TITLE

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

AUTHORS

Emily R Ebel[1,2], Frans A Kuypers[3], Carrie Lin[2], Dmitri A Petrov[1*], Elizabeth S Egan[2,4*]

[1]Department of Biology, Stanford University, Stanford, CA 94305

[2]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305

[3]Children's Hospital Oakland Research Institute, Oakland, CA, 94609

[4]Department of Microbiology & Immunology, Stanford University School of Medicine, Stanford, CA 94305

[*]co-corresponding authors: eegan@stanford.edu; dpetrov@stanford.edu

SUMMARY

The replication of *Plasmodium falciparum* parasites within red blood cells (RBCs) causes severe disease in humans, especially in Africa. The influence of host RBC variation on parasite replication is largely uncharted, aside from a handful of deleterious alleles like sickle cell. Here, we integrated analyses of exome sequencing, RBC phenotyping, and parasite fitness assays on blood from 122 individuals, most with African ancestry. In donors lacking alleles for hemoglobinopathies or G6PD deficiency, RBC phenotypes including size, deformability, and hydration status explained 21-38% of the variation in parasite growth rate. With the addition of non-pathogenic polymorphisms in 14 RBC proteins including *SPTA1*, *PIEZO1,* and *ATP2B4,* our model explained 73-82% of the variation in parasite growth rate. Interestingly, we observed little evidence for divergent selection on this variation between Africans and Europeans. These findings suggest a model in which widespread, non-pathogenic variation in a moderate number of genes strongly modulates *P. falciparum* fitness in RBCs.

KEY WORDS: malaria, *Plasmodium falciparum*, red blood cells, parasite fitness, natural variation

32 **INTRODUCTION**

33 Malaria caused by the replication of *Plasmodium falciparum* parasites in red blood
34 cells (RBCs) kills hundreds of thousands of children each year (WHO, 2019). In each 48-
35 hour cycle of blood stage malaria, parasites must deform RBC membranes to invade
36 them(Koch *et al.*, 2017; Kariuki *et al.*, 2020); consume hemoglobin and tolerate the
37 resulting oxidative stress (Francis, Sullivan and Goldberg, 1997); multiply to displace half
38 the RBC volume (Hanssen *et al.*, 2012); and remodel the RBC membrane to avoid immune
39 detection (Zhang *et al.*, 2015). Consequently, genetic disorders that alter aspects of RBC
40 biology are well known to influence malaria susceptibility (Kwiatkowski, 2005). For
41 example, sickle cell trait impairs parasite growth by altering hemoglobin polymerization at
42 low oxygen tension (Pasvol, Weatherall and Wilson, 1978; Archer *et al.*, 2018), while
43 deficiency of the G6PD enzyme involved in oxidative stress tolerance is thought to make
44 parasitized RBCs more susceptible to breakdown (Ruwende and Hill, 1998). Aside from
45 these diseases, however, the genetic basis of malaria susceptibility remains mostly
46 unknown. Several large genome-wide association studies (GWAS) have recently made
47 progress on this question by identifying a few dozen loci that collectively explain up to 11%
48 of the heritability of the risk of severe versus uncomplicated malaria (Timmann *et al.*, 2012;
49 Rockett *et al.*, 2014; Band *et al.*, 2015; Leffler *et al.*, 2017; Ndila *et al.*, 2018; MGE Network,
50 2019a). Eleven of these loci are in or near genes expressed predominantly in RBCs, and one
51 new GWAS variant has since been shown to regulate expression of the *ATP2B4* calcium
52 channel (Zámbó *et al.*, 2017) and to be associated with RBC dehydration (Li *et al.*, 2013).
53 However, most other RBC loci that have been identified by GWAS are not novel, and a
54 functional link between *ATP2B4* and *P. falciparum* replication has yet to be demonstrated.
55 More discoveries about the influence of RBC variation on malaria susceptibility are not
56 expected from similar GWAS (Boyle, Li and Pritchard, 2017; MGE Network, 2019b), in part
57 because severe malaria is a complex phenotype that depends on a combination of factors
58 from RBCs, the vascular endothelium, the immune system, and the environment
59 (Mackinnon *et al.*, 2005; De Mendonça, Goncalves and Barral-Netto, 2012). Alternate
60 approaches are therefore needed to discover more genetic variation that impacts the
61 replication of malaria parasites in human RBCs.

62 RBC phenotypes like mean cell volume (MCV), hemoglobin content (HGB/MCH), and
63 antigenic blood type vary widely within and between human populations (Lo *et al.*, 2011;
64 Cooling, 2015; Canela-Xandri, Rawlik and Tenesa, 2018). 40-90% of the variation in
65 common RBC indices is heritable (Whitfield, Martin and Rao, 1985; Evans, Frazer and
66 Martin, 1999; Pilia *et al.*, 2006), and GWAS have demonstrated that these phenotypes have
67 polygenic bases (van der Harst *et al.*, 2012; Astle *et al.*, 2016; Chami *et al.*, 2016; Chen *et al.*,
68 2020; Vuckovic *et al.*, 2020). Average hemoglobin levels, hematocrit, and especially RBC
69 membrane fragility have also been shown to differ between African and European
70 populations, likely for genetic reasons (Garn, 1981; Perry *et al.*, 1992; Beutler and West,
71 2005; Kanias *et al.*, 2017; Page *et al.*, 2017). Although African populations are highly
72 genetically diverse(Campbell and Tishkoff, 2008), some of these differences between
73 Africans and Europeans are explained by RBC disease alleles (Beutler and West, 2005; Lo *et*
74 *al.*, 2011; Kanias *et al.*, 2017) that have been widely selected across Africa for their
75 protective effects on malaria. It remains untested whether other extensive phenotypic and
76 genetic diversity in RBCs, aside from these disease alleles, also influences malaria
77 susceptibility and has been similarly shaped by malaria selection.
78 Here, we approach these questions by performing exome sequencing and extensive
79 RBC phenotyping on blood samples from a diverse human cohort of 122 individuals. We
80 show that *P. falciparum* fitness varies widely among donor cells *in vitro*, with the
81 distributions of parasite phenotypes in 'healthy' RBCs overlapping those from RBCs
82 carrying classic disease alleles. We apply Lasso variable selection to identify a small set of
83 alleles and phenotypes that strongly predict parasite fitness outside of the context of RBC
84 disease, highlighting RBC membrane and dehydration genes and phenotypes as key to
85 modulating *P. falciparum* fitness. However, we find no evidence that non-pathogenic alleles
86 or phenotypes that confer parasite protection are associated with African ancestry.
87 Overall, these findings advance our understanding of the origin and function of common
88 RBC variation and suggest new targets for therapeutic intervention for malaria.
89
90 **RESULTS**
91 **Many healthy blood donors with African ancestry carry alleles for RBC disease**
92 We collected blood samples from 121 donors with no known history of blood

