JCI The Journal of Clinical Investigation

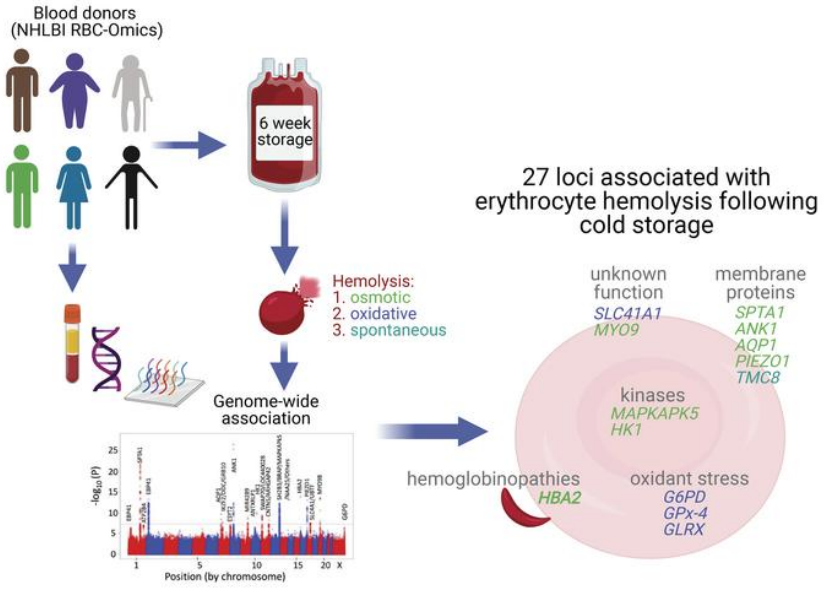
Multi-ancestry genome-wide association study identifies 27 loci associated with measures of hemolysis following blood storage

Grier P. Page, Tamir Kanias, Yuelong John Guo, Marion C. Lanteri, Xu Zhang, Alan E. Mast, Ritchard G. Cable, Bryan R. Spencer, Joseph E. Kiss, Fang Fang, Stacy M. Endres-Dighe, Donald Brambilla, Mehdi Nouraie, Victor R. Gordeuk, Steve Kleinman, Michael P. Busch, Mark T. Gladwin

J Clin Invest. 2021. https://doi.org/10.1172/JCI146077.

Clinical Research and Public Health | In-Press Preview | Genetics

Graphical abstract



Multi-ancestry genome-wide association study identifies 27 loci associated with measures of hemolysis following blood storage

Grier P. Page,[1]* Tamir Kanias,[2]* Yuelong John Guo,[3] Marion C. Lanteri,[4] Xu Zhang,[5] Alan E. Mast,[6] Ritchard G. Cable,[7] Bryan R. Spencer,[8] Joseph E. Kiss,[9] Fang Fang,[3] Stacy M. Endres-Dighe,[10] Don Brambilla,[10], Mehdi Nouraie,[11] Victor R. Gordeuk,[12] Steve Kleinman,[13] Michael P. Busch,[14]** Mark T. Gladwin,[15]** for the National Heart, Lung, and Blood Institute (NHLBI) Recipient Epidemiology Donor Evaluation Study (REDS)-III Program

* Shared first authors

** Shared senior authors

[1]Division of Biostatistics and Epidemiology, RTI International, Atlanta, GA

[2]Vitalant Research Institute, Denver, CO

[3]Division of Biostatistics and Epidemiology, RTI International, Durham, NC

[4] Vitalant Research Institute, San Francisco, CA

[5]Department of Medicine, University of Illinois, Chicago, IL

[6]Blood Research Institute, Blood Center of Wisconsin, and Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI

[7]American Red Cross, Farmington, CT

[8]American Red Cross, Dedham, MA

[9]Vitalant Northeast Division, Pittsburgh, PA

[10]Division of Biostatistics and Epidemiology, RTI International, Bethesda, MD

[11]Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Pittsburgh, PA

[12]Department of Medicine, University of Illinois, Chicago, IL

[13]University of British Columbia, Victoria, BC, Canada

[14]Vitalant Research Institute, San Francisco, CA

[15]Pittsburgh Heart, Lung, and Blood Vascular Medicine Institute, University of Pittsburgh; Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, UPMC

Correspondence

Grier P. Page, PhD
RTI International
2987 Clairmont Rd Ste 400
Atlanta, GA 30340
Phone: 770-407-4907
Email: gpage@rti.org
and
Mark T. Gladwin, MD
University of Pittsburgh
1218 Scaife Hall, 3550 Terrace Street
Pittsburgh, PA 15213
Phone: 412-648-9641
Email: gladwinmt@upmc.edu

Disclosures: All authors have declared any conflicts of interest according to journal and ICMJE policy (see disclosure forms for individual authors). The following authors declare no relevant conflicts: DB, RC, SE, FF, YG, TK, SK, ML, AM, MN, BS, and XZ. MB declares NHLBI REDS salary support. MTG declares industry consulting relationships with Pfizer, Bayer, Fulcrum, and Novartis (broadly relevant to SCD or pulmonary/vascular disease). VG declares institutional grants or contracts from CSL Behring, Novartis, Global Blood Therapeutics, Imara, and Ironwood, and DSMB or advisory board participation with CSL Behring, Novartis, Global Blood Therapeutics, and Forma. JK declares NHLBI salary/study support and meeting expenses. GPP declares honorarium or payment for study section service for NIH and DSMB service for Arbor Research and University of Michigan.

ABSTRACT

The evolutionary pressure of endemic malaria and other erythrocytic pathogens has shaped variation in genes encoding erythrocyte structural and functional proteins, influencing responses to hemolytic stress during transfusion and disease. We sought to identify such genetic variants in blood donors by conducting a genome-wide association study (GWAS) of 12,353 volunteer donors, including 1,483 African Americans, 1,477 Asians, and 960 Hispanics, whose stored erythrocytes were characterized by quantitative assays of in vitro osmotic, oxidative, and cold-storage hemolysis. GWAS revealed 27 significant loci ($p<5\times10^{-8}$), many in candidate genes known to modulate erythrocyte structure, metabolism, and ion channels, including *SPTA1*, *ALDH2*, *ANK1*, *HK1*, *MAPKAPK5*, *AQP1*, *PIEZO1*, and *SLC4A1/Band 3*. GWAS of oxidative hemolysis identified variants in antioxidant enzymes including *GLRX, GPX4, G6PD*, and a novel golgi-transport protein *SEC14L4*. Genome wide significant loci were also tested for association with the severity of steady state (baseline) in vivo hemolytic anemia in patients with sickle cell disease, with confirmation of identified SNPs in *HBA2*, *G6PD*, *PIEZO1*, *AQP1* and S*EC14L4*. Many of the identified variants, such as those in *G6PD*, have previously been shown to impair erythrocyte recovery after transfusion, associate with anemia, or cause rare Mendelian human hemolytic diseases. Candidate SNPs in these genes, especially in polygenic combinations, may affect RBC recovery after transfusion and modulate disease severity in hemolytic diseases, such as sickle cell disease and malaria.

INTRODUCTION

Blood transfusion is one of the most common procedures during hospital stays, with more than 36,000 red blood cell (RBC) transfusions performed daily in the United States. Clinically, RBC transfusions are largely considered to be homogeneous. However, a growing number of studies have evaluated the potential impact of unique donor characteristics, such as sex, age and body mass index, on RBC storage integrity (1-3), post-transfusion recovery and survival of RBCs, and consequent clinical outcomes (4-7). In addition, the U.S. donor population is ethnically diverse, with hundreds of functionally and immunologically relevant RBC single nucleotide polymorphisms (SNPs) (8, 9).

Studies evaluating inbred mouse strains demonstrated strong heritable determinants of RBC susceptibility to canonical in vitro stressors such as cold storage hemolysis, osmotic hemolysis, and oxidative hemolysis; importantly, these in vitro responses also correlated with post-transfusion RBC recovery and function (4, 10, 11). In humans, in vitro hemolysis of donor RBCs in response to osmotic or oxidative stress is a reproducible and heritable trait that can be further modulated by factors such as donation history, ancestry, age, and sex (12, 13). Human studies of RBC recovery and survival following blood bank storage have demonstrated variability among donors that is reproducible over time, suggesting donor specific factors such as sickle cell trait (7) and glucose-6-phosphate dehydrogenase (G6PD) deficiency (14-16), can reduce post-transfusion RBC recovery (17). Indeed, in a recent study, the post transfusion RBC recovery was evaluated in 10 volunteers with G6PD deficiency using chromium-51 cell labeling. Recovery was 78.5% in G6PD-deficient subjects versus 85.3% for transfusion in 27 control subjects without G6PD ($P = 0.0009$) (16).

Genetic variability also contributes to the intensity of hemolysis observed in Mendelian hemolytic diseases, such as sickle cell anemia. In patients with homozygous hemoglobin (Hb) SS disease, there is a significant variability in the intensity of steady state or baseline hemolysis (18-21). Co-inheritance of α- and β-thalassemia and mutations modulating the expression of fetal hemoglobin (HbF) influence hemoglobin levels and hemolysis in patients with sickle cell disease (21, 22). Furthermore, the variability in severity of hemolysis influences clinical outcomes (23), promoting vasculopathy and the development of end-organ complications, such as pulmonary

hypertension, cutaneous leg ulceration and chronic kidney injury. We and others have demonstrated that cell-free hemoglobin released during hemolysis in the setting of sickle cell disease and transfusion of aged stored blood is toxic, driving nitric oxide depletion, oxidative injury, heme-mediated inflammation, and iron overload (19-21, 23, 24).

These findings inform a hypothesis that rare and common genetic variants modulate various characteristics of erythrocytes leading to altered susceptibility to hemolysis that may influence erythrocyte storage in blood banks, transfusion outcomes, and potentially the severity of hemolytic diseases. Considering this hypothesis, the aim of this study was to identify genes that modulate hemolysis in cold storage and hemolytic disorders by conducting a genome-wide association (GWA) study in RBC donors enrolled in the National Heart, Lung, and Blood Institute (NHLBI) RBC-Omics project (2, 25). We tested the associations between in vitro measures of stress hemolysis in cold stored RBCs (spontaneous storage hemolysis, osmotic fragility, and oxidative hemolysis) and high-density GWA SNPs (26) to discover candidate loci that regulate the function of human RBCs and their resilience to stress. This GWA cohort of 12,353 volunteer donors was enriched for groups with African, Hispanic, and Asian ancestry. Collected and stored RBCs were characterized by quantitative assays for in vitro osmotic, oxidative and cold-storage hemolysis. Consistent with the anticipated genetic variability in donor RBCs, our analysis identified 27 GWA significant loci ($p < 5\times10^{-8}$), many in candidate genes known to modulate erythrocyte structure, metabolism, and ion channels. We further verified whether the SNPs identified from our in vitro hemolytic stress phenotyping have relevance to hemolytic disease by analyzing their association with in vivo measurements of the severity of steady-state (outpatient) hemolytic anemia (low hemoglobin values and high indices of RBC hemolysis) in two cohorts of sickle cell disease (SCD) patients.

