

# Common host variation drives malaria parasite fitness in healthy human red cells

Emily R Ebel<sup>1,2</sup>, Frans A Kuypers<sup>3</sup>, Carrie Lin<sup>2</sup>, Dmitri A Petrov<sup>1\*</sup>, Elizabeth S Egan<sup>2,4\*</sup>

<sup>1</sup>Department of Biology, Stanford University, Stanford, United States; <sup>2</sup>Department of Pediatrics, Stanford University School of Medicine, Stanford, United States;

<sup>3</sup>Children's Hospital Oakland Research Institute, Oakland, United States;

<sup>4</sup>Department of Microbiology & Immunology, Stanford University School of Medicine, Stanford, United States

**Abstract** The replication of *Plasmodium falciparum* parasites within red blood cells (RBCs) causes severe disease in humans, especially in Africa. Deleterious alleles like hemoglobin S are well-known to confer strong resistance to malaria, but the effects of common RBC variation are largely undetermined. Here, we collected fresh blood samples from 121 healthy donors, most with African ancestry, and performed exome sequencing, detailed RBC phenotyping, and parasite fitness assays. Over one-third of healthy donors unknowingly carried alleles for G6PD deficiency or hemoglobinopathies, which were associated with characteristic RBC phenotypes. Among non-carriers alone, variation in RBC hydration, membrane deformability, and volume was strongly associated with *P. falciparum* growth rate. Common genetic variants in *PIEZO1*, *SPTA1/SPTB*, and several *P. falciparum* invasion receptors were also associated with parasite growth rate. Interestingly, we observed little or negative evidence for divergent selection on non-pathogenic RBC variation between Africans and Europeans. These findings suggest a model in which globally widespread variation in a moderate number of genes and phenotypes modulates *P. falciparum* fitness in RBCs.

\*For correspondence: dpetrov@stanford.edu (DAP); eegan@stanford.edu (ESE)

**Competing interest:** The authors declare that no competing interests exist.

**Funding:** See page 20

**Preprinted:** 09 October 2020

**Received:** 27 April 2021

**Accepted:** 22 September 2021

**Published:** 23 September 2021

**Reviewing Editor:** Jenny Tung, Duke University, United States

© Copyright Ebel et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Malaria caused by the replication of *Plasmodium falciparum* parasites in red blood cells (RBCs) kills hundreds of thousands of children each year ([WHO, 2019](#)). In each 48-hr cycle of blood-stage malaria, parasites must deform RBC membranes to invade them ([Koch, 2017](#); [Kariuki et al., 2020](#)); consume hemoglobin and tolerate the resulting oxidative stress ([Francis et al., 1997](#)); multiply to displace half the RBC volume ([Hanssen et al., 2012](#)); and remodel the RBC membrane to avoid immune detection ([Zhang, 2015](#)). Consequently, genetic disorders that alter aspects of RBC biology are well-known to influence malaria susceptibility ([Kwiatkowski, 2005](#)). For example, sickle cell trait impairs parasite growth by altering hemoglobin polymerization at low oxygen tension ([Pasvol et al., 1978](#); [Archer et al., 2018](#)), while deficiency of the G6PD enzyme involved in oxidative stress tolerance is thought to make parasitized RBCs more susceptible to breakdown ([Ruwende and Hill, 1998](#)). Aside from these diseases, however, the genetic basis of RBC susceptibility to malaria remains mostly unknown.

Large genome-wide association studies (GWAS) have identified a few dozen loci that collectively explain up to 11% of the heritability of the risk of severe versus uncomplicated malaria ([Timmann et al., 2012](#); [Malaria Genomic Epidemiology Network, 2014](#); [Band et al., 2015](#); [Leffler et al., 2017](#); [Ndila et al., 2018](#); [Malaria Genomic Epidemiology Network, 2019](#)). About 10 of the highest-confidence GWAS signals, including 6 loci known from earlier methods ([Allison, 1954](#); [Field et al., 1994](#); [Ruwende and Hill, 1998](#); [Lell et al., 1999](#); [Rowe, 2007](#); [Cao and Galanello, 2010](#); [Galanello and Cao, 2011](#)), are in or near genes expressed predominantly in RBCs. One new GWAS variant has

since been shown to regulate expression of the *ATP2B4* calcium channel (Zámbó *et al.*, 2017) and to be associated with RBC dehydration (Li *et al.*, 2013b), although a functional link between *ATP2B4* and *P. falciparum* replication has yet to be demonstrated. Additional GWAS discoveries of RBC variation important for malaria are not expected without massive increases in sample size (Boyle *et al.*, 2017; *Malaria Genomic Epidemiology Network*, 2019), in part because of the large number of hypotheses tested. Severe malaria is a complex phenotype that combines many factors from RBCs, the vascular endothelium, the immune system, the parasite, and the environment (Mackinnon *et al.*, 2005; de Mendonça *et al.*, 2012). Alternate approaches are therefore needed to discover more genetic variation that impacts the replication of malaria parasites in human RBCs.

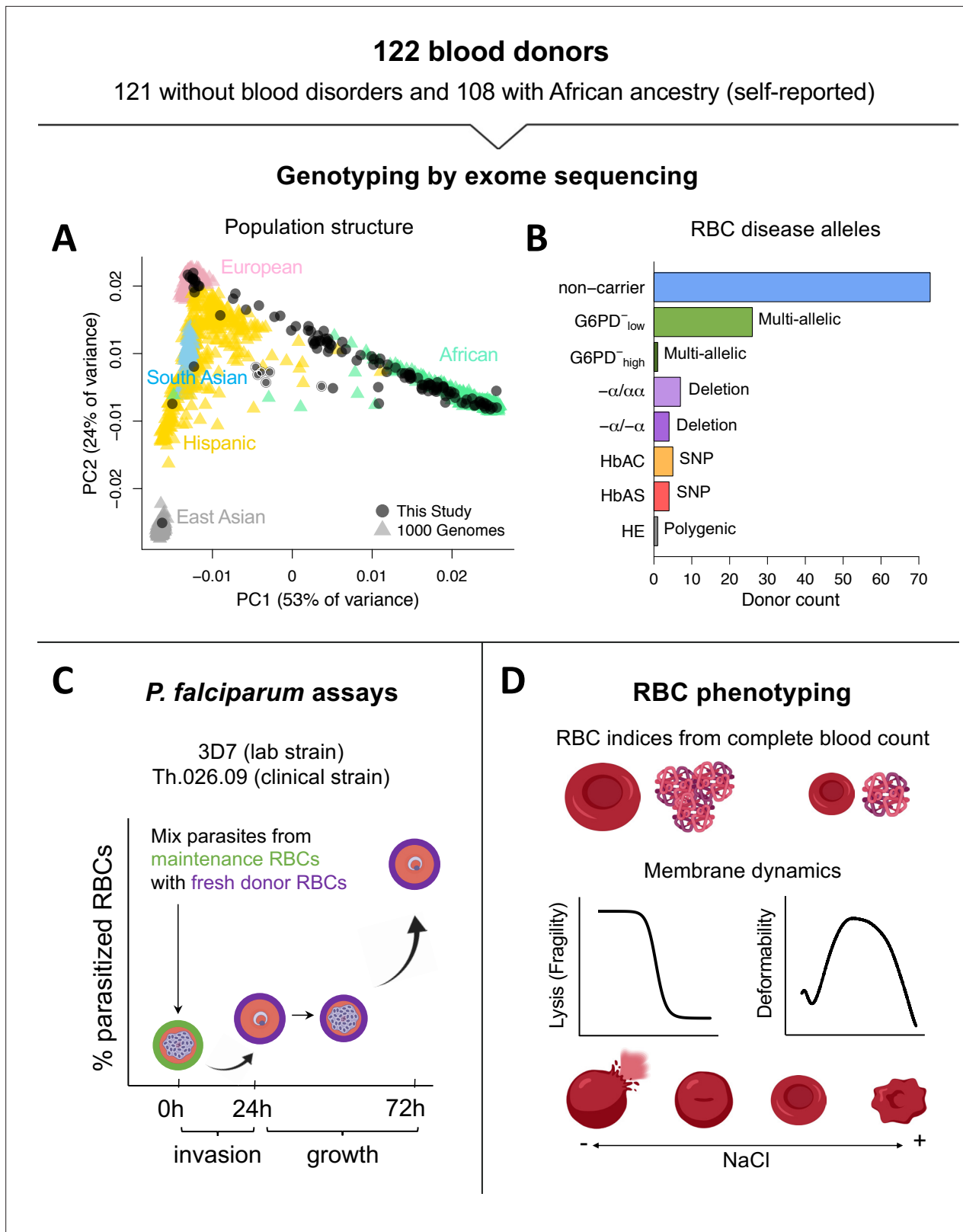
Heritable RBC phenotypes like mean cell volume (MCV), hemoglobin content (HGB/MCH), and antigenic blood type vary widely within and between human populations (Whitfield *et al.*, 1985; Evans *et al.*, 1999; Pilia *et al.*, 2006; Lo *et al.*, 2011; Cooling, 2015; Canela-Xandri *et al.*, 2018). Large GWAS conducted mostly in Europeans have demonstrated that many blood cell phenotypes are shaped by hundreds of small-effect loci distributed throughout the genome, consistent with polygenic or omnigenic models of complex trait genetics (van der Harst *et al.*, 2012; Astle *et al.*, 2016; Chami *et al.*, 2016; Boyle *et al.*, 2017; Chen *et al.*, 2020; Vuckovic *et al.*, 2020). Certain blood phenotypes like average hemoglobin levels, hematocrit, and RBC membrane fragility are also known to differ between African and European populations, although the differences are typically small in magnitude (Garn, 1981; Perry *et al.*, 1992; Beutler and West, 2005; Kaniyas *et al.*, 2017; Page *et al.*, 2021). This variation across populations can largely be explained by a few RBC disease alleles that have been widely selected across Africa for their protective effects on malaria (Beutler and West, 2005; Lo *et al.*, 2011; Kaniyas *et al.*, 2017). Despite the importance of these population-specific variants, a much larger number of common variants with small individual effects on RBC phenotypes are expected to be globally widespread (Biddanda *et al.*, 2020; Chen *et al.*, 2020). It remains untested whether this extensive phenotypic and genetic diversity in RBCs influences malaria susceptibility and, if so, whether it has been shaped by malaria selection.

Here, we approach these questions by performing exome sequencing and extensive RBC phenotyping on blood samples from a diverse human cohort of 122 individuals. We show that *P. falciparum* fitness varies widely among donor cells in vitro, with the distribution of parasite phenotypes in 'healthy' RBCs overlapping those from RBCs carrying classic disease alleles. We apply LASSO variable selection to identify a small set of genes and phenotypes that strongly predict parasite fitness outside of the context of RBC disease, highlighting RBC dehydration and membrane properties as key to modulating *P. falciparum* fitness. We find little evidence that non-pathogenic alleles or phenotypes that confer parasite protection are associated with African ancestry, perhaps because *P. falciparum* is susceptible to RBC variation that exists for other selective or demographic reasons. Overall, these findings advance our understanding of the origin and function of common RBC variation and suggest new targets for therapeutic intervention for malaria.

## Results

### Many healthy blood donors with African ancestry carry alleles for RBC disease

We collected blood samples from 121 donors with no known history of blood disorders, most of whom self-identified as having recent African ancestry (Figure 1A). As a positive control, we also sampled a patient with hereditary elliptocytosis (HE), a polygenic condition characterized by extremely fragile RBC membranes that strongly inhibit *P. falciparum* growth (Schulman *et al.*, 1990; Facer, 1995; Dhermy *et al.*, 2007; Gallagher, 2013). We performed whole-exome sequencing (Figure 1—source data 1), both to check for the presence of known RBC disease alleles and to confirm the population genetic ancestry of our donors. A principal component analysis of more than 35,000 exomic single-nucleotide polymorphisms (SNPs) showed that most donors fell along a continuum from African to European ancestry, as defined by data from the 1000 Genomes Project (Figure 1A). Pairwise kinship coefficients demonstrated that all donors were unrelated, apart from a six-member family with unique ancestry (Figure 1A, light borders). We found that 16% of the healthy donors carried pathogenic hemoglobin alleles (Figure 1B), including 5 heterozygotes for hemoglobin S (HbAS), 4 heterozygotes for hemoglobin C (HbAC), and 11 individuals with one or two copies of an *HBA2* deletion causing



**Figure 1.** Overview of blood donors and study design. (A) PCA of genetic variation across 35,759 unlinked exome SNPs. Donors from this study are plotted on coordinate space derived from 1000 Genomes reference populations. Points with white borders represent six related individuals, five of whom were excluded from the study. All exome variants passing quality filters are available in **Figure 1—source data 1**. (B) Over a third of donors carried alleles for RBC disorders linked to *Plasmodium falciparum* resistance. Individuals with >1 disease allele were classified by their most severe

Figure 1 continued on next page

## Figure 1 continued

condition. non-carrier: Donor without any of the following alleles or conditions.  $G6PD_{low}^-$ : Mild to medium G6PD deficiency (<42% loss of function).  $G6PD_{high}^-$ : Severe G6PD deficiency (>60% loss of function).  $-\alpha/\alpha\alpha$ : heterozygous HBA2 deletion, or alpha thalassemia minima.  $-\alpha/-\alpha$ : homozygous HBA2 deletion, or  $\alpha$ -thalassemia trait. HbAC: heterozygous HBB:E7K, or hemoglobin C trait. HbAS: heterozygous HBB:E7V, or sickle cell trait. HE: hereditary elliptocytosis. (C) Two components of *P. falciparum* fitness were measured with flow cytometry at three timepoints. Invasion is the change in parasitemia as schizonts egress from maintenance RBCs (green) and invade fresh acceptor RBCs from the blood donors (purple). Growth is the multiplication rate from a complete parasite cycle in the fresh acceptor RBCs. (D) RBC phenotypes were measured using complete blood counts with RBC indices, osmotic fragility tests, and ektacytometry on fresh samples. This figure was partially created with [Biorender.com](https://biorender.com). RBC, red blood cell; SNP, single-nucleotide polymorphism.

The online version of this article includes the following figure supplement(s) for figure 1:

**Source data 1.** Individual genotypes, population frequencies, and protein annotations for exome variants passing quality filters (N~160,000).