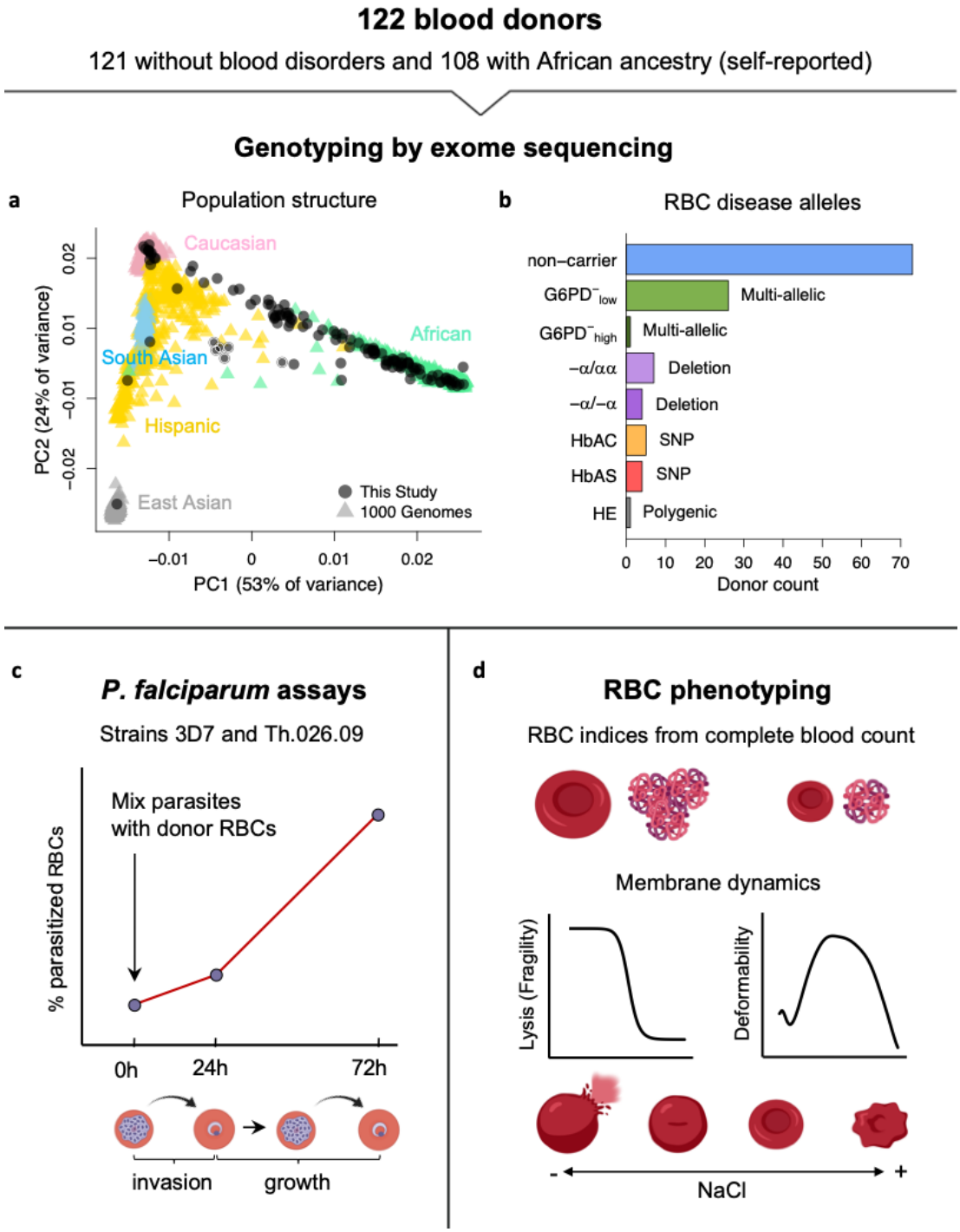


Figure 1. Overview of study design and blood donors. (a) PCA of genetic variation across 35,759 unlinked exome SNPs. Samples collected in this study are plotted on coordinate space derived from 1000 Genomes reference populations. Points with white borders represent the six related individuals in this study. **(b)** Exome sequencing revealed that 37% of donors carried alleles for RBC disorders linked to *P. falciparum* resistance. Individuals with >1 disease allele were assigned to the category of the most severe 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). **-α/αα:** heterozygous HBA2 deletion, or alpha thalassemia minima. **-α/-α:** 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)** Parasite replication rate in donor RBCs was measured via flow cytometry at three timepoints over 72 hours. 3D7 is a lab-adapted strain used in routine culture. Th.026.09 is a recent clinical isolate from Senegal. **(d)** RBC phenotypes were measured using complete blood counts with RBC indices, osmotic fragility tests, and ektacytometry.

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, Schrével and Lecomte, 2007; Gallagher, 2013). We performed whole-exome sequencing, 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 unlinked, exomic 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). 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 α-thalassemia (Galanello and Cao, 2011). We also scored eight polymorphisms in *G6PD* that have been functionally associated with various degrees of G6PD deficiency (Yoshida, Beutler and Motulsky, 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 one 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). To our knowledge, we detected no alleles linked to other monogenic RBC disorders, including β-thalassemia or xerocytosis (Cao and Galanello, 2010; Glogowska *et al.*, 2017). We therefore classified the remaining 73 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 strains of parasite: a