These studies suggest that the identification of genetic variables that modulate the stability of RBCs in storage, response to stressors, and the functional integrity of RBCs after transfusion could advance donor selection criteria and procedures and storage policies. Identification and removal of genetically susceptible RBC donors/units that rapidly degrade in storage (exclusion of “fragile” RBC donors/units) and selection of profiled “super donors” that might be stable for longer periods of storage or survive longer after transfusion could provide for

a precision transfusion medicine strategy, more advanced than current random sampling of donors and transfusion of RBC units irrespective of recipient disease status or short- or long-term transfusion requirements. In addition, the variants could provide information about risk and severity of hemolytic anemia in patients with hemolytic diseases, such as sickle cell, thalassemia and malaria, as well as advance the discovery of proteins and enzymes that modulate RBC function.

RESULTS

Population Ancestry of REDS-III RBC-Omics cohort

The RBC-Omics cohort included a diverse group of U.S. blood donors born in many (n=71) countries. Initially, groups were divided into continent ancestry groups; however, we have followed recent recommendations to divide the Hispanic (27, 28) and Asian ancestry groups into multiple subgroups based upon country of birth. Donors of Hispanic ancestry were divided into two groups: Mexican and Central American Hispanics (Supplemental Figures 1 and 2) and Caribbean Island Hispanics (Supplemental Figures 1 and 3). Donors of Asian ancestry were divided into East Asians and South Asians to reflect the diversity of these RBC-Omics subpopulations (27, 28). In total the REDS-III RBC-Omics populations (Figure 1A) were divided into seven ancestry groups that included non-Hispanic Whites (n = 7,586), East Asians (n = 1,049), South Asians (n = 257), Mexican and Central American Hispanics (n = 456), Caribbean Island Hispanics (n = 489), African Americans (n = 1,046) and "Other" participants (n = 1,336). "Other" participants is a heterogeneous group including all individuals that did not cluster within the other groups, but included people who self-identified as Native Americans, Native Hawaiians, Native Alaskans, multiple races, or were from countries like Iran and the Philippines. We also considered the entire RBC-Omics as a single group referred to as ALL Ancestries.

Genome Wide Association Studies of Osmotic, Oxidative, and Storage Hemolysis in Mega Analysis

The SNP-based heritability from LDScore regression for osmotic hemolysis was 0.348 (SE=0.062), and for oxidative hemolysis was 0.156 (SE=0.073). The heritability score for storage hemolysis was not different from zero. Genome wide analysis of 12,353 subjects from the REDS-III RBC-Omics cohort was conducted between 14.1 million genotyped and imputed SNPs for osmotic (Figure 1B), oxidative (Figure 1C), and cold storage hemolysis (Figure 1D). GWA analyses using ALL Ancestries samples identified 14, 4, and 2 genome-wide significant regions that were associated with osmotic, oxidative, and spontaneous cold storage hemolysis, respectively (Table 1). Q-Q plots (Supplemental Figure 4) did not exhibit any p-value inflation.

Genome wide analysis of osmotic hemolysis in the entire data set (ALL Ancestries) revealed that the genome-wide significant variants were in or close to several logical candidate genes known to modulate RBC structure and function, such as spectrin alpha chain, erythrocytic 1 (*SPTA1*/Band 1; p<1.01E-22), Ankyrin 1 (*ANK1*/Band2.1; p<5.85E-28), Aquaporin 1 (*AQP1*; p<4.23E-10), and Solute Carrier Family 4 Member 1 (*SLC4A1*/Band 3; p<3.62E-08) (Table 1). In addition, a number of novel GWA significant variations were found in metabolic enzymes (Hexokinase 1 - *HK1;* p<4.90E-11*),* stress kinases (*MAPKAPK5;* p<2.24E-13), ion channels (Piezo type mechanosensitive ion channel component 1; *PIEZO1;* p<4.04E-14), and other proteins, such as Myosin IXB (*MYO9B;* p<9.88E-15). Supporting the internal validity of these findings, many of these SNPs are in proteins known to cause RBC disorders such as spherocytosis (23), elliptocytosis (29), xerocytosis (30), and alpha-thalassemia (31).

GWA analysis of oxidative hemolysis identified genome wide significant SNPs in glucose-6-phosphate dehydrogenase (*G6PD*; p<2.66E-17), SEC14 Like 4 (*SEC14L4*; p<9.85E-10), glutaredoxin (*GLRX*; p<1.15E-12) and glutathione peroxidase 4 (*GPX4*; p<3.80E-14). *G6PD*, *GLRX*, and *GPX4* are all known to have roles in protecting cells from oxidative damage. Analysis of storage hemolysis (Figure 1D) identified only two genome wide significant loci; one on chromosome 8 distant more than 500 Kb from the nearest genes, and another on chromosome 17 (TMC8; P< 1.34E-08).

Ancestry-Specific GWA Results

Individual principal component analysis (PCA) defined ancestry group GWA revealed a high degree of overlap with the ALL Ancestries analysis; however, seven additional genome-wide significant loci were observed in genes such as *EYS* (P<3.20E-09), *HBB (*P<3.66E-10), *HBA2 (*P<2.90E-14), and *G6PD (*P<2.66E-17) within specific ancestry groups (Table 1) and in only some cases (*G6PD* and *HBA2*) were the results significant in the ALL Ancestries analysis. Several loci such as *GPX4* and *SEC14L4* were only significant when considered with ALL Ancestries groups together. Only studying hemolysis in ancestry specific and in combined analysis enabled the discovery of all 27 of these loci.

Identification and Bioinformatics Analysis of Variation

We identified 12 directly genotyped genome-wide significant ($p<5X10^{-8}$) non-synonymous variants (NSVs) for hemolysis measures in the entire population or in the ancestry specific groups, predicted using SIFT or PolyPhen2. *SPTA1* contains a NSV rs857725 (Lys1693Gln p<8.75E-21, Figure 2A). Notably, the marker for alpha-thalassemia (Figure 2B) deletion (chr16:223678) and the HbS variant modulated osmotic, oxidative, and spontaneous storage hemolysis (7). In *HBB*, the HbS variant (rs334, Glu7Val, p<3.66E-10) was significantly associated with osmotic hemolysis in the African American ancestry group (Figure 2C). For oxidative hemolysis, *SEC14L4* AX-83171224/rs9606739 (Arg124Gly, p<3.07E-09, Figure 2D) and *G6PD* rs1050828 (Val68Me, p<2.66E-17, Figure 2E) were significant NSVs, whereas for spontaneous storage hemolysis, *TMC8* rs7208422 (Asn306Ile p<1.23E-08, Figure 2F) was GWA significant.

Chromosome 8 had two non-overlapping genome-wide significant loci for osmotic hemolysis within *ANK1* (Figure 3A-D). The first locus is centered on rs4737010 (Figure 3A), and the second is 87 kb away and centered on the NSV rs34664882 (Ala114Val; Figure 3B). PolyPhen2and SIFT suggested that rs34664882 is deleterious. The SNP appears to have a large quantitative effect on osmotic hemolysis across multiple ancestry groups, accounting for 3.2% of the variation in osmotic hemolysis in the combined data set. The second GWA-significant locus near *ANK1* is centered on rs4737009, which is in the canonical binding motif for the MAZ and STAT5A transcription factors (Supplemental Figure 5). It is likely that both rs34664882 and rs4737009 are independent and functionally consequential mutations for osmotic hemolysis. Conditional GWA showed these loci (rs34664882 and rs4737009) are fully independent and each is genome wide significant conditional on the other locus. Additional conditional GWA suggested there may be two or more independent locus at *SEC14L4* and *PIEZO1* (data not shown).

Within *G6PD*, the rs1050828 Val68Met variant associated with oxidative hemolysis in this study is a common class III variant, also referred to as *G6PD* A-. Individuals with class III *G6PD* variants are susceptible to acute hemolytic anemia when their RBCs are exposed to oxidant stress (32). G6PD deficiency is a chromosome X linked disorder. Figure 2E shows that female heterozygotes have intermediate phenotypes for oxidant-induced hemolysis between the female homozygote groups, and the few (n=4) females A- homozygotes are similar to the male

hemizygote groups. This supports the observation that heterozygotes for many disorders potentially have altered or intermediary phenotypes (33).

Pathway and gene-set enrichment analysis identified three Bonferroni corrected significant groups for osmotic hemolysis: Spectrin associated cytoskeleton ($P_{bon} = 6.77 \times 10^{-4}$), Steiner erythrocyte membrane genes ($P_{bon} = 2.58 \times 10^{-3}$), and Nikolsky breast cancer 19p13 amplicon ($P_{bon} = 0.028$). For oxidative hemolysis, there were no gene sets significantly enriched after the Bonferroni correction.

Inference of Differential Expression

MetaXcan was used to infer expression patterns for all genes based on the genotypes that have been identified by GTEX as eQTLs. The inferred gene expression was correlated with spontaneous storage, osmotic, and oxidative hemolysis in the RBC-Omics cohort. Thirteen genes were predicted to be significantly ($p<0.05$) differentially expressed and significantly ($p<0.05$) associated with osmotic (n=11) or oxidative (n=2) hemolysis but not spontaneous storage hemolysis (n=0) (Table 2). Of these, ten were situated within one of the genome-wide significant regions, and two others were close (<700 kb). Most of the genes (*SLC4A1*, *SWAP70*, MFSD2B) found by MetaXcan were kinases, channels, and metabolic genes whose mechanisms could be affected by changes in gene expression (34-36). MetaXcan did not identify RBC membrane structural genes, such as *ANK1* and *SPTA1*, which is consistent with the previous observations that disease causative variations in genes coding for structural genes tend to be to gain or loss of function mutations, as opposed to changes in gene expression levels (37-39). The most significant SNP in *GLRX* (rs72785409; p=6.14E-48) is an eQTL for *GLRX* in whole blood based upon 15 cohorts in the eQTLGen database (40).

Polygenic Scores

We modeled the polygenic scores (PGS) by using data from 2/3 of the population, whereas data from the remaining third was used for validation. We found the pruning and thresholding model in osmotic hemolysis at $p<10^{-7}$ and $r^2<0.4$ to validate better than the best

LDPred score (correction of best LD Pruning = 0.173 versus best LDPred model = 0.0904; Supplemental Figures 6-9). According to these data for osmotic and oxidative hemolysis, pruning and thresholding is a more precise method of developing PGS scores than LDPred.

Table 3 highlights the correlation of each of the three hemolysis PGS within each ancestry group with the observed hemolysis measures. Within non-Hispanic White samples, the correlation with osmotic hemolysis was 0.221, which explained more of the variability in osmotic hemolysis than any single marker. The best model for oxidative hemolysis was in African American and Mexican/Central American Hispanics (MCAH) samples, where the PGS correlation is approximately 0.260. Some ancestry groups did not yield PGS scores because of small sample sizes or lack of markers with a p-value less than 10^{-7} when split for cross validation. To develop predictors within these groups, hemolysis measures by ancestry group were correlated with the non-Hispanic White PGS score. This revealed that an ancestry specific PGS score was more precise than those developed in other ancestry groups, even if the latter sample size is larger. Therefore, when possible, PGS should be developed in ancestry-appropriate groups; if not applicable, scores from other ancestry groups can be used but will give diminished precision.