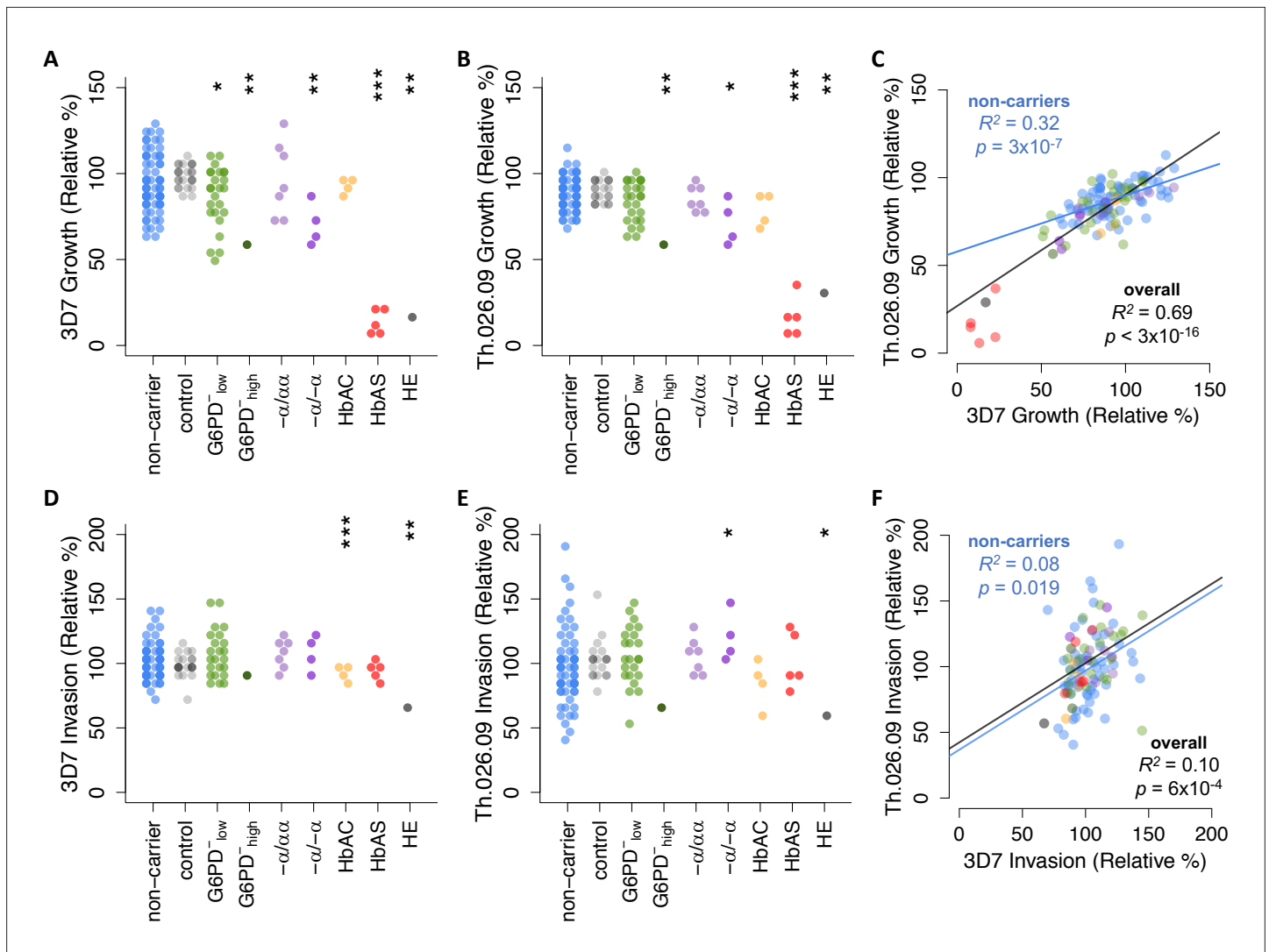
$\alpha$ -thalassemia (Galanello and Cao, 2011). We also scored eight polymorphisms in *G6PD* that have been functionally associated with various degrees of G6PD deficiency (Yoshida et al., 1971; Clarke et al., 2017) and found that 32% of the study population carried at least one, including 12 of the 20 donors with hemoglobinopathies. Among those with wild-type hemoglobin, we identified 1 individual with polymorphisms associated with severe G6PD deficiency (>60% loss of function) and 23 with polymorphisms associated with mild to medium deficiency (<42% loss of function). We detected no alleles linked to other monogenic RBC disorders, including  $\beta$ -thalassemia or xerocytosis (Cao and Galanello, 2010; Glogowska et al., 2017). We therefore classified the remaining 68 unrelated donors as 'non-carriers' of known disease alleles for the purposes of this work.

### *P. falciparum* replication rates vary widely among non-carrier RBCs

To determine the variation in *P. falciparum* fitness among samples with different genotypes, we performed invasion and growth assays with two parasite strains. The genome reference strain 3D7, which was originally isolated from a European, has been continuously cultured in academic labs for at least 40 years (Walliker et al., 1987; Moser, 2020). Th.026.09 is a drug-resistant strain collected from Senegal in 2009 that is minimally adapted to lab culture (Daniels et al., 2012). These divergent strains were selected in an attempt to balance biological realism with reliable in vitro data.

We observed a wide range of *P. falciparum* growth rates among RBC samples, especially among non-carriers that lacked known disease alleles (Figure 2A–C). Each strain's growth rate is defined here as parasite multiplication over a full 48-hr cycle in donor RBCs (Figure 1C), with the mean value for non-carriers set to 100 % after normalization. Briefly, we used a repeated control RBC sample (Figure 2, gray points) and other batch-specific factors to correct for variation in parasite growth across multiple experiments (Figure 2—figure supplement 1). Among non-carriers, growth rates ranged from 64% to 136% for 3D7 (SD=17.7%) and 76% to 128% for Th.026.09 (SD=10.6%) (Figure 2A–B). Per-sample growth rates were strongly correlated between the two strains (Figure 2C,  $R^2=0.69$ ,  $p<3\times 10^{-16}$ ) and positively correlated when measured in different weeks ( $p=0.35$ , Figure 2—figure supplement 2), demonstrating that these data capture meaningful variation among donor RBCs. Furthermore, as expected (Friedman, 1978; Ifediba et al., 1985; Greene, 1993; Facer, 1995), we detected reductions in mean growth rate for both strains in RBCs carrying known disease alleles. These included individuals with  $\alpha$ -thalassemia trait (3D7  $p=0.027$ ; Th.026.09  $p=0.077$ ), HbAS (3D7  $p=1.05\times 10^{-7}$ ; Th.026.09  $p=1.2\times 10^{-4}$ ), and the single carriers of HE and severe G6PD deficiency. Notably, the wide distribution of growth rates for non-carrier RBCs had considerable overlap with the growth rates in carrier RBCs. Only the HbAS and HE samples fell entirely outside the non-carrier range. This observation implies the existence of previously unknown RBC variation that impacts *P. falciparum* growth, which may have cumulative effect sizes comparable to known disease alleles.

We observed a similarly wide range in the efficiency of *P. falciparum* invasion into donor RBCs (Figure 2D–F). Invasion is defined here as the fold-change in parasitemia over the first 24 hours of the assay, when parasites previously maintained in standard culture conditions egressed and invaded new donor RBCs (Figure 1C). Among non-carriers, invasion rates ranged from 70% to 143% for 3D7 (SD=14.9%) and 41% to 193% for Th.026.09 (SD=29.1%) (Figure 2D–E). Compared to growth rates, no disease alleles conferred protection against invasion that was extreme enough to fall outside the broad non-carrier range. HbAC was associated with an 11% decrease in 3D7 invasion ( $p=0.008$ ), while  $\alpha$ -thalassemia trait was associated with a 22% increase in Th.026.09 invasion ( $p=0.091$ ). Only HE



**Figure 2.** *Plasmodium falciparum* replication rate varies widely among donor RBCs. (A, B) Growth of *P. falciparum* lab strain 3D7 (A) or clinical isolate Th.026.09 (B) over a full 48-hr cycle in donor RBCs (see Figure 1C). Growth is presented relative to the average non-carrier rate after correction for batch effects (Figure 2—figure supplement 1; see Materials and methods), including comparison to a repeated RBC control shown in gray. Each carrier group was compared to unrelated non-carriers using Student’s t-test, except in cases where N=1, where asterisks instead indicate the percentile of the non-carrier distribution. Repeated measurements of 11 donors are shown in Figure 2—figure supplement 2. (C) Per-sample growth rates are correlated between the two *P. falciparum* strains. (D–F) As in (A–C) but for *P. falciparum* invasion efficiency (see Figure 1C).  $R^2$  and p-values are derived from OLS regression. \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$ . RBC, red blood cell.

The online version of this article includes the following figure supplement(s) for figure 2:

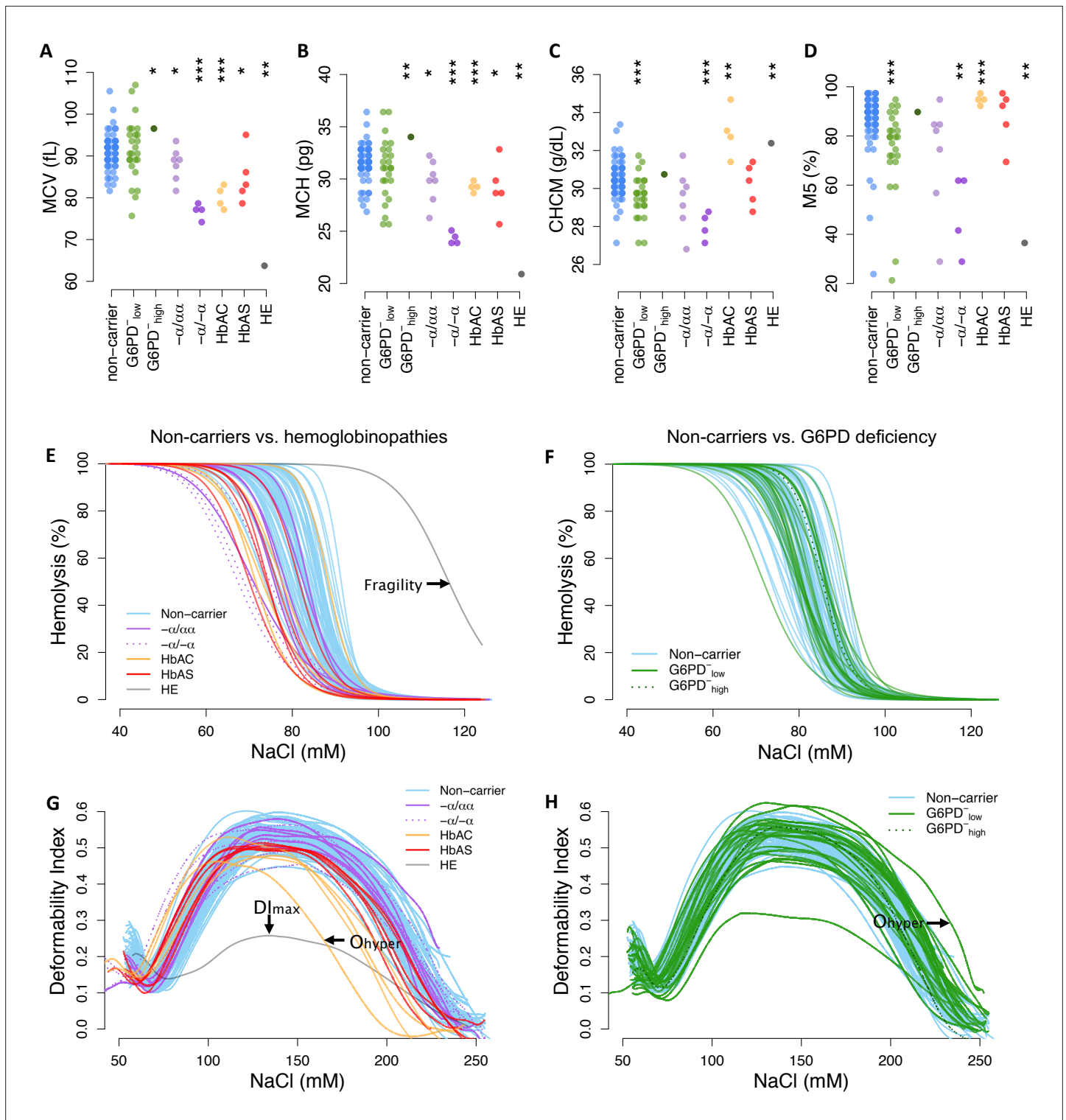
**Figure supplement 1.** Linear models of batch effects on parasite fitness.

**Figure supplement 2.** Repeatability of parasite assays in the same donors over time.

had a strong effect on the invasion efficiency of both strains. The correlation of invasion efficiencies between strains was weaker than for growth (Figure 2F,  $R^2=0.10$ ,  $p=6 \times 10^{-4}$ ), potentially reflecting strain-specific differences in the pathways used for invasion (Wright and Rayner, 2014). However, we also observed greater batch effects (Figure 2—figure supplement 1) and greater variability between repeated samples (Figure 2—figure supplement 2) for invasion than for growth, suggesting that invasion is influenced by greater experimental noise.

### RBC phenotypes vary widely among non-carriers

To assess phenotypic variation across donor RBCs, we measured 22 common indices of RBC size and hemoglobin content from complete blood counts using an ADVIA hematology analyzer (Figure 3A–D;



**Figure 3.** Red cell phenotypes that are abnormal in carriers also vary widely among non-carriers. (A–D) Red cell indices were measured by an ADVIA hematology analyzer. Additional indices are shown in **Figure 3—figure supplement 1**. MCV: mean corpuscular (RBC) volume; MCH: mean cellular hemoglobin; CHCM: cellular hemoglobin concentration; M5: fraction of RBCs with normal volume and normal hemoglobin (see **Figure 3—figure supplement 2**). Statistical tests as in **Figure 2**. (E, F) Osmotic fragility curves. Fragility is defined as the NaCl concentration at which 50% of RBCs lyse (see **Figure 3—figure supplement 4**). (G, H) Ektacytometry curves characterize RBC deformability and dehydration under salt stress (**Figure 3—figure supplement 5**). A heatmap of all phenotypes by carrier status is available in **Figure 3—figure supplement 3**. RBC, red blood cell.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure 3 continued on next page

Figure 3 continued

**Figure supplement 1.** RBC indices data from complete blood counts in donor RBCs.

**Figure supplement 2.** RBC Matrix data from complete blood counts in donor RBCs.

**Figure supplement 3.** Heatmap of RBC phenotypes by carrier status.

**Figure supplement 4.** Osmotic fragility diagram and summary data.

**Figure supplement 5.** Ektacytometry diagram and summary data.

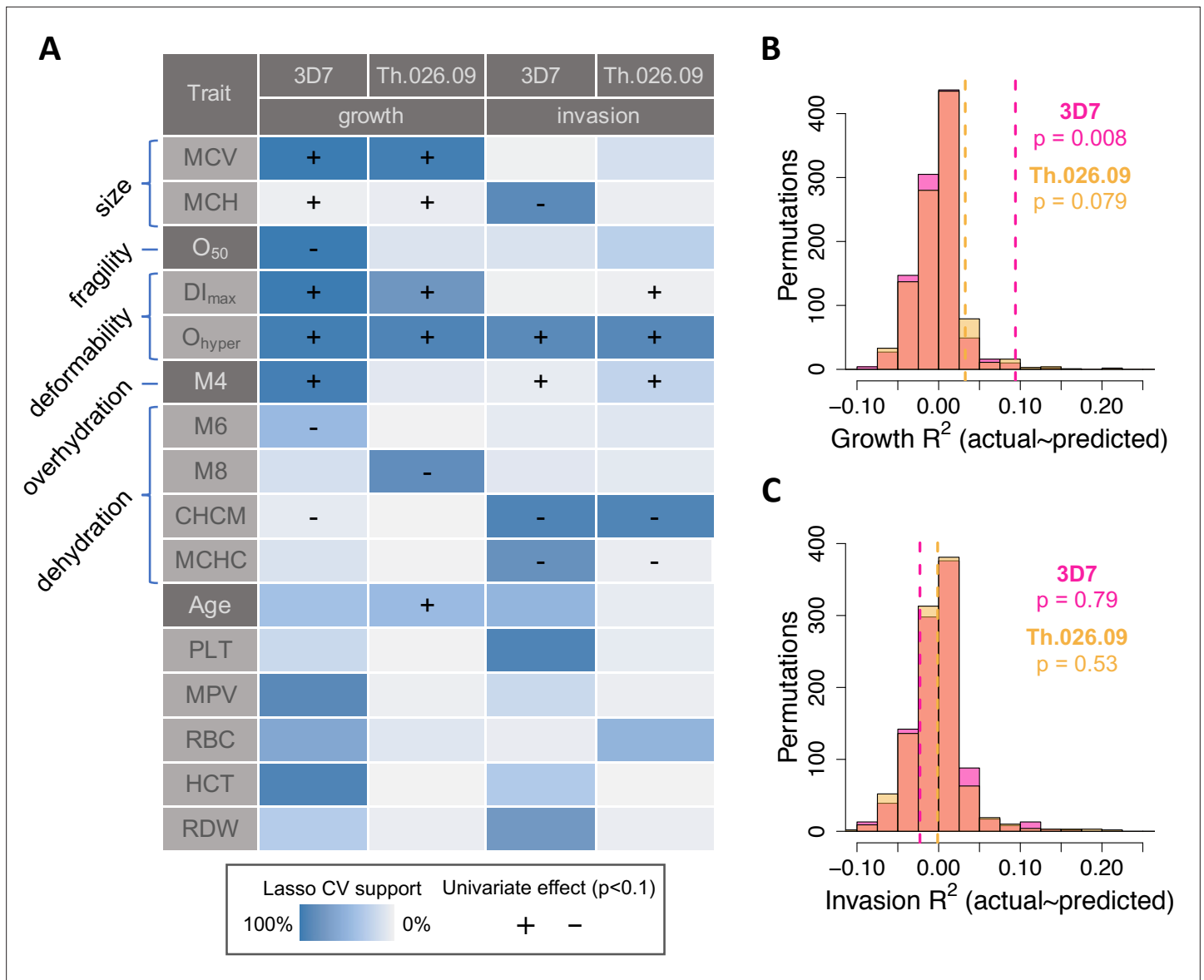
**Figure 3—figure supplements 1–2.** Mean cellular volume (MCV) and hemoglobin mass (MCH) are closely related traits, which can be represented together as cellular hemoglobin concentration (CHCM) or the fraction of RBCs with ‘normal’ hemoglobin and volume indices (M5). As expected, each known disease allele was associated with a distinct set of RBC abnormalities (**Figure 3—figure supplement 3**). These included elevated CHCM for HbAC ( $p=0.033$ ), consistent with dehydration, and very low MCV ( $p=6.8\times 10^{-5}$ ) and MCH ( $p=2.5\times 10^{-7}$ ) for  $\alpha$ -thalassemia trait ( $-\alpha/-\alpha$ ), consistent with microcytic anemia (*Galanello and Cao, 2011*). RBCs from the HE patient also had very low MCV and MCH, reflecting the membrane breakage and volume loss characteristic of this disease. For all these phenotypic measures, we also observed broad distributions in non-carriers that overlapped the distributions of most carriers (**Figure 3A–D**; **Figure 3—figure supplements 1–2**). Notably, the breadth of the non-carrier distribution for each phenotype was large (e.g., 24 fl range for MCV) compared to the average difference between Africans and Europeans (e.g., 3–5 fl; *Beutler and West, 2005*; *Lo et al., 2011*). This wide diversity and substantial overlap between non-carrier and carrier traits suggest that healthy RBCs exist on the same phenotypic continuum as RBCs carrying known disease alleles.

