000	0.00000	0.00000
P1PK:22:42693843T> C	0.27438	0.00060	0.00011	0.00045	0.13594	0.00000	1.00000	0.00000	0.00000

**Table 5.** Tabulation of blood group variants which were found significantly distinct in the global populations along with their p-values.

## Supplementary Datasets

**Supplementary Table 1.** Locus genomic reference (LRG) coordinates of list of genes encoding human blood group, neutrophil and platelet antigens.

**Supplementary Table 2.** Tabulation of variant counts in each blood gene used in the analysis.

## References



- 27th Regional Congress of the ISBT, Copenhagen, Denmark, June 17-21, 2017. (2017). *Vox Sanguinis*, 112 Suppl 1(Suppl 1), 5–295.
- Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.], Chapter 7, Unit7.20.*
- AlBilali, A., Taibah University., AlFraidy, O., AlHarbi, H., Mhros., S., AlSayed., H., Tolah., K., AlHarbi, A., Kurban., A., Taibah University., Taibah University., Taibah University., Taibah University., Taibah University., Taibah University., & Taibah University. (2017). THE DISTRIBUTION OF ABO BLOOD GROUPS AND RHESUS FACTOR AMONG REGULAR BLOOD DONORS IN AL-MADINAH AL-MONAWARAH, SAUDI ARABIA 2016. In *International Journal of Advanced Research* (Vol. 5, Issue 1, pp. 2917–2920).  
<https://doi.org/10.21474/ijar01/3092>
- AlSuhaibani, E. S., Kizilbash, N. A., & Malik, S. (2015). Heterogeneity and diversity of ABO and Rh blood group genes in select Saudi Arabian populations. *Genetics and Molecular Research: GMR*, 14(3), 7850–7863.
- Anstee, D. J. (2010). The relationship between blood groups and disease. *Blood*, 115(23), 4635–4643.
- Balci, Y. (2010). ABO and Rh Blood Groups Frequency in Denizli Province. In *International Journal of Hematology and Oncology* (Vol. 20, Issue 2, pp. 103–105).  
<https://doi.org/10.4999/uhod.09141>
- Bashwari, L. A., Al-Mulhim, A. A., Ahmad, M. S., & Ahmed, M. A. (2001). Frequency of ABO blood groups in the Eastern region of Saudi Arabia. *Saudi Medical Journal*, 22(11), 1008–1012.
- Bisht, A., Marwaha, N., Kaur, R., Gupta, D., & Chhabra, R. (2020). Haemovigilance Programme

of India: Comparative analysis of transfusion reactions reported over a 5-year period through two reporting formats and key recommendations for blood safety. *Asian Journal of Transfusion Science*, 14(2), 103–116.

Canizalez-Román, A., Campos-Romero, A., Castro-Sánchez, J. A., López-Martínez, M. A., Andrade-Muñoz, F. J., Cruz-Zamudio, C. K., Ortiz-Espinoza, T. G., León-Sicairos, N., Gaudrón Llanos, A. M., Velázquez-Román, J., Flores-Villaseñor, H., Muro-Amador, S., Martínez-García, J. J., & Alcántar-Fernández, J. (2018). Blood Groups Distribution and Gene Diversity of the ABO and Rh (D) in the Mexican Population. *BioMed Research International*, 2018, 1925619.

Chandra, T., & Gupta, A. (2012). Frequency of ABO and rhesus blood groups in blood donors. *Asian Journal of Transfusion Science*, 6(1), 52–53.

Chen, Q., & Flegel, W. A. (2005). Random survey for RHD alleles among D+ European persons. *Transfusion*, 45(7), 1183–1191.

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* , 31(16), 2745–2747.

Complications of blood transfusion. (2006). *Continuing Education in Anaesthesia Critical Care & Pain*, 6(6), 225–229.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Computational Biology*, 6(12), e1001025.

Dean, L. (2005). Blood group antigens are surface markers on the red blood cell membrane. In *Blood Groups and Red Cell Antigens [Internet]*. National Center for Biotechnology Information (US).