Unlike single gene disorders where only a few people contain causal loci, for polygenic traits such as hemolysis everyone has a combination of alleles which increase or decrease hemolysis across all identified loci. For example, for the top 50 loci identified in the non-Hispanic White PRS score for osmotic hemolysis, all RBC-Omic donors are heterozygous for between 7 and 34 of the loci (mean$\pm$SD = 18.3$\pm$4.6). Thus, genetic factors modulated osmotic and oxidative hemolysis in all individuals.

Genetic Analysis of In vivo Hemolysis in the WALK-PhASST and PUSH SCD Cohorts

To test the hypothesis that the genetic findings obtained from in vitro stress hemolysis perturbations of cold-stored RBCs from healthy blood donors may also be relevant to the in vivo severity of steady-state hemolytic anemia in human diseases, the genome-wide significant SNPs identified in the 27 loci for each hemolysis GWA were then tested in two cohorts of patients with SCD (Walk-PHaSST and PUSH). Note that there were 232 significant SNPs within these 27 loci. The same SNPs were tested for association using an in vivo measure of intensity of steady-state

hemolytic anemia as a quantitative trait in the SCD patient cohorts. Considering our small sample size of SCD cohorts, we restricted our analysis to the 1000 SNPs that were common (MAF >0.05) in SCD cohorts and were imputed with good quality (r^2 >0.8). Results between in vitro and in vivo hemolysis were considered consistent if the initial GWA p-value was significant at the genome level ($p<5X10^{-8}$) and the p-value for the association in the two SCD cohorts was also significant ($p<0.05$).

Consistent results were found in seven regions, including four regions for osmotic hemolysis GWA and three of four regions from the oxidative hemolysis GWA ($p<0.05$; Table 4). Significant results were found for osmotic hemolysis on chromosomes 7 (*AQP1*), 12 (several genes), and 16 (*HBA2*, *PIEZO1*). Oxidative hemolysis was concordant for three of the four genome-wide significant loci including on chromosome 5 (*GLRX*), 22 (*SEC14L4*), and X (*G6PD*). Even using more conservative assessments, the *HBA2 and G6PD* loci were significant in the sickle cell cohorts with Bonferroni testing correction.

DISCUSSION

This study is the first genome-wide evaluation of in vitro RBC stress hemolysis in cold-stored samples from blood donors, with secondary assessment of GWA significant findings on the in vivo severity of baseline (steady-state) hemolytic anemia in SCD patients. Increased hemolysis is a hallmark of several diseases, including SCD, and is associated with worse transfusion outcomes, such as poor RBC recovery and increased rates of post-transfusion sepsis. This notion is supported by recent murine studies demonstrating mouse strain-specific susceptibility to RBC cold storage injury that correlate with post-transfusion RBC recovery and function (4, 7, 11). In addition to limiting storage time and reducing post-transfusion RBC recovery, hemolysis drives endothelial dysfunction and vascular injury. We and others have demonstrated that cell-free hemoglobin released during hemolysis in the setting of SCD and transfusion of aged stored blood is toxic, driving nitric oxide depletion, oxidative injury, heme-mediated inflammation, and iron overload (19-21, 23, 24).

We identified twenty loci that were genome-wide significant in all sample analysis ($p<5x10^{-8}$) for at least one of the hemolysis measures (Table 1). Many of the identified variants were concentrated in proteins known to cause human RBC disorders characterized RBC fragility such as dehydrated hereditary stomatocytosis (*PIEZO1*) (41, 42), spherocytosis *ANK1, SPTA1, SLC4A1* (23, 43), ellipto-poikilocytosis (*SPTA1*) (44), xerocytosis (*PIEZO1*) (30), alpha-thalassemia (*HBB*) (31), and spontaneous and oxidant stress–induced hemolytic anemia (*HK1 and G6PD*) (16, 45). Providing additional validity, many of the implicated SNPs have been associated with laboratory complete blood cell count measurements, such as reticulocyte counts (*SPTA1 and PIEZO*1) (46)*,* and other complete blood count indices *(G6PD)* (47). Consistent with relevance of our in vitro quantitative measures of stress hemolysis, the identified SNPs from the RBC donor GWAS cohort in alpha-thalassemia, *G6PD*, *PIEZO-1*, Aquaporin-1, *SEC14 Like 4 (SEC14L4),* and glutaredoxin *(GLRX)* were found to GWA significantly associate with hemoglobin and hemolytic lab indices in the blood of SCD patients.

In addition to genes known to alter RBC function and hemolytic propensity and promote disorders (e.g., spherocytosis and xerocytosis), we identified a number of genes not previously known to impact RBC function including *MYO9B*. We also identified seven loci, including HBB,

HBA, G6PD and EYS2, that were genome-wide significant in at least one non-Hispanic white ancestry group (African American, East Asian, South Asian, Caribbean Island Hispanic (CIH), Mexican/Central American Hispanic (MCAH), and Others, which includes multi-racial people, Alaska/Hawaiian/Native Americans, and Pacific islanders (Table 1) highlighting the importance of studying diverse populations to provide a more comprehensive evaluation of genetic factors which affect RBC hemolysis. The number of discoveries in the specific ancestry groups is fewer than in the non-Hispanic white population in part due to the lower power from reduced sample sizes in these populations. For some of the loci such as G6PD and HBB in African Americans, the effect is likely due to known variation in these genes such as the A- or HbS variant that are mostly absent in non-Hispanic white populations. This was not always the case; the specific variants identified in MCAH in EYS or rs118149920 on chromosome 13 are unlikely to be the causative variants but are likely to be in linkage disequilibrium with actual causative variants that could be on chromosomes of Native American ancestry, especially since the G allele at rs118149920 is absent in European and African populations, but common in Native American and Asian populations.

The validity of the identified regions in the current study of hemolysis in the RBC-Omics cohort is supported by four observations: a) the biological plausibility of the identified SNPs, with most in proteins known to cause RBC disorders such as dehydrated hereditary stomatocytosis (*PIEZO1*) (41), spherocytosis (*ANK1, SPTA1, SLC4A1* (23, 43), ellipto-poikilocytosis (*SPTA1*) (44), xerocytosis (*PIEZO1*) (30), alpha-thalassemia(*HBB*) (31), and spontaneous and severe non-spherocytic hemolytic anemia (*HK1*) (45); b) some of the SNPs have been associated with laboratory complete blood cell count measurements, such as reticulocyte counts (*SPTA1* and *PIEZO*1) (46)*,* and other complete blood count indices *(G6PD)* (47); c) MetaXcan (Table 2) finds that the variation in a number of GWA significant genes contain eQTL for those genes and that the genes' expression are associated with hemolysis measures; and d) the consistency of GWA findings with significant SNPs that modulate the severity of in vivo hemolysis in patients with SCD (Table 4).

There were a number of variants identified in RBC antioxidant enzymes. For example, the finding that genetic variations in the *GPX4* gene modulated oxidative hemolysis is of interest

because this enzyme has been linked to key regulatory pathways in erythropoiesis including erythroblast enucleation and reticulocyte maturation (48-50). With regard to antioxidant activity, *GPX4* neutralizes bioactive lipid hydroperoxides to lipid alcohols, thereby preventing iron dependent cell death or ferroptosis (51, 52). Metabolomics studies of the RBC storage lesion have demonstrated the formation and accumulation of inflammatory bioactive lipids (oxylipins; e.g. 12-hydroxyeicosatetraenoic acid) during cold storage (53, 54). Therefore, genetic mutations that compromise RBC *GPX4* function may contribute to transfusion-related oxidative injury and inflammatory reactions. There were also significant associations between X-linked G6PD A- (the V68M variant) and both in vitro oxidative hemolysis and the in vivo severity of hemolytic anemia in patients with sickle cell disease. The enzyme G6PD controls the pentose phosphate pathway dependent generation of reduced NADPH, necessary for reduction of intracellular glutathione. *G6PD* A- (V68M) is common in African Americans; approximately 11% of African American men are hemizygous for this SNP. The enzyme activity of G6PD A- in erythrocytes is moderately decreased, 10-23% of normal activity. Hemizygotes do not have chronic hemolysis but can undergo acute hemolysis if exposed to oxidant stress (32). The finding of more marked hemolysis in G6PD A- hemizygous and homozygous sickle cell disease persons under basal circumstances in this study would reflect the ongoing oxidative stress that sickle cell erythrocytes experience (55). As mentioned in the introduction, this variant has also been shown to directly relate to post-transfusion RBC recovery (16), highlighting the potential relevance of our GWAS findings to transfusion medicine outcomes.

Polygenic scores were developed for oxidative and osmotic hemolysis in several of the ancestral groups (Table 3) that were able to predict far more of the variance in hemolysis than any single SNP or gene locus alone. The application of Non-Hispanic Whites-developed PGS to other ancestry groups has enabled the calculation of a PGS when there is not sufficient power to develop an ancestry-specific PRS (56), although the transferability of PRS scores across ancestry groups should be viewed with caution (56-58). For example, in African Americans the correlation for PGS with oxidative hemolysis with an ancestry specific PRS was 0.259 but with the Non-Hispanic Whites PRS was only 0.103. When possible, ancestry specific PGS scores should be developed and used appropriately. PGS will be useful for leveraging the combined genetic effect

on individuals and can be combined with other clinical and omics data to gain insights into the pathways leading to RBC function. All individuals in the cohort have some combination of alleles across the loci contribution to the PGS score. For the top 50 loci identified in the Non-Hispanic Whites PRS score for osmotic hemolysis, no one, across all racial groups, contains minor alleles at fewer than 7 of the loci or more than 34. Thus, genetic variation contributes to variation in oxidative and osmotic hemolysis in all individuals.

We were unable to identify a true replication cohort for the in vitro hemolysis measures of the RBC-Omics cohorts since this is the first such study to explore stress hemolysis as a quantitative trait in a large donor population. Instead, we chose to test whether the RBC-Omics results generalize to in vivo levels of anemia and hemolysis in the Walk-PHaSST and PUSH SCD cohorts. We found that several variants and regions associated with in vitro hemolysis measures in the REDS-III RBC-Omics donor population were also significant for in vivo hemolysis measures within the SCD WALK-PhASST and PUSH cohorts, such as *HBA2, HBB, GLRX, AQP1, SEC14L4* (Table 4). These observations suggest that the in vitro stress measurements identified known and new variants that under the stress of human disease may modulate RBC biology. Such findings could lead to identifying rare variants that may modulate the outcomes of many hemolytic diseases. Consistent with this hypothesis, one of the variants identified, the A-G6PD deficiency, has been recently shown to reduce post-transfusion RBC recovery (16).

We propose that the identification of genetic variables that modulate the stability of RBCs in storage after response to stressors and the functional integrity of RBCs after transfusion could advance donor selection and storage policies and improve transfusion outcomes. Identification and removal of genetically susceptible “fragile” RBC donors/units that rapidly degrade in storage and selection of profiled “super donor” blood components that might be stable for longer periods of storage or survive longer after transfusion could provide for a precision transfusion medicine strategy, more advanced than current random sampling of donors and transfusion of RBC units irrespective of donor genotypes, recipient disease status or short- or long-term transfusion requirements.

Further studies are needed to understand the manner in which the genetic variation leads to changes in expression, protein, epigenome, metabolome and to understand the interaction

network that led to interindividual differences in hemolysis (59, 60). For example, we have studied the metabolomic changes induced by the G6PD A- variant associated with oxidative hemolysis and identified significant effects on the NADPH and glutathione-dependent detoxification pathways of oxidized lipids (61), which could lead to alterations in the dynamics of the RBC membrane.