We observed similar patterns of variation in RBC membrane fragility (**Figure 3E–F**; **Figure 3—figure supplement 4**) and membrane deformability (**Figure 3G–H**; **Figure 3—figure supplement 5**), as measured with osmotic fragility tests and osmotic gradient ektacytometry. Both sets of curves represent RBC tolerance to osmotic stress, which can result in swelling and lysis (fragility, **Figure 3E–F**) or dehydration and decreased deformability ( $O_{\text{hyper}}$ , **Figure 3G–H**). Specific hemoglobinopathies were associated with moderate to strong reductions in fragility, deformability, and/or resistance to loss of deformability when dehydrated (**Figure 3—figure supplements 3–5**). HE cells were both extremely fragile and extremely non-deformable. In non-carriers, the distributions for all membrane measures were wide, continuous, and overlapped the distribution of most carriers (**Figure 3E–H**). Overall, these data demonstrate that multiple phenotypic alterations associated with RBC disease alleles are also present in non-carrier RBCs.

### Non-carrier variation in RBC phenotypes predicts *P. falciparum* replication rate

To identify sets of phenotypes associated with *P. falciparum* replication in non-carrier RBCs, we used a machine learning method called LASSO (Least Absolute Shrinkage and Selection Operator) that performs regularization and variable selection (*Tibshirani, 1994*). Briefly, LASSO shrinks the regression coefficients for some possible predictors to zero to obtain a subset of predictors (in this case, phenotypes) that minimizes prediction error. This method is well-suited for data sets where predictors are correlated, as are RBC size, hemoglobin, and membrane dynamics; and for cases where the number of possible predictors is large compared to the number of measurements. To validate RBC phenotypes associated with *P. falciparum* replication by LASSO, we performed k-folds cross-validation (CV) on train and test sets derived from 10,000 divisions of the non-carrier data in 10-folds (see Materials and methods). To further control for overfitting, we also applied the same procedure to 1000 random permutations of the parasite data. Finally, for each trait selected by LASSO in at least 40% of training sets, we applied univariate OLS regression to estimate the sign of its effect on all measured components of parasite fitness. The highest-confidence results from this analysis are summarized in **Figure 4A**, with complete details provided in **Figure 4—source data 1**.

*P. falciparum* fitness in non-carrier RBCs was strongly predicted by variation in traits related to volume, hemoglobin, deformability, and dehydration (**Figure 4A**). Among 25 tested phenotypes, the most strongly predictive was the ektacytometry parameter  $O_{\text{hyper}}$ , which represents a cell’s tendency to retain deformability in the face of dehydration (**Figure 3—figure supplement 4**). In univariate models, non-carrier RBCs with the largest  $O_{\text{hyper}}$  values—that is, those that retained more deformability when



**Figure 4.** RBC phenotypes predict *Plasmodium falciparum* fitness in non-carriers. **(A)** Phenotypes selected by LASSO in at least 40% of train data sets (blue shading; see Materials and methods) in at least one of four models of parasite replication (columns). Each model was trained on ~90% of the data **(B, C)** and tested on the remaining 10% **(B, C)**. (+/-) shows the direction of effect if the phenotype was significantly correlated ( $p < 0.1$ ) with the parasite fitness component in a separate, univariate linear model (**Figure 4—figure supplement 1**; ). MCV: mean RBC volume (fl). MCH: mean corpuscular hemoglobin (pg/RBC). O<sub>50</sub>: Osmotic fragility (mM NaCl; see **Figure 3—figure supplement 4**). DI<sub>max</sub>: Maximum membrane deformability (arbitrary units; see **Figure 3—figure supplement 5**). O<sub>hyper</sub>: Tendency to resist osmotic dehydration and loss of deformability. M4: fraction of RBCs with normal volume and low hemoglobin (see **Figure 3—figure supplement 2**). M6: fraction of RBCs with normal volume and high hemoglobin. M8: fraction of RBCs with low volume and normal hemoglobin. CHCM: cellular hemoglobin concentration mean (g/dl). MCHC: mean corpuscular hemoglobin concentration (g/dl). PLT: platelet number ( $\times 10^3/\mu\text{l}$ ). MPV: mean platelet volume (fl). RBC: red cell number ( $\times 10^6/\mu\text{l}$ ). HCT: hematocrit, or the fraction of blood volume composed of RBCs. RDW: red cell distribution width (%). **(B, C)** Variance in parasite fitness explained by RBC phenotypes in LASSO models. Dashed lines indicate average R<sup>2</sup> for the measured test data. Each histogram shows the same procedure on 1000 permutations of the measured test data. RBC, red blood cell.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

**Source data 1.** Association statistics for individual phenotypic predictors with non-zero LASSO support.

**Figure supplement 1.** Scatterplots of RBC phenotypes against parasite fitness in non-carriers.



dehydrated—supported 22–46% faster parasite growth (3D7  $p=0.003$ , Th.026.09  $p=0.007$ ) and 31–83% more effective invasion (3D7  $p=0.008$ , Th.026.09  $p=0.005$ ) than RBCs with the smallest  $O_{\text{hyper}}$  values, which quickly lost deformability. Consistent with this result, *P. falciparum* replication was inhibited in RBCs that were more dehydrated at baseline (e.g., with higher CHCM; 3D7 invasion  $p=0.006$ , 3D7 growth  $p=0.024$ , Th.026.09 invasion  $p=0.004$ , Th.026.09 growth  $p=0.26$ ). Parasites also grew faster in RBCs with larger mean volume (MCV; 3D7  $p=0.001$ , Th.026.09  $p=0.0009$ ); a greater mass of hemoglobin per cell (MCH; 3D7  $p=0.071$ , Th.026.09  $p=0.016$ ); and more deformable membranes ( $DI_{\text{max}}$ ; 3D7  $p=0.0005$ , Th.026.09  $p=0.008$ ). 3D7 growth was also reduced in RBCs with more fragile membranes ( $O_{50}$ ;  $p=0.005$ ). Additional phenotypes related to platelets and RBC density were selected for some models, but the direction of their effects was unclear when they were considered individually (**Figure 4A**, **Figure 4—source data 1**). These results indicate that common, non-pathogenic variation in RBC size, membrane dynamics, and other correlated traits have meaningful effects on *P. falciparum* replication rate in RBCs.

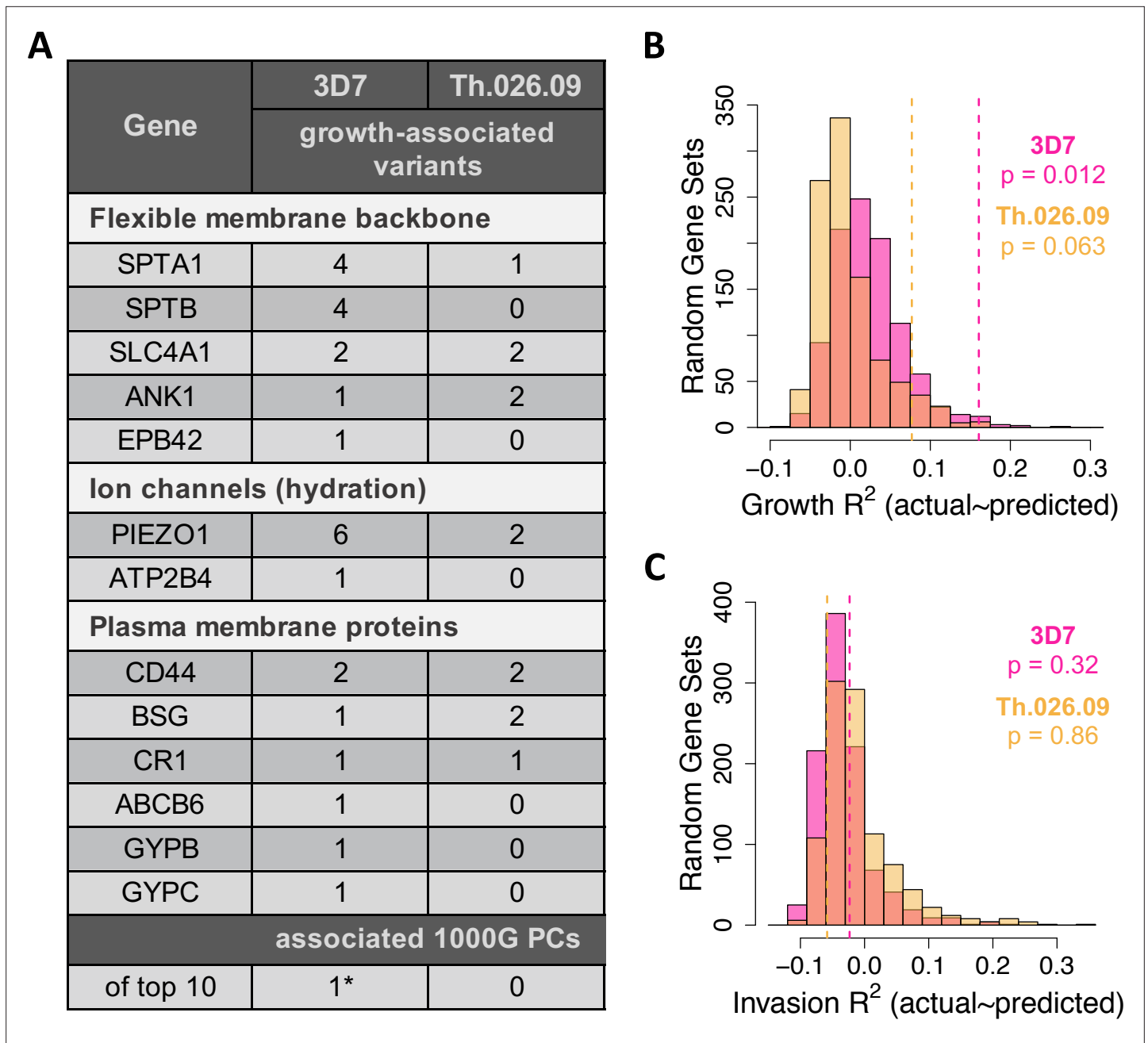
Taken together, the non-carrier phenotypes selected by LASSO from training data ( $N\sim 61$ , **Figure 4A**) explained 3–9% of the variation in parasite growth in separate test data ( $N\sim 7$ , **Figure 4B**; 3D7  $p=0.008$  and RMSE=18.0%; Th.026.09  $p=0.079$  and RMSE=10.8%). This fraction was significantly greater than expected from random permutations, which were centered on  $R^2=0$  in the test data (**Figure 4B**). Notably, prediction error was greater for individuals with parasite growth values farther away from the mean. In contrast, for invasion, RBC phenotypes did not explain more variation in the test data than expected from permutations (**Figure 4C**; 3D7  $p=0.79$  and RMSE=15.0%; Th.026.09  $p=0.53$  and RMSE=29.5%). All phenotype models were less predictive for the clinical isolate Th.026.09 than the lab strain 3D7, perhaps because clinical isolates are less adapted to laboratory conditions. Overall, these results demonstrate that multiple, variable phenotypes impact *P. falciparum* susceptibility in healthy RBCs. Non-carrier cells that are less hospitable to parasites share specific traits with RBCs that carry disease alleles, including smaller size, decreased deformability, and an increased tendency to lose deformability when dehydrating.

### Common RBC alleles predict *P. falciparum* replication in non-carriers

Next, we tested whether non-carrier genotypes derived from exome sequencing could improve our predictions of *P. falciparum* replication rate. With a sample of 68 unrelated non-carriers, we lacked the power to perform the many thousands of tests that are typical in large genetic association studies (**Fadista et al., 2016**). Instead, our study design focused on 23 RBC proteins previously associated with malaria (**Figure 5—source data 1**), which we hypothesized are enriched for common variants impacting *P. falciparum* fitness, as compared to random control sets of RBC proteins. We used the same LASSO procedure described above to test 106 unlinked genetic variants (pairwise  $r^2<0.1$ ) in these 23 RBC proteins, along with RBC phenotypes, for association with *P. falciparum* fitness in non-carriers. To test for the effects of population structure, we also included the top 10 principal components (PCs) from 1000 Genomes as possible predictors. Notably, PC1 is equivalent to the exome-wide fraction of African ancestry, as determined by ADMIXTURE with  $K=4$  from the 1000 Genomes reference populations (see Materials and methods). We again compared these results to permuted data, as well as to 1,000 sets of 23 genes drawn at random from the RBC proteome (**Figure 5—source data 2**).

Taken together, genotypes and phenotypes selected by LASSO explained 7–15% of the variation in parasite growth rate in the test data (**Figure 5B**; 3D7  $p=0.012$  and RMSE=16.5%; Th.026.09  $p=0.063$  and RMSE=10.6%). Prediction error was greater for donors with parasite values farther away from the mean, though this trend was weaker than for phenotype-only models. The variance explained by models using real genotype and phenotype data was significantly larger than expected from permutation (**Figure 5—figure supplement 1A**) and random sets of RBC genes (**Figure 5B**), suggesting that the 23 malaria-related genes contain variation that influences *P. falciparum* development.

Nearly all of the 32 polymorphisms selected by LASSO in growth models occurred in (1) ion channel proteins, which regulate RBC hydration; (2) components of the flexible RBC membrane backbone; or (3) red cell plasma membrane proteins, including known invasion receptors (**Figure 5A**). In the first category, the highly polymorphic ion channel *PIEZO1* contained seven polymorphisms associated with small (<3.7%) to moderate (31%) reductions of *P. falciparum* growth rate. In practice, the smallest effect size that could be reliably determined for an allele with our data was  $\pm 3.7\%$  (**Figure 5—source data 3**). The microsatellite variant *PIEZO1*-E756del, which has been a focus of several recent studies



**Figure 5.** Common variation in malaria-associated genes predicts *Plasmodium falciparum* fitness in non-carrier RBCs. **(A)** Variants in 23 malaria-related genes (**Figure 5—source data 1**) and genetic PCs selected by LASSO in at least >40% of train data sets. Each model was trained on ~90% of the measured data (**B C**) and tested on the remaining 10% (**B C**). The following genes had no associated variants in non-carriers: *CD55*, *EPB41*, *FPN*, *G6PD*, *GYP A*, *GYP E*, *HBA1/2*, *HBB*, and *HP*. \*The only significant PC association was driven by a single East Asian donor (**Figure 5—figure supplement 5**). **(B, C)** Variance in parasite fitness explained by LASSO models including 23 malaria-related genes, the top 10 PCs, and RBC phenotypes. Dashed lines indicate average  $R^2$  for models using the measured test data. Each histogram shows  $R^2$  for models including variants from 23 random genes in the RBC proteome (**Figure 5—source data 2**) instead of malaria-related genes. All predictors with non-zero LASSO support are shown in **Figure 5—source data 3**. Additional histograms from permuted data are shown in **Figure 5—figure supplement 1**. The variance explained by variants undiscovered by previous GWAS is shown in **Figure 5—figure supplement 4**. GWAS, genome-wide association studies; PC, principal component; RBC, red blood cell.