136 laboratory adapted strain (3D7) and a recent clinical isolate from Senegal (Th.026.09).
137 We observed a wide range of *P. falciparum* growth rates among RBC samples, especially
138 among non-carriers that lacked known disease alleles (Figure 2A-C). Each strain's growth
139 rate is defined here as parasite multiplication over a full 48-hour cycle in donor RBCs
140 (Figure 1C), with the mean value for non-carriers set to 100% after normalization. Briefly,
141 we used a repeated control RBC sample (Figure 2, gray circles) and other batch-specific
142 factors to correct for variation in parasite growth across multiple experiments (see
143 Methods). Among non-carriers, growth rates ranged from 64-136% for 3D7 (SD = 17.7%)
144 and 76-128% for Th.026.09 (SD = 10.7%) (Figure 2A-B). Per-sample growth rates were
145 strongly correlated between the two strains (Figure 2C, $p < 3x10^{-16}$) and positively
146 correlated when measured in different weeks (Figure S1A, ρ=0.34), demonstrating that
147 these data capture real variation among donor RBCs. Furthermore, as expected (Friedman,
148 1978; Ifediba *et al.*, 1985; Greene, 1993; Facer, 1995), we detected reductions in mean
149 growth rate for both strains in RBCs carrying known disease alleles. These included severe
150 G6PD deficiency (3D7, 60% of the non-carrier average, p<0.014; Th.026.09, 64%,
151 p<0.014), α-thalassemia trait (3D7, 75%, p=0.027; Th.026.09, 81%, p = 0.077), HbAS (3D7,
152 16%, $p=1.05x10^{-7}$; Th.026.09, 19%, $p =1.2x10^{-4}$), and HE (3D7, 18%, p<0.014; Th.026.09,
153 33%, p<0.014). Notably, the wide distribution of growth rates for non-carrier RBCs had
154 considerable overlap with the growth rates in carrier RBCs. Only the HbAS and HE samples
155 fell entirely outside the non-carrier range. This observation implies the existence of
156 previously unknown RBC variation that impacts *P. falciparum* growth, which may have
157 effect sizes comparable to the milder known disease alleles.
158 We observed a similarly wide range in the efficiency of *P. falciparum* invasion into
159 donor RBCs (Figure 2D-F). Invasion is defined here as the fold-change in parasitemia over
160 the first 24 hours of the assay, when parasites previously maintained in standard culture
161 conditions egressed and invaded new donor RBCs (Figure 1C). Among non-carriers,
162 invasion rates ranged from 66-173% for 3D7 (SD=15.8%) and 41-193% for Th.026.09
163 (SD=28.9%) (Figure 2D-E). Compared to growth rates, no disease alleles conferred
164 protection against invasion that was extreme enough to fall outside the broad non-carrier
165 range. In 3D7 only, HbAC was associated with a 9% decrease in invasion (p=0.019) and

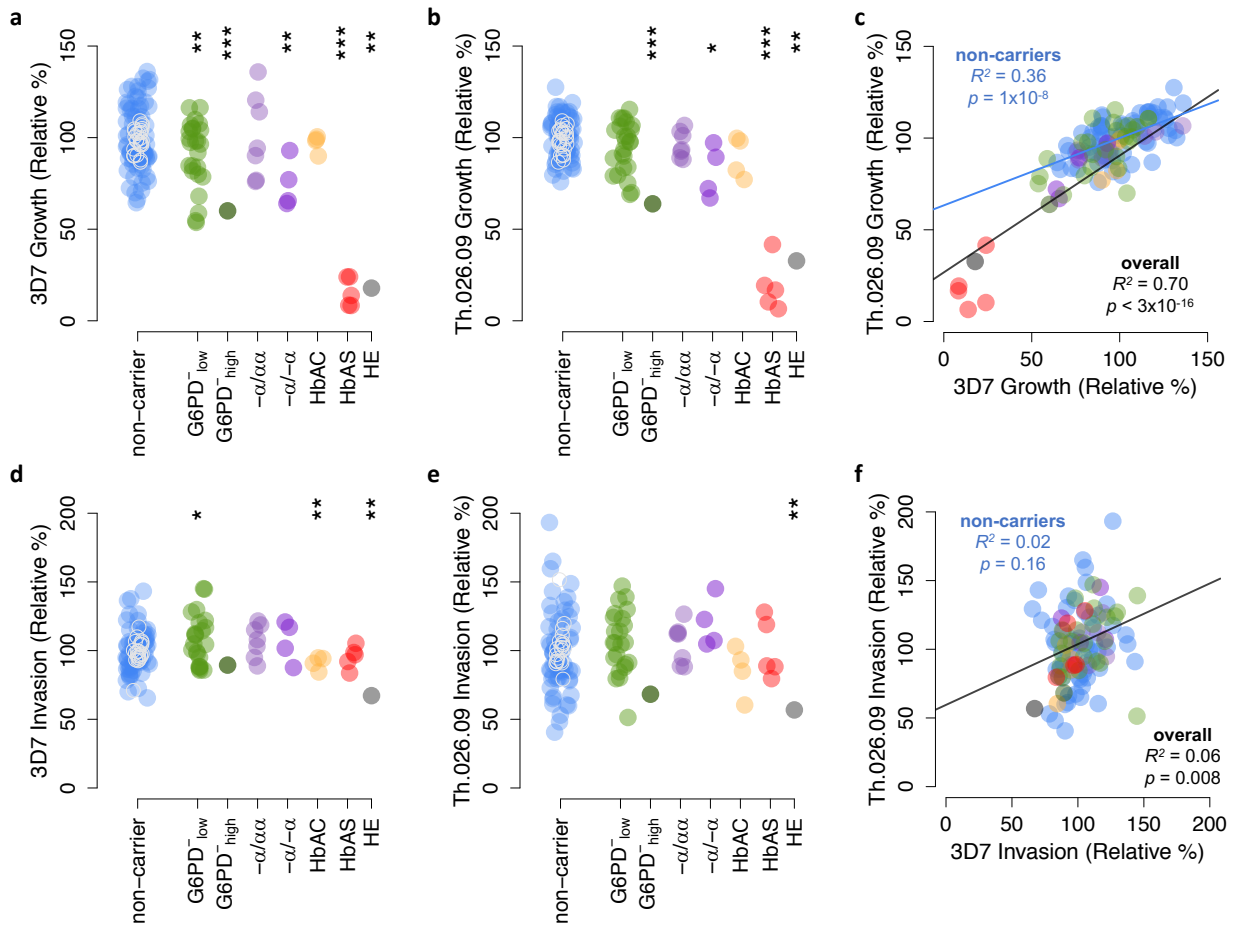


Figure 2. *P. falciparum* replication rate varies widely among donor RBCs. (a-b) Growth of *P. falciparum* strain 3D7 **(a)** or clinical isolate Th026.09 **(b)** over a full 48-hour cycle in donor RBCs (hours 24-72 of assay). Growth is presented relative to the average non-carrier rate after correction for batch effects, including comparison to a repeated RBC control (Methods). Repeated measurements of the same RBC control over the course of the experiment are shown as open gray circles. **(c)** Correlation in per-sample growth rates between the two *P. falciparum* strains. **(d-e)** Invasion efficiency of *P. falciparum* strain 3D7 **(d)** or clinical isolate Th026.09 **(e)** after a 24-hour incubation with donor RBCs, normalized in the same way as growth. Repeated measurements of the RBC control over time are shown as gray circles. **(f)** Correlation in per-sample invasion success for the two *P. falciparum* strains. Colors in **(c)** and **(f)** correspond to the classifications in other panels; R^2 and *p*-values are derived from OLS regression. *p<0.1; **p<0.05; ***p<0.01.

mild G6PD deficiency was associated with a 7% increase (p=0.072). Only HE had a significant effect on the invasion efficiency of both strains (3D7, 67% of the non-carrier average, p<0.014; Th.026.09, 57%, p<0.014). There was little correlation between the invasion efficiencies of the two parasite strains in individual donor samples (Figure 2F), potentially reflecting strain-specific differences in the pathways used for invasion (Wright and Rayner, 2014). However, as we also observed greater variability between repeated