The genetic information developed in this study is being used in new studies of donor-blood component-recipient outcomes to evaluate the consequences of some of the reported gene variants on transfusion efficacy in patients (62, 63). Additional effort is aimed at evaluating non-genetic factors (64, 65) that influence RBC recipient outcomes. Current studies are underway to advance the field of precision transfusion medicine via the development of a transfusion-specific microarray that would provide enhanced tools for the screening of blood donors.

METHODS

RBC-Omics Cohort:

The REDS-III RBC-Omics cohort donor recruitment and study design are described in detail in Endres-Dighe et al (25). Briefly, 13,403 whole blood donors over the age of 18 were recruited from December 2013 to December 2015 at four REDS-III blood centers. All subjects were healthy allogeneic blood donors who passed screening and were not anemic. Samples were excluded because of duplicate enrollment, low call rate (< 97%), sample swap, if blood donation quantity was not sufficient, and if markers of infectious disease were reactive. We analyzed only one relative per family, selected based on having the most complete data. The final informative sample size was 12,353. Institutional review board approvals were obtained at all institutions.

Evaluation of Donor Predisposition to In vitro Hemolysis:

Stored (39-42 days) leukocyte-reduced RBCs were evaluated for spontaneous (cold storage) and two stress hemolysis assays including osmotic fragility and oxidative hemolysis using 2'-azobis(2-amidinopropane) dihydrochloride (AAPH) as detailed elsewhere (2, 66). Each of the hemolysis measures is a quantitative trait on the range from 0% to 100% (osmotic mean = 28%, oxidative mean = 36%, and storage = 0.4%) (2).

Genotyping:

Samples were genotyped on a Transfusion Medicine microarray (TM-Array) (26) that contained a total 879,000 SNPs (genotype coverage of >90% for SNPs with Minor Allele Frequency (MAF) > 5%) for Non-Hispanic White, Hispanics, African Americans, and Asians. The data from this study is available in dbGAP as accession phs001955.v1.p1 (67). We used PLINK (68) to perform quality control for genotyped data to eliminate potential biases. Individuals for whom calculated genetic sex and self-reported gender differed as well as individuals with more than 3% missing genetic data were excluded. SNPs with genotype missing rates higher than 3% or failing a Hardy-Weinberg Equilibrium (HWE) validation ($p<1E-4$) in any ancestry group were excluded from the study. 811,782 SNPs passed these steps and were used for the imputation.

Imputation:

Statistical phasing was conducted by Shape-IT (69). Imputation was then conducted for each 2 Mb interval with 1Mb flanking regions on each side using Impute2 (70) and 1000 Genomes Project Phase 3 as reference haplotypes. Imputation results were further filtered by using an INFO score > 0.8 before conducting association analyses (70). The final high-quality dataset had 8.1 million SNPs at >5% MAF and 14 million SNPs at >1% MAF. We ran GWA analysis in all ancestry groups for the 14 million SNPs.

Ancestry:

Ancestry analysis was conducted in the RBC-Omics cohort with 1000 Genomes Phase 3 samples (Figure 1A, supplemental figures 1-3). Ancestry principle components (PCs) were calculated using Bioconductor package SNPRelate (71) in the entire data set and separately for participants within each genetic ancestry group. The RBC-Omics subjects were then divided into seven ancestry groups: non-Hispanic White, African American, East Asian, South Asian, Caribbean Island Hispanic (CIH), Mexican/Central American Hispanic (MCAH), and Others, which includes multi-racial people, Alaska/Hawaiian/Native Americans, and Pacific islanders for GWAS analyses.

GWA Study of Common Genetic Variation:

Association analyses were conducted using the software ProbABEL (72). In previous studies, we determined that sex, age, ancestry, and donation frequency were associated with the levels of storage, osmotic, and oxidative hemolysis and were thus were used as covariates as well as the first 10 ancestry PCs in our genetic analyses (2, 73). The distributions of osmotic and oxidative hemolysis were normally distributed, but storage hemolysis was skewed, thus we used log-transformed storage hemolysis values (supplemental figure 10).

We conducted two types of analysis: The first was an analysis of all subjects, called ‘ALL Ancestries’ in the Tables and Figures. In the second, we conducted individual ancestry GWA analyses for Non-Hispanic Whites, East Asians, South Asians, CIH, MCAH, and African Americans. A GWA analysis for this study is a multivariable linear model with p-value threshold of $5x10^{-8}$ defined as statistically significant for all GWA analyses (Table 1). Conditional GWA analysis

verified whether any of the significant loci were independent by incorporating the SNPs with the smallest p-value in a region as a covariate in the GWA model and testing the region 50 kb on each side of this SNP. Linkage disequilibrium score (LDSC) regression (74) was used to estimate the SNP-based heritability (h^2) of both osmotic and oxidative hemolysis.

Bioinformatic Analyses:

HaploReg v 4.1 (75) was used to annotate the genes nearest to the index SNPs. Version 1.3 of LocusZoom (76) was used, with 1000 Genome Phase 3 linkage disequilibrium (LD) estimation. MetaXcan was implemented to infer gene expression patterns based on genotyped and imputed SNPs from the REDS-RBC-Omics in 922 whole blood expression profiles from the Depression Gene Network (Table 2) (71). MetaXcan uses a combination of linear and multivariate linear models with a Bonferroni (based on number of genes) correct p-value of < 0.05 being the significance cut-off. FUMA (77) was used to explore the biological pathways and enriched gene-set related to osmotic and oxidative hemolysis using the p-values from the GWA results for all subject analyses. The curated gene-sets and go-terms tested were from the Molecular Signatures Database (MSigDB) in GSEA (78).

Polygenic Score for Hemolysis:

Polygenic scores (PGS) provide a quantitative metric of the magnitude of an individual's inherited factors on a trait based on the cumulative impact of many common polymorphisms (79). Several methods for calculating PGS exist including linkage disequilibrium (LD) pruning, p-value thresholding, and LDPred (80), which were applied to calculate PGS for oxidative, osmotic, and storage hemolysis. Models were built in two-thirds of the samples selected at random without reference for various p-value and linkage disequilibrium prunes that were validated independently in the remaining third. The 2/3 and 1/3 split was chosen as an intermediate of the possible splits suggested by different machine learning approaches (81). The set of LD pruning and p-value thresholding (supplemental figures 6-9) model that was provided the best estimate as measured by the r^2 between the hemolysis measure in the left out 1/3 and the PRS model built in the 2/3 was LD Pruning (r^2<0.2) and p-value thresholding ($P<10^{-7}$). This set of thresholds was

then used in the entire dataset to estimate the final PGS (Table 3). All markers in the final PGS had a p-value of $<10^{-7}$.

In vivo Studies in Sickle Cell Disease Cohorts WALK-PHaSST and PUSH Cohorts:

The SNPs that were genome wide significant from the REDS-III cohort were tested in the Treatment of Pulmonary Hypertension and SCD with Sildenafil Therapy (WALK-PHaSST dbGAP accession PHS001513.v1.p1) and Pulmonary Hypertension and Hypoxic Response in SCD (PUSH dbGAP accession PHS001682.v1.p1). This included 232 SNPs in 27 loci common in SCD cohorts (MAF >0.05) and imputed with $r^2>0.8$, all with p values less than 5×10^{-8} in REDS-III cohort. We evaluated the association of SNPs with a mathematical measure of the severity of in vivo hemolysis at steady state (baseline not during a vaso-occlusive event). The end point used is the first factor of a previously validated principal component (PC) measure of severity of steady state hemolysis in SCD patients (21). The PC is derived from clinically available standard lab measures that reflect RBC hemolysis. These measures include log transformed serum lactic acid dehydrogenase, aspartate aminotransferase, and total bilirubin, as well as the square root–transformed percent reticulocytes, and venous hemoglobin levels (21). These measures were adjusted for clinical site of blood collection and were standardized. This estimate of the severity of hemolytic anemia has been previously validated in patients with sickle cell disease and shown to significantly correlate with plasma hemoglobin and plasma RBC microparticles, as well as associate with clinical measures that modulate the intensity of hemolysis (fetal hemoglobin level and a-thalassemia) (21, 82). Genetic association of the severity of hemolysis by PCA adjusted for age, sex, hemoglobin genotype severity (SS and S-β^0 versus SC and S-β+), cohort, use of hydroxyurea validated by Hb F level, recent transfusion, and population stratification. Recent transfusion was defined by hemoglobin A level >50% in Walk-PHaSST and transfusion within the past two months in PUSH. Results between in vitro and in vivo hemolysis were considered consistent if the initial GWA p-value was significant ($p < 5\times10^{-8}$) and the p-value for the association in the two SCD cohorts was significant ($p<0.05$). We also show combined meta-analysis p-values for all results in Table 4.

The Walk-PHaSST study has 429 analyzable informative patients at least 12 years of age from nine U.S. Centers and one UK Center (83, 84). The PUSH study was conducted at four tertiary

medical centers in the United States and contains 282 analyzable patients 3 to 20 years of age (82). These SCD samples were genotyped on the Illumina Human 610-Quad SNP Array, which covers 588,451 genome-wide SNPs. Sample and SNP quality control was described previously (85). Genotypes were phased (85) and imputed (86) to 1000 Genomes Phase 3 data using African reference population samples. PCs of autosomal SNPs were estimated using the GCTA software (86).

Study approvals:

RBC-Omics was conducted under regulations applicable to all human subject research supported by federal agencies. The Data Coordinating Center (RTI International, Rockville, MD) of REDS-III was responsible for the overall compliance of human subjects to regulator protocols, including institutional review board approval from each participating blood center, from the REDS-III Central Laboratory (Vitalant, San Francisco, CA) and the Data Coordinating Center. Approval of the Walk-PHaSST study protocol (clinical trial # NCT00492531) was obtained from local institutional review boards or ethics committees, and written informed consent was obtained from all study subjects in accordance with the Declaration of Helsinki.

Author Contributions: Study concept and design: GPP, MB, JG, VG, ML, AM, RC, BS, JK, SE, DB, SG, SK MTG. Acquisition and preparation of data and samples: ML, AM, RC, BS, JK, SE, DB, SG, SK, MTG. Hemolysis and Lab Assays: MTG, TK, Statistical analyses: GPP, JG, FF, XZ, MTG. Interpretation of data: GP, JG, FF, XZ, MB, TK, MTG. Manuscript writing: GPP, JG, MB, TK, XZ, VG, SG, MTG. All authors contributed to the critical revision of the manuscript for important intellectual content. GPP and TK are joint first authors with GPP in lead role since GPP led the writing of the manuscript since he is a geneticist, and the paper is focused on genetics. TK is the expert in hemolysis assays and contributed extensively to the writing.

Grant support:

The NHLBI REDS-III was supported by NHLBI contracts NHLBI HHSN2682011-00001I, 00002I, 00003I, 00004I, 00005I, 00006I, 00007I, 00008I, and 00009I.

Acknowledgments:

The authors thank the RBC-Omics research staff at all participating blood centers, the testing laboratories for performing tests and for their contribution to this project, and all blood donors who agreed to participate in this study. The authors also thank Dr. Qinzi Xu for his assistance with figure preparation.

