Published in final edited form as:

Transfusion. 2019 January ; 59(1): 101–111. doi:10.1111/trf.15012.

Development and evaluation of a transfusion medicine genome wide genotyping array

Yuelong Guo¹, Michael P. Busch^{2,3}, Mark Seielstad^{2,3}, Stacy Endres-Dighe⁴, Connie M. Westhoff⁵, Brendan Keating⁶, Carolyn Hoppe⁷, Aarash Bordbar⁸, Brian Custer³, Adam S. Butterworth^{9,10}, Tamir Kanias^{11,12}, Alan E. Mast^{13,14}, Steve Kleinman¹⁵, Yontao Lu¹⁶, Grier P. Page¹⁷ National Heart, Lung, and Blood Institute Recipient Epidemiology Donor Evaluation Study (REDS)-III

¹RTI International, Research Triangle Park, North Carolina ²Vitalant Research Institute (formerly Blood Systems Research Institute), San Francisco, California ³Department of Laboratory Medicine, University of California San Francisco, San Francisco, California ⁴RTI International, Rockville, Maryland ⁵New York Blood Center, New York, New York ⁶Penn Transplant Institute, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania ⁷UCSF Benioff Children's Hospital Oakland, Oakland, California ⁸Sinopia Biosciences, San Diego, California ⁹MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK ¹⁰NIHR Blood and Transplant Unit in Donor Health and Genomics, University of Cambridge, Cambridge, UK ¹¹Pittsburgh Heart, Lung, Blood, and Vascular Medicine Institute, University of Pittsburgh, Pittsburgh, Pennsylvania ¹²Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania ¹³Blood Research Institute, Blood Center of Wisconsin, Milwaukee, Wisconsin ¹⁴Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, Wisconsin ¹⁵University of British Columbia, Victoria, British Columbia, Canada ¹⁶Affymetrix Incorporated, Santa Clara, California ¹⁷RTI International, Atlanta, Georgia

Abstract

BACKGROUND: Many aspects of transfusion medicine are affected by genetics. Current single-nucleotide polymorphism (SNP) arrays are limited in the number of targets that can be interrogated and cannot detect all variation of interest. We designed a transfusion medicine array (TM-Array) for study of both common and rare transfusion-relevant variations in genetically diverse donor and recipient populations.

STUDY DESIGN AND METHODS: The array was designed by conducting extensive bioinformatics mining and consulting experts to identify genes and genetic variation related to a wide range of transfusion medicine clinical relevant and research-related topics. Copy number polymorphisms were added in the alpha globin, beta globin, and Rh gene clusters.

Address reprint requests to: John Guo, RTI International, 3040 E. Cornwallis Road, Research Triangle Park, NC 27709; jyguo@rti.org.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

RESULTS: The final array contains approximately 879,000 SNP and copy number polymorphism markers. Over 99% of SNPs were called reliably. Technical replication showed the array to be robust and reproducible, with an error rate less than 0.03%. The array also had a very low Mendelian error rate (average parent–child trio accuracy of 0.9997). Blood group results were in concordance with serology testing results, and the array accurately identifies rare variants (minor allele frequency of 0.5%). The array achieved high genome-wide imputation coverage for African-American (97.5%), Hispanic (96.1%), East Asian (94.6%), and white (96.1%) genomes at a minor allele frequency of 5%.

CONCLUSIONS: A custom array for transfusion medicine research has been designed and evaluated. It gives wide coverage and accurate identification of rare SNPs in diverse populations. The TM-Array will be useful for future genetic studies in the diverse fields of transfusion medicine research.

Transfusion medicine encompasses diverse medical services and fields of research including blood donation, component processing, immunohematology, infectious disease, recipient outcomes, therapeutic apheresis, stem cell collections, cellular therapy, and hemostasis. Genetics play a central role in transfusion research and transfusion practice.^{1–3} The early discovery of human blood groups, particularly the ABO and Rh systems,^{4–7} and the recognition of the need to ensure compatibility between donors and recipients has led to a high degree of safety such that blood transfusions became one of the most common medical procedures. In contrast, despite the early discovery of key genes and mutations causal to hemoglobinopathies, hemochromatosis, thrombosis, and coagulation,^{8–11} genome-wide understanding of the effects of genetics on many other aspects of transfusion medicine practice, such as storage hemolysis and posttransfusion reactions, is lacking.

Genome-wide association studies (GWASs) simultaneously test for associations between phenotypic variables and millions of genetic markers.^{12,13} GWASs have been applied to numerous hematologic endpoints and have advanced our knowledge of the genetic underpinnings of modifiers of complete blood count, hematopoiesis, cardiovascular diseases, and sickle cell anemia.^{14–18} Transplant research has benefitted from GWASs by identifying novel genes and alleles that are associated with kidney transplant rejection and clinical outcomes in patients with renal and lung transplants.^{19–24} Although advances in whole genome sequencing have made it possible to interrogate causal variants directly as opposed to evaluating variants that are in linkage disequilibrium with causal variants,²⁵ array-based genotyping remains a cost-effective method for genotyping large cohorts, such as the UK Biobank cohort, in which half a million participants have been enrolled and genotyped.²⁶ Additionally, conducting array-based GWASs on samples prior to whole genome sequencing or whole exome sequencing is of value for detecting gender mismatches, sample swaps, and relatedness in the study population. Array-based genotyping is highly compatible with whole exome sequencing/whole genome sequencing to ensure concordance and fidelity especially in high-priority regions of the genome.

Modern genetics will continue to play a critical role in blood donor safety and transfusion outcome research and practices.^{27,28} A customized Affymetrix Axiom genome-wide single-nucleotide polymorphism (SNP) array, called the transfusion medicine array (TM-Array),

was developed to facilitate transfusion genetics research. It was developed with two goals in mind: 1) to allow the genome-wide study of genetically diverse populations, and 2) to study transfusion medicine-specific variants and genes in detail. The TM-Array includes genes involved in red blood cell (RBC) and platelet structure and function, human leukocyte antigen (HLA), blood groups including human platelet antigen, iron metabolism, sickle cell trait, glucose-6-phosphate-dehydrogenase deficiency, pica, and restless leg syndrome, among others. SNPs were selected to be informative in diverse populations. The array has been used to genotype nearly 16,000 participants: 13,168 blood donors enrolled in the National Heart, Lung, and Blood Institute (NHLBI) Recipient Epidemiology Donor Evaluation Study (REDS)-III Red Blood Cell Omics (RBC-Omics) study and 2,700 patients with sickle cell disease (SCD) from six hemocenters in Brazil enrolled in the REDS-III Brazil SCD cohort.