185 samples for invasion than for growth (Figure S1), the invasion results may be influenced by
186 greater experimental noise.
187
188 **RBC phenotypes vary widely among non-carriers**
189 To assess phenotypic variation across donor RBCs, we measured 19 common
190 indices of RBC size and hemoglobin content from complete blood counts using an Advia
191 hematology analyzer (Figure 3A-D; Figure S2). Mean cellular volume (MCV) and
192 hemoglobin mass (MCH) are closely related traits, which can be represented together as
193 cellular hemoglobin concentration (CHCM) or the fraction of RBCs with 'normal'
194 hemoglobin and volume indices (M5). As expected, each known disease allele was
195 associated with a distinct set of RBC abnormalities. These included elevated CHCM for
196 HbAC ($p=0.033$), consistent with dehydration, and very low MCV ($p=8.7x10^{-5}$) and MCH
197 ($p=4.6x10^{-7}$) for α-thalassemia trait (αα/--), consistent with microcytic anemia (Galanello
198 and Cao, 2011). RBCs from the HE patient also had very low MCV and MCH (both $p<0.014$),
199 reflecting the membrane breakage and volume loss characteristic of this disease. However,
200 for all of these phenotypic measures, we also observed broad distributions in non-carriers
201 that overlapped the distributions of most carriers (Figure 3A-D; Figure S2). Notably, the
202 breadth of the non-carrier distribution for each phenotype was large (e.g., 24 fL range for
203 MCV) compared to the average difference between Africans and Caucasians (e.g., 3-5
204 fL(Beutler and West, 2005; Lo *et al.*, 2011)). This wide diversity and substantial overlap
205 between non-carrier and carrier traits suggest that 'healthy' RBCs exist on the same
206 phenotypic continuum as RBCs carrying known disease alleles.
207 We observed similar patterns of variation in RBC membrane fragility (Figure 3E-F)
208 and membrane deformability (Figure 3G-H), as measured with osmotic fragility tests and
209 osmotic gradient ektacytometry (see Methods). Both sets of curves represent RBC
210 tolerance to osmotic stress, which can result in swelling and lysis (fragility, Figure 3E-F) or
211 dehydration and decreased deformability (O_{hyper}, Figure 3G-H). Specific
212 hemoglobinopathies were associated with moderate to strong reductions in fragility (αα/α-
213 $p=0.039$; αα/-- $p=0.003$; HbAS $p=0.014$), deformability (DI_{max}; HbAC $p=0.007$; HbAS
214 $p=4.4x10^{-6}$), and/or resistance to loss of deformability when dehydrated (O_{hyper}; HbAC

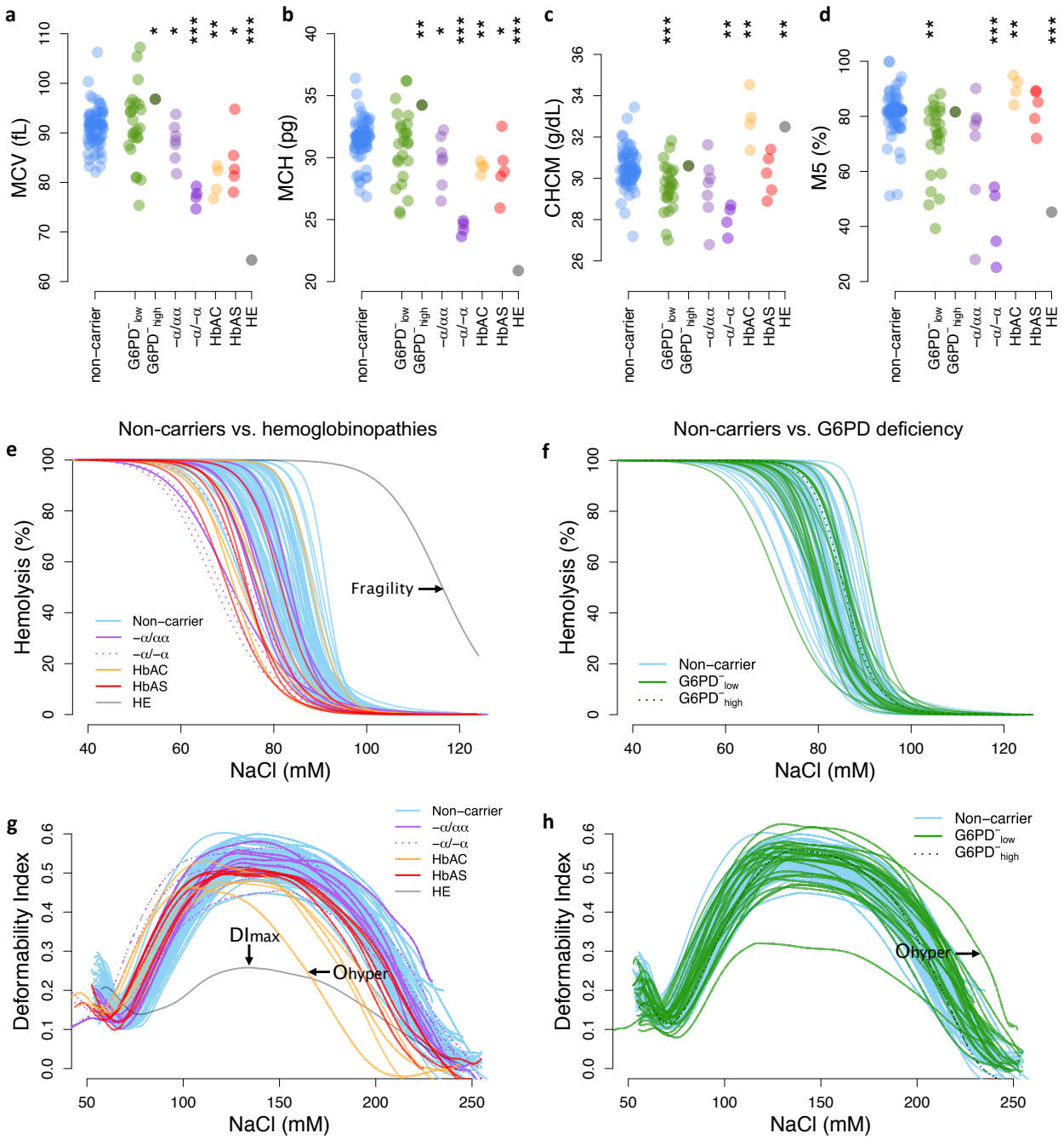


Figure 3. Red cell phenotypes that are abnormal in carriers also vary widely among non-carriers. **MCV:** mean RBC volume. **MCH:** mean cellular hemoglobin. **CHCM:** cellular hemoglobin concentration. **M5:** fraction of RBCs with 'normal' indices of hemoglobin concentration (28-41 g/dL) and volume (60-120 fL). **Fragility:** Tendency to lyse when osmotically overhydrated, defined as [NaCl] causing 50% lysis. **DI_{max}:** Maximum membrane deformability under physiological salt conditions. **O_{hyper}:** Tendency to resist osmotic dehydration and loss of deformability, defined as the hypertonic [NaCl] at which deformability is half maximum. *p<0.1; **p<0.05; ***p<0.01; tests and samples as in **Figure 2.**