Doku, G. N., Agbozo, W. K., Annor, R. A., Kisseh, G. D., & Owusu, M. A. (2019). Frequency of

ABO/Rhesus (D) blood groupings and ethnic distribution in the Greater-Accra region of Ghana, towards effective blood bank inventory. *International Journal of Immunogenetics*, 46(2), 67–73.

E.f.b.t.m., R. M. A. E.-F. M. D. ; T. M. A. E.-G., Abou El-Fetouh, R. M., M.D.; TAREK M.A.

EL-GEMMEZI, E F B T, Fouda, M. A., M D ; Faten, & C. T. M-R. C. P. S. of Edinburgh. (2018). Prevalence of Red Cell Blood Group Antigens Among Egyptian Population in Comparison with Other Ethnic Groups. In *The Medical Journal of Cairo University* (Vol. 86, Issue 3, pp. 989–994). <https://doi.org/10.21608/mjcu.2018.55771>

Fichou, Y., Parchure, D., Gogri, H., Gopalkrishnan, V., Le Maréchal, C., Chen, J.-M., Férec, C., Madkaikar, M., Ghosh, K., & Kulkarni, S. (2018). Molecular basis of weak D expression in the Indian population and report of a novel, predominant variant RHD allele. *Transfusion*, 58(6), 1540–1549.

Flegel, W. A., von Zabern, I., Doescher, A., Wagner, F. F., Strathmann, K. P., Geisen, C., Palfi, M., Písacka, M., Poole, J., Polin, H., Gabriel, C., & Avent, N. D. (2009). D variants at the RhD vestibule in the weak D type 4 and Eurasian D clusters. *Transfusion*, 49(6), 1059–1069.

Garg, P., Upadhyay, S., Chufal, S. S., Hasan, Y., & Tayal, I. (2014). Prevalance of ABO and Rhesus Blood Groups in Blood Donors: A Study from a Tertiary Care Teaching Hospital of Kumaon Region of Uttarakhand. *Journal of Clinical and Diagnostic Research: JCDR*, 8(12), FC16–FC19.

Gates, M. A., Wolpin, B. M., Cramer, D. W., Hankinson, S. E., & Tworoger, S. S. (2011). ABO blood group and incidence of epithelial ovarian cancer. *International Journal of Cancer. Journal International Du Cancer*, 128(2), 482–486.

Gnad, F., Baucom, A., Mukhyala, K., Manning, G., & Zhang, Z. (2013). Assessment of computational methods for predicting the effects of missense mutations in human

cancers. *BMC Genomics*, 14 Suppl 3, S7.

Golassa, L., Tsegaye, A., Erko, B., & Mamo, H. (2017). High rhesus (Rh(D)) negative frequency and ethnic-group based ABO blood group distribution in Ethiopia. *BMC Research Notes*, 10(1), 330.

Grootkerk-Tax, M. G. H. M., van Wintershoven, J. D., Ligthart, P. C., van Rhenen, D. J., van der Schoot, C. E., & Maaskant-van Wijk, P. A. (2006). RHD(T201R, F223V) cluster analysis in five different ethnic groups and serologic characterization of a new Ethiopian variant DARE, the DIII type 6, and the RHD(F223V). *Transfusion*, 46(4), 606–615.

Hiltunen, L. M., Laivuori, H., Rautanen, A., Kaaja, R., Kere, J., Krusius, T., Paunio, M., & Rasi, V. (2009). Blood group AB and factor V Leiden as risk factors for pre-eclampsia: a population-based nested case-control study. *Thrombosis Research*, 124(2), 167–173.

Höher, G., Fiegenbaum, M., & Almeida, S. (2018). Molecular basis of the Duffy blood group system. *Blood Transfusion = Trasfusione Del Sangue*, 16(1), 93–100.

Howes, R. E., Patil, A. P., Piel, F. B., Nyangiri, O. A., Kabaria, C. W., Gething, P. W., Zimmerman, P. A., Barnadas, C., Beall, C. M., Gebremedhin, A., Ménard, D., Williams, T. N., Weatherall, D. J., & Hay, S. I. (2011). The global distribution of the Duffy blood group. *Nature Communications*, 2, 266.