See Supplemental Acknowledgments for study program details.

REFERENCES

1. Tzounakas VL, Georgatzakou HT, Kriebardis AG, Voulgaridou AI, Stamoulis KE, Foudoulaki-Paparizos LE, et al. Donor variation effect on red blood cell storage lesion: a multivariable, yet consistent, story. *Transfusion.* 2016;56(6):1274-86.
2. Kanias T, Lanteri MC, Page GP, Guo Y, Endres SM, Stone M, 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(15):1132-41.
3. Hazegh K, Fang F, Bravo MD, Tran JQ, Muench MO, Jackman RP, et al. Blood donor obesity is associated with changes in red blood cell metabolism and susceptibility to hemolysis in cold storage and in response to osmotic and oxidative stress. *Transfusion.* 2021;61(2):435-48.
4. Kanias T, Sinchar D, Osei-Hwedieh D, Baust JJ, Jordan A, Zimring JC, et al. Testosterone-dependent sex differences in red blood cell hemolysis in storage, stress, and disease. *Transfusion.* 2016;56(10):2571-83.
5. Chasse M, McIntyre L, English SW, Tinmouth A, Knoll G, Wolfe D, et al. Effect of blood donor characteristics on transfusion outcomes: A systematic review and meta-analysis. *Transfusion medicine reviews.* 2016;30(2):69-80.
6. Chasse M, Tinmouth A, English SW, Acker JP, Wilson K, Knoll G, et al. Association of Blood Donor Age and Sex With Recipient Survival After Red Blood Cell Transfusion. *JAMA Intern Med.* 2016;176(9):1307-14.
7. Osei-Hwedieh DO, Kanias T, Croix CS, Jessup M, Xiong Z, Sinchar D, et al. Sickle Cell Trait Increases Red Blood Cell Storage Hemolysis and Post-Transfusion Clearance in Mice. *EBioMedicine.* 2016;11:239-48.
8. Kanias T, and Gladwin MT. Nitric oxide, hemolysis, and the red blood cell storage lesion: interactions between transfusion, donor, and recipient. *Transfusion.* 2012;52(7):1388-92.
9. Lopez C, Saravia C, Gomez A, Hoebeke J, and Patarroyo MA. Mechanisms of genetically-based resistance to malaria. *Gene.* 2010;467(1-2):1-12.

10. Zimring JC, Smith N, Stowell SR, Johnsen JM, Bell LN, Francis RO, et al. Strain-specific red blood cell storage, metabolism, and eicosanoid generation in a mouse model. *Transfusion.* 2014;54(1):137-48.
11. de Wolski K, Fu X, Dumont LJ, Roback JD, Waterman H, Odem-Davis K, et al. Metabolic pathways that correlate with post-transfusion circulation of stored murine red blood cells. *Haematologica.* 2016;101(5):578-86.
12. Kanias T, Stone M, Page GP, Guo Y, Endres-Dighe SM, Lanteri MC, et al. Frequent blood donations alter susceptibility of red blood cells to storage- and stress-induced hemolysis. *Transfusion.* 2019;59(1):67-78.
13. Lanteri MC, Kanias T, Keating S, Stone M, Guo Y, Page GP, et al. Intradonor reproducibility and changes in hemolytic variables during red blood cell storage: results of recall phase of the REDS-III RBC-Omics study. *Transfusion.* 2019;59(1):79-88.
14. Francis RO, Jhang J, Hendrickson JE, Zimring JC, Hod EA, and Spitalnik SL. Frequency of glucose-6-phosphate dehydrogenase-deficient red blood cell units in a metropolitan transfusion service. *Transfusion.* 2013;53(3):606-11.
15. Raciti PM, Francis RO, Spitalnik PF, Schwartz J, and Jhang JS. Acquired hemoglobin variants and exposure to glucose-6-phosphate dehydrogenase deficient red blood cell units during exchange transfusion for sickle cell disease in a patient requiring antigen-matched blood. *J Clin Apher.* 2013;28(4):325-9.
16. Francis RO, D'Alessandro A, Eisenberger A, Soffing M, Yeh R, Coronel E, et al. Donor glucose-6-phosphate dehydrogenase deficiency decreases blood quality for transfusion. *J Clin Invest.* 2020;130(5):2270-85.
17. Dern RJ, Gwinn RP, and Wiorkowski JJ. Studies on the preservation of human blood. I. Variability in erythrocyte storage characteristics among healthy donors. *J Lab Clin Med.* 1966;67(6):955-65.
18. Gordeuk VR, Shah BN, Zhang X, Thuma PE, Zulu S, Moono R, et al. The CYB5R3(c) (.350C>G) and G6PD A alleles modify severity of anemia in malaria and sickle cell disease. *Am J Hematol.* 2020;95(11):1269-79.

19. Kato GJ, Nouraie SM, and Gladwin MT. Lactate dehydrogenase and hemolysis in sickle cell disease. *Blood.* 2013;122(6):1091-2.
20. Milton JN, Rooks H, Drasar E, McCabe EL, Baldwin CT, Melista E, et al. Genetic determinants of haemolysis in sickle cell anaemia. *Br J Haematol.* 2013;161(2):270-8.
21. Nouraie M, Lee JS, Zhang Y, Kanias T, Zhao X, Xiong Z, et al. The relationship between the severity of hemolysis, clinical manifestations and risk of death in 415 patients with sickle cell anemia in the US and Europe. *Haematologica.* 2013;98(3):464-72.
22. Saraf SL, Akingbola TS, Shah BN, Ezekekwu CA, Sonubi O, Zhang X, et al. Associations of alpha-thalassemia and BCL11A with stroke in Nigerian, United States, and United Kingdom sickle cell anemia cohorts. *Blood Adv.* 2017;1(11):693-8.
23. Kato GJ, Steinberg MH, and Gladwin MT. Intravascular hemolysis and the pathophysiology of sickle cell disease. *J Clin Invest.* 2017;127(3):750-60.
24. Donadee C, Raat NJ, Kanias T, Tejero J, Lee JS, Kelley EE, et al. Nitric oxide scavenging by red blood cell microparticles and cell-free hemoglobin as a mechanism for the red cell storage lesion. *Circulation.* 2011;124(4):465-76.
25. Endres-Dighe SM, Guo Y, Kanias T, Lanteri M, Stone M, Spencer B, et al. Blood, sweat, and tears: Red Blood Cell-Omics study objectives, design, and recruitment activities. *Transfusion.* 2019;59(1):46-56.
26. Guo Y, Busch MP, Seielstad M, Endres-Dighe S, Westhoff CM, Keating B, et al. Development and evaluation of a transfusion medicine genome wide genotyping array. *Transfusion.* 2019;59(1):101-11.
27. Moreno-Estrada A, Gignoux CR, Fernandez-Lopez JC, Zakharia F, Sikora M, Contreras AV, et al. Human genetics. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. *Science.* 2014;344(6189):1280-5.
28. Nelson SC, Stilp AM, Papanicolaou GJ, Taylor KD, Rotter JI, Thornton TA, et al. Improved imputation accuracy in Hispanic/Latino populations with larger and more diverse reference panels: applications in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Hum Mol Genet.* 2016;25(15):3245-54.

29. van der Harst P, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature.* 2012;492(7429):369-75.
30. Fortugno C, Galea E, Cantaffa R, Gigliotti F, Fabiano RL, Talarico V, et al. Hereditary red blood cell membrane defects. Detection of PIEZO1 mutations associated with SPTA1 mutations. An unusual clinical case of hereditary xerocytosis. *Pediatr Hematol Oncol.* 2020:1-7.
31. Tatu T, and Sweatman D. Hemolysis area: A new parameter of erythrocyte osmotic fragility for screening of thalassemia trait. *J Lab Physicians.* 2018;10(2):214-20.
32. Luzzatto L, Ally M, and Notaro R. Glucose-6-phosphate dehydrogenase deficiency. *Blood.* 2020;136(11):1225-40.
33. Vogel F. Clinical consequences of heterozygosity for autosomal-recessive diseases. *Clin Genet.* 1984;25(5):381-415.
34. Niss O, Chonat S, Dagaonkar N, Almansoori MO, Kerr K, Rogers ZR, et al. Genotype-phenotype correlations in hereditary elliptocytosis and hereditary pyropoikilocytosis. *Blood cells, molecules & diseases.* 2016;61:4-9.
35. Ripich T, and Jessberger R. SWAP-70 regulates erythropoiesis by controlling alpha4 integrin. *Haematologica.* 2011;96(12):1743-52.
36. Ferru E, Giger K, Pantaleo A, Campanella E, Grey J, Ritchie K, et al. Regulation of membrane-cytoskeletal interactions by tyrosine phosphorylation of erythrocyte band 3. *Blood.* 2011;117(22):5998-6006.
37. Hellwege JN, Jeff JM, Wise LA, Gallagher CS, Wellons M, Hartmann KE, et al. A multi-stage genome-wide association study of uterine fibroids in African Americans. *Hum Genet.* 2017;136(10):1363-73.
38. Gao G, Pierce BL, Olopade OI, Im HK, and Huo D. Trans-ethnic predicted expression genome-wide association analysis identifies a gene for estrogen receptor-negative breast cancer. *PLoS Genet.* 2017;13(9):e1006727.
39. Fryett JJ, Inshaw J, Morris AP, and Cordell HJ. Comparison of methods for transcriptome imputation through application to two common complex diseases. *Eur J Hum Genet.* 2018;26(11):1658-67.

40. Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv.* 2018:447367.
41. Zama D, Giulietti G, Muratore E, Andolfo I, Russo R, Iolascon A, et al. A novel PIEZO1 mutation in a patient with dehydrated hereditary stomatocytosis: a case report and a brief review of literature. *Ital J Pediatr.* 2020;46(1):102.
42. Ma S, Dubin AE, Zhang Y, Mousavi SAR, Wang Y, Coombs AM, et al. A role of PIEZO1 in iron metabolism in mice and humans. *Cell.* 2021;184(4):969-82 e13.
43. Wang X, Zhang A, Huang M, Chen L, Hu Q, Lu Y, et al. Genetic and Clinical Characteristics of Patients With Hereditary Spherocytosis in Hubei Province of China. *Front Genet.* 2020;11:953.
44. Iolascon A, King MJ, Robertson S, Avvisati RA, Vitiello F, Asci R, et al. A genomic deletion causes truncation of alpha-spectrin and ellipto-poikilocytosis. *Blood cells, molecules & diseases.* 2011;46(3):195-200.
45. Jamwal M, Aggarwal A, Palodi A, Sharma P, Bansal D, Maitra A, et al. A nonsense variant in the Hexokinase 1 gene (HK1) causing severe non-spherocytic haemolytic anaemia: genetic analysis exemplifies ambiguity due to multiple Isoforms. *Br J Haematol.* 2019;186(5):e142-e5.
46. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell.* 2016;167(5):1415-29 e19.
47. Ding K, de Andrade M, Manolio TA, Crawford DC, Rasmussen-Torvik LJ, Ritchie MD, et al. Genetic variants that confer resistance to malaria are associated with red blood cell traits in African-Americans: an electronic medical record-based genome-wide association study. *G3 (Bethesda).* 2013;3(7):1061-8.
48. Canli O, Alankus YB, Grootjans S, Vegi N, Hultner L, Hoppe PS, et al. Glutathione peroxidase 4 prevents necroptosis in mouse erythroid precursors. *Blood.* 2016;127(1):139-48.