225 p=0.019; HbAS p=0.041). HE cells were both extremely fragile (p<0.014) and extremely
226 non-deformable (p<0.014). In non-carriers, the distributions for all membrane measures
227 were wide, continuous, and overlapped the distribution of most carriers (Figure 3E-H;
228 Figure S3). Overall, these data demonstrate that multiple phenotypic alterations associated
229 with RBC disease alleles are also present in non-carrier RBCs.
230
231 **Non-carrier variation in RBC phenotypes predicts *P. falciparum* replication rate**
232 To identify sets of phenotypes associated with *P. falciparum* replication in non-
233 carrier RBCs, we used a machine learning method called Lasso (least absolute shrinkage
234 and selection operator) that performs regularization and variable selection (Tibshirani and
235 Tibshirani, 1994) (see Methods). Briefly, Lasso shrinks the regression coefficients for some
236 possible predictors (in this case, phenotypes) to zero, in order to obtain a subset of
237 predictors that minimizes prediction error. This method is well-suited for datasets in which
238 possible predictors are correlated, as are standard measurements of RBC size, hemoglobin,
239 and membrane dynamics. For each trait in each predictive subset selected by Lasso, we
240 then applied univariate OLS regression to more accurately estimate the sign of its effect on
241 all measured components of parasite fitness. The highest-confidence results are
242 summarized in Figure 4A, with complete details provided in Table S2.
243 *P. falciparum* fitness in non-carrier RBCs was strongly predicted by variation in
244 several RBC traits related to deformability, dehydration, volume, and hemoglobin content.
245 Among 24 tested phenotypes, the most important trait was the ektacytometry parameter
246 O_{hyper}, which represents a cell's tendency to retain deformability in the face of dehydration
247 (Figure 3G). In univariate models, non-carrier RBCs with the largest O_{hyper} values supported
248 1.2-1.5X faster parasite growth (3D7 1.47X, p=0.003; Th.026.09 1.24X, p=0.005) and 1.4-
249 2.0X more effective invasion (3D7 1.36X, p=0.018; Th.026.09 1.94X, p=0.006) than RBCs
250 with the smallest O_{hyper} values (Table S2; Figure S5A). Consistent with this result, *P.*
251 *falciparum* replication was relatively inhibited in RBCs that were more dehydrated at
252 baseline, as indicated by higher CHCM (3D7 invasion p=0.007; 3D7 growth p=0.042;
253 Th.026.09 invasion p=0.006; Th.026.09 growth p=0.36). Parasites also grew faster in RBCs
254 with larger mean volume, or MCV (3D7 p=0.001; Th.026.09, p=0.001); a greater mass of
255 hemoglobin per cell, or MCH (3D7 p=0.073; Th.026.09 p=0.019); and more deformable

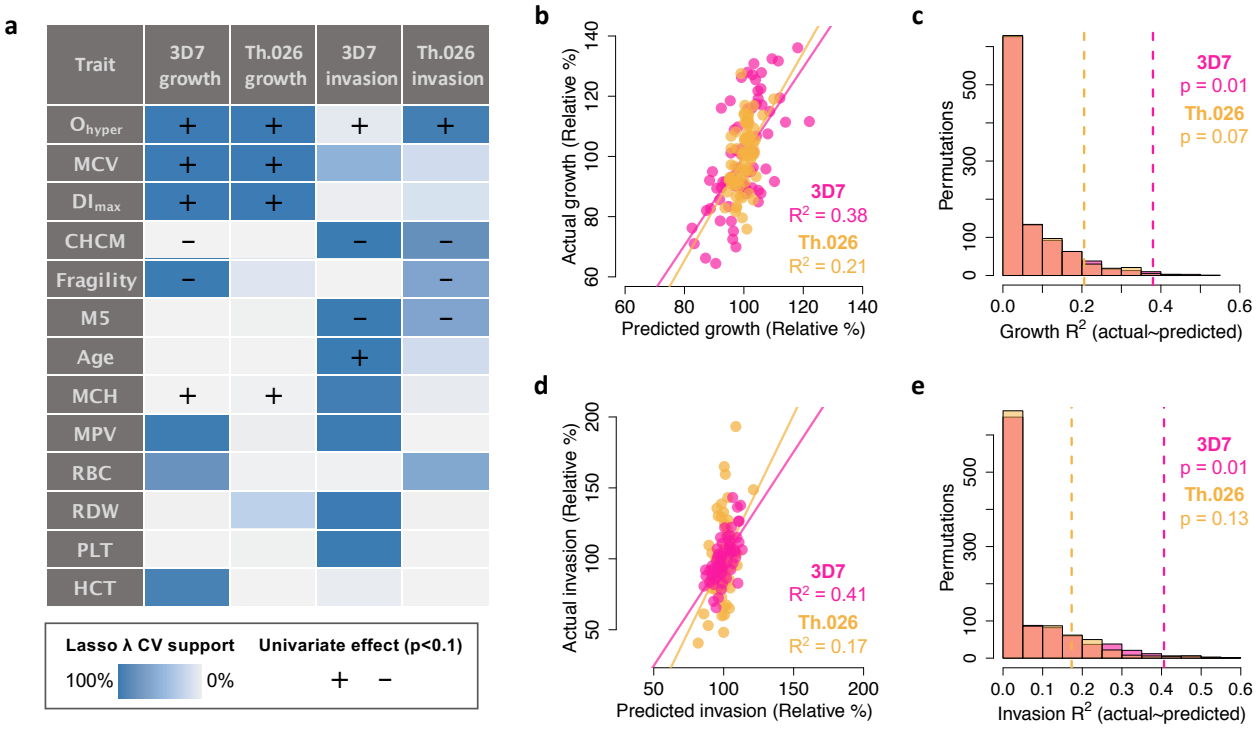
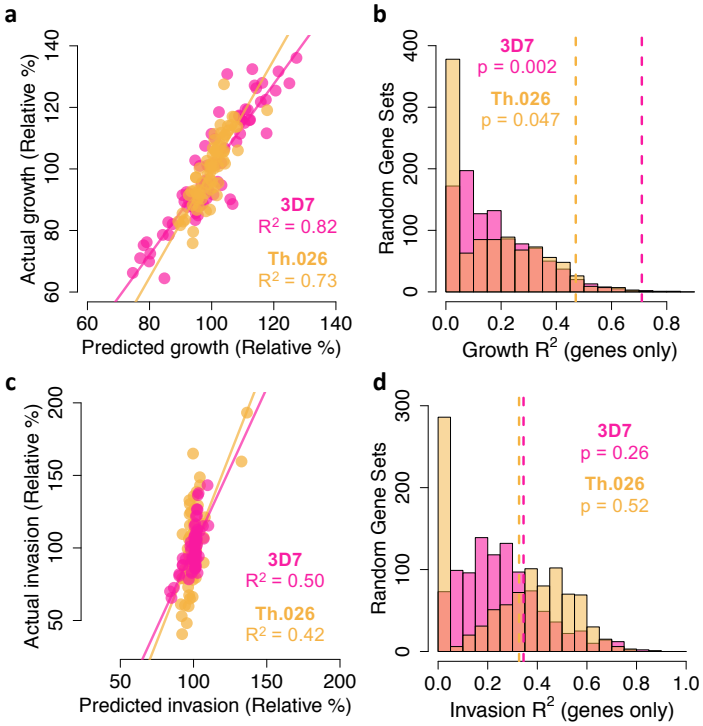