The online version of this article includes the following figure supplement(s) for figure 5:

**Source data 1.** Twenty-three RBC genes with strong links to malaria in the literature.

**Source data 2.** Proteins present in mature RBCs.

**Source data 3.** All genetic and phenotypic predictors with non-zero LASSO support.

Figure 5 continued on next page

Figure 5 continued

**Figure supplement 1.** Variance in parasite fitness explained by permuted data in LASSO models.

**Figure supplement 2.** Lack of association between RBC dehydration phenotypes and PIEZO1 rs59446030 or ATP2B4 rs1419114.

**Figure supplement 3.** Three non-carrier variants with potentially overdominant effects on 3D7 growth.

**Figure supplement 4.** Variants undiscovered by previous GWAS drive most of the association signal between parasite replication rate and the 23 malaria-related genes.

**Figure supplement 5.** An outlier individual for PC2 drives an apparent association between PC2 and 3D7 growth.

**Figure supplement 6.** A six-member family has unique ancestry and parasite susceptibility compared to other non-carrier donors.

(*Ilboudo et al., 2018; Ma et al., 2018; Rooks et al., 2019; Nguetse et al., 2020*), predicted a moderate reduction in Th.026.09 growth ( $-7.9\%$ ,  $p=0.01$ ) but was not related to RBC dehydration in these data (**Figure 5—figure supplement 2**). For 3D7, we also detected one growth-associated variant in *ATP2B4* ( $-5.9\%$ ,  $p=0.075$ ), which encodes the primary RBC calcium channel PMCA4b. This variant tags an *ATP2B4* haplotype implicated by GWAS in protection from severe malaria and many RBC phenotypes (*van der Harst et al., 2012; Li et al., 2013b; Lessard et al., 2017; Lin et al., 2020, Timmann et al., 2012, Zámbo et al., 2017*). Notably, however, this variant has never before been functionally demonstrated to be associated with *P. falciparum* fitness.

*SPTA1* and *SPTB*, which encode the flexible spectrin backbone of RBCs, contained several variants associated with the growth of at least one *P. falciparum* strain, as did the structural linker genes *ANK1*, *SLC4A1*, and *EPB42* (**Figure 5A**). We also identified a total of 10 polymorphisms in *ABCB6*, *GYPB*, *GYPC*, *CR1*, *CD44*, and basigin (*BSG*) that were associated with *P. falciparum* growth. These plasma membrane proteins have all been previously implicated in *P. falciparum* invasion by genetic deficiency studies (*Mayer et al., 2009; Crosnier et al., 2011; Egan et al., 2015; Egan et al., 2018*), and in some cases, studies of natural polymorphisms (*Nagayasu et al., 2001; Leffler et al., 2017*). Notably, two of the variants identified here are synonymous quantitative trait loci (QTL) for *CD44* splicing (rs35356320) and *BSG* expression (rs4682) (*GTE Consortium et al., 2017*), further supporting the possibility that they are functional. No associated variants were detected in the other 10 tested genes, including 3 hemoglobin proteins, *G6PD*, 2 glycoporphins, *CD55*, *EPB41*, *FPN*, and *HP*. Taken together, these data demonstrate that dozens of host genetic variants shape the phenotypic distribution of red cell susceptibility to *P. falciparum* in non-carriers.

Eighteen of the 32 variants selected by LASSO were synonymous, which was not significantly different from the input set of 106 variants ( $p=0.72$ , two-sided binomial test). Over half of the growth-associated variants have previously been associated with gene expression traits, GWAS traits, or GWAS loci through linkage (**Figure 5—source data 3**), suggesting that they indeed tag functional polymorphisms. Novel variants nonetheless contribute substantially to the predictive power of these models (**Figure 5—figure supplement 4**), and nearly all the variants are novel in terms of association with *P. falciparum* growth rate.

In contrast to growth, models of invasion that included genotypic predictors were no more accurate than expected by chance (**Figure 5C**,  $p \geq 0.3$ ; **Figure 5—figure supplement 1B**,  $p \geq 0.15$ ). However, six of the nine RBC invasion receptors contained variants associated with growth (**Figure 5A**), including a SNP in glycoporphin B (*GYPB*) that has been linked to malaria risk in Brazil (*Tarazona-Santos et al., 2011*). These patterns likely stem from experimental noise in our measure of invasion (**Figure 1C; Figure 2C and F; Figure 2—figure supplements 1–2**), though we note that our definition of growth involves a re-invasion event (**Figure 1C**).

No PCs of population structure were significantly associated with *P. falciparum* growth rate (**Figure 5—source data 3**), including the PC that distinguishes Africans from other populations (PC1, **Figure 1A**). One PC was selected by LASSO for 3D7 growth, but this association was driven by a single donor with East Asian ancestry and relatively high susceptibility (**Figure 5—figure supplement 5**). We note that the unique ancestry and extreme phenotypes of the six-member family (**Figure 5—figure supplement 6**) would have driven additional correlations if family members had not been excluded from the LASSO models. Although the present study is limited by sample size, these associations between global genetic PCs and *P. falciparum* growth suggest that additional functional variants remain to be discovered in many populations.

## African ancestry does not predict *P. falciparum* resistance in red cells

Based on evidence from balanced disease alleles like HbAS, it has been suggested that anti-malarial selection has shaped polygenic red cell phenotypes in African populations as a whole (Goheen et al., 2016; Kaniyas et al., 2017; Ma et al., 2018; Page et al., 2021). We tested this hypothesis by examining the correlation between African ancestry and *P. falciparum* fitness in non-carrier RBCs (Figure 6A–D). Surprisingly, we found no evidence that these traits were related, apart from a positive relationship between African ancestry and invasion rate of Th.026.09, the clinical Senegalese strain ( $p=0.004$ ,  $R^2=0.13$ , Figure 6D). To understand this result, we next examined how key RBC phenotypes identified in this study (Figure 4A) vary with African ancestry (Figure 6F–H; Figure 6—figure supplement 1). We found that greater African ancestry predicts reduced osmotic fragility ( $p=1.2\times 10^{-6}$ ), reduced RBC dehydration (CHCM  $p=0.009$ ; MCHC  $p=0.089$ ), and a greater fraction of ‘overhydrated’ RBCs with normal volume and low hemoglobin (M4  $p=0.041$ ). All of these traits actually predict greater red cell susceptibility to *P. falciparum* (Figure 4A), although together they explain less than 13% of the non-carrier variation in 3D7 growth. The remaining key phenotypes do not vary with African ancestry, which may explain why African ancestry itself is only weakly associated with *P. falciparum* fitness in non-carrier RBCs (Figure 6A–D).

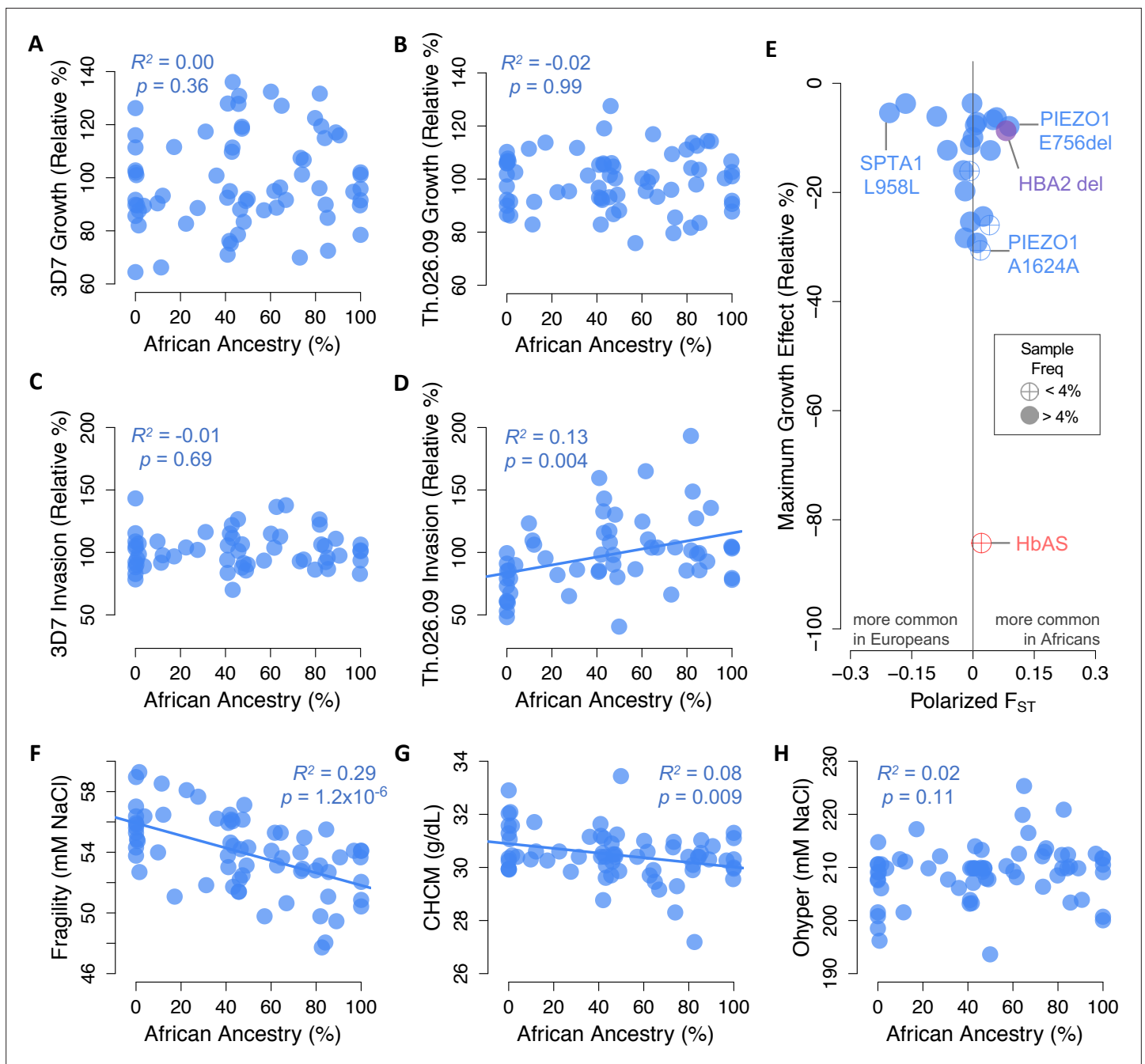
Next, we used allele frequency data from over 54,000 individuals in the gnomAD collection (Karczewski et al., 2020) to test whether the polymorphisms we associated with *P. falciparum* growth occur at different frequencies in African and European populations. Geographical differences in malaria selection are sometimes hypothesized to have increased the frequency of hundreds or thousands of undiscovered anti-malarial alleles in Africa (Mackinnon et al., 2005; Williams, 2006), as has been shown for several variants causing common RBC disorders (Kariuki and Williams, 2020). To address this hypothesis for non-carrier variation, we calculated  $F_{ST}$  between Africans and Europeans for 22 alleles with protective effects large enough to be specified in our sample ( $\geq 3.7\%$ ; Figure 5—source data 3). We found that 11 of these protective alleles (50%) are more common in Africans, which is not more than expected by chance ( $p=0.5$ , one-sided binomial test). The three protective variants with the largest absolute  $F_{ST}$  values are all more common in Europeans, including a synonymous *SPTA1* allele with GWAS associations to several RBC and white blood cell traits. Two protective *PIEZO1* variants are more common in Africans, including E756del and a synonymous variant of large effect. Overall, however, we find no evidence that African populations are enriched for non-pathogenic RBC polymorphisms or phenotypes associated with impaired *P. falciparum* growth in vitro.

## Discussion

Healthy RBCs harbor extensive phenotypic and proteomic variation, both within and between human populations. In this study, we demonstrate that this variation modulates a wide range of RBC susceptibility to *P. falciparum* parasites. Our findings add to a growing understanding of the genetic and phenotypic basis of RBC resistance to *P. falciparum*, especially for RBCs that lack population-specific disease alleles. These findings suggest new targets for future malaria interventions, in addition to challenging assumptions about the role of malaria selection in shaping human RBC diversity.

Exponential replication of *P. falciparum* is a significant driver of malaria disease progression (Bejon et al., 2007). Therefore, the ample variation that we observed in this trait in vitro could be relevant for clinical outcomes in endemic regions. Growth inhibition from HbAS, for example, reduces the risk of death from malaria by reducing parasite density in the blood (Allison, 1954; Luzzatto, 2012). While HbAS has a uniquely extreme effect size, we found a threefold range of parasite replication rates among non-carrier RBCs that share substantial overlap with RBCs carrying other protective variants. Although the physiologically complex basis of severe malaria (Okwa, 2012) makes it difficult to estimate the precise contribution of RBC factors to severe malaria risk, the genotypes and phenotypes we have associated with *P. falciparum* fitness may contribute to malaria susceptibility.

We have shown here that widespread, ‘normal’ variation in RBC hydration and deformability traits are associated with *P. falciparum* fitness in non-carrier RBCs. Interestingly, the protective phenotypes we detect in non-carrier RBCs are also present in carriers, albeit to a stronger degree (Clark et al., 1983; Mockenhaupt, 2000; Pengon et al., 2018). These results are consistent with experimental manipulations that reduce *P. falciparum* growth, such as chemical or genetic dehydration of RBCs (Tiffert et al., 2005; Ma et al., 2018). They are also consistent with the protective effect conferred



**Figure 6.** Little evidence of widespread selection in Africa for slower *Plasmodium falciparum* replication, protective alleles, or protective phenotypes in non-carriers. (A–D) Parasite replication versus the exome-wide fraction of African ancestry in non-carriers, determined with ADMIXTURE by comparison to 1000 Genomes reference populations.  $R^2$  and  $p$ -values are shown for OLS regression. (E) Alleles in 23 malaria-related genes that predict slower *P. falciparum* growth in non-carriers (Figure 5A) are not enriched for higher frequencies in Africa versus Europe. Effect sizes are shown for one allele copy for 3D7 or Th.026.09 growth, whichever was greater. Effect sizes were determined from additive models except for three alleles that appeared overdominant (Figure 5—figure supplement 3).  $F_{ST}$  was calculated from African and European samples in gnomAD (see Materials and methods). HbAS and the HBA2 deletion are shown for comparison. (F–H) RBC phenotypes associated with *P. falciparum* growth versus the exome-wide fraction of African ancestry in non-carriers. Slower *P. falciparum* growth in RBCs is predicted by greater fragility (F), greater dehydration (G), and lower  $O_{hyper}$  (H) (Figure 4A). Additional phenotypes are shown in Figure 6—figure supplement 1.

The online version of this article includes the following figure supplement(s) for figure 6:

**Figure supplement 1.** Scatterplots of RBC phenotypes versus African ancestry in non-carriers.

by Dantu, a rare glycoprotein variant associated with increased membrane tension (*Field et al., 1994; Leffler et al., 2017; Kariuki et al., 2020*). Our data expand upon these prior findings by demonstrating for the first time that common, healthy phenotypic variation in RBC traits contributes meaningfully to *P. falciparum* growth.

In the last decade, several association studies have explored the genetic basis of common variation in RBC traits using large, mostly European cohorts (*van der Harst et al., 2012; Astle et al., 2016; Chami et al., 2016; Canela-Xandri et al., 2018; Chen et al., 2020; Vuckovic et al., 2020*). These studies agree that the broad distribution of RBC phenotypes in humans is shaped by a large number of common alleles, similar to other complex traits (*Boyle et al., 2017*). Although the effects of most individual alleles are likely too small to be considered pathogenic on their own, different combinations of alleles may underlie the broad phenotypic variation observed in non-carriers. We cannot rule out the possibility that some extreme phenotypes could be better explained by the presence of large-effect 'disease' alleles that remain undiscovered. In particular, our study was not powered to detect rare alleles, which could be an important source of missing heritability (*Génin, 2020; Kierczak, 2021*). Some RBC phenotypes are also shaped by environmental variation, such as diet and time of day (*England et al., 1976; Sennels et al., 2011*), which likely diminishes correlations between repeated samples. Although this study cannot distinguish among these explanations for phenotypic variation among non-carrier RBCs, it does suggest that this broad variation is both healthy and functional.