49. Altamura S, Vegi NM, Hoppe PS, Schroeder T, Aichler M, Walch A, et al. Glutathione peroxidase 4 and vitamin E control reticulocyte maturation, stress erythropoiesis and iron homeostasis. *Haematologica.* 2020;105(4):937-50.
50. Ouled-Haddou H, Messaoudi K, Demont Y, Lopes Dos Santos R, Carola C, Caulier A, et al. A new role of glutathione peroxidase 4 during human erythroblast enucleation. *Blood Adv.* 2020;4(22):5666-80.
51. Forcina GC, and Dixon SJ. GPX4 at the Crossroads of Lipid Homeostasis and Ferroptosis. *Proteomics.* 2019;19(18):e1800311.
52. Seibt TM, Proneth B, and Conrad M. Role of GPX4 in ferroptosis and its pharmacological implication. *Free Radic Biol Med.* 2019;133:144-52.
53. D'Alessandro A, Reisz JA, Zhang Y, Gehrke S, Alexander K, Kanias T, et al. Effects of aged stored autologous red blood cells on human plasma metabolome. *Blood Adv.* 2019;3(6):884-96.
54. Fu X, Felcyn JR, Odem-Davis K, and Zimring JC. Bioactive lipids accumulate in stored red blood cells despite leukoreduction: a targeted metabolomics study. *Transfusion.* 2016;56(10):2560-70.
55. Nader E, Romana M, and Connes P. The Red Blood Cell-Inflammation Vicious Circle in Sickle Cell Disease. *Front Immunol.* 2020;11:454.
56. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature.* 2019;570(7762):514-8.
57. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet.* 2017;100(4):635-49.
58. Cavazos TB, and Witte JS. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG Adv.* 2021;2(1).
59. Leopold JA, Maron BA, and Loscalzo J. The application of big data to cardiovascular disease: paths to precision medicine. *J Clin Invest.* 2020;130(1):29-38.

60. Maron BA, Altucci L, Balligand JL, Baumbach J, Ferdinandy P, Filetti S, et al. A global network for network medicine. *NPJ Syst Biol Appl.* 2020;6(1):29.
61. D'Alessandro A, Fu X, Kanias T, Reisz JA, Culp-Hill R, Guo Y, et al. Donor sex, age and ethnicity impact stored red blood cell antioxidant metabolism through mechanisms in part explained by glucose 6-phosphate dehydrogenase levels and activity. *Haematologica.* 2020.
62. Roubinian N, Kleinman S, Murphy EL, Glynn SA, and Edgren G. Methodological considerations for linked blood donor-component-recipient analyses in transfusion medicine research. *ISBT Sci Ser.* 2020;15(1):185-93.
63. Roubinian NH, and Kanias T. Blood donor component-recipient linkages: is there fire where there is smoke? *Transfusion.* 2019;59(8):2485-8.
64. Roubinian NH, Plimier C, Woo JP, Lee C, Bruhn R, Liu VX, et al. Effect of donor, component, and recipient characteristics on hemoglobin increments following red blood cell transfusion. *Blood.* 2019;134(13):1003-13.
65. Edgren G, Murphy EL, Brambilla DJ, Westlake M, Rostgaard K, Lee C, et al. Association of Blood Donor Sex and Prior Pregnancy With Mortality Among Red Blood Cell Transfusion Recipients. *JAMA.* 2019;321(22):2183-92.
66. Stone M, Keating SM, Kanias T, Lanteri MC, Lebedeva M, Sinchar D, et al. Piloting and implementation of quality assessment and quality control procedures in RBC-Omics: a large multi-center study of red blood cell hemolysis during storage. *Transfusion.* 2019;59(1):57-66.
67. NIH. NHLBI Recipient Epidemiology Donor Evaluation Study (REDS)-III - Red Blood Cell Omics (RBC-Omics) Study. dbGaP Study Accession: phs001955.v1.p1 https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001955.v1.p1. Accessed 2/7/2021, 2021.
68. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-75.

69. Delaneau O, Coulonges C, and Zagury JF. Shape-IT: new rapid and accurate algorithm for haplotype inference. *BMC Bioinformatics.* 2008;9:540.
70. Howie B, Marchini J, and Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda).* 2011;1(6):457-70.
71. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, and Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics.* 2012;28(24):3326-8.
72. Aulchenko YS, Struchalin MV, and van Duijn CM. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics.* 2010;11:134.
73. Kanias T, Stone M, Page GP, Guo Y, Endres-Dighe SM, Lanteri MC, et al. Frequent blood donations alter susceptibility of red blood cells to storage- and stress-induced hemolysis. *Transfusion.* 2019;59(1):67-78.
74. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015;47(3):291-5.
75. Ward LD, and Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 2012;40(Database issue):D930-4.
76. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics.* 2010;26(18):2336-7.
77. Watanabe K, Taskesen E, van Bochoven A, and Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826.
78. GSEA. Molecular Signatures Database v7.2. https://www.gsea-msigdb.org/gsea/msigdb/index.jsp.
79. Torkamani A, Wineinger NE, and Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet.* 2018;19(9):581-90.

80. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet.* 2018;50(9):1219-24.
81. Kuhn M, and Johnson K. *Applied Predictive Modeling.* Springer; 2013.
82. Minniti CP, Sable C, Campbell A, Rana S, Ensing G, Dham N, et al. Elevated tricuspid regurgitant jet velocity in children and adolescents with sickle cell disease: association with hemolysis and hemoglobin oxygen desaturation. *Haematologica.* 2009;94(3):340-7.
83. Sachdev V, Kato GJ, Gibbs JS, Barst RJ, Machado RF, Nouraie M, et al. Echocardiographic markers of elevated pulmonary pressure and left ventricular diastolic dysfunction are associated with exercise intolerance in adults and adolescents with homozygous sickle cell anemia in the United States and United Kingdom. *Circulation.* 2011;124(13):1452-60.
84. Machado RF, Barst RJ, Yovetich NA, Hassell KL, Kato GJ, Gordeuk VR, et al. Hospitalization for pain in patients with sickle cell disease treated with sildenafil for elevated TRV and low exercise capacity. *Blood.* 2011;118(4):855-64.
85. Zhang X, Shah BN, Zhang W, Saraf SL, Miasnikova G, Sergueeva A, et al. A genetic variation associated with plasma erythropoietin and a non-coding transcript of PRKAR1A in sickle cell disease. *Hum Mol Genet.* 2016;25(20):4601-9.
86. Howie BN, Donnelly P, and Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009;5(6):e1000529.

Figure 1: Ancestry of RBC-Omics population and Manhattan plots. Figure 1A: First two principal component (PC) plot of the extended RBC-Omics population overlain on the 1000 Genomes Phase 3 samples. Individuals are labeled by genetic ancestry (AFR: African American, EAS: East Asian, SAS: South Asian, EUR: Non-Hispanic White, AMR: admixed American, CIH: Caribbean Island Hispanics, MCAH: Mexican and Central American Hispanics, OTH: Other/multi ancestry) overlain by ancestry groups from 1000 Genomes v3. Figure 1B–D: Manhattan plots summarizing the mega analysis results for osmotic hemolysis (n = 12,215, lambda = 1.003, Figure 1B), oxidative hemolysis (n = 10,007, lambda = 1.048, Figure 1C), and storage hemolysis (n = 12,177, lambda = 1.002, Figure 1D). Each point corresponds to a -log10 (p-value) from a multi-variant linear regression model's p-value for a SNP. The black horizontal line represents an accepted p-value level of genome-wide significance ($p = 5X10^{-8}$). Circles represent non-coding variants, and triangles are coding variants.

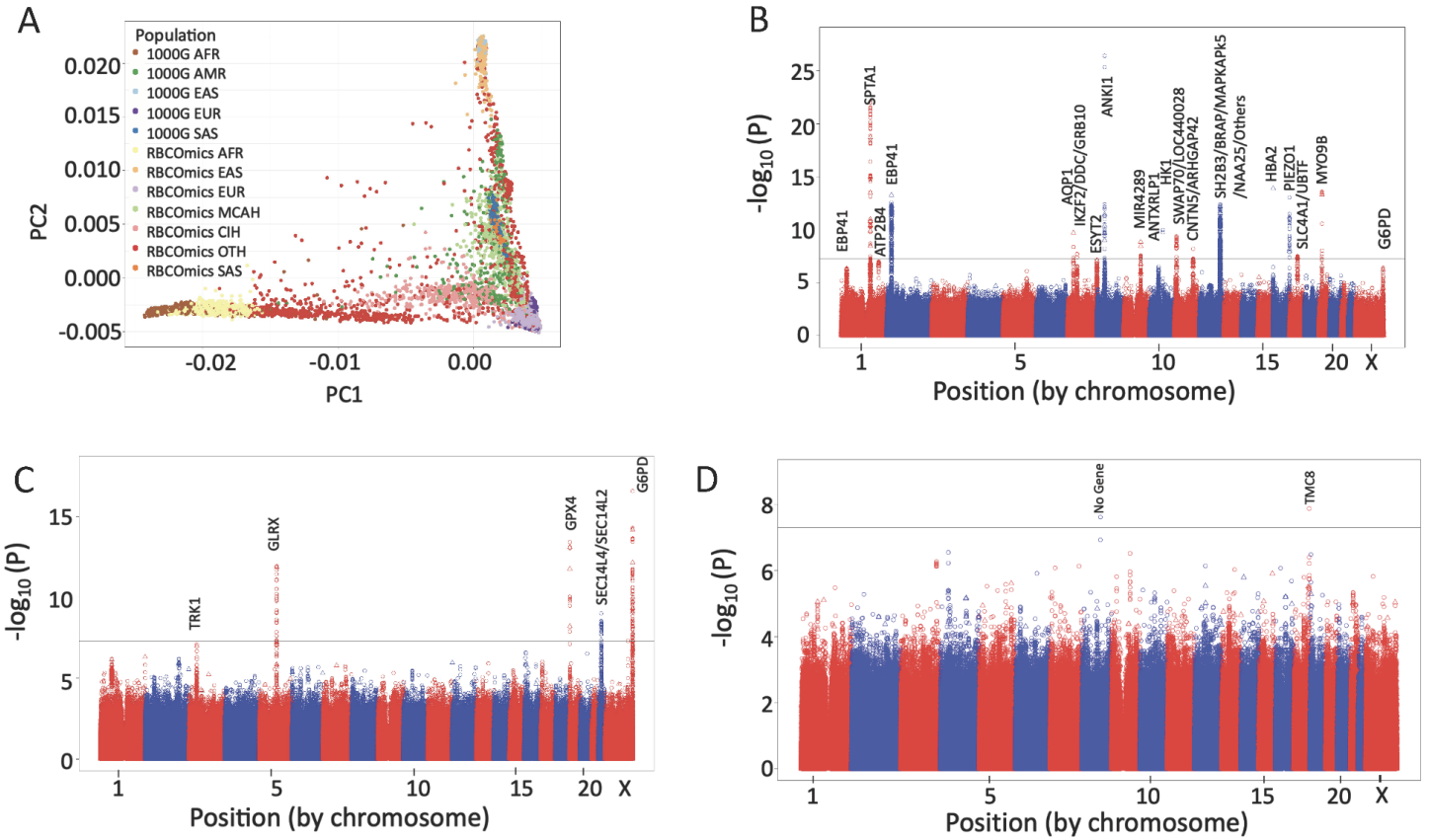


Figure 2: Box plots of various hemolysis levels by genotype for GWA-significant non-synonymous (except HBA2) variants by ancestry group. Osmotic hemolysis: Figure 2A, Osmotic *SPTA1* rs857725/Lys1693Gln; Figure 2B, Osmotic *HBA2* chr16:223678; Figure 2C, Osmotic *HBB* rs334/Gul7Val (Hb S). N=12,219 for all osmotic analyses. Oxidative hemolysis: Figure 2D, Oxidative *SEC14L4* (AX-83171224/rs9606739) Arg112Gly; Figure 2E, Oxidative *G6PD* (rs1050828) Val68Met is on the X chromosome, therefore male (M) and female (F) sample members are displayed separately. N=10,007 for all oxidative analyses. Spontaneous (Storage) hemolysis: Figure 2F, Storage *TMC8* (rs7208422) Asn306Ile. Minor allele homozygotes are in shades of red, heterozygotes in green, and reference allele homozygotes in shades of blue. N=12,219 for all storage analyses. Ancestry groups AFR: African Americans, EUR: Non-Hispanic Whites, EAS: East Asians, SAS: South Asians, CIH: Caribbean Island Hispanics, MCAH: Mexican/Central American Hispanics, and OTH: Other.