Figure 4. RBC phenotypes predict *P. falciparum* fitness in non-carriers. (a) Phenotypes shown were selected by Lasso in at least one of four models (columns) with at least 60% cross-validation (CV) support, as indicated by blue shading. **(+/-)** shows the direction of effect if the phenotype was significantly correlated (p<0.1) with the parasite fitness component in a separate, univariate linear model. **(b)** Lasso model predictions of parasite growth as a function of RBC phenotypes are plotted against measured growth values. Each point represents one non-carrier. The solid line shows a perfect 1:1 fit; R^2 indicates the proportion of variance explained by Lasso predictions. **(c)** For 1000 permutations of the growth data, the histograms show the proportion of variance explained by the Lasso modeling procedure. Dashed lines indicate R^2 for the actual data. p-values indicate the fraction of permutations that explain at least as much variance as the real data. **(d,e)** As in **(b,c)**, but for *P. falciparum* invasion. **O_{hyper}:** Tendency to resist osmotic dehydration and loss of deformability, defined as the hypertonic [NaCl] at which deformability is half maximum (mM). **MCV:** mean RBC volume (fL). **DI_{max}:** Maximum membrane deformability under physiological salt conditions. **CHCM:** cellular hemoglobin concentration (g/dL). **Fragility:** Tendency to lyse when osmotically overhydrated, defined as [NaCl] causing 50% lysis (mM). **M5:** fraction of RBCs with 'normal' indices of hemoglobin concentration (28-41 g/dL) and volume (60-120 fL) (%). **MCH:** mean corpuscular hemoglobin (pg/RBC). **MPV:** mean platelet volume (fL). **RBC**: red cell number (x10^6/uL). **PLT:** platelet number (x10^3/uL).). **HCT:** hematocrit (%).

membranes, measured by DI_{max} (3D7 p=0.0006; Th.026.09 p=0.006). Some measures of parasite fitness were also reduced in RBCs with more fragile membranes (3D7 invasion p=0.65; 3D7 growth p=0.003; Th.026.09 invasion p=0.067; Th.026.09 growth p=0.14). Additional phenotypes related to platelets and RBC density were selected for some models, but the direction of their effects was unclear since they were not significant when considered individually. These results indicate that common, non-pathogenic variation in

283 RBC size, membrane dynamics, and other correlated traits have meaningful effects on *P.*
284 *falciparum* replication rate in RBCs.
285 Together, 13 phenotype measurements explained approximately 40% of the
286 variation in 3D7 growth and invasion in non-carriers (Figure 4B,D). To test for overfitting,
287 we applied the same model-fitting procedure many times to randomly shuffled parasite
288 data, which usually failed to produce predictive models (i.e., R^2=0, Figure 4C,E). These
289 permutation tests found that real data explained, on average, 6.5X the variance in 3D7
290 growth and 6.3X the variance in 3D7 invasion than expected by chance (both p=0.01,
291 Figure 4C,E). For Th.026.09, the phenotypes selected by Lasso explained 17-21% of the
292 total variation in parasite fitness (Figure 4B), which was marginally better than expected
293 by chance for growth (3.7X, p=0.07, Figure 4C) and marginally worse than expected for
294 invasion (3.0X, p=0.13, Figure 4E). The weaker result for Th.026.09 may be explained by its
295 smaller dynamic growth range (Figure 2A-B), perhaps because clinical isolates are
296 relatively poorly adapted to laboratory conditions. Overall, these results demonstrate that
297 multiple, variable phenotypes impact *P. falciparum* susceptibility in healthy RBCs. Non-
298 carrier cells that are less hospitable to parasites share specific traits with RBCs that carry
299 disease alleles, including smaller size, decreased deformability, and an increased tendency
300 to lose deformability when dehydrating.
301
302 **A small set of RBC genotypes predicts *P. falciparum* replication in non-carriers**
303 Next, we tested whether non-carrier genotypes derived from exome sequencing could
304 improve our predictions of *P. falciparum* replication rate. To limit false positives and
305 increase statistical power (Ioannidis, 2005; Flynn, Hurvich and Simonoff, 2017), we
306 focused on 23 candidate RBC proteins that have been previously associated with *P.*
307 *falciparum* resistance in the literature (Table S3). These proteins contained 106 unlinked
308 genetic variants (pairwise $r^2 < 0.1$) among the 73 non-carriers. Along with these
309 polymorphisms and phenotypes described above, we also treated the top 10 principal
310 components (PCs) of 1000 Genomes population structure as possible predictors of *P.*
311 *falciparum* fitness.
312 The addition of genetic factors substantially improved all four models of parasite
313 fitness. Across strains, 73-82% of the parasite variation in non-carriers was explained by a



e

	3D7 growth	Th.026 growth	3D7 invasion	Th.026 invasion
Gene	**associated variants**			
Flexible membrane backbone				
SPTA1	3	3	0	0
SPTB	4	1	0	0
ANK1	0	2	1	0
SLC4A1	1	2	0	0
EPB42	1	0	0	0
Ion channels (hydration)				
PIEZO1	5	2	1	1
ATP2B4	1	0	0	0
Plasma membrane proteins				
CD44	2	2	0	0
CR1	1	1	1	1
BSG	1	2	0	0
ABCB6	1	0	0	0
GYPB	1	0	0	0
Other				
FPN	0	1	0	0
G6PD	0	1	0	0
	associated 1000G PCs			
of top 10	1	1	1	1