In our linear models of *P. falciparum* growth, phenotypic variation among RBCs was outperformed by genetic variation in a small number of RBC proteins. This result implies the existence of protective RBC phenotypes that we did not measure (or did not measure with sufficient accuracy), such as quantitative proteomic, transcriptomic, and metabolomic traits that could be addressed by future studies. Approximately half of the polymorphisms we identified are non-synonymous and may therefore exert direct effects on phenotypes like RBC membrane structure or ion transport. The other half of associated polymorphisms were synonymous, which could be linked to coding variants but could also have direct effects on mRNA transcription, splicing, and stability (*Sauna and Kimchi-Sarfaty, 2011*). Indeed, silent and coding SNPs are equally likely to be associated with human disease (*Chen et al., 2010*), and many synonymous sites experience strong selection (*Supek et al., 2014; Machado et al., 2020*). Synonymous SNPs that impact splicing, like rs35356320 in CD44, may also impact protein structure. Some other conceivable RBC phenotypes, such as the dynamics of membrane modification during *P. falciparum* development, may only become evident in more detailed time course experiments. The true number of RBC phenotypes that impact *P. falciparum* may be effectively infinite (*Kinsler et al., 2020*), making it useful in practice that genetic variation is more predictive of parasite growth.

One reason that our study could identify genetic associations with a modest sample size was because we focused on a relatively well-defined component of a larger disease that lends itself to controlled, in vitro experiments. Another important explanation is our use of LASSO variable selection on a restricted set of polymorphisms in genes with strong existing links to malaria (*Flynn et al., 2017*). Focusing our tests on a limited number of hypotheses obviated the need to meet an exome-wide significance threshold, while still allowing for the discovery of novel alleles. This approach relies directly on prior knowledge (*Figure 5—source data 1*) and cannot readily be expanded to explicitly test large numbers of anonymous genetic variants. However, testing fewer hypotheses that are more likely to be true helps ensure that 'significant' results are reliable (*Ioannidis, 2005*). Exome-wide data were still critical in this study for assessing population structure, as well as for performing permutation tests that confirmed an enrichment of signal in our 23 focal genes. However, future studies with many more than 68 non-carriers will be required to discover additional associations in unknown genes, non-genic regulatory variation, and any alleles with smaller effects. It is also important to note that genetic linkage complicates the identification of the exact functional polymorphisms in any population sample (*Sohail et al., 2019*); as in GWAS, we cannot rule out that some associated variants are merely linked to the true functional variants. Indeed, about half of our associated variants occur in linkage blocks containing other SNPs associated with RBC traits by GWAS. In this way, our evidence most strongly supports the conclusion that 13 specific RBC genes are strongly enriched for polymorphisms with impacts on *P. falciparum* growth.

The associations we observed for parasite growth were stronger and more significant than the associations for parasite invasion. While batch effects clearly played a role, this may also be due to missing invasion data in 10 non-carrier samples (see Materials and methods) that reduced statistical

power. Both technical and biological reasons may drive the relatively greater noise observed in our invasion data. For example, invasion success may depend on the length of time spent outside the incubator during assay set-up as well as the genotypes of both donor and acceptor red cells. The reproducibility of our invasion data is also constrained by low and variable starting parasitemia and a 24hr time point, which could be substantially improved in future studies using live-cell imaging focused on invasion. Despite these limitations, our 'growth' measurement includes a 're-invasion' event and our growth and invasion measurements are correlated. RBC deformability and dehydration are associated with both fitness components, and SNPs in several canonical invasion receptors are only associated with growth. The invasion data also allow us to highlight unique and interesting trends in the Senegalese clinical strain and in carriers of hemoglobin C.

We also observed weaker associations for the clinical strain Th.026.09 than for the lab strain 3D7. These strains display large differences in absolute growth rate, possibly because Th.026.09 carries costly alleles for drug resistance and 3D7 has had decades longer to adapt to lab conditions (**Walliker et al., 1987; Daniels et al., 2012; Moser, 2020**). Interestingly, African ancestry predicted higher invasion only for Th.026.09, which might indicate that this strain is better adapted to African RBCs. Despite these differences, we showed that normalized fitness values were significantly correlated between the two strains across donors. Several RBC phenotypes and genotypes that predicted fitness in one strain were also replicated in the other. These results suggest that our findings may be generalizable across divergent strains of *P. falciparum*, although future studies would benefit by testing many more lab strains and clinical isolates.

One of the unique aspects of our study is the participation of individuals with a range of African ancestry, defined by similarity to donors from five 1000 Genomes reference populations. We found that African ancestry was unexpectedly associated with RBC phenotypes that improved parasite fitness, particularly for Th.026.09. In the future, it would be very interesting to test for local parasite adaptation to human RBCs using *P. falciparum* strains and RBC samples from around the globe. We also found that the total set of polymorphisms associated with *P. falciparum* growth by LASSO are not enriched in African populations included in the gnomAD database of human variation. Notably, a recent test of data from a large GWAS for severe malaria (**Malaria Genomic Epidemiology Network, 2019**) was also unable to demonstrate that natural selection has driven many malaria-protective alleles to higher frequencies in African versus European populations. Therefore, for the total set of alleles detectable in this study, we offer at least four possible explanations for this unexpected result. First, compared to large-effect disease alleles, the majority of non-pathogenic variants may not have had sufficient time to increase in frequency since *P. falciparum* began expanding in humans some 5000–10,000 years ago (**Sundararaman et al., 2016; Otto et al., 2018**). Second, the complexity of severe malaria could mean that the variants discovered here do not substantially impact disease outcome, especially relative to known disease variants. Third, the variants discovered here may have pleiotropic effects on other phenotypes, which are themselves subject to other selective pressures besides malaria resistance. Finally, human adaptation may be too local to detect with coarse-grain sampling of sub-Saharan African genetic diversity (e.g., **Pankratov et al., 2020**). Overall, however, our data suggest that few RBC alleles remain to be discovered that are both particularly common in Africa and have large effects on *P. falciparum* proliferation in RBCs.

More broadly, these data show that it may be inaccurate to make assumptions about RBC susceptibility to *P. falciparum* based on a person's race or continental ancestry. These kinds of hypotheses (**Williams, 2006; Goheen et al., 2016; Kaniyas et al., 2017; Ma et al., 2018**) are based on well-known examples of balanced disease alleles, which are notable exceptions to the overwhelming genetic similarity of all human populations (**Rosenberg et al., 2002; Novembre and Di Rienzo, 2009**). In our data, RBC variation that is associated with reduced *P. falciparum* fitness is clearly not limited to individuals with recent African ancestry. This result is an important reminder that >90% of the total genetic variation among humans occurs within populations, rather than across them (**Lewontin, 1972; Rosenberg, 2011**); and that the majority of common genetic variation is shared among all human populations (**Biddanda et al., 2020**).

In conclusion, this study demonstrates that substantial phenotypic and genetic diversity in healthy human RBCs impacts the replication of malaria parasites. Whether or not this diversity is shaped by malaria selection, a better understanding of how *P. falciparum* biology is impacted by natural RBC variation could help lead to new therapies for one of humanity's most important infectious diseases.

## Materials and methods

### Key resources table

#### Reagent type (species) or resource

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Biological sample ( <i>Homo sapiens</i> )	Primary whole blood samples	This paper		Freshly drawn from de-identified human subjects into CPDA tubes (IRB #40479)
Strain, strain background ( <i>Plasmodium falciparum</i> )	3D7	PMID:3299700; Obtained from Walter and Eliza Hall Institute, Melbourne, Australia		
Strain, strain background ( <i>P. falciparum</i> )	Th026.09	PMID:22430961; Gift from Daouda Ndiaye and Sarah Volkman, Senegal		
Commercial assay or kit	DNeasy Blood and Tissue Kit	QIAGEN		
Commercial assay or kit	KAPA Hyperplus Kit	Roche		
Commercial assay or kit	SeqCap EZ Prime Exome Kit	Roche		
Sequence-based reagent	Primers amplifying <i>PIEZO1</i> exon 17	PMID:32265284		
Software, algorithm	bwa mem	<a href="http://arxiv.org/abs/1303.3997">http://arxiv.org/abs/1303.3997</a>	0.7.17-r1188	
Software, algorithm	GATK	<a href="https://gatk.broadinstitute.org/hc/en-us">https://gatk.broadinstitute.org/hc/en-us</a>	4.0.0.0	
Software, algorithm	vcftools	doi:10.1093/bioinformatics/btr330	0.1.15	
Software, algorithm	ANNOVAR	PMID:20601685	2018-04-16	
Software, algorithm	PLINK	PMID:17701901	v1.90b6.8 64-bit	
Software, algorithm	ADMIXTURE	PMID:21682921	1.3.0	
Software, algorithm	R	<a href="https://www.R-project.org/">https://www.R-project.org/</a>	3.5.1	
Other	SYBR Green I nucleic acid stain	Invitrogen	S7563	
Other	Drabkin's Reagent	Ricca Chemical	2660-32	

### Sample collection and preparation

One-hundred and twenty-one subjects with no known history of RBC disorders were recruited to donate blood at the Stanford Clinical and Translational Research Unit. This study size was designed to sample multiple individuals carrying alleles of moderate frequency (5% or higher). Written informed consent was obtained from each subject and/or their parent as part of a protocol approved by the Stanford University Institutional Review Board (#40479). To help control for weekly batch effects, subject 1111 donated fresh blood for each parasite assay. Eleven other subjects donated blood on at least 2 different weeks, constituting biological replicates. Whole blood samples from a HE patient were obtained from Dr. Bertil Glader under a separate approved protocol (Stanford IRB #14004) that permitted sample sharing among researchers. All samples were de-identified upon collection by labeling with a random four-to-six digit code. Two samples were eventually removed from analysis based on a failed sequencing library (6449KD) and history of stem cell transplant (8715).

Whole blood was drawn into CPDA tubes and spun down within 36 hr to separate serum, buffy coat, and RBCs. RBCs were washed and stored in RPMI-1640 medium (Sigma-Aldrich) supplemented with 25 mM HEPES, 50 mg/L hypoxanthine, and 2.42 mM sodium bicarbonate at 4°C. Buffy coat was transferred directly to cryotubes and stored at -80°C.



## Exome sequencing and genotype calling

Genomic DNA was isolated from frozen buffy coats using a DNeasy Blood and Tissue Kit (QIAGEN). Libraries were prepared using a KAPA Hyperplus Kit (Roche) and hybridized to human exome probes using the SeqCap EZ Prime Exome Kit (Roche). The resulting exome libraries were sequenced with paired-end 150 bp Illumina reads on the HiSeq or NextSeq platforms at Admera Health (South Plainfield, NJ).

Reads were aligned to the hg38 human reference genome using `bwa mem` (Li, 2013a), yielding an average coverage of 42 X across targeted exome regions (excluding sample 6449KD). Variants were called using GATK best practices (Van der Auwera et al., 2013) and hard filtered with the following parameters:  $QD < 2.0$ ,  $FS > 60.0$ ,  $ReadPosRankSum < -2.5$ ,  $SOR > 2.5$ ,  $MQ < 55.0$ ,  $MQRankSum < -1.0$ , and  $DP < 500$ . To minimize the effects of sequencing errors, variants not present in 1000 Genomes, dbSNP\_138, or the Mills indel collection (Mills et al., 2006) were discarded. Variants that were significantly more frequent in our sample than in gnomAD African and European populations (Karczewski et al., 2020) were also discarded, in order to avoid false associations from miscalled variants. We also excluded singleton variants from all association analyses, potentially including some variants unique to other populations. With the remaining variants, we calculated kinship coefficients among all pairs of donors using `vcftools --relatedness2`. Only the six members of the known family had pairwise coefficients  $> 0.044$ , confirming that no other donors were related.

PIEZO1 E756del was genotyped via PCR and Sanger sequencing according to a previously published protocol (Nguetse et al., 2020). To call deletion variants that cause  $\alpha$ -thalassemia in the paralogous genes HBA2 and HBA1, we extracted reads from each .bam file that lacked any mismatches or soft-clipping and had  $MAPQ \geq 13$  (i.e.,  $< 5\%$  chance of mapping error). Coverage with these well-mapped reads was calculated over the 73 and 81 bp of unique sequence in HBA2 and HBA1 and normalized to each sample's exome-wide coverage. To determine which samples has unusually low coverage, we formed an ad hoc reference panel of seven donors who were unlikely to carry deletion alleles based on their normal MCH, MCV, and HGB and  $> 96\%$  exome-wide European ancestry (Weatherall, 2001). We called heterozygous HBA2 deletions when normalized coverage across three unique regions of the HBA2 gene was below the minimum reference value. Similarly, we called homozygous HBA2 deletions when normalized coverage across three unique regions of the HBA2 gene was less than half of the minimum reference value. This approach resulted in an estimated HBA2 copy number of 2.0 in the reference panel, 0.95 in eight putative heterozygotes and 0.12 in four putative homozygotes. The same method produced no evidence of HBA1 deletion in any sample.

## Variant classification and linkage pruning

Exonic variants in RefSeq genes were identified using ANNOVAR (Wang et al., 2010). Variants were classified into three categories: those within 23 malaria-related genes (Figure 5—source data 1); those within 887 other RBC proteins (Figure 5—source data 2) derived with a medium-confidence filter from the Red Blood Cell Collection database ([rbcc.hegelab.org](http://rbcc.hegelab.org)); and those within any other gene.

Linkage between all pairs of bi-allelic, exonic variants in our 121 genotyped samples was calculated using the `--geno-r2` and `--interchrom-geno-r2` functions in `vcftools` (Danecek et al., 2011). Variants in RBC genes that shared  $r^2 > 0.1$  with any variant in the 23-gene set were removed. Within the 23-gene set and RBC-gene set separately, non-carrier variants were ranked by the p-values of their OLS regression with all four parasite measures. Then, one variant was removed from each pair with  $r^2 > 0.1$ , prioritizing retention in the following order: greater significance across models; non-synonymous protein change; higher frequency in our sample; and finally by random sampling. We report results from additive genetic models (genotypes coded 0/1/2), which performed as well or better than recessive (0/0/2) and dominant (0/2/2) models. For three variants, overdominant models (0/1/-) provided the best fit and were used to estimate effect sizes (Figure 5—figure supplement 3).

## Population analysis

The population ancestry of our donors was assessed by comparison with African, European, East Asian, and South Asian reference populations from the 1000 Genomes Project (Auton et al., 2015). Briefly, variants called from an hg38 alignment of the 1000 Genomes data (Lowy-Gallego et al., 2019) were filtered for concordance with the variants genotyped in this study. The `--indep-pairwise`

command in PLINK (Purcell et al., 2007) was used to prune SNPs with  $r^2 > 0.1$  with any other SNP in a 50-SNP sliding window, producing 35,759 unlinked variants. These variants were analyzed in both PLINK `--pca` and in ADMIXTURE (Alexander and Lange, 2011) with  $K=4$  for the 121 genotyped individuals in this study, alongside 2458 individuals from 1000 Genomes. Pan-African and pan-European allele frequencies were obtained from gnomAD v3 (Karczewski et al., 2020).  $F_{ST}$  for specific alleles was calculated as  $(H_T - H_S)/H_T$  and then polarized, such that positive values indicate variants more common in Africa.

### ***P. falciparum* culture and assays**

Our 3D7 strain of *P. falciparum* was obtained from the Walter and Eliza Hall Institute (Melbourne, Australia) and routinely cultured in human erythrocytes obtained from the Stanford Blood Center. Th.026.09 is a clinical strain isolated from a patient in Senegal in 2009 and kindly provided by Daouda Ndiaye and Sarah Volkman. 3D7 is drug-sensitive and has been lab-adapted for over 40 years, whereas Th.026.09 is drug-resistant and minimally lab-adapted (Walliker et al., 1987; Daniels et al., 2012; Moser, 2020). 3D7 was maintained at 2% hematocrit in RPMI-1640 supplemented with 25 mM HEPES, 50 mg/L hypoxanthine, 2.42 mM sodium bicarbonate, and 4.31 mg/ml Albumax (Invitrogen) at 37°C in 5% CO<sub>2</sub> and 1% O<sub>2</sub>. Th.026.09 was maintained in the same conditions, except that half the Albumax was replaced with heat-inactivated human AB serum.