METHODS

The Affymetrix axiom genotyping platform

The Affymetrix Axiom platform is a hybridization-based genotyping array that offers a flexible choice of predesigned modules and customizable contents. Its utility has been shown in large cohorts, such as the UK Biobank²⁶ and Million Veteran Program,²⁹ multirace/ethnic cohorts,³⁰ and in focused research areas such as transplant research.²⁴ There are approximately 1.38 million features available. Some SNPs and copy number polymorphisms (CNPs) take more than one feature to be genotyped with high quality. As a result, the TM-Array contains approximately 879,000 SNPs and CNPs.

TM array design and variant selection

The TM-Array's core imputation sets of SNPs was built starting with the UK Biobank Array European GWAS grid content and supplemented with SNPs to increase coverage in Eastern Asians, African Americans, and Brazilians with the aim of providing near full genome imputed coverage down to minor allele frequency (MAF) of 5% in European, African, and East Asian populations. In addition to the core genome-wide imputation grid, the design included customized transfusion medicine content. Experts in the areas of RBCs, platelets, blood groups, iron metabolism, SCD, pica, restless leg syndrome, and other hematologic disorders recommended the addition of relevant genes and genetic variation. Extensive bioinformatics mining of resources such as PubMed, Genotype-Tissue Expression project,³¹ and the GWA SNP catalog¹³ was performed. Additionally, we drew on the Affymetrix catalog of variation and existing arrays such as the transplant array.²⁴ The design is detailed below and illustrated in Fig. 1.

Genome-wide coverage module (approximately 559,000 markers)

The genome-wide coverage module consists of three main components:

1. UK Biobank common GWAS grid (approximately 348,000 markers): This content is shared with the Affymetrix UK Biobank Array. It provides high-density genome-wide coverage across European populations.

2. African American GWAS booster (approximately 163,000 markers): This module was designed using an Affymetrix variant selection algorithm that maximizes genome-wide coverage for African American populations while balancing the number of features used.
3. East Asian GWAS booster (approximately 48,000 markers): This module was also designed using the Affymetrix variant selection algorithm.

African American and East Asian GWAS boosters were designed to provide over 90% imputation coverage for common variants (MAF $\geq$ 5%).

Affymetrix commercial modules

Predesign module contents were added to the design to capture markers identified based on currently identified GWAS signals, targeted regions, pharmacokinetics, and pharmacology-relevant signals. This consists of the following six modules:

1. Chromosome Y and mitochondria module (approximately 1.2 thousand markers)
2. Pharmacokinetics and pharmacology “absorption, distribution, metabolism, and excretion” markers module (approximately 2000 markers)
3. The HLA and the killer-cell immunoglobulin-like receptors region module (approximately 14,000 markers)
4. GWAS catalog¹³ markers module (approximately 17,000 markers)
5. Nonsynonymous SNPs module (approximately 80,000 markers)
6. Expression quantitative trait loci markers both from Affymetrix and the Genotype-Tissue Expression project³¹ (approximately 23,000 markers)

Transfusion medicine-specific SNPs, CNPs, and genes

Through comprehensive literature searches, bioinformatics mining, and expert panel review, the design team curated a list of variants and genes that are transfusion medicine related. Clinical endpoints being considered included blood groups, blood and iron disorders, platelet physiology and platelet related disorders, RBC genetics, SCD, and coagulation genetics. The full list of variants and genes are available at https://reds-iii.rti.org/publish/Axiom_TM_Array.na35.annot.csv.gz.

All known blood systems, including ABO, Rh, and 32 additional blood group loci were considered in the design of this array. SNPs and insertion/deletions (indels) were initially extracted from two sources: (1) the latest (third edition) AABB standards for accreditation of laboratories testing for blood groups, platelet, and neutrophil antigens by DNA³²; and (2) the European Bioinformatics Institute Immuno Polymorphism-human platelet antigens database.³³ The lists were further expanded using the blood group antigen gene mutation (BGMUT) database.^{34,35} The BGMUT database documented 1777 alleles associated with blood group phenotypes. We extracted all nucleotide changes for these alleles and mapped them to human genome (GRCh37) chromosome and position location whenever possible. This list consists of 729 SNPs and 2 CNPs (the 37 bp indel defining the RHD pseudogene and the 109 bp indel defining RHC/c).

Bioinformatics mining and comprehensive literature review were performed for variants and genes associated with hemolysis [search term “(hemolysis) AND (SNP OR polymorphism OR variant) AND (association OR associated)”]; 222 results], iron disorders [search term “iron AND disorder AND (SNP OR polymorphism OR variant) AND (association OR associated)”]; 422 results) and various specific blood- and iron-related disorders. Examples of blood disorder queries include anemia, leukemia, clotting, and bleeding. Examples of iron disorder queries include hemochromatosis, restless leg syndrome, atherosclerosis, metabolic syndrome, and liver disease; 3,885 variants and 141 genes associated with these disorders were extracted from research articles and reviews. The list of genes was further appended by adding 936 proximal genes.

A search of relevant genetic modulators of inflammation included cytokines, chemokines, interleukins, interferons, growth factors, tumor necrosis factor superfamily, anti-inflammatory cytokines, and their receptors. Additional comprehensive literature review was performed by searching PubMed using query “(coagulation OR cytokines) AND platelet AND (SNP OR polymorphism OR variant) AND (association OR associated)” (search date 10/19/2015; 256 results).

This list was further augmented by genes that encode for proteins previously detected in proteomics studies of red blood cells³⁶ and platelets (33 studies aggregated in Thomas et al.³⁷), as well proteins in the cell-scale metabolic networks for those cells.^{37,38} Additional blood group variants and SCD genes manually curated by the expert panel were appended to the lists.