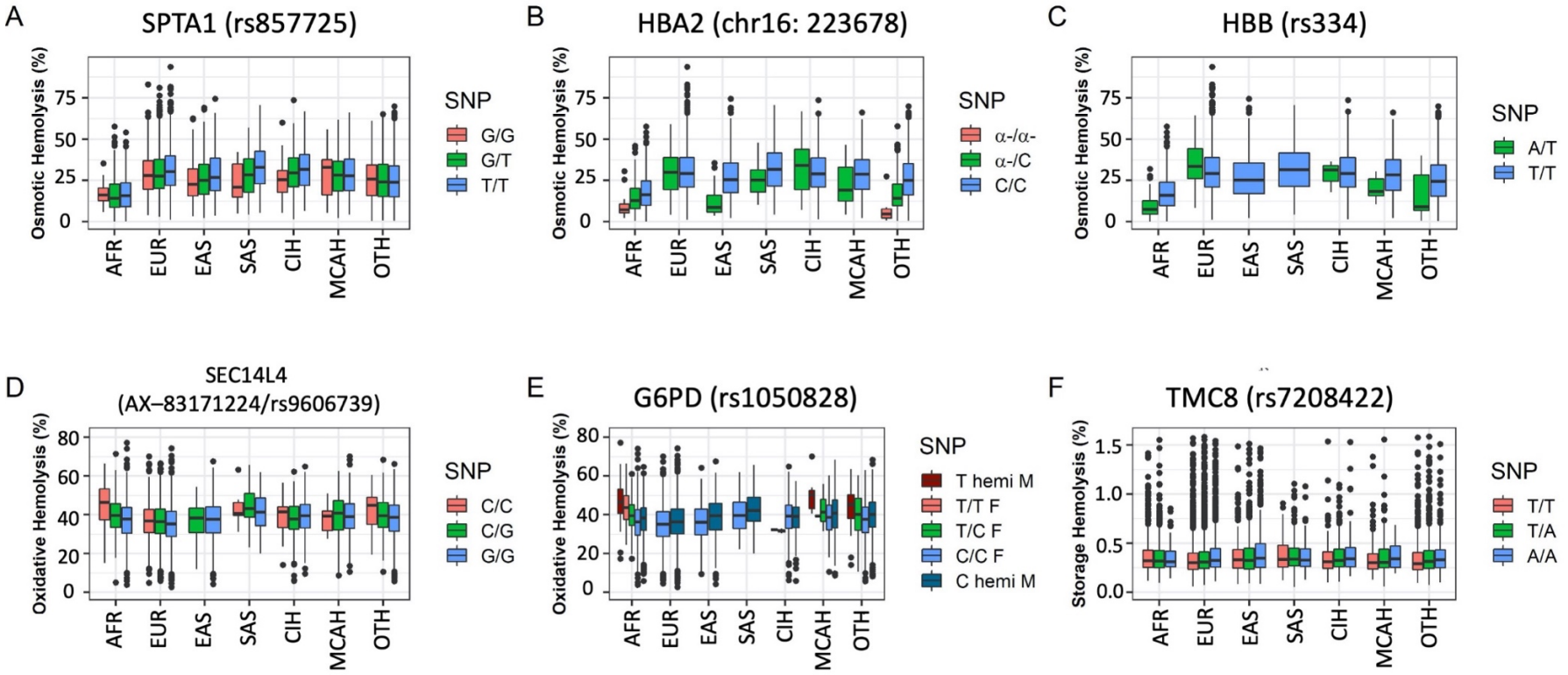


Figure 3: LocusZoom and box plots for two non-overlapping genome-wide significant loci in *ANK1*. Figure 3A: LocusZoom plots centered on rs4737010 in *ANK1*. Figure 3B: LocusZoom plot rs34664882 in *ANK1*. In each plot, each point represents a SNP passing quality control in the linear regression analysis of imputed dosage plotted with its p value as a function of genomic position (GRCh38 Assembly). The lead SNP is represented by the purple symbol. The color coding of all other SNPs indicates LD with the lead SNP (estimated by Phase II HapMap CEU r^2 values): red, $r^2 \geq 0.8$; gold, $0.6 \leq r^2 < 0.8$; green, $0.4 \leq r^2 < 0.6$; cyan, $0.2 \leq r^2 < 0.4$; blue, $r^2 < 0.2$; gray, r^2 unknown. Recombination rates are estimated from 1000 Genomes Phase 3 data. Figure 3C: Box plots of osmotic hemolysis measure by genotype and genetic ancestry group for rs4737010. Figure 3D: Box plots of osmotic hemolysis measure by genotype and genetic ancestry group for rs34664882. Ancestry groups AFR: African Americans, EUR: Non-Hispanic Whites, EAS: East Asians, SAS: South Asians, CIH: Caribbean Island Hispanics, MCAH: Mexican/Central American Hispanics, and OTH: Other.

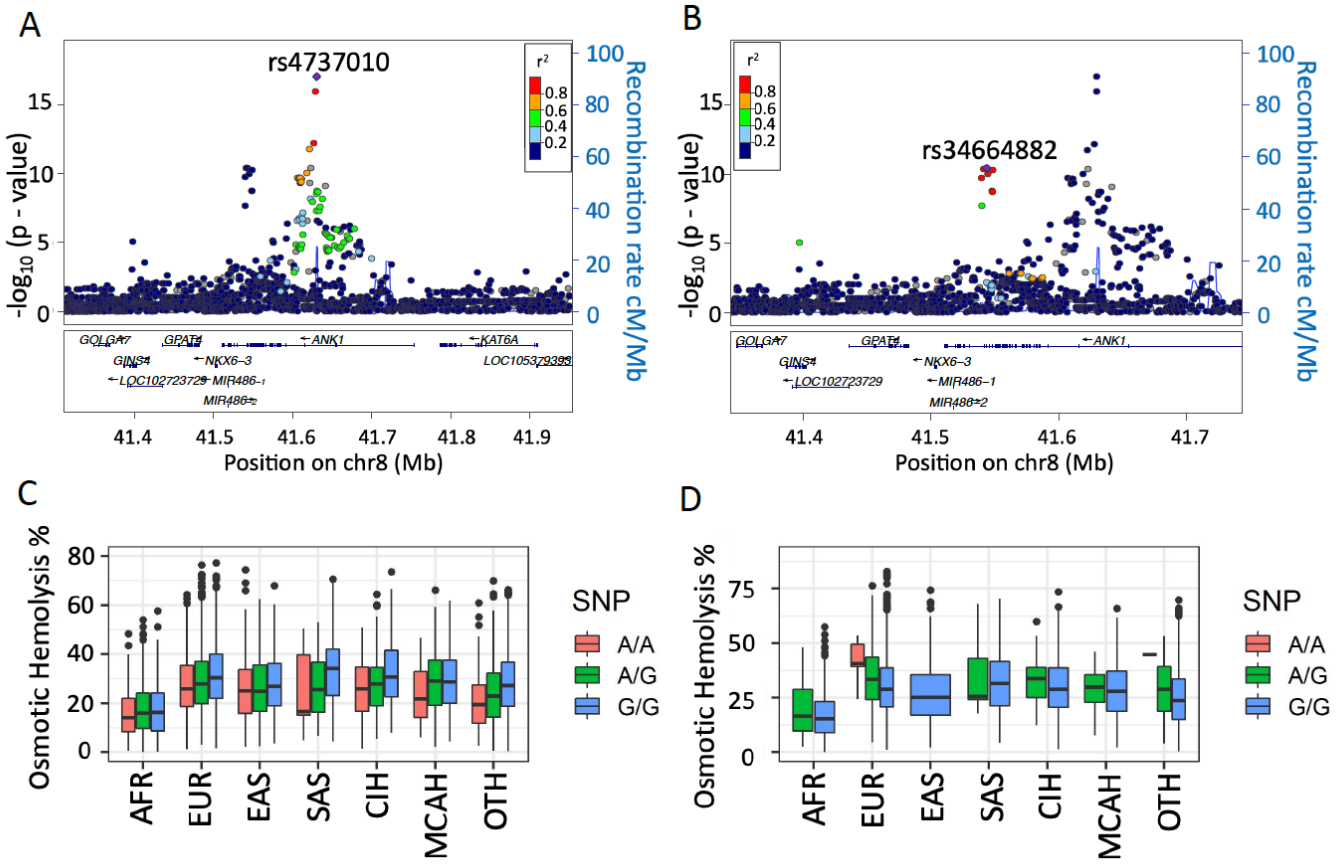


Table 1: Genome-wide significant results for hemolysis in all samples or within individual ancestry groups. Results are presented for the single most-significant SNP under a GWA peak. Directly genotyped SNPs that are also GWA significant are presented as well. Information includes chromosome location, RS IDs for markers, the p-value, population(s) in which the SNP was significant, the nucleotide for the minor allele, the minor allele frequency in the non-Hispanic White population, the r^2 for imputation accuracy, the sample size of the most GWA-significant population, calculated beta, nearest gene(s) to the most significant SNP (# represents no close gene), and any functional information on the SNPs. The r^2 is a measure of the imputation accuracy. Minor allele frequencies in various populations based on TOPMED frequencies (* missing/not mapped in TOPMED release 8). If $r^2 = 1$, the marker was directly genotyped. INT – intronic, NSM – non-synonymous mutation.