Figure 5. RBC genotypes substantially improve predictions of *P. falciparum* fitness in non-carriers. (a) Lasso model predictions of parasite growth as a function of RBC phenotypes, PCs of population structure, and variants in 23 malaria-related proteins (Table S3) are plotted against measured growth values. Each point represents one non-carrier. The solid line shows a perfect 1:1 fit; R^2 indicates the proportion of variance explained by Lasso-selected predictors. **(b)** For 1000 random sets of 23 genes drawn from the RBC proteome, the histograms show the variance explained *by variants within those genes* after applying the same Lasso modeling procedure as (a). Dashed lines indicate the variance explained by variants within the 23 malaria-related genes. *p*-values indicate the fraction of random gene sets that explain at least as much variance as the real data. **(c,d)** As in (a,b), but for *P. falciparum* invasion**. (e)** Variants and PCs counted in each cell were selected in the designated Lasso model with at least 70% cross-validation (CV) support. Several tested genes had 0 associated variants: *CD55*, *EPB41*, *HP*, *HBA1/2, GYPA, GYPC, GYPB,* and *HBB.* Details on lower-confidence variants are available in Table S4.

total of 47 genetic and phenotypic variables selected by Lasso (Figure 5A; Table S4). For 3D7 and Th.026.09 respectively, polymorphisms in 14 of 23 malaria-related genes explained 71% and 41% of the growth variation in non-carriers (Figure 5B). These models were significant by permutation (both p=0.011; Figure S6AB), indicating that the results were unlikely to be due to overfitting. Nonetheless, to confirm that the genetic signal was specific to our set of 23 malaria-related genes, we used polymorphisms from 1000 sets of random genes from the RBC proteome (Table S6) to build 1000 alternative models (see Methods). On average, variants from random RBC genes explained only 17.9% of the

338 observed variation in *P. falciparum* growth for 3D7 (p=0.002) and 16.4% for Th.026.09
339 (p=0.047) (Figure 5B). In contrast to growth, the genotype/phenotype models of *P.*
340 *falciparum* invasion (Figure 5C) were only marginally significant by permutation (3D7
341 p=0.09, Th.026.09 p=0.11, Figure S6C-D) and the genetic signal was not specific to the set of
342 23 malaria-related genes (3D7 p=0.26, Th.026.09 p=0.52, Figure 5D). Nonetheless, 4 of the
343 8 variants selected by Lasso for 3D7 invasion were also selected for 3D7 growth. Together,
344 these results indicate that *P. falciparum* growth rate in RBCs is strongly determined by host
345 genetic variation concentrated in a small number of genes.
346 Of the top 31 variants associated by Lasso with *P. falciparum* growth, seventeen
347 (55%) were synonymous (Table S4). This fraction did not significantly differ from the input
348 set of 106 variants (p=0.64, 2-tailed z-score). Although 15/17 of these synonymous
349 variants have no known, direct function, 7/15 are in linkage disequilibrium ($r^2 > 0.1$) with
350 other variants that have been associated with RBC phenotypes by GWAS (Table S4).
351 Similarly, 6 of the top 14 non-synonymous polymorphisms associated with *P. falciparum*
352 growth have been previously linked to RBC or other traits (Table S4). This overlap with
353 previous studies suggests that the variants associated with *P. falciparum* growth in this
354 study are likely to tag functional genetic loci.
355 Nearly all of the top 31 growth-associated polymorphisms occurred in either (1) ion
356 channel proteins, which regulate RBC hydration; (2) components of the flexible RBC
357 membrane backbone; or (3) red cell plasma membrane proteins, including known invasion
358 receptors (Figure 5E). In the first category, the highly polymorphic ion channel *PIEZO1*
359 contained six polymorphisms associated with small (-3%) to moderate (-32%) reductions
360 of *P. falciparum* growth rate (Table S4). The microsatellite variant *PIEZO1*-E756del, which
361 has been a focus of several recent studies (Ilboudo *et al.*, 2018; Ma *et al.*, 2018; Rooks *et al.*,
362 2019; Nguetse *et al.*, 2020), predicted a moderate reduction in Th.026.09 growth (-8.2%,
363 p=0.011) but was not related to RBC dehydration in these data (Table S5). We also detected
364 one growth-associated variant in *ATP2B4,* which encodes the primary RBC calcium channel
365 PMCA4b (-13% 3D7, p=0.006). This variant tags an *ATP2B4* haplotype implicated by GWAS
366 in protection from severe malaria (Timmann *et al.*, 2012; Lessard *et al.*, 2017; Zámbó *et al.*,
367 2017) and many RBC phenotypes (van der Harst *et al.*, 2012; Li *et al.*, 2013; Lessard *et al.*,
368 2017; Lin, Brown and Machiela, 2020) including RDW, which was replicated in our data set

369 (Table S4). *SPTA1* and *SPTB,* which encode the flexible spectrin backbone of RBCs, also
370 contained several variants associated with the growth of at least one *P. falciparum* strain,
371 as did the structural linker proteins ANK1, SLC4A1, and EPB42 (Figure 5E; Table S4).
372 Finally, we identified one polymorphism in ABCB6, two in CD44, and two in basigin (BSG)
373 that were strongly associated with *P. falciparum* growth (Figure 5E; Table S4). These
374 plasma membrane proteins have previously been implicated in *P. falciparum* invasion by *in*
375 *vitro* genetic deficiency studies (Crosnier *et al.*, 2011; Egan *et al.*, 2015, 2018). Notably, two
376 of the natural polymorphisms identified here are synonymous quantitative trait loci (QTL)
377 for CD44 splicing (rs35356320) and BSG expression (rs4682) (GTEx Consortium, 2017).
378 We also detected one variant in the RBC iron exporter *FPN* and one in *G6PD* with small
379 effects on Th.026.09 growth (2.4%, p=0.015 and 3.4%, p=0.010 respectively; Table S4). No
380 associated variants were detected in the other nine tested genes (Table S2), including three
381 hemoglobin proteins, three glycophorins, *CD55*, *EPB41*, and *HP*. Together, these data
382 demonstrate that *P. falciparum* growth rate in non-carrier RBCs is strongly determined by
383 dozens of host genetic variants, which together shape the phenotypic distribution of red
384 cell susceptibility.
385 In addition to these specific genetic variants, PCs of population structure explained
386 22-26% of the variation in non-carrier invasion and 10-12% of the variation in growth.
387 Interestingly, neither PC1 nor PC2—which distinguish Africans from other populations
388 (Figure 1A)—were related to parasite fitness. Instead, this signal of correlated variation
389 was partially driven by a six-member family with unique ancestry in our sample (Figure
390 1A, white borders) and extreme parasite fitness values (Figure S6). Although the present
391 study is limited by sample size, the association between global genetic PCs and *P.*
392 *falciparum* fitness suggests that additional functional variants remain to be discovered in
393 many populations.
394
395 **African ancestry does not predict *P. falciparum* resistance in red cells**
396 Based on evidence from balanced disease alleles like HbAS, it has been suggested
397 that anti-malarial selection has shaped polygenic red cell phenotypes in African
398 populations as a whole (Goheen *et al.*, 2016; Kanias *et al.*, 2017; Ma *et al.*, 2018). We tested
399 this hypothesis by examining the correlation between African ancestry and *P. falciparum*