In summary, prior to the design of the genome-wide booster grid, we included 145 transplant-relevant SNPs, 13 restless leg syndrome SNPs, 3885 SNPs in blood and iron disorder genes, 729 SNPs in blood group genes, 125 SNPs in coagulation and cytokine genes, and all 34,000 SNPs previously associated with a disorder in a GWAS. After the design achieved the desired genome-wide coverage by including the genome-wide booster modules, the array was filled by SNPs down to 1% in gene regions of 141 iron-related genes, 3856 megakaryocytes and platelets proteins, 1285 genes with the corresponding proteins known to occur in or on the red cells, 238 cytokine genes, 155 transforming growth factor genes, and 48 SCD-related genes. We tiled the alpha globin, beta globin, and RHD/RHCE loci with approximately 1000 SNPs to allow the detection of copy number changes.

These lists have overlap with each other, and some trimming was performed on loci in high linkage disequilibrium. In total, the TM-Array includes 879,348 SNPs and CNPs (Fig. 1).

Study populations

The major goal of the NHLBI REDS-III multicenter program is to conduct research that advances blood storage and TM practices and outcomes. One of the REDS-III components, referred to as the RBC-Omics study, enrolled 13,403 blood donors from many races/ethnicities.^{39,40} The study characterized fresh blood from the donated blood units (using complete blood count and ferritin assays) and end of storage (i.e., approx. 42-day stored) leukoreduced RBC component-derived samples derived from their blood donation (using spontaneous and stress hemolysis assays). While recruitment targeted high-intensity donors

and minority populations, including Asian, African American, and Hispanic donors, the US donor population most closely resembles this study population. Additionally, REDS-III has also conducted a cohort study of 2798 patients with SCD from six hemocenters in Brazil who are being followed longitudinally.⁴¹ High-quality DNA was obtained from both cohorts using white blood cells (WBCs) eluted from leukoreduced filters for the RBC-Omics study and whole blood samples for the Brazil SCD cohort.

RESULTS

Genotyping quality control

A total of 16,417 samples (13,168 of 13,403 blood donors, 2700 of 2798 patients with SCD, 36 biological replicates, and control samples including those from the 1000 Genomes Project⁴² were run on the TM-Array generating 14,436,256,116 genotypes. On average, more than 99.2% of biallelic markers were successfully called as high-resolution SNPs using the Affymetrix Axiom analysis pipeline. Only a few samples (<0.5%) were called with less than 95% of markers (Table 1).

The quality of genotypes was assessed by calling them twice using different permutations of sample batches. Inconsistent genotype calls were present in 0.08% of high-quality markers, of which 97.6% were no-call versus call. Only 5704 markers (0.7%) showed inconsistencies among five or more people. These markers were excluded from subsequent evaluation. Further examination of genotype calling metrics using SNPfisher off-target variant (OTV) caller⁴³ revealed that the cluster plots provided good cluster resolution for 833,872 biallelic SNP markers (95.2% of all designed markers).

Genome-wide marker coverage evaluation

Imputation was conducted using IMPUTE2⁴⁴ using the 1000 Genomes project phase 3 (1000G_p3) panel as the reference. We evaluated the postimputation genome-wide coverage for each race/ethnicity bin. Evaluations were conducted in two subsets of all 1000G_p3 variants: 1) variants with MAF of 5% or greater, and 2) variants with MAF of 1% or greater (MAFs estimated in the study population restricted to the corresponding race/ethnicity categories). Markers were considered covered through imputation if the IMPUTE2 info score was greater than 0.8. Overall, the array attained 94.7%, 97.5%, 96.1%, and 96.1% genomewide coverage of variants down to MAF of 5% for Asian, African American, white, and Hispanic, respectively (Table 2A). By chromosome, coverage ranged from 91.9% to 96.1% for Asian, from 95.2% to 98.4% for African American, from 93.7% to 97.5% for white, and from 94.0% to 97.8% for Hispanic (Table 2A). For variants with MAF down to 1%, the array attained 88.8%, 96.2%, 92.4%, and 94.6% genome-wide coverage for Asian, African American, white, and Hispanic populations, respectively (Table 2B). By chromosome, coverage ranged from 84.0% to 91.1% for Asian, from 93.91% to 97.4% for African American, from 86.6% to 94.7% for white, and from 92.3% to 96.0% for Hispanic. Data are also reported down to MAF of 1% (Table 2B).

Reproducibility evaluation

Within-array reproducibility—DNA extraction and genotyping were conducted twice on 36 subjects to evaluate the performance of the array and the reproducibility of genotype calling. For a single SNP, genotype calls were deemed as mismatch if the call results differed between the two replicates and neither genotyping resulted in no call. The median mismatch rate among the 36 replications was 0.029% (range, 0.013%-0.37%; 29 of the 36 samples lower than 0.1%). Additionally, the rate of difference between no call and call in the replicated samples was low (median rate, 0.41%, 3366 markers; 28 of the 36 samples lower than 1%).

Mendelian error checking—One trio from the 1000 Genomes Project (HG00418, HG00419, HG00420) and four retrospectively genetically identified RBC-Omics trios were used to assess Mendelian error rates. On average, fewer than 200 (0.024%) Mendelian errors were identified in each trio. A total of 1306 (0.16%) markers failed Mendelian checking in any trios.

Marker-based CNP calling for D antigen

The Rh blood group system contains many antigens. The D-negative phenotype when it occurs in a European population most often results from the homozygous deletion of the RHD gene, although other mutations are also known. Traditionally, it has been hard to tag the deletion directly using marker-based techniques. Therefore, determination of D-negative status by DNA-based methods includes the use of SNP markers that are in linkage disequilibrium with this deletion or the use of polymerase chain reaction-based methods.^{45–47} In the TM-Array, probes were designed specifically for typing the deletion. Approximately 17% of D-negative participants were observed in the RBC-Omics population, a close resemblance of the US donation population. Furthermore, the typing method achieved 99.6% accuracy compared to serology testing (Fig. 2). Thus, the TM-Array represents a potential low-cost, high-throughput D typing method.