Kappler-Gratias, S., Auxerre, C., Dubeaux, I., Beolet, M., Ripaux, M., Le Pennec, P.-Y., & Pham, B.-N. (2014). Systematic RH genotyping and variant identification in French donors of African origin. *Blood Transfusion = Trasfusione Del Sangue*, 12 Suppl 1(Suppl 1), s264–s272.

Kaur, H., Khanna, A., Manjari, M., & Khanna, M. (2013). Prevalence of ABO blood groups and rhesus (Rh) factor in the population residing in and around Amritsar, Punjab (a 4-year study from June 2007 to June 2011). *Asian Journal of Transfusion Science*, 7(2), 159.

Lamba, D. S., Kaur, R., & Basu, S. (2013). Clinically Significant Minor Blood Group Antigens amongst North Indian Donor Population. *Advances in Hematology*, 2013.

<https://doi.org/10.1155/2013/215454>

Lane, W. J., Westhoff, C. M., Uy, J. M., Aguad, M., Smeland-Wagman, R., Kaufman, R. M., Rehm, H. L., Green, R. C., Silberstein, L. E., & MedSeq Project. (2016). Comprehensive red blood cell and platelet antigen prediction from whole genome sequencing: proof of principle.

*Transfusion*, 56(3), 743–754.

Lawicki, S., Covin, R. B., & Powers, A. A. (2017). The Kidd (JK) Blood Group System.

*Transfusion Medicine Reviews*, 31(3), 165–172.

Mastana, S. S. (2014). Unity in diversity: an overview of the genomic anthropology of India.

*Annals of Human Biology*, 41(4), 287–299.

Mitra, R., Mishra, N., & Rath, G. P. (2014). Blood groups systems. *Indian Journal of*

*Anaesthesia*, 58(5), 524–528.

Möller, M., Jöud, M., Storry, J. R., & Olsson, M. L. (2016). ErythroGene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes

Project. *Blood Advances*, 1(3), 240–249.

Mondal, B., Maiti, S., Biswas, B. K., Ghosh, D., & Paul, S. (2012). Prevalence of

hemoglobinopathy, ABO and rhesus blood groups in rural areas of West Bengal, India.

*Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences*, 17(8), 772–776.

Nanu, A., & Thapliyal, R. M. (1997). Blood group gene frequency in a selected north Indian population. *The Indian Journal of Medical Research*, 106, 242–246.

Nazli, R., Haider, J., Khan, M. A., Akhtar, T., & Aslam, H. (2015). Frequency of ABO blood groups and RhD factor in the female population of District Peshawar. *Pakistan Journal of*

*Medical Sciences Quarterly*, 31(4), 984–986.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), 3812–3814.

Noizat-Pirenne, F., Lee, K., Pennec, P.-Y. L., Simon, P., Kazup, P., Bachir, D., Rouzaud, A.-M., Roussel, M., Juszczak, G., Ménanteau, C., Rouger, P., Kotb, R., Cartron, J.-P., & Ansart-Pirenne, H. (2002). Rare RHCE phenotypes in black individuals of Afro-Caribbean origin: identification and transfusion safety. *Blood*, 100(12), 4223–4231.

Ohto, H., Uchikawa, M., Ito, S., Wada, I., Nollet, K. E., Omae, Y., Ogasawara, K., & Tokunaga, K. (2022). The KANNO blood group system. *Immunohematology / American Red Cross*, 38(4), 119–122.

Owaidah, A. Y., Naffaa, N. M., Alumran, A., & Alzahrani, F. (2020). Phenotype Frequencies of Major Blood Group Systems (Rh, Kell, Kidd, Duffy, MNS, P, Lewis, and Lutheran) Among Blood Donors in the Eastern Region of Saudi Arabia. *Journal of Blood Medicine*, 11, 59–65.

Periyavan, S., Sangeetha, S. K., Marimuthu, P., Manjunath, B. K., & Seema, D. M. (2010). Distribution of ABO and Rhesus-D blood groups in and around Bangalore. *Asian Journal of Transfusion Science*, 4(1), 41.