Parasite growth and invasion assays were performed using schizont-stage parasites isolated from routine culture using a MACS magnet (Miltenyi). Parasites were added at ~0.5% initial parasitemia to fresh erythrocytes suspended at 1% hematocrit in complete RPMI, as above. Parasites were cultured in each erythrocyte sample for 3–5 days in triplicate 100 µl wells. Parasitemia was determined as the average of the three technical replicates, excluding single outlier points, on day 0, day 1 (24 hr), day 3 (72 hr), and in some cases day 5 (120 hr). The fraction of infected RBCs was measured by staining with SYBR Green one nucleic acid stain (Invitrogen, Thermo Fisher Scientific, Eugene, OR) at 1:2000 dilution in PBS/0.3% BSA for 20 min, followed by flow cytometry analysis on a MACSQuant flow cytometer (Miltenyi). Raw invasion rate was defined as the day 1 parasitemia divided by the day 0 parasitemia; raw growth rate was defined as the day 3 (or day 5) parasitemia divided by the day 1 (or day 3) parasitemia. Day 0 parasitemia was not measured in weeks 1–3, so invasion rate estimates are absent for these samples (N=58 unrelated non-carriers with invasion data). The parasite assays failed for both strains in week 9 and for Th.026.09 in week 10, and so were repeated in weeks 10 and 11 with RBCs that had been stored for 1 or 2 weeks.

To correct for batch effects, including substantial week-to-week variation in *P. falciparum* replication rate, we extracted the residuals from a linear regression of the raw parasite values against up to four significantly related batch variables: (1) the raw values for control donor 1111 each week; (2) the parasitemia measured at the previous time point; (3) the age in weeks of the RBCs being measured; and (4) the experimenter performing the assays. Notably, there was no additional effect of ‘Week’ or the length of the experiment (i.e., 3 or 5 days) once the above variables were regressed out. To convert these residuals (mean 0%) to relative percentages (mean 100%), we first trained linear models for growth and invasion in each strain with data from control donor 1111 and carriers with extreme parasite values (HbAS and HE for growth; G6PD<sup>high</sup> and HE for invasion). For these models, relative percentages were calculated by normalizing the raw multiplication rates in these samples to the raw multiplication rate in the 1111 control from that week. These linear models were used to convert residuals to relative percentages for all samples. Finally, the relative percentages were arithmetically adjusted so that the mean invasion and growth values for non-carriers was 100%. Code for this normalization is available at <https://github.com/emily-ebel/RBC> (copy archived at [swh:1:rev-31f953428a4ec5f0fa83201085ada0a0995facb2](https://doi.org/10.7554/eLife.69808)), Ebel, 2021.

### **Red cell phenotyping and normalization**

Complete blood count (CBC) data for RBCs, reticulocytes, and platelets were obtained with an ADVIA 120 hematology analyzer (Siemens, Laguna Hills, CA) at the Red Cell Laboratory at Children’s Hospital Oakland Research Institute. These data were: RBC, HGB, HCT, MCV, MCH, MCHC, CHCM, RDW, HDW, PLT, MPV, Reticulocyte number and percentage, and the fraction of RBCs in each of nine cells of the RBC matrix (see [Figure 3—figure supplement 2](#)). Systematic biases were evident for some

measures in certain weeks, but data from control donor 1111 were not available for all weeks. Therefore, CBC data were normalized such that the median value for non-carrier samples was equal across weeks.

Osmotic fragility tests were performed in duplicate by incubating 20  $\mu$ l of washed erythrocytes for 5 min in 500  $\mu$ l solutions of NaCl in 14 concentrations: 7.17, 6.14, 5.73, 5.32, 4.91, 4.50, 4.30, 4.09, 3.89, 3.68, 3.27, 3.07, 2.66, and 2.46 g/L. Tubes were spun for 5 min at 1000 g and 100  $\mu$ l of supernatant was transferred to a 96-well plate. Hemoglobin concentration was determined by adding 100  $\mu$ l of Drabkin's reagent (Ricca Chemical) to each well and measuring absorbance at OD<sub>540nm</sub> with a Synergy H1 Plate Reader (Biotek). Relative lysis was determined by normalizing to the maximum hemoglobin concentration in the 14-tube series for each sample. After outlier points were manually removed, sigmoidal osmotic fragility curves were estimated under a self-starting logistic model in the *nls* package in R. Curves were summarized by the relative tonicity at which 50% lysis occurred (see **Figure 3—figure supplement 4**) and normalized within weekly batches, such that this value was equal for control sample 1111 across weeks.

Osmotic gradient ektacytometry (Clark et al., 1983; Kuypers, 1990) was performed at the Red Cell Laboratory at Children's Hospital Oakland Research Institute. Red cell deformability estimates across a gradient of NaCl concentrations were fitted to a 20-parameter polynomial model to generate a smooth curve, which was manually verified to closely fit the data. Each curve was summarized with three standard points (**Figure 3—figure supplement 5**; Clark et al., 1983), which were normalized such that the median x- and y-values of the three points was equal for non-carrier samples across weeks.

## Statistical analysis

Student's t-test was used to compare trait values between non-carriers and carriers where  $N > 1$ . Given our modest sample sizes and the expected noise in parasite data, we defined statistical significance as  $p < 0.1$ . Where  $N = 1$  (i.e., for G6PD<sup>high</sup> and HE), significance was assessed with the percentile of the non-carrier distribution. For all comparisons of two continuous variables, OLS linear regression was performed with the *lm* function in R unless otherwise specified. Adjusted  $R^2$  values are reported.

LASSO regression (Tibshirani, 1994; Chatterjee, 2013) was performed in a k-folds CV framework with the *glmnet* and *caret* packages in R. For each of 1000 iterations, we used the *createFolds* function with  $k = 10$  to split the non-carrier data into 10-folds of roughly equal size. Each fold was used as a 'test set' for a LASSO model trained on the remaining nine folds. For each of the 1000 iterations in which 10-folds were created, we collected 10 sets of predictors from the 10 train sets; one average  $R^2$  value for the 10 train sets; and one average  $R^2$  for the 10 test sets. Each set of 1000 resulting  $R^2$  values were normally distributed, and their average is reported in **Figures 4 and 5**. The fraction of k-folds CV support per predictor is based on 10,000 total train models (1000 iterations\*10 folds each) and is reported in **Figure 4A** and **Figure 5—source data 3**.

To perform LASSO with each training set, we used the *cv.glmnet* function with  $\alpha = 1$ . This function split the train data into 10 folds 11 times, the first to estimate a lambda sequence and the rest to compute the fit with one fold omitted. The lambda value that produced minimal error in the training data was then used to predict values in the independent test data described above. Since *cv.glmnet* selects folds at random, we performed this procedure five times for each train/test set (which we term 'internal cross-validation'). We retained  $R^2$  values and selected predictors from the median model of these five internal CVs. Internal CVs did not otherwise contribute to the k-folds CV support reported in the main text.

To assess the significance of each LASSO result, we applied the same modeling procedure to 1000 data sets with randomly permuted parasite values, which preserved the original correlations among RBC predictors. We performed 10 iterations of fold creation for each permuted data set and retained the average  $R^2$  for each set of 10-folds, which generated 1000 fold-averaged  $R^2$  values for train sets and 1000 fold-averaged  $R^2$  values for test sets. Significance was determined by the percentile of the permuted distribution in which the real data fell. We also applied this same procedure to 1000 sets of 23 genes chosen at random from the RBC proteome (**Figure 5—source data 2**).

We noticed that LASSO effect size estimates for each predictor varied considerably across models. Therefore, we used univariate OLS regression on all non-carrier data (excluding five of the six family members) to estimate the effect size of each predictor selected at least once by LASSO. OLS p-values

are reported as a measure of confidence in these effect size estimates, with  $p < 0.1$  considered sufficient evidence to report the effect size. However, because OLS regression was only performed for variants pre-selected by LASSO, these p-values cannot be interpreted on their own as evidence of significant associations.

We compared groups of selected genetic variants using the `binom.test` function in R. For synonymous alleles, we used the proportion of synonymous alleles in the input set of 106 variants (53%) as the null hypothesis. For allele frequencies in Africa and Europe, we categorized protective variants as more common (to any absolute degree) among Africans (N=21,042) or non-Finnish Europeans (N=32,399) in the gnomAD database. The null hypothesis was that 50% of the alleles would be more common in Africans.

## Acknowledgements

The authors gratefully acknowledge the invaluable participation of all volunteer blood donors. Nick Bondy, Bertil Glader, Sandra Larkin, Brian Fleischer, Ashley Dunn, Talal Seddik, Trung Pham, David Vu, and Spectrum Child Health provided crucial assistance in donor coordination and sample processing. *P. falciparum* strain Th.026.09 was kindly provided by Daouda Ndiaye and Sarah Volkman. For quantitative advice, the authors thank Grant Kinsler, Jonathan Pritchard, Susan Holmes, and the Stanford Statistics Consulting Group. This study was primarily supported by a Pilot Early Career award from the Stanford Maternal Child Health Research Institute and a Gabilan Faculty Award from the Stanford University School of Medicine Office of Faculty Development and Diversity (ESE). ERE was an NSF Graduate Research Fellow (DGE-1247312) and received additional support from the Stanford Center for Computational, Evolutionary, and Human Genomics. DAP was funded through an NIH MIRA award 5R35GM118165-05. ESE is a Tashia and John Morgridge Endowed Faculty Scholar in Pediatric Translational Medicine through the Stanford Maternal Child Health Research Institute. Local blood samples were drawn at the Stanford Clinical and Translational Research Unit, which is supported by CTSA Grant UL1 TR001085.

## Additional information

### Funding

Funder	Grant reference number	Author
Stanford University School of Medicine		Elizabeth S Egan
Stanford University School of Medicine		Elizabeth S Egan
Stanford Center for Computational, Evolutionary and Human Genomics		Emily R Ebel
National Institute of General Medical Sciences	5R35GM118165-05	Dmitri A Petrov
National Science Foundation	DGE-1247312	Emily R Ebel

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Emily R Ebel, Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing; Frans A Kuypers, Investigation, Methodology, Resources, Writing – review and editing; Carrie Lin, Investigation; Dmitri A Petrov, Conceptualization, Methodology, Resources, Supervision, Writing – review and editing; Elizabeth S Egan, Conceptualization, Investigation, Methodology, Resources, Supervision, Writing – review and editing

**Author ORCIDs**Emily R Ebel  <http://orcid.org/0000-0002-3244-4250>Dmitri A Petrov  <http://orcid.org/0000-0002-3664-9130>Elizabeth S Egan  <http://orcid.org/0000-0002-2112-7700>**Ethics**

Human subjects: Written informed consent and consent to publish was obtained from each subject and/or their parent as part of a protocol approved by the Stanford University Institutional Review Board (#40479).

**Decision letter and Author response**Decision letter <https://doi.org/10.7554/eLife.69808.sa1>Author response <https://doi.org/10.7554/eLife.69808.sa2>**Additional files****Supplementary files**

- Transparent reporting form

**Data availability**

All data generated or analyzed during this study are included in the manuscript and supporting files. Source data files have been provided for Figures 1, 4, and 5 and other raw data and normalization scripts are available at <https://github.com/emily-ebel/RBC> (copy archived at <https://archive.software-heritage.org/swh:1:rev:31f953428a4ec5f0fa83201085ada0a0995facb2>).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Ebel ER, Kuypers FA, Lin C, Petrov DA, Egan ES	2020	Exome Sequencing from Participants in RBC/Malaria Study	<a href="https://www.ncbi.nlm.nih.gov/bioproject/PRJNA683732/">https://www.ncbi.nlm.nih.gov/bioproject/PRJNA683732/</a>	NCBI BioProject, PRJNA683732

**References**

- Alexander DH, Lange K. 2011. Enhancements to the admixture algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**: 246. DOI: <https://doi.org/10.1186/1471-2105-12-246>, PMID: 21682921
- Allison AC. 1954. Protection afforded by sickle-cell trait against subtertian malarial infection. *British Medical Journal* **1**: 290–294. DOI: <https://doi.org/10.1136/bmj.1.4857.290>, PMID: 13115700
- Archer NM, Petersen N, Clark MA, Buckee CO, Childs LM, Duraisingh MT. 2018. Resistance to *Plasmodium falciparum* in sickle cell trait erythrocytes is driven by oxygen-dependent growth inhibition. *PNAS* **115**: 7350–7355. DOI: <https://doi.org/10.1073/pnas.1804388115>, PMID: 29946035
- Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, Mead D, Bouman H, Riveros-Mckay F, Kostadima MA, Lambourne JJ, Sivapalaratnam S, Downes K, Kundu K, Bomba L, Berentsen K, Bradley JR, Daugherty LC, Delaneau O, Freson K, et al. 2016. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**: 1415–1429. DOI: <https://doi.org/10.1016/j.cell.2016.10.042>, PMID: 27863252
- Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526**: 68–74. DOI: <https://doi.org/10.1038/nature15393>, PMID: 26432245
- Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Malaria Genomic Epidemiology Network. 2015. A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature* **526**: 253–257. DOI: <https://doi.org/10.1038/nature15390>, PMID: 26416757
- Bejon P, Berkley JA, Mwangi T, Ogada E, Mwangi I, Maitland K, Williams T, Scott JAG, English M, Lowe BS, Peshu N, Newton CRJC, Marsh K. 2007. Defining childhood severe falciparum malaria for intervention studies. *PLOS Medicine* **4**: e251. DOI: <https://doi.org/10.1371/journal.pmed.0040251>, PMID: 17713980
- Beutler E, West C. 2005. Hematologic differences between African-Americans and whites: the roles of iron deficiency and alpha-thalassemia on hemoglobin levels and mean corpuscular volume. *Blood* **106**: 740–745. DOI: <https://doi.org/10.1182/blood-2005-02-0713>, PMID: 15790781
- Biddanda A, Rice DP, Novembre J. 2020. A variant-centric perspective on geographic patterns of human allele frequency variation. *eLife* **9**: e60107. DOI: <https://doi.org/10.7554/eLife.60107>, PMID: 33350384
- Boyle EA, Li Yi, Pritchard JK. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**: 1177–1186. DOI: <https://doi.org/10.1016/j.cell.2017.05.038>, PMID: 28622505