Blood typing variant assessment

More than 700 variants are documented in the BGMUT database for 33 blood group systems.^{34,35} Blood donors enrolled in the RBC-Omics study had serologic ABO typing and D typing available. Numerous other serology typing results were available for the Brazil SCD cohort patients. High concordance (98.18%) was observed comparing genetically inferred ABO typing based on the five most informative variants (c.261delG, 467C > T, 703G > A, 802G > A, 1096G > A) to the serology testing results (Table 3). Allele frequencies are close to the expected population frequency for Brazilian and US donor populations for blood group markers (Table 4). Even though rare variants are hard to call due to the lack of rare homozygotes or heterozygotes around which to base a cluster, variants with low (<1%) MAF on the TM-Array were genotyped and clustered well, assessed based on clustering statistics and congruency with expected population frequencies (Table 4). A full list of blood group variants included in the array design are documented in Table S1, available as supporting information in the online version of this paper.

Sickle cell anemia variants assessment

A total of 12 pairs of probes were designed to capture HbS and HbC alleles, which are 1 bp apart (rs334, chr11:5227002, T = sickle S; rs33930165, chr11:5227003, A = sickle C). A supervised clustering method with cross validation resulted in an approximately 96% concordance rate within the Brazil SCD population. An unsupervised k-means clustering method with K = 3 applied to genotype calling and signal intensity data resulted in an approximately 93% concordance rate compared to clinical diagnosis. An example probe intensity plot is shown in Fig. S1, available as supporting information in the online version of this paper. The supervised clustering model was applied to RBC-Omics to predict the population prevalence of sickle cell trait (person who carries an HbS mutation) in the United States. Approximately 0.69% of the RBC-Omics participants carry a sickle cell trait.

DISCUSSION

A TM-Array for transfusion medicine discovery research has been tailor-made following extensive literature review, consultation with transfusion experts, and bioinformatics mining. The array includes transfusion medicine-relevant variants with low MAF (<5%). This overcomes a major limitation of array-based SNP assays, which is the limited number of variants an array can contain; thus, rare or uncommon variants or multiple markers needed for comprehensive determination of, for example, ABO and RH are not well covered. GWAS-based discovery of a causal variant not directly represented on the array is made by association of markers in linkage disequilibrium (LD) with the causal variant and the phenotype of interest. This imputation not only reduces statistical power to detect the variant, but also biases analyses toward identification of causal variants with small effect size that are in high LD with markers on the array, versus causal variants with large effect size but in low LD with markers on the array.⁴⁸

The design of the TM-Array puts emphasis on transfusion medicine-relevant genetics and provides dense variant coverage down to MAF of 1% in transfusion medicine-relevant gene regions such as blood group genes (e.g., ABO, MNS, and RHD/RHCE). For example, KEL 578 T > C (rs8176058) associates with the Kell system K+/- phenotype. However, because of the low T allele frequency (approx. 0.006), this SNP is omitted in most genome-wide genotyping arrays, and cannot be imputed accurately for the same reason. Another example is the sickle cell HbS mutation (rs334), whose MAF ranges from 0.004 to 0.027 depending on the reference population. These variants are important to transfusion medicine but would not be available in any array design unless specifically targeted.

In addition to known variants, low-MAF variants are covered in high density among gene regions that may have transfusion implications. This design empowers the discovery of causal variants, even for the ones with low MAF, for transfusion medicine endpoints. Specifically, the array included variants designed to tag CNPs in the RHD region, which genetically determined a subject's D typing in high concordance with serology results. Additionally, a commercialized Axiom HLA module that provides 2- and 4-digit resolution, with 4-digit resolution accuracy of 88.6% to 98.4%,⁴⁹ was integrated into the array design because HLA antigens and antibodies are known to cause some adverse transfusion outcomes.⁵⁰ Recent studies have also shown that mitochondrial genetics may be responsible

for some transfusion outcomes.^{51,52} Therefore, a mitochondrial module of 472 SNPs was also integrated into the design of the TM-Array.

Although the TM-Array was carefully designed and tailored for transfusion medicine, there is room for further refinement. First, multiallelic markers were incorporated into the design of the array. However, we have not yet been able to perform a reliable global genotype calling for those multiallelic markers. A few important multiallelic variants were called manually using ad hoc methodologies through examination of the clustering plot and signal intensity, but the majority of multiallelic markers in the array have been excluded from analyses to date. Excluding these multiallelic markers would free up space for any additional content on the TM-Array. Second, variants for the GWAS booster were selected using 1000 Genomes Phase 1 release as the reference population. Imputation coverage for both East Asian and African American were higher than our design (Table 2) when using the 1000G_p3 release as imputation reference, notably a boost from 91% in design to over 97% for African American, because of the overrepresentation of the populations added in the reference. Future redesigning the GWAS booster based on 1000G_p3 may free up space for populations not specifically targeted by this array, such as South Asians and Native Americans, or to further boost overall imputation coverage. Third, the 12 probes designed to capture HbS and HbC variants that are 1 bp apart achieved only a 93% to 96% concordance rate, suggesting the current design is not yet suitable for classification of SCD genotypes in clinical settings. However, as shown in Fig. S1, array technology is capable of distinguishing two clinically significant variants by adding signal intensity to the genotype clustering; thus, it is suitable as an alternative low-cost and high-throughput screening method, although further refinement of the design and development of genotype calling algorithms would be necessary.