Chr	Position	Nearest Gene	Rsid	Hemolysis	p-Value	GWA-Significant Populations	Minor Allele	r^2	beta SNP	Functional Annotation	MAF African	MAF Native American	MAF East Asian	MAF European	MF South Asian	MAF TOPMED Freeze 8
1	158586966	SPTA1	rs2022003	Osmotic	1.01E-22	All, EAU	A	0.99	1.7	**INT**	0.22	0.33	0.43	0.30	0.21	0.29
1	158607935	SPTA1	rs857725	Osmotic	8.75E-21	All, EAU	G	1	-1.6	**NSM**	0.10	0.30	0.44	0.28	0.17	0.25
1	195275704	#	rs17661899	Oxidative	3.54E-08	MCAH	G	0.98	-6.0	**No gene**	0.05	0.09	0.00	0.16	0.04	0.07
1	246711077	TFB2M	rs35558093	Osmotic	3.87E-08	East Asian	A	0.99	-21.2	**INT**	0.18	0.22	0.00	0.29	0.11	0.16
2	24029928	MFSD2B	rs55707417	Osmotic	1.07E-13	All, EAU	G	1.00	-1.8	**INT**	0.05	0.08	0.05	0.14	0.16	0.09
3	34083520	#	rs148642995	Osmotic	1.09E-08	African American	T	0.89	-11.2	**NA**	0.02	0.01	0.00	0.00	0.00	0.01
5	95166235	GLRX	rs72785409	Oxidative	1.15E-12	All, EAU	C	0.98	1.5	**INT**	0.21	0.09	0.04	0.12	0.08	0.12
6	66045716	EYS	rs78484557	Osmotic	3.20E-09	MCAH	C	0.90	-25.7	**INT**	0.00	0.01	0.00	0.03	0.00	0.02
7	30954977	AQP1	rs5883264	Osmotic	4.23E-10	All, EAU	GC	0.98	-1.3	**INT**	0.31	0.18	0.01	0.16	0.13	0.17
7	50429442	IKZF1	rs6976036	Osmotic	1.86E-08	All, EAU	T	1.00	0.9	**INT**	0.47	0.60	0.44	0.50	0.49	0.49
8	41630447	ANK1	rs4737010	Osmotic	5.85E-28	All, EAU	A	0.97	-1.9	**INT**	0.54	0.27	0.53	0.24	0.18	0.37
8	89839432	#	rs189900952	Storage	2.42E-08	All, EAU	G	0.67	0.002	**No Gene**	0.00	0.02	0.12	0.00	0.01	0.02
9	91408347	MIR4289	chr9:91408347	Osmotic	6.44E-10	All, EAU	C	0.96	-2.3	**NA**	*	*	*	*	*	*
10	71099109	HK1	rs72805692	Osmotic	4.90E-11	All, EAU	A	0.99	1.8	**INT**	0.002	0.07	0.00	0.11	0.02	0.04
11	5248232	HBB	rs334	Osmotic	3.66E-10	African American	A	0.82	-8.2	**NSM**	0.10	0.01	0.00	0.00	0.00	0.03
11	9762274	SWAP70	rs360153	Osmotic	1.28E-10	All, EAU	T	1.00	-1.0	**INT**	0.48	0.52	0.38	0.60	0.51	0.50
11	100493995	ARHGAP42	rs717662	Osmotic	4.00E-09	All, EAU	C	0.97	-1.5	**INT**	0.01	0.06	0.27	0.11	0.09	0.11
12	112061723	BRAP/ALDH2/ MAKKAPK5	rs77684561	Osmotic	2.24E-13	All, EAU	C	0.97	1.5	**INT**	0.15	0.15	0.35	0.17	0.22	0.21

13	76938832	#	rs118149920	Osmotic	9.74E-09	MCAH	T	0.93	-12.5	**INT**	0.00	0.09	0.08	0.01	0.06	0.04
16	223678	HBA2	chr 16:223678	Osmotic	2.90E-14	All, EAU, African American	α-/del	0.80	-4.4	**INT**	*	*	*	*	*	*
16	226229	HBA2	chr 16:226229	Osmotic	1.06E-12	East Asian	α-/del	0.90	-17.9	**INT**	0.00	0.00	0.02	0.00	0.00	0.00
16	88856084	PIEZO1	rs551118	Osmotic	4.04E-14	All, EAU	C	0.91	-1.3	**INT**	0.74	0.50	0.62	0.59	0.60	0.62
17	42304644	UBTF/SLC4A1 (Band 3)	rs7222349	Osmotic	3.62E-08	All, EAU	**G**	1.00	-1.0	INT	0.96	0.45	0.75	0.32	0.31	0.55
17	76130575	TMC8; TMC6	rs7208422	Storage	1.34E-08	All	**T**	1.00	-0.004	NSM	0.71	0.59	0.42	0.49	0.51	0.56
19	1103230	GPX4	rs8178962	Oxidative	3.80E-14	All	**A**	0.96	-1.1	INT	0.62	0.58	0.28	0.45	0.42	0.49
19	17252151	MYO9B	rs35365035	Osmotic	9.88E-15	All, EAU	**T**	1.00	1.2	INT	0.74	0.43	0.36	0.41	0.35	0.49
22	30891859	SEC14L4	rs9608944	Oxidative	9.85E-10	All	**G**	1.00	1.1	INT	0.15	0.29	0.04	0.23	0.14	0.20
23	153764217	G6PD	rs1050828	Oxidative	2.66E-17	All, African American	**T**	1.00	3.8	NSM	0.10	0.01	0.00	0.00	0.00	0.04

Table 2: MetaXcan analysis of genes whose expression is modeled to be associated with osmotic and oxidative hemolysis. Presented is the gene name, chromosome location of the gene, Bonferroni (BF) corrected p-value, predicted performance (r^2) of the models of the gene's expression, predicted performance q-value of the model, number of SNPs in the gene used to estimate the gene's expression level, and whether the gene is under one of the genome-wide significant peaks.

Gene Name	Hemolysis	Chromosome Location	BF p-Value	r^2	Q-Value	Number of SNPs	Under GWA Hit?
PFN4	Osmotic	*chr2:24,338,241-24,346,347*	5.0E-04	0.04	1.23E-08	8	Yes
MFSD2B	Osmotic	*chr2:24,232,951-24,286,191*	1.3E-02	0.08	1.27E-17	28	Yes
ESYT2	Osmotic	*chr7:158,523,686-158,622,944*	4.5E-02	0.40	3.92E-102	49	Yes
C8orf40(SMIM19)	Osmotic	*chr8:42,396,298-42,409,603*	7.9E-03	0.26	1.73E-61	31	No
SLC20A2	Osmotic	*chr8:42,273,993-42,397,069*	3.4E-02	0.23	1.28E-53	24	No
SWAP70	Osmotic	*chr11:9,685,624-9,774,538*	1.8E-02	0.08	5.51E-17	35	Yes
NAA25	Osmotic	*chr12:112,464,493-112,546,826*	1.6E-06	0.03	6.09E-07	17	Yes
SH2B3	Osmotic	*chr12:111,843,752-111,889,427*	1.1E-04	0.04	1.06E-09	21	Yes
FAM109A	Osmotic	*chr12:111,798,455-111,806,925*	1.0E-03	0.06	1.13E-14	18	Yes
TMEM116	Osmotic	*chr12:112,369,086-112,451,023*	1.4E-02	0.16	3.33E-37	30	Yes
SLC4A1	Osmotic	*chr17:42,325,753-42,345,509*	9.7E-04	0.01	1.14E-03	7	Yes
C17orf59(BORCS6)	Oxidative	*chr17: 8,091,651-8,093,564*	2.7E-02	0.02	4.90E-05	27	No
GPX4	Oxidative	*chr19: 1,103,936-1,106,787*	3.2E-06	0.31	8.49E-76	74	Yes

Table 3: Ancestry and cross-ancestry polygenic risk scores. Summary of correlation of PRS score calculated in ancestry-specific groups with each hemolysis measure in the entire ancestry-specific group. The ancestry groups are defined by principal component analysis based on genetic data (Figure 1). **The correlation within ancestry is calculated between the polygenic risk scores trained within each ancestry and the measured osmotic or oxidative hemolysis in each ancestry group. NA -no PRS could be calculated that was different from 0. *** The correlation from non-Hispanic white (EAU) is calculated between the polygenic risk score trained in this group and the measured osmotic or oxidative hemolysis within each ancestry group.

Genetically Defined Ancestry*	**Sample Size**	**Correlation with Osmotic hemolysis ****	**Correlation with EAU Osmotic PGS*****	**Correlation with Oxidative hemolysis****	**Correlation with EAU Oxidative PGS*****
Non-Hispanic White (EAU)	7,757	0.221	0.221	0.0834	0.0834
African American	1,052	NA	0.117	0.259	0.103
East Asian	1,112	0.180	0.134	NA	0.0126
South Asian	265	NA	0.184	NA	0.125
Caribbean Island Hispanics	497	NA	0.182	NA	0.0635
Mexican Central American Hispanics	459	0.251	0.184	0.263	0.0901
Other	598	NA	0.208	NA	0.0855

Table 4: Testing of osmotic, oxidative, and storage hits from the REDS-III RBC-Omics full data in the combined Walk-PHaSST and PUSH cohorts. Significant SNPs with nominal P value <0.05 in the SCD study were pruned so that linkage disequilibrium r^2 <0.3 in SCD cohorts. Information includes RS IDs for markers, the nearest genes, chromosome location, the nucleotide for the minor allele, the minor allele frequency from the REDS-III RBC-Omics full data and from the combined Walk-PHaSST and PUSH cohorts, the r^2 for imputation accuracy, and p-value and beta estimation of the association between minor allele and hemolysis trait. Given different measures of hemolysis the directions of the betas are not necessarily consistent between the in vivo and in vitro measures. RBC-Omics n = 12,219 for oxidative hemolysis and 10,017 for osmotic hemolysis. SCD n = 711 (Walk-PHaSST n 429; PUSH n = 282).

Nearest Gene	Hemolysis	RSid	Chr	Position	Minor Allele	REDS-III				SCD			
						MAF	r^2	beta SNP	p-Value	MAF	r^2	beta SNP	p-Value
GLRX	oxidative	rs10067881	5	95162475	A	0.10	0.99	1.5	9.02E-12	0.065	0.91	-0.36	0.0211
AQP1	osmotic	rs73305784	7	30990948	A	0.17	0.99	1.2	4.48E-08	0.13	0.98	0.24	0.0350
Several	osmotic	rs7967238	12	112378371	A	0.17	0.92	-1.4	9.13E-10	0.31	0.93	0.19	0.0241
Several	osmotic	rs10850001	12	112553032	A	0.44	0.95	1.0	6.80E-10	0.11	0.98	-0.26	0.0297
HBA2	osmotic	chr 16:223678	16	223678	C	0.024	0.80	-4.4	1.27E-14	0.16	0.84	-0.54	8.08E-07
PIEZO1	osmotic	rs34383297	16	88845444	CT	0.47	0.86	0.98	1.02E-08	0.39	0.87	0.17	0.0338
SEC14L4	oxidative	rs9606739	22	30891294	C	0.19	1.0	1.0	3.07E-09	0.13	0.99	0.27	0.0138
G6PD	oxidative	rs78751796	23	153416537	A	0.017	0.49	2.5	2.08E-08	0.11	0.83	0.25	0.0210
G6PD	oxidative	rs115202723	23	153677778	A	0.036	0.93	1.8	1.98E-08	0.25	0.94	0.18	0.0119
G6PD	oxidative	rs28844711	23	153726824	T	0.032	0.83	2.2	6.50E-11	0.24	0.91	0.29	8.96E-05