- Canela-Xandri O**, Rawlik K, Tenesa A. 2018. An atlas of genetic associations in UK Biobank. *Nature Genetics* **50**: 1593–1599. DOI: <https://doi.org/10.1038/s41588-018-0248-z>, PMID: 30349118
- Cao A**, Galanello R. 2010. Beta-thalassemia. *Genetics in Medicine* **12**: 61–76. DOI: <https://doi.org/10.1097/GIM.0b013e3181cd68ed>, PMID: 20098328
- Chami N**, Chen MH, Slater AJ, Eicher JD, Evangelou E, Tajuddin SM, Love-Gregory L, Kacprowski T, Schick UM, Nomura A, Giri A, Lessard S, Brody JA, Schurmann C, Pankratz N, Yanek LR, Manichaikul A, Pazoki R, Mihailov E, Hill WD, et al. 2016. EXOME genotyping identifies pleiotropic variants associated with red blood cell traits. *American Journal of Human Genetics* **99**: 8–21. DOI: <https://doi.org/10.1016/j.ajhg.2016.05.007>, PMID: 27346685
- Chatterjee S**. 2013. Assumptionless Consistency of the Lasso. *arXiv*. <http://arxiv.org/abs/1303.5817>
- Chen R**, Davydov EV, Sirota M, Butte AJ. 2010. Non-synonymous and synonymous coding snps show similar likelihood and effect size of human disease association. *PLOS ONE* **5**: e13574. DOI: <https://doi.org/10.1371/journal.pone.0013574>, PMID: 21042586
- Chen MH**, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, Trivedi B, Jiang T, Akbari P, Vuckovic D, Bao EL, Zhong X, Manansala R, Laplante V, Chen M, Lo KS, Qian H, Lareau CA, Beaudoin M, Hunt KA, et al. 2020. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**: 1198–1213. DOI: <https://doi.org/10.1016/j.cell.2020.06.045>, PMID: 32888493
- Clark MR**, Mohandas N, Shohet SB. 1983. Osmotic gradient ektacytometry: Comprehensive characterization of red cell volume and surface maintenance. *Blood* **61**: 899–910. DOI: <https://doi.org/10.1182/blood.V61.5.899.899>, PMID: 6831052
- Clarke GM**, Rockett K, Kivinen K, Hubbart C, Jeffreys AE, Rowlands K, Jallow M, Conway DJ, Bojang KA, Pinder M, Usen S, Sisay-Joof F, Sirugo G, Toure O, Thera MA, Konate S, Sissoko S, Niangaly A, Poudiougou B, Mangano VD, et al. 2017. Characterisation of the opposing effects of g6pd deficiency on cerebral malaria and severe malarial anaemia. *eLife* **6**: e15085. DOI: <https://doi.org/10.7554/eLife.15085>, PMID: 28067620
- Cooling L**. 2015. Blood groups in infection and host susceptibility. *Clinical Microbiology Reviews* **28**: 801–870. DOI: <https://doi.org/10.1128/CMR.00109-14>, PMID: 26085552
- Crosnier C**, Bustamante LY, Bartholdson SJ, Bei AK, Theron M, Uchikawa M, Mboup S, Ndir O, Kwiatkowski DP, Duraisingh MT, Rayner JC, Wright GJ. 2011. Basigin is a receptor essential for erythrocyte invasion by *plasmodium falciparum*. *Nature* **480**: 534–537. DOI: <https://doi.org/10.1038/nature10606>, PMID: 22080952
- Danecek P**, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, 1000 Genomes Project Analysis Group. 2011. The variant call format and vcf tools. *Bioinformatics* **27**: 2156–2158. DOI: <https://doi.org/10.1093/bioinformatics/btr330>, PMID: 21653522
- Daniels R**, Ndiaye D, Wall M, McKinney J, Séne PD, Sabeti PC, Volkman SK, Mboup S, Wirth DF. 2012. Rapid, field-deployable method for genotyping and discovery of single-nucleotide polymorphisms associated with drug resistance in *Plasmodium falciparum*. *Antimicrobial Agents and Chemotherapy* **56**: 2976–2986. DOI: <https://doi.org/10.1128/AAC.05737-11>, PMID: 22430961
- de Mendonça VRR**, Goncalves MS, Barral-Netto M. 2012. The host genetic diversity in malaria infection. *Journal of Tropical Medicine* **2012**: 940616. DOI: <https://doi.org/10.1155/2012/940616>, PMID: 23316245
- Dhermy D**, Schrével J, Lecomte MC. 2007. Spectrin-based skeleton in red blood cells and malaria. *Current Opinion in Hematology* **14**: 198–202. DOI: <https://doi.org/10.1097/MOH.0b013e3180d21afd>, PMID: 17414207
- Ebel ER**. 2021. RBC. swh:1:rev:31f953428a4ec5f0fa83201085ada0a0995facb2. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:936eeefd426d26efb71dfd49bea1ccafaa03ac3f;origin=https://github.com/emily-ebel/RBC;visit=swh:1:snp:4e53482ed5ff50b01ed3179d405ec1ac9387b7a5;anchor=swh:1:rev:31f953428a4ec5f0fa83201085ada0a0995facb2>
- Egan ES**, Jiang RHY, Moechtar MA, Barteneva NS, Weekes MP, Nobre LV, Gygi SP, Paulo JA, Frantzreb C, Tani Y, Takahashi J, Watanabe S, Goldberg J, Paul AS, Brugnara C, Root DE, Wiegand RC, Doench JG, Duraisingh MT. 2015. Malaria. A forward genetic screen identifies erythrocyte CD55 as essential for *plasmodium falciparum* invasion. *Science* **348**: 711–714. DOI: <https://doi.org/10.1126/science.aaa3526>, PMID: 25954012
- Egan ES**, Weekes MP, Kanjee U, Manzo J, Srinivasan A, Lomas-Francis C, Westhoff C, Takahashi J, Tanaka M, Watanabe S, Brugnara C, Gygi SP, Tani Y, Duraisingh MT. 2018. Erythrocytes lacking the langereis blood group protein abcb6 are resistant to the malaria parasite *plasmodium falciparum*. *Communications Biology* **1**: 45. DOI: <https://doi.org/10.1038/s42003-018-0046-2>, PMID: 30271928
- England JM**, Ward SM, Down MC. 1976. Microcytosis, anisocytosis and the red cell indices in iron deficiency. *British Journal of Haematology* **34**: 589–597. DOI: <https://doi.org/10.1111/j.1365-2141.1976.tb03605.x>, PMID: 990192
- Evans DM**, Frazer IH, Martin NG. 1999. Genetic and environmental causes of variation in basal levels of blood cells. *Twin Research* **2**: 250–257. DOI: <https://doi.org/10.1375/136905299320565735>, PMID: 10723803
- Facer CA**. 1995. Erythrocytes carrying mutations in spectrin and protein 4.1 show differing sensitivities to invasion by *plasmodium falciparum*. *Parasitology Research* **81**: 52–57. DOI: <https://doi.org/10.1007/BF00932417>, PMID: 7724514
- Fadista J**, Manning AK, Florez JC, Groop L. 2016. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European Journal of Human Genetics* **24**: 1202–1205. DOI: <https://doi.org/10.1038/ejhg.2015.269>, PMID: 26733288
- Field SP**, Hempelmann E, Mendelow BV, Fleming AF. 1994. Glycophorin variants and *plasmodium falciparum*: Protective effect of the dantu phenotype in vitro. *Human Genetics* **93**: 148–150. DOI: <https://doi.org/10.1007/BF00210600>, PMID: 8112738

- Flynn CJ, Hurvich CM, Simonoff JS. 2017. On the sensitivity of the lasso to the number of predictor variables. *Statistical Science. Institute of Mathematical Statistics* **32**: 88–105. DOI: <https://doi.org/10.1214/16-STS586>
- Francis SE, Sullivan DJ, Goldberg D. E. 1997. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. *Annual Review of Microbiology* **51**: 97–123. DOI: <https://doi.org/10.1146/annurev.micro.51.1.97>, PMID: 9343345
- Friedman MJ. 1978. Erythrocytic mechanism of sickle cell resistance to malaria. *PNAS* **75**: 1994–1997. DOI: <https://doi.org/10.1073/pnas.75.4.1994>, PMID: 347452
- Galanello R, Cao A. 2011. Alpha-thalassemia. *Genetics in Medicine* **13**: 83–88. DOI: <https://doi.org/10.1097/GIM.0b013e3181fcb468>
- Gallagher PG. 2013. Abnormalities of the erythrocyte membrane. *Pediatric Clinics of North America* **60**: 1349–1362. DOI: <https://doi.org/10.1016/j.pcl.2013.09.001>, PMID: 24237975
- Garn SM. 1981. Lower hematocrit levels in blacks are not due to diet or socioeconomic factors. *Pediatrics* **67**: 580. PMID: 7254983.
- Génin E. 2020. Missing heritability of complex diseases: case solved? *Human Genetics* **139**: 103–113. DOI: <https://doi.org/10.1007/s00439-019-02034-4>, PMID: 31165258
- Glogowska E, Schneider ER, Maksimova Y, Schulz VP, Lezon-Geyda K, Wu J, Radhakrishnan K, Keel SB, Mahoney D, Freidmann AM, Altura RA, Gracheva EO, Bagriantsev SN, Kalfa TA, Gallagher PG. 2017. Novel mechanisms of piezo1 dysfunction in hereditary xerocytosis. *Blood* **130**: 1845–1856. DOI: <https://doi.org/10.1182/blood-2017-05-786004>, PMID: 28716860
- Goheen MM, Wegmüller R, Bah A, Darboe B, Danso E, Affara M, Gardner D, Patel JC, Prentice AM, Cerami C. 2016. Anemia offers stronger protection than sickle cell trait against the erythrocytic stage of falciparum malaria and this protection is reversed by iron supplementation. *EBioMedicine* **14**: 123–130. DOI: <https://doi.org/10.1016/j.ebiom.2016.11.011>, PMID: 27852523
- Greene LS. 1993. G6pd deficiency as protection against falciparum malaria: An epidemiologic critique of population and experimental studies. *American Journal of Physical Anthropology* **36**: 153–178. DOI: <https://doi.org/10.1002/ajpa.1330360609>
- GTEX Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEX (eGTEX) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration & Visualization—EBI, Genome Browser Data Integration & Visualization—UCSC Genomics Institute, University of California Santa Cruz, Lead analysts, Laboratory, Data Analysis & Coordinating Center (LDACC); NIH program management; et al. 2017. Genetic effects on gene expression across human tissues. *Nature* **550**: 204–213. DOI: <https://doi.org/10.1038/nature24277>, PMID: 29022597
- Hansen E, Knoechel C, Dearnley M, Dixon MWA, Le Gros M, Larabell C, Tilley L. 2012. Soft x-ray microscopy analysis of cell volume and hemoglobin content in erythrocytes infected with asexual and sexual stages of *Plasmodium falciparum*. *Journal of Structural Biology* **177**: 224–232. DOI: <https://doi.org/10.1016/j.jsb.2011.09.003>, PMID: 21945653
- Ifediba TC, Stern A, Ibrahim A, Rieder RF. 1985. *Plasmodium falciparum* in vitro: Diminished growth in hemoglobin h disease erythrocytes. *Blood* **65**: 452–455. DOI: <https://doi.org/10.1182/blood.V65.2.452.452>, PMID: 3881144
- Ilboudo Y, Bartolucci P, Garrett ME, Ashley-Koch A, Telen M, Brugnara C, Galactéros F, Lettre G. 2018. A common functional piezo1 deletion allele associates with red blood cell density in sickle cell disease patients. *American Journal of Hematology* **93**: E362–E365. DOI: <https://doi.org/10.1002/ajh.25245>, PMID: 30105803
- Ioannidis JPA. 2005. Why most published research findings are false. *PLOS Medicine* **2**: e124. DOI: <https://doi.org/10.1371/journal.pmed.0020124>, PMID: 16060722
- Kanias T, Lanteri MC, Page GP, Guo Y, Endres SM, Stone M, Keating S, Mast AE, Cable RG, Triulzi DJ, Kiss JE, Murphy EL, Kleinman S, Busch MP, Gladwin MT. 2017. Ethnicity, sex, and age are determinants of red blood cell storage and stress hemolysis: Results of the REDS-III Rbc-omics study. *Blood Advances* **1**: 1132–1141. DOI: <https://doi.org/10.1182/bloodadvances.2017004820>, PMID: 29034365
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**: 434–443. DOI: <https://doi.org/10.1038/s41586-020-2308-7>, PMID: 32461654
- Kariuki SN, Marin-Menendez A, Introvini V, Ravenhill BJ, Lin Y-C, Macharia A, Makale J, Tendwa M, Nyamu W, Kotar J, Carrasquilla M, Rowe JA, Rockett K, Kwiatkowski D, Weekes MP, Cicuta P, Williams TN, Rayner JC. 2020. Red blood cell tension protects against severe malaria in the Dantu blood group. *Nature* **585**: 579–583. DOI: <https://doi.org/10.1038/s41586-020-2726-6>, PMID: 32939086
- Kariuki SN, Williams TN. 2020. Human genetics and malaria resistance. *Human Genetics* **139**: 801–811. DOI: <https://doi.org/10.1007/s00439-020-02142-6>, PMID: 32130487
- Kierczak M. 2021. The contribution of rare whole genome sequencing variants to plasma protein levels and to the missing heritability. *Research Square* **1**: 625433/v1. DOI: <https://doi.org/10.21203/rs.3.rs-625433/v1>
- Kinsler G, Geiler-Samerotte K, Petrov D. 2020. A Genotype-Phenotype-Fitness Map Reveals Local Modularity and Global Pleiotropy of Adaptation. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.06.25.172197>
- Koch M. 2017. *Plasmodium falciparum* erythrocyte-binding antigen 175 triggers a biophysical change in the red blood cell that facilitates invasion. *PNAS* **114**: 4225–4230. DOI: <https://doi.org/10.1073/pnas.1620843114>

- Kuypers FA.** 1990. Use of ektacytometry to determine red cell susceptibility to oxidative stress. *The Journal of Laboratory and Clinical Medicine* **116**: 527–534.
- Kwiatkowski DP.** 2005. How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics* **77**: 171–192. DOI: <https://doi.org/10.1086/432519>, PMID: 16001361
- Leffler EM,** Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, Bojang KA, Conway DJ, Jallow M, Sisay-Joof F, Bougouma EC, Mangano VD, Modiano D, Sirima SB, Achidi E, Apinjoh TO, Marsh K, Ndila CM, Peshu N, Williams TN, et al. 2017. Resistance to malaria through structural variation of red blood cell invasion receptors. *Science* **356**: 1140–1152. DOI: <https://doi.org/10.1126/science.aam6393>, PMID: 28522690
- Lell B,** May J, Schmidt-Ott RJ, Lehman LG, Luckner D, Greve B, Matousek P, Schmid D, Herlich K, Mockenhaupt FP, Meyer CG, Bienzle U, Kremsner PG. 1999. The role of red blood cell polymorphisms in resistance and susceptibility to malaria. *Clinical Infectious Diseases* **28**: 794–799. DOI: <https://doi.org/10.1086/515193>, PMID: 10825041
- Lessard S,** Gatof ES, Beaudoin M, Schupp PG, Sher F, Ali A, Prehar S, Kurita R, Nakamura Y, Baena E, Ledoux J, Oceandy D, Bauer DE, Lettre G. 2017. An erythroid-specific atp2b4 enhancer mediates red blood cell hydration and malaria susceptibility. *The Journal of Clinical Investigation* **127**: 3065–3074. DOI: <https://doi.org/10.1172/JCI94378>, PMID: 28714864
- Lewontin RC.** 1972. The apportionment of human diversity', in evolutionary biology. *Taylor and Francis* **6**: 381–398. DOI: [https://doi.org/10.1007/978-1-4684-9063-3\\_14](https://doi.org/10.1007/978-1-4684-9063-3_14)
- Li H.** 2013a. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv*. <http://arxiv.org/abs/1303.3997>
- Li J,** Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, Mentch FD, Guo Y, Kim C, Xia Q, Chiavacci RM, Thomas KA, Qiu H, Grant SFA, Furth SL, Hakonarson H, Sleiman PMA. 2013b. GWAS of blood cell Traits identifies novel associated loci and epistatic interactions in caucasian and african-american children. *Human Molecular Genetics* **22**: 1457–1464. DOI: <https://doi.org/10.1093/hmg/dd534>, PMID: 23263863
- Lin S-H,** Brown DW, Machiela MJ. 2020. LDTRAIT: An online tool for identifying published phenotype associations in linkage disequilibrium. *Cancer Research* **80**: 3443–3446. DOI: <https://doi.org/10.1158/0008-5472.CAN-20-0985>, PMID: 32606005
- Lo KS,** Wilson JG, Lange LA, Folsom AR, Galarnau G, Ganesh SK, Grant SFA, Keating BJ, McCarroll SA, O'Donnell CJ, Palmas W, Tang W, Tracy RP, Reiner AP, Lettre G. 2011. Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and african Americans. *Human Genetics* **129**: 307–317. DOI: <https://doi.org/10.1007/s00439-010-0925-1>, PMID: 21153663
- Lowy-Gallego E,** Fairley S, Zheng-Bradley X, Ruffier M, Clarke L, Flicek P, 1000 Genomes Project Consortium. 2019. Variant calling on the GRCH38 assembly with the data from phase three of the 1000 Genomes project. *Wellcome Open Research* **4**: 50. DOI: <https://doi.org/10.12688/wellcomeopenres.15126.2>, PMID: 32175479
- Luzzatto L.** 2012. Sickle cell anaemia and malaria. *Mediterranean Journal of Hematology and Infectious Diseases* **4**: e2012065. DOI: <https://doi.org/10.4084/MJHID.2012.065>, PMID: 23170194
- Ma S,** Cahalan S, LaMonte G, Grubaugh ND, Zeng W, Murthy SE, Paytas E, Gamini R, Lukacs V, Whitwam T, Loud M, Lohia R, Berry L, Khan SM, Janse CJ, Bandell M, Schmedt C, Wengelnik K, Su AI, Honore E, et al. 2018. Common piezo1 allele in african populations causes rbc dehydration and attenuates plasmodium infection. *Cell* **173**: 443–455. DOI: <https://doi.org/10.1016/j.cell.2018.02.047>, PMID: 29576450
- Machado HE,** Lawrie DS, Petrov DA. 2020. Pervasive strong selection at the level of codon usage bias in *Drosophila melanogaster*. *Genetics* **214**: 511–528. DOI: <https://doi.org/10.1534/genetics.119.302542>, PMID: 31871131
- Mackinnon MJ,** Mwangi TW, Snow RW, Marsh K, Williams TN. 2005. Heritability of malaria in Africa. *PLOS Medicine* **2**: 20340. DOI: <https://doi.org/10.1371/journal.pmed.0020340>, PMID: 16259530
- Malaria Genomic Epidemiology Network.** 2014. Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics* **46**: 1197–1204. DOI: <https://doi.org/10.1038/ng.3107>, PMID: 25261933
- Malaria Genomic Epidemiology Network.** 2019. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nature Communications* **10**: 5732. DOI: <https://doi.org/10.1038/s41467-019-13480-z>, PMID: 31844061
- Mayer DCG,** Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, Mendoza LH, Miller LH. 2009. Glycophorin B is the erythrocyte receptor of *Plasmodium falciparum* erythrocyte-binding ligand, EBL-1. *PNAS* **106**: 5348–5352. DOI: <https://doi.org/10.1073/pnas.0900878106>, PMID: 19279206
- Mills RE,** Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE. 2006. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Research* **16**: 1182–1190. DOI: <https://doi.org/10.1101/gr.4565806>, PMID: 16902084
- Mockenhaupt FP.** 2000. Anaemia in pregnant Ghanaian women: importance of malaria, iron deficiency, and haemoglobinopathies', *Transactions of the Royal Society of Tropical Medicine and Hygiene*. *Royal Society of Tropical Medicine and Hygiene* **94**: 477–483. DOI: [https://doi.org/10.1016/S0035-9203\(00\)90057-9](https://doi.org/10.1016/S0035-9203(00)90057-9)
- Moser KA.** 2020. 'strains used in whole organism *Plasmodium falciparum* vaccine trials differ in genome structure, sequence, and immunogenic potential', *genome medicine* 2020 12:1. *BioMed Central* **12**: 1–17. DOI: <https://doi.org/10.1186/S13073-019-0708-9>
- Nagayasu E,** Ito M, Akaki M, Nakano Y, Kimura M, Looareesuwan S, Aikawa M. 2001. CR1 density polymorphism on erythrocytes of falciparum malaria patients in Thailand. *The American Journal of Tropical Medicine and Hygiene* **64**: 1–5. DOI: <https://doi.org/10.4269/ajtmh.2001.64.1.11425154>, PMID: 11425154