This array shared a common GWAS grid backbone with the UK Biobank Array, which was employed in the UK Interval Study of blood donors to investigate factors relevant to safe donation frequency, donor complications, and health outcomes.¹⁴ This makes it possible to combine data for imputation. The TM-Array provided over 96% genome-wide coverage for African American, Hispanic, and white populations and over 94% for Asians. This not only enables genomics profiling of minority (and majority) populations at scale, but also makes it suitable for genotyping admixture populations of both blood donors and transfusion recipients. Although it is generally believed that racial and ethnic differences between a donor and a recipient do not impact transfusion outcome(s), donor genetics likely impact RBC storage properties^{39,40} and potentially RBC transfusion efficacy, and mismatches of genetic variants between a donor and cognate recipient as a result of differing haplotype background might contribute to undesirable outcomes. For example, in patients with SCD, who often require chronic blood transfusion, induction of alloantibodies from exposure to genetically dissimilar RBCs complicates their future blood group matching.^{53,54} Using a single platform for transfusion medicine-related GWAS will make experimental design easier for the investigation of potential risk factors or to adjust for any variants in statistical tests. Additionally, recent studies have shown that transethnic meta-analyses increase the power to detect novel loci and improve resolution of fine mapping, not only through increased sample size but also by leveraging differences in local linkage disequilibrium structure between ethnic groups.^{55,56}

In addition to being used in hypothesis generation studies such as GWAS, the array can be used in a diverse set of transfusion medicine-related studies. Because it is constructed to be broadly informative across diverse populations, it can be applied to study sickle cell anemia even in a racially mixed population such as in the Brazil SCD cohort. The INTERVAL trial,²⁸ which enrolled approximately 45,000 blood donors, identified novel variants associated with blood cell traits⁵⁷ demonstrating the use of blood donors as healthy controls.^{58,59} This array can be used to identify novel variants among diverse populations or to prioritize the European signals for functional study through replication effort. Additionally, this array could be developed into a clinical diagnostic tool, potentially through slight modification or subsetting, that leverages existing and ongoing studies to assess donor or recipient risks. For example, it may be proven feasible to use subsets of the probes on the array to construct a low-cost genotyping assay with utility for clinical application to either enhance serologic testing or potentially replace it.

CONCLUSIONS

The TM-Array is performing well, with over 99% of SNPs being called across 13,168 blood US donors in the REDS-III/RBC-Omics study and 2700 Brazilian patients in the REDS-III Brazil SCD cohort. Individually, 99.8% of samples have greater than 95% of SNPs being called. Of the biallelic SNPs, 95.2% (833,872 of 875,837) are being called with high confidence; this includes the ability to type rare markers accurately. The TM-Array is a powerful tool for studying blood- and transfusion-related disorders in many ethnic groups. The array has produced high-quality genotype calls across the first 16,000 individuals screened. The number of genotypes per sample called is high, and genotype calling is accurate. We have allowed the content of this array to be made available to the wider community as a commercially available tool that is supported for quality control/quality assurance by the manufacturer (Thermo Fisher Scientific). Most of the genotyped data will become available through dbGaP.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors wish to acknowledge NHLBI Recipient Epidemiology and Donor Evaluation Study-III (REDS-III), which was supported by NHLBI contracts NHLBI HHSN2682011-00001I, -00002I, -00003 L, -00004 L, -00005I, -00006I, -00007I, -00008I, and - 00009I. The authors would like to express their deep gratitude to Dr. Simone Glynn of NHLBI for her outstanding support throughout this study, to the RBC-Omics research staff at all participating blood centers and testing labs for their exceptional performance and contribution to this project, and to all blood donors and patients with sickle cell disease who agreed to participate in this study.

Y.G. and G.P.P. contributed to the design and evaluation of the array, wrote the manuscript, and performed data analyses and interpretation; M.B. and M.S. contributed to the design of the array, performed DNA extraction and genotyping, and edited and finalized the manuscript; S.M.E. assisted with oversight of data coordination from all participating sites; C.M.W., B.K., T. K., A.E.M., and S.K. contributed to the design of the array, specifically, provided expert opinion on selection of blood group antigen markers, RBC genes, and iron metabolism and pica and restless leg syndrome markers, and edited and finalized the manuscript. C.H. and B.C. contributed to selection of SCD markers design, A. B. and A.S.B. contributed to RBC genes design, and Y.L. contributed to the overall marker selection of the array

The REDS-III, RBC-Omics Study is the responsibility of the following persons: Hubs: A. E. Mast, J. L. Gottschall, W. Bialkowski, L. Anderson, J. Miller, A. Hall, Z. Udee, and V. Johnson, BloodCenter of Wisconsin, Milwaukee, WI; D. J. Triulzi, J. E. Kiss, and P. A. D'Andrea, Institute for Transfusion Medicine, Pittsburgh, PA; E. L. Murphy and A. M. Gultinan, University of California, San Francisco, San Francisco, CA; R. G. Cable, B. R. Spencer, and S. T. Johnson, American Red Cross Blood Services, Farmington, CT; data coordinating center: D. J. Brambilla, M. T. Sullivan, S. M. Endres, G. P. Page, Y. Guo, N. Haywood, D. Ringer, and B. C. Siegel, RTI International, Rockville, MD; central and testing laboratories: M. P. Busch, M. C. Lanteri, M. Stone, and S. Keating, Blood Systems Research Institute, San Francisco, CA; T. Kanas and M. Gladwin, Pittsburgh Heart, Lung, Blood, and Vascular Medicine Institute, Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA; steering committee chairman: S. H. Kleinman, University of British Columbia, Victoria, BC, Canada; National Heart, Lung, and Blood Institute, National Institutes of Health: S. A. Glynn, K. B. Malkin, and A. M. Cristman. The Brazil SCD cohort is the responsibility of the following persons: Hemocenters: A. B. Carneiro-Proietti, C. M. Teixeira, T. P. S. Silva, and R. A. Mota, Hemominas Foundation, Brazil; E. C. Sabino, Faculdade de Medicina, Brazil; L. Capuani, Pró-Sangue Foundation, Brazil; A. Araujo, and P. Loureiro, Hemope Foundation, Brazil; C. Máximo, and C. Ilobo, Hemorio Foundation, Brazil; M. V. Flor Park, ITACI, São Paulo, Brazil; data coordinating center: D. J. Brambilla, M. T. Sullivan, G. P. Page, Y. Guo, L. Preiss, and C. McClure, RTI International, Rockville, MD; central and testing laboratories: B. Custer and S. Kelly, Blood Systems Research Institute, San Francisco, CA.

Supported by NHLBI contracts NHLBI HHSN2682011-00001I, -00002I, -00003I, -00004I, -00005I, -00006I, -00007I, -00008I, and - 00009I.

CONFLICT OF INTEREST

A.E.M. receives research grant funding from Novo Nordisk. The other authors have disclosed no conflicts of interest.