Peyrard, T., Lam, Y., Saison, C., Arnaud, L., Babinet, J., Rouger, P., Bierling, P., & Janvier, D. (2012). Anti-U-like as an alloantibody in S-s-U- and S-s-U+(var) black people. *Transfusion*, 52(3), 622–628.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20(1), 110–121.

Reid, M. E. (2009). MNS blood group system: a review. *Immunohematology / American Red Cross*, 25(3), 95–101.

- Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J., & Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 47(D1), D886–D894.
- Sarhan, M. A., Saleh, K. A., & Bin-Dajem, S. M. (2009). Distribution of ABO blood groups and rhesus factor in Southwest Saudi Arabia. *Saudi Medical Journal*, 30(1), 116–119.
- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., Day, I. N. M., & Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, 34(1), 57–65.
- Shimizu, Y., Ao, H., Soemantri, A., Tiwawech, D., Settheetham-Ishida, W., Kayame, O. W., Kimura, M., Nishioka, T., & Ishida, T. (2000). Sero- and molecular typing of Duffy blood group in Southeast Asians and Oceanians. *Human Biology*, 72(3), 511–518.
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), 1034–1050.
- Singh, A., Srivastava, R. K., Deogharia, K. S., & Singh, K. K. (2016). Distribution of ABO and Rh types in voluntary Blood donors in Jharkhand area as a study conducted by RIMS, Ranchi. *Journal of Family Medicine and Primary Care*, 5(3), 631–636.
- Storry, J. R., Clausen, F. B., Castilho, L., Chen, Q., Daniels, G., Denomme, G., Flegel, W. A., Gassner, C., de Haas, M., Hyland, C., Yanli, J., Keller, M., Lomas-Francis, C., Nogues, N., Olsson, M. L., Peyrard, T., van der Schoot, E., Tani, Y., Thornton, N., ... Yahalom, V. (2019). International Society of Blood Transfusion Working Party on Red Cell Immunogenetics and Blood Group Terminology: Report of the Dubai, Copenhagen and Toronto meetings.

*Vox Sanguinis*, 114(1), 95–102.

Thakral, B., Agrawal, S. K., Dhawan, H. K., Saluja, K., Dutta, S., & Marwaha, N. (2007). First report from India of haemolytic disease of newborn by anti-c and anti-E in Rh (D) positive mothers. *Hematology* , 12(5), 377–380.

Tufano, A., Coppola, A., Nardo, A., Bonfanti, C., Crestani, S., Cerbone, A. M., & Franchini, M. (2013). Non-O blood group as a risk factor for cerebral vein thrombosis. *Thrombosis and Haemostasis*, 110(1), 197–199.

Van Buren, N. L., Gillen, B., Lomas-Francis, C., Burgos, A., Pease, D., Beaver, M., Imbryk, A., Dugger, J., Fugate, A., Hebel, E., Lodermeier, M., Vege, S., & Westhoff, C. M. (2022). YTGT: A new high-prevalence antigen in the Yt blood group system in two unrelated Native Americans and transfusion management. *Transfusion*, 62(9), 1917–1922.

Wang, D.-S., Chen, D.-L., Ren, C., Wang, Z.-Q., Qiu, M.-Z., Luo, H.-Y., Zhang, D.-S., Wang, F.-H., Li, Y.-H., & Xu, R.-H. (2012). ABO blood group, hepatitis B viral infection and risk of pancreatic cancer. *International Journal of Cancer. Journal International Du Cancer*, 131(2), 461–468.

Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 38(16), e164.

Wang, S.-S., Zhang, H., Qu, L., Zhao, Z., & Li, L. (2021). A renewed understanding of anti-human globulin reagents: interference constraints using an optimization method in pretransfusion compatibility tests. *Journal of Clinical Laboratory Analysis*, 35(3), e23695.

Yu, Y., Ma, C., Sun, X., Guan, X., Zhang, X., Saldanha, J., Chen, L., & Wang, D. (2016). Frequencies of red blood cell major blood group antigens and phenotypes in the Chinese Han population from Mainland China. *International Journal of Immunogenetics*,



43(4), 226–235.