- Ndila CM**, Uyoga S, Macharia AW, Nyutu G, Peshu N, Ojal J, Shebe M, Awuondo KO, Mhuri N, Tsofa B, Sepúlveda N, Clark TG, Band G, Clarke G, Rowlands K, Hubbart C, Jeffreys A, Kariuki S, Marsh K, Mackinnon M, et al. 2018. Human candidate gene polymorphisms and risk of severe malaria in children in Kilifi, Kenya: A case-control association study. *The Lancet* **5**: e333–e345. DOI: [https://doi.org/10.1016/S2352-3026\(18\)30107-8](https://doi.org/10.1016/S2352-3026(18)30107-8), PMID: 30033078
- Nguetse CN**, Purington N, Ebel ER, Shakya B, Tetard M, Kreamsner PG, Velavan TP, Egan ES. 2020. A common polymorphism in the mechanosensitive ion channel *piezo1* is associated with protection from severe malaria in humans. *PNAS* **117**: 9074–9081. DOI: <https://doi.org/10.1073/pnas.1919843117>, PMID: 32265284
- Novembre J**, Di Rienzo A. 2009. Spatial patterns of variation due to natural selection in humans. *Nature Reviews. Genetics* **10**: 745–755. DOI: <https://doi.org/10.1038/nrg2632>, PMID: 19823195
- Okwa OO**. 2012. Malaria parasites, malaria parasites. InTech. DOI: <https://doi.org/10.5772/1477>
- Otto TD**, Gilabert A, Crellen T, Böhme U, Arnathau C, Sanders M, Oyola SO, Okouga AP, Boundenga L, Willaume E, Ngoubangoye B, Moukoudoum ND, Paupy C, Durand P, Rougeron V, Ollomo B, Renaud F, Newbold C, Berriman M, Prugnolle F. 2018. Genomes of all known members of a *Plasmodium* subgenus reveal paths to virulent human malaria. *Nature Microbiology* **3**: 687–697. DOI: <https://doi.org/10.1038/s41564-018-0162-2>, PMID: 29784978
- Page GP**, Kaniyas T, Guo YJ, Lanteri MC, Zhang X, Mast AE, Cable RG, Spencer BR, Kiss JE, Fang F, Endres-Dighe SM, Brambilla D, Nourai M, Gordeuk VR, Kleinman S, Busch MP, Gladwin MT, National Heart, Lung, and Blood Institute (NHLBI) Recipient Epidemiology Donor Evaluation Study–III (REDS-III) program. 2021. Multiple-ancestry genome-wide association study identifies 27 loci associated with measures of hemolysis following blood storage. *The Journal of Clinical Investigation* **131**: 146077. DOI: <https://doi.org/10.1172/JCI146077>, PMID: 34014839
- Pankratov V**, Montinaro F, Kushniarevich A, Hudjashov G, Jay F, Saag L, Flores R, Marnetto D, Seppel M, Kals M, Vösa U, Taccioli C, Möls M, Milani L, Aasa A, Lawson DJ, Esko T, Mägi R, Pagani L, Metspalu A, et al. 2020. Differences in local population history at the finest level: The case of the Estonian population. *European Journal of Human Genetics* **28**: 1580–1591. DOI: <https://doi.org/10.1038/s41431-020-0699-4>, PMID: 32712624
- Pasvol G**, Weatherall DJ, Wilson RJM. 1978. Cellular mechanism for the protective effect of haemoglobin S against *P. falciparum* malaria. *Nature* **274**: 701–703. DOI: <https://doi.org/10.1038/274701a0>, PMID: 353566
- Pengon J**, Svasti S, Kamchonwongpaisan S, Vattanaviboon P. 2018. Hematological parameters and red blood cell morphological abnormality of glucose-6-phosphate dehydrogenase deficiency co-inherited with thalassemia. *Hematology/Oncology and Stem Cell Therapy* **11**: 18–24. DOI: <https://doi.org/10.1016/j.hemonc.2017.05.029>, PMID: 28641093
- Perry GS**, Byers T, Yip R, Margen S. 1992. Iron nutrition does not account for the hemoglobin differences between blacks and whites. *The Journal of Nutrition* **122**: 1417–1424. DOI: <https://doi.org/10.1093/jn/122.7.1417>, PMID: 1619469
- Pilia G**, Chen W-M, Scuteri A, Orrú M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, Mameli C, Vacca L, Deiana M, Olla N, Masala M, Cao A, Najjar SS, Terracciano A, Nedorezov T, Sharov A, et al. 2006. Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLOS Genetics* **2**: e132. DOI: <https://doi.org/10.1371/journal.pgen.0020132>, PMID: 16934002
- Purcell S**, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**: 559–575. DOI: <https://doi.org/10.1086/519795>, PMID: 17701901
- Rooks H**, Brewin J, Gardner K, Chakravorty S, Menzel S, Hannemann A, Gibson J, Rees DC. 2019. A gain of function variant in *piezo1* (E756DEL) and sickle cell disease. *Haematologica* **104**: e91–e93. DOI: <https://doi.org/10.3324/haematol.2018.202697>, PMID: 30237267
- Rosenberg NA**, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. 2002. Genetic structure of human populations. *Science* **298**: 2381–2385. DOI: <https://doi.org/10.1126/science.1078311>, PMID: 12493913
- Rosenberg NA**. 2011. A population-genetic perspective on the similarities and differences among worldwide human populations. *Human Biology* **83**: 659–684. DOI: <https://doi.org/10.3378/027.083.0601>, PMID: 22276967
- Rowe JA**. 2007. Blood group O protects against severe *Plasmodium falciparum* malaria through the mechanism of reduced rosetting. *PNAS* **104**: 17471–17476. DOI: <https://doi.org/10.1073/pnas.0705390104>
- Ruwende C**, Hill A. 1998. Glucose-6-phosphate dehydrogenase deficiency and malaria. *Journal of Molecular Medicine* **76**: 581–588. DOI: <https://doi.org/10.1007/s001090050253>, PMID: 9694435
- Sauna ZE**, Kimchi-Sarfaty C. 2011. Understanding the contribution of synonymous mutations to human disease. *Nature Reviews. Genetics* **12**: 683–691. DOI: <https://doi.org/10.1038/nrg3051>, PMID: 21878961
- Schulman S**, Roth EF Jr, Cheng B, Rybicki AC, Sussman II, Wong M, Wang W, Ranney HM, Nagel RL, Schwartz RS. 1990. Growth of *Plasmodium falciparum* in human erythrocytes containing abnormal membrane proteins. *PNAS* **87**: 7339–7343. DOI: <https://doi.org/10.1073/pnas.87.18.7339>, PMID: 2205856
- Sennels HP**, Jørgensen HL, Hansen ALS, Goetze JP, Fahrenkrug J. 2011. Diurnal variation of hematology parameters in healthy young males: The Bispebjerg study of diurnal variations. *Scandinavian Journal of Clinical and Laboratory Investigation* **71**: 532–541. DOI: <https://doi.org/10.3109/00365513.2011.602422>, PMID: 21988588
- Sohail M**, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CW, Hirschhorn J, Daly MJ, Patterson N, Neale B, Mathieson I, Reich D, Sunyaev SR. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *eLife* **8**: e39702. DOI: <https://doi.org/10.7554/eLife.39702>, PMID: 30895926

- Sundararaman SA**, Plenderleith LJ, Liu W, Loy DE, Learn GH, Li Y, Shaw KS, Ayouba A, Peeters M, Speede S, Shaw GM, Bushman FD, Brisson D, Rayner JC, Sharp PM, Hahn BH. 2016. Genomes of cryptic chimpanzee plasmodium species reveal key evolutionary events leading to human malaria. *Nature Communications* **7**: 11078. DOI: <https://doi.org/10.1038/ncomms11078>, PMID: 27002652
- Supek F**, Miñana B, Valcárcel J, Gabaldón T, Lehner B. 2014. Synonymous mutations frequently act as driver mutations in human cancers. *Cell* **156**: 1324–1335. DOI: <https://doi.org/10.1016/j.cell.2014.01.051>, PMID: 24630730
- Tarazona-Santos E**, Castilho L, Amaral DRT, Costa DC, Furlani NG, Zuccherato LW, Machado M, Reid ME, Zalis MG, Rossit AR, Santos SEB, Machado RL, Lustigman S. 2011. Population genetics of GYPB and association study between GYPB\*S/s polymorphism and susceptibility to *P. falciparum* infection in the Brazilian Amazon. *PLOS ONE* **6**: e16123. DOI: <https://doi.org/10.1371/journal.pone.0016123>, PMID: 21283638
- Tibshirani R**. 1994. Regression shrinkage and selection via the lasso. *JSTOR* **58**: 267–288. DOI: <https://doi.org/10.1.1.35.7574>
- Tiffert T**, Lew VL, Ginsburg H, Krugliak M, Croisille L, Mohandas N. 2005. The hydration state of human red blood cells and their susceptibility to invasion by *Plasmodium falciparum*. *Blood* **105**: 4853–4860. DOI: <https://doi.org/10.1182/blood-2004-12-4948>, PMID: 15728121
- Timmann C**, Thye T, Vens M, Evans J, May J, Ehmen C, Sievertsen J, Muntau B, Ruge G, Loag W, Ansong D, Antwi S, Asafo-Adjei E, Nguah SB, Kwakye KO, Akoto AOY, Sylverken J, Brendel M, Schuldt K, Loley C, et al. 2012. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature* **489**: 443–446. DOI: <https://doi.org/10.1038/nature11334>, PMID: 22895189
- Van der Auwera GA**, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43**: bi1110s43. DOI: <https://doi.org/10.1002/0471250953.bi1110s43>, PMID: 25431634
- van der Harst P**, Zhang W, Mateo Leach I, Rendon A, Verweij N, Sehmi J, Paul DS, Elling U, Allayee H, Li X, Radhakrishnan A, Tan S-T, Voss K, Weichenberger CX, Albers CA, Al-Hussani A, Asselbergs FW, Ciullo M, Danjou F, Dina C, et al. 2012. Seventy-five genetic loci influencing the human red blood cell. *Nature* **492**: 369–375. DOI: <https://doi.org/10.1038/nature11677>, PMID: 23222517
- Vuckovic D**, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, Chen MH, Raffield LM, Tardaguila M, Huffman JE, Ritchie SC, Megy K, Ponstingl H, Penkett CJ, Albers PK, Wigdor EM, Sakaue S, Moscati A, Manansala R, Lo KS, et al. 2020. The polygenic and monogenic basis of blood traits and diseases. *Cell* **182**: 1214–1231. DOI: <https://doi.org/10.1016/j.cell.2020.08.008>, PMID: 32888494
- Walliker D**, Quakyi IA, Wellems TE, McCutchan TF, Szarfman A, London WT, Corcoran LM, Burkot TR, Carter R. 1987. Genetic analysis of the human malaria parasite *Plasmodium falciparum*. *Science* **236**: 1661–1666. DOI: <https://doi.org/10.1126/SCIENCE.3299700>, PMID: 3299700
- Wang K**, Li M, Hakonarson H. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* **38**: e164. DOI: <https://doi.org/10.1093/nar/gkq603>, PMID: 20601685
- Weatherall DJ**. 2001. Phenotype-genotype relationships in monogenic disease: Lessons from the thalassaemias. *Nature Reviews. Genetics* **2**: 245–255. DOI: <https://doi.org/10.1038/35066048>, PMID: 11283697
- Whitfield JB**, Martin NG, Rao DC. 1985. Genetic and environmental influences on the size and number of cells in the blood. *Genetic Epidemiology* **2**: 133–144. DOI: <https://doi.org/10.1002/gepi.1370020204>, PMID: 4054596
- WHO**. 2019. Malaria eradication: Benefits, future scenarios and feasibility. Executive summary, who strategic advisory group on malaria eradication. <https://www.who.int/publications/i/item/who-cds-gmp-2019-10> [Accessed August 07, 2021].
- Williams TN**. 2006. Human red blood cell polymorphisms and malaria. *Current Opinion in Microbiology* **9**: 388–394. DOI: <https://doi.org/10.1016/j.mib.2006.06.009>, PMID: 16815736
- Wright GJ**, Rayner JC. 2014. *Plasmodium falciparum* erythrocyte invasion: Combining function with immune evasion. *PLOS Pathogens* **10**: e1003943. DOI: <https://doi.org/10.1371/journal.ppat.1003943>, PMID: 24651270
- Yoshida A**, Beutler E, Motulsky AG. 1971. Human glucose-6-phosphate dehydrogenase variants bulletin of the World Health Organization. *World Health Organization* **45**: 243–253 PMID: 5316621.,
- Zámbó B**, Várady G, Padányi R, Szabó E, Németh A, Langó T, Enyedi Á, Sarkadi B. 2017. Decreased calcium pump expression in human erythrocytes is connected to a minor haplotype in the atp2b4 gene. *Cell Calcium* **65**: 73–79. DOI: <https://doi.org/10.1016/j.ceca.2017.02.001>, PMID: 28216081
- Zhang Y**. 2015. Multiple stiffening effects of nanoscale knobs on human red blood cells infected with *Plasmodium falciparum* malaria parasite. *PNAS* **112**: 6068–6073. DOI: <https://doi.org/10.1073/pnas.1505584112>