ABBREVIATIONS:

BGMUT	blood group antigen gene mutation
CNPs	copy number polymorphisms
GWASs	genome-wide association studies
MAF	minor allele frequency
NHLBI	National Heart, Lung, and Blood Institute
RBC-Omics	Red Blood Cell Omics
REDS-III	Recipient Epidemiology Donor Evaluation Study-III
SCD	sickle cell disease
SNP	single-nucleotide polymorphism

REFERENCES

1. Assal A, Coste J, Barlet V, et al. Application of molecular biology to blood transfusion safety: the nucleic acid testing. *Transfus Clin Biol* 2003;10:217–26. [PubMed: 12798861]
2. Peugh WN, Wood KJ, Morris PJ. Genetic aspects of the blood transfusion effect. *Transplantation* 1988;46:438–43. [PubMed: 3047933]
3. Reid ME. Transfusion in the age of molecular diagnostics. *Hematol Am Soc Hematol Educ Program* 2009;2009:171–7.
4. von Dungern E, Hirschfeld L. Ueber vererbung gruppenspezifischer strukturen des blutes. *Zeits Immunforsch* 1910;1:284–92.
5. Landsteiner K Agglutination phenomena in normal human blood. *Wien Klin Wochenschr* 1901;14:1132–4.

6. Landsteiner K, Wiener AS. An agglutinable factor in human blood recognized by immune sera for rhesus blood. *Proc Soc Exp Biol Med* 1940;43:223–3.
7. Landsteiner K, Wiener AS. Studies on an agglutinin (Rh) in human blood reacting with anti-rhesus sera and with human isoantibodies. *J Exp Med* 1941;74:309–20. [PubMed: 19871137]
8. Feder JN, Tsuchihashi Z, Irrinki A, et al. The hemochromatosis founder mutation in HLA-H disrupts β 2-microglobulin interaction and cell surface expression. *J Biol Chem* 1997;272:14025–8. [PubMed: 9162021]
9. Howard TE, Marusa M, Boisa J, et al. The prothrombin gene 3'-untranslated region mutation is frequently associated with factor V leiden in thrombophilic patients and shows ethnic-specific variation in allele frequency. *Blood* 1998;91:1092–3. [PubMed: 9446674]
10. Kangsadalampai S, Board PG. The Val34Leu polymorphism in the A subunit of coagulation factor XIII contributes to the large normal range in activity and demonstrates that the activation peptide plays a role in catalytic activity. *Blood* 1998;92:2766–70. [PubMed: 9763561]
11. Taliaferro WH, Huck JG. The inheritance of sickle-cell anaemia in man. *Genetics* 1923;8:594–8. [PubMed: 17246028]
12. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science* 2005; 308:385–9. [PubMed: 15761122]
13. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42(D1):D1001–6. [PubMed: 24316577]
14. Astle WJ, Team UBTG. Epigenome CB. A GWAS of 170,000 individuals identifies thousands of alleles perturbing blood cell traits, many of which are in super enhancers setting cell identity. *Blood* 2016;128:2652.
15. Mtatiro SN, Singh T, Rooks H, et al. Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania. *PLoS One* 2014;9:e111464. [PubMed: 25372704]
16. Sankaran VG, Orkin SH. Genome-wide association studies of hematologic phenotypes: a window into human hematopoiesis. *Curr Opin Genet Dev* 2013;23:339–44. [PubMed: 23477921]
17. Sebastiani P, Solovieff N, Hartley SW, et al. Genetic modifiers of the severity of sickle cell anemia identified through a genome-wide association study. *Am J Hematol* 2010;85:29–35. [PubMed: 20029952]
18. Warren HR, Evangelou E, Cabrera CP, et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat Genet* 2017;49:403–15. [PubMed: 28135244]
19. Oetting WS, Schladt DP, Guan W, et al. Genomewide association study of tacrolimus concentrations in African American kidney transplant recipients identifies multiple CYP3A5 alleles. *Am J Transplant* 2016;16:574–82. [PubMed: 26485092]
20. Sanghavi K, Brundage R, Miller M, et al. Genotype-guided tacrolimus dosing in African American kidney transplant recipients. *Pharmacogenomics J* 2017;17:61–8. [PubMed: 26667830]
21. Stapleton CP, Conlon PJ, Phelan PJ. Using omics to explore complications of kidney transplantation. *Transpl Int* 2017;31:251–62. [PubMed: 28892567]
22. Yang JY, Sarwal MM. Transplant genetics and genomics. *Nat Rev Genet* 2017;18:309–26. [PubMed: 28286337]
23. International Genetics & Translational Research in Transplantation Network (iGeneTRAiN). Design and implementation of the international genetics and translational research in transplantation network. *Transplantation* 2015;99:2401–12. [PubMed: 26479416]
24. Li YR, van Setten J, Verma SS, et al. Concept and design of a genome-wide association genotyping array tailored for transplantation-specific studies. *Genome Med* 2015;7:90. [PubMed: 26423053]
25. Rabbani B, Tekin M, Mahdih N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet* 2014;59:5–15. [PubMed: 24196381]
26. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779. [PubMed: 25826379]
27. Harmening DM, editors. *Modern blood banking & transfusion practices*. 6th ed. Philadelphia: F.A. Davis Company; 2012 p. 608.

28. Di Angelantonio E, Thompson SG, Kaptoge S, et al. Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. *Lancet* 2017;390:2360–71. [PubMed: 28941948]
29. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol* 2016;70:214–23. [PubMed: 26441289]
30. Hoffmann TJ, Zhan Y, Kvale MN, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics* 2011;98:422–30. [PubMed: 21903159]
31. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45:580–5. [PubMed: 23715323]
32. The AABB Standards Program Committee. Standards for molecular testing for red cell, platelet, and neutrophil antigens. 3rd ed. Bethesda, MD: American Association of Blood Banking;2016.
33. Robinson J, Halliwell JA, Hayhurst JD, et al. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Res* 2015;43(D1):D423–31. [PubMed: 25414341]
34. Patnaik SK, Helmberg W, Blumenfeld OO. BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. *Nucleic Acids Res* 2012;40(D1):D1023–9. [PubMed: 22084196]
35. Patnaik SK, Helmberg W, Blumenfeld OO. BGMUT database of allelic variants of genes encoding human blood group antigens. *Transfus Med Hemother* 2014;41:346–51. [PubMed: 25538536]
36. D'Alessandro A, Righetti PG, Zolla L. The red blood cell proteome and interactome: an update. *J Proteome Res* 2010;9:144–63. [PubMed: 19886704]
37. Thomas A, Rahmanian S, Bordbar A, et al. Network reconstruction of platelet metabolism identifies metabolic signature for aspirin resistance. *Sci Rep* 2014;4:3925. [PubMed: 24473230]
38. Bordbar A, Jamshidi N, Palsson BO. iAB-RBC-283: a proteomically derived knowledge-base of erythrocyte metabolism that can be used to simulate its physiological and pathophysiological states. *BMC Syst Biol* 2011;5:110. [PubMed: 21749716]
39. Kanas T, Lanteri MC, Page GP, et al. Ethnicity, sex, and age are determinants of red blood cell storage and stress hemolysis: results of the REDS-III RBC-Omics study. *Blood Adv* 2017;1:1132–41. [PubMed: 29034365]
40. Endres-Dighe SM, Guo Y, Kanas T, et al. Blood, sweat and tears: Red Blood Cell-Omics study objectives, design, and recruitment activities. *Transfusion* 2019;59:46–56. [PubMed: 30267427]
41. Carneiro-Proietti AB, Kelly S, Teixeira CM, et al. Clinical and genetic ancestry profile of a large multi-centre sickle cell disease cohort in Brazil. *Br J Haematol* 2018;182:895–908. [PubMed: 30027669]
42. Siva N 1000 Genomes Project. *Nat Biotechnol* 2008;26:256.
43. Didion JP, Yang H, Sheppard K, et al. Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. *BMC Genomics* 2012;13:34. [PubMed: 22260749]
44. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome wide association studies. *PLoS Genet* 2009;5:e1000529. [PubMed: 19543373]
45. Tounta G, Vrettou C, Kolialexi A, et al. A multiplex PCR for non-invasive fetal RHD genotyping using cell-free fetal DNA. *In Vivo* 2011;25:411–7. [PubMed: 21576416]
46. Wagner FF. RHD PCR of D-negative blood donors. *Transfus Med Hemother* 2013;40:172–81. [PubMed: 23922542]
47. Yankowitz J, Li S, Murray JC. Polymerase chain reaction determination of RhD blood type: an evaluation of accuracy. *Obstet Gynecol* 1995;86:214–7. [PubMed: 7617351]
48. Visscher PM, Brown MA, McCarthy MI, et al. Five years of GWAS discovery. *Am J Hum Genet* 2012;90:7–24. [PubMed: 22243964]
49. Khor S-S, Yang W, Kawashima M, et al. High-accuracy imputation for HLA class I and II genes based on high-resolution SNP data of population-specific references. *Pharmacogenomics J* 2015;15:530–7. [PubMed: 25707395]

50. Navarrete CV. The HLA system in blood transfusion. *Baillieres Best Pract Res Clin Haematol* 2000;13:511–32. [PubMed: 11102274]
51. Lee Y-L, King MB, Gonzalez RP, et al. Blood transfusion products contain mitochondrial DNA damage-associated molecular patterns: a potential effector of transfusion-related acute lung injury. *J Surg Res* 2014;191:286–9. [PubMed: 25039013]
52. Yasui K, Matsuyama N, Kuroishi A, et al. Mitochondrial damage-associated molecular patterns as potential proinflammatory mediators in post-platelet transfusion adverse effects. *Transfusion* 2016;56:1201–12. [PubMed: 26920340]
53. Zimring JC, Welniak L, Semple JW, et al. Current problems and future directions of transfusion-induced alloimmunization: summary of an NHLBI working group. *Transfusion* 2011;51:435–41. [PubMed: 21251006]
54. Williamson LM, Devine DV. Challenges in the management of the blood supply. *Lancet* 2013;381:1866–75. [PubMed: 23706803]
55. Mägi R, Horikoshi M, Sofer T, et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet* 2017;26:3639–50. [PubMed: 28911207]
56. Morris AP. Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* 2011;35:809–22. [PubMed: 22125221]
57. Astle WJ, Elding H, Jiang T, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* 2016;167:1415–1429.e19. [PubMed: 27863252]
58. Ji S-G, Juran BD, Mucha S, et al. Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. *Nat Genet* 2017;49:269–73. [PubMed: 27992413]
59. Singh T, Walters JT, Johnstone M, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. *Nat Genet* 2017;49:1167–73. [PubMed: 28650482]

Fig. 1.

Key content modules included in the TM-Array. Marker modules are divided into those in common with the UK Biobank Axiom Array, genome-wide booster for Asian and African American population that is customized for this array, Affymetrix commercially available content, and those transfusion medicine-related SNPs, CNPs, and genes along with module descriptions and counts in each category. AA, African American; Affy, Affymetrix; GWA, genome-wide associated; Mt, mitochondria; RBCs, red blood cells; TGF, transforming growth factor. [Color figure can be viewed at wileyonlinelibrary.com]

Fig. 2.

Evaluation of CNP calling through comparison with serological records of Rh antigens. D antigen and the principle component analysis (PCA) abstracted RHD gene CNP results. Forty-one D-negative and 13 D-positive subjects were classified differently comparing marker-based and serologic testing (concordance rate 99.6%).

TABLE 1.

Comparison of genotype calling of biallelic markers stratified by race/ethnicity

Study	Group ID	Average number of SNPs called (percentage [*])	Number of Samples < 98% SNPs called	Number of Samples < 95% SNPs called	Number of Samples < 90% SNPs called
RBC-Omics	AFR AMRCN	833,339.46 (99.24%)	63 (4.08%)	1 (0.06%)	0 (0%)
	ASIAN	834,021.29 (99.32%)	58 (3.65%)	5 (0.31%)	0 (0%)
	CAUC	833,788.97 (99.29%)	294 (3.59%)	17 (0.21%)	0 (0%)
	HISP	833,547.8 (99.27%)	35 (3.51%)	0 (0%)	0 (0%)
	OTHER	833,684.78 (99.28%)	19 (3%)	0 (0%)	0 (0%)
SCD	SCD	833,740.24 (99.29%)	469 (17.09%)	23 (0.84%)	0 (0%)

Over 99% of samples have call rates greater than 95% among all 875,837 biallelic markers.

^{*} SNP call rate generated using PLINK, where a total of 839,711 markers were included.

SNPs = single-nucleotide polymorphisms.

TABLE 2.

Chromosome imputation coverage (with $r^2 \geq 0.8$) of the array stratified by racial group. 2A. Imputation coverage for variants with MAF greater than or equal to 5%. 2B. Imputation coverage for variants with MAF down to 1%

2A						
MAF	0.05	East & South Asian	African American	White	Hispanic	Other
chr1	94.62%	97.35%	96.04%	96.48%	96.07%	
chr2	95.30%	98.11%	96.98%	97.37%	96.99%	
chr3	95.67%	98.28%	97.29%	97.59%	97.23%	
chr4	95.44%	97.75%	96.83%	97.18%	96.79%	
chr5	95.93%	98.44%	97.45%	97.79%	97.39%	
chr6	96.07%	98.36%	97.49%	97.71%	97.47%	
chr7	95.05%	97.55%	96.47%	96.99%	96.49%	
chr8	95.74%	98.14%	97.13%	97.46%	97.12%	
chr9	94.31%	96.72%	95.67%	96.12%	95.67%	
chr10	95.26%	97.82%	96.83%	97.21%	96.78%	
chr11	95.68%	98.36%	97.32%	97.51%	97.20%	
chr12	95.29%	98.17%	96.93%	97.38%	96.91%	
chr13	95.87%	98.40%	97.48%	97.84%	97.38%	
chr14	94.36%	96.81%	95.63%	96.08%	95.68%	
chr15	94.05%	96.70%	95.44%	95.95%	95.44%	
chr16	94.31%	97.33%	95.77%	96.40%	95.90%	
chr17	94.01%	97.24%	95.50%	96.25%	95.61%	
chr18	95.32%	97.90%	96.86%	97.34%	96.82%	
chr19	93.83%	96.85%	95.36%	96.02%	95.28%	
chr20	95.24%	98.29%	96.96%	97.43%	96.97%	
chr21	92.48%	95.77%	94.00%	94.47%	94.02%	
chr22	91.89%	95.18%	93.66%	94.01%	93.59%	
chr23	93.25%	97.44%	94.19%	95.34%	95.22%	
Overall	94.65%	97.50%	96.17%	96.65%	96.20%	
2B						
MAF	0.01	East & South Asian	African American	Caucasian	Hispanic	Other

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

2A		MAF 0.05	East & South Asian	African American	White	Hispanic	Other
	chr1		88.57%	96.00%	92.25%	94.23%	93.46%
	chr2		89.28%	97.00%	93.56%	95.36%	94.49%
	chr3		90.17%	97.30%	94.11%	95.71%	94.92%
	chr4		89.73%	96.69%	93.55%	95.20%	94.25%
	chr5		89.96%	97.36%	94.15%	95.86%	94.93%
	chr6		91.13%	97.38%	94.66%	96.08%	95.27%
	chr7		89.47%	96.35%	93.15%	94.94%	93.97%
	chr8		89.62%	97.02%	93.71%	95.50%	94.61%
	chr9		88.39%	95.18%	92.22%	94.01%	93.05%
	chr10		89.60%	96.52%	93.54%	95.28%	94.29%
	chr11		90.05%	97.23%	94.10%	95.67%	94.88%
	chr12		89.78%	96.94%	93.59%	95.50%	94.50%
	chr13		90.13%	97.11%	94.07%	95.92%	94.75%
	chr14		89.14%	95.69%	92.35%	94.36%	93.51%
	chr15		88.66%	95.39%	91.59%	93.89%	93.05%
	chr16		88.15%	95.85%	91.70%	94.10%	93.23%
	chr17		88.12%	95.84%	91.47%	94.23%	93.04%
	chr18		89.49%	96.28%	93.24%	95.15%	94.03%
	chr19		88.12%	95.54%	91.52%	93.86%	92.80%
	chr20		89.21%	96.96%	92.88%	95.08%	94.13%
	chr21		87.09%	94.69%	89.89%	92.42%	91.64%
	chr22		87.07%	93.91%	90.24%	92.33%	91.52%
	chr23		83.98%	95.87%	86.62%	92.45%	91.57%
	Overall		88.80%	96.22%	92.42%	94.58%	93.64%

MAF = minor allele frequency.

TABLE 3.

Cross-tabulation of genetically inferred ABO typing with serologically derived ABO blood types show high concordance (> 98%) between the two methods, suggesting high genotyping quality (n = 12,879). All RBC-Omics participants were included in this analysis. Haplotype from five variants (c.261delG, 467C > T, 703G > A, 802G > A, 1096G > A) were used for genetical inference

Genetically inferred ABO	Typed ABO			
	A	AB	B	O
A	4578	0	160	34
AB	3	526	1	1
B	5	3	1789	7
O	14	1	6	5751

Example blood typing markers and their minor allele frequencies show high concordance with expected population frequency, even for variants with low MAF, suggesting the array is performing well for both common and rare variants

TABLE 4.

Gene	MAF Brazil	MAF RBC-Omits	1000G MAF	SNP rsID	Variant	Blood Group
ABO	0.114	0.122	0.1530	rs8176743	c.703G > A	B blood group
ABO	0.0133	0.0179	0.0106	rs41302905	c.802G > A	B blood group
ABO	0.301	0.351	0.344	rs8176719	c.261delG	O blood group
ACKR1	0.461	0.114	0.2664	rs2814778	c.-67 T > C	Duffy Null
FUT1	0.0210	0.0453	0.0998	rs2071699	c.35C > T	Bombay
SLC14A1	0.378	0.462	0.4109	rs1058396	c.838G > A	Kidd
KEL	0.00510	0.00664	0.0052	rs8176059	c.841C > T	Kel
SLC4A1	0.0131	0.00440	0.0130	rs2285644	c.2561C > T	Diego

MAF = minor allele frequency; SNP = single-nucleotide polymorphism.