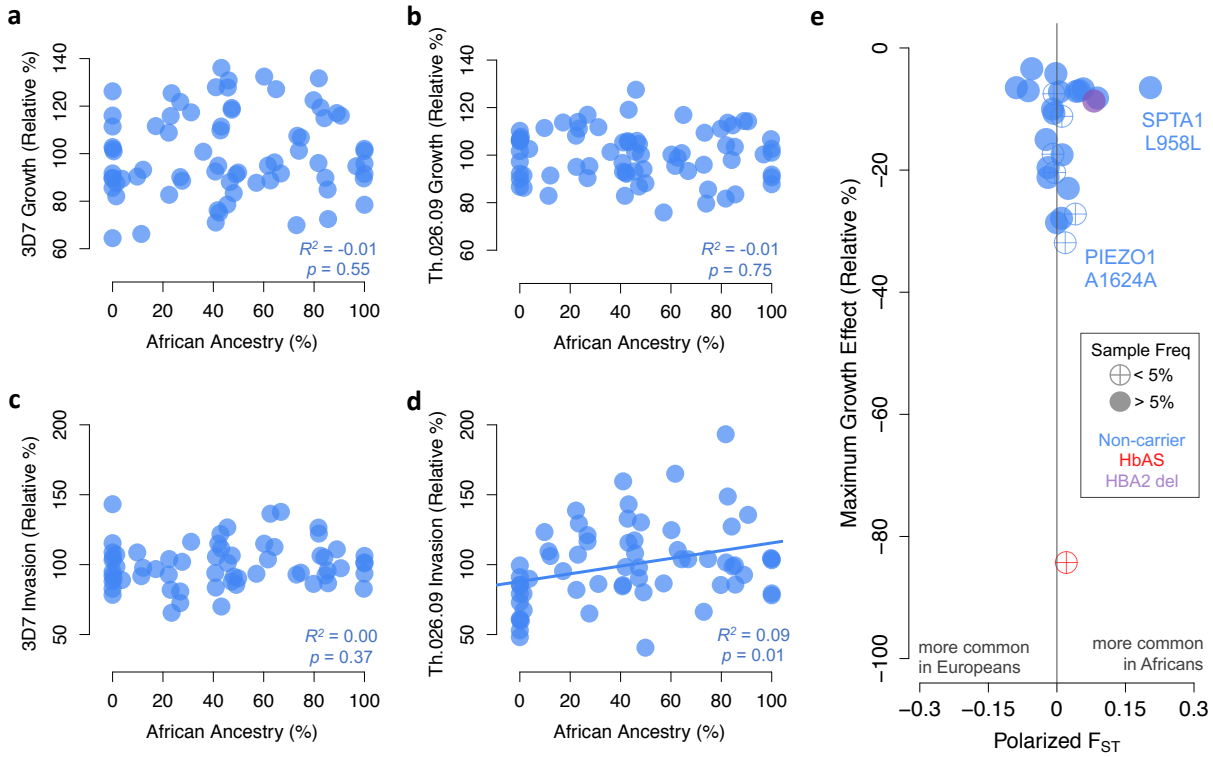


Figure 6. No evidence of widespread selection in Africa for slower *P. falciparum* replication or protective alleles in non-carriers. (a-d) Parasite growth and invasion measurements plotted against the exome-wide fraction of African ancestry, determined by comparison to 1000 Genomes reference populations. Adjusted-R^2 and *p*-values are shown for OLS regression. **(e)** Blue points represent non-carrier alleles selected by Lasso that also have significant effects on parasite growth ($p<0.1$) in univariate linear models (Table S4). The maximum estimated effect across two *P. falciparum* strains is shown. Red and purple points represent HbAS and the HBA2 deletion in an additive model, respectively. F_{ST} was calculated between pan-African and pan-European samples in gnomAD (Methods). Other disease alleles are not shown because they do not function as single loci (G6PD$^-$, HE) or do not affect growth (HbAC).

replication rate in RBCs from non-carriers (Figure 6A-D). We found no evidence that these traits were related, apart from an weakly positive relationship between African ancestry and the invasion rate of Th.026.09, a clinical Senegalese strain (p=0.01, R^2=0.09, Figure 6D). To understand this result, we next examined how key RBC phenotypes identified in this study (Figure 4A) vary with African ancestry. We found that greater African ancestry predicts reduced osmotic fragility (p= 9.03 x 10^{-6}), increased hypotonic deformability (DI_{min}, p=0.001), lower cellular hemoglobin concentration (CHCM, reflecting reduced dehydration, p=0.007), and a reduced fraction of RBCs with normal volume and hemoglobin content (M5, p=0.033). All of these traits actually predict *greater* red cell

susceptibility to *P. falciparum* (Figure 4A), although together they explain less than 9% of the non-carrier variation in parasite fitness. The remaining key phenotypes (Figure 4A) do not vary with African ancestry, which may explain why African ancestry itself is mostly unassociated with *P. falciparum* fitness in non-carrier RBCs (Figure 6A-D).

Next, we used allele frequency data from the global gnomAD collection (Karczewski *et al.*, 2020) to test whether polymorphisms associated with *P. falciparum* fitness in RBCs (Table S4) vary between African and European populations. Geographical differences in malaria selection are 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 (Kwiatkowski, 2005). Here, aside from well-known disease variants, we find no evidence that African populations as a whole are enriched for RBC polymorphisms that impair *P. falciparum* growth in at least one strain (Figure 6E). Of 25 alleles with detectable protective effects, 52% are in fact more common in Europeans than Africans (p=0.95, binomial test). Besides HbAS, the polymorphism with the largest effect is a synonymous variant in PIEZO1 (-31.9%, p=0.001), which is found at <4% frequency in both Africans and Europeans. A few other protective polymorphisms, including two in *PIEZO1* and two in *SPTB*, are nearly fixed in both populations. We do observe one F_{ST} outlier variant, a synonymous allele at SPTA1 L958, which is substantially more common in Africa (70% frequency) than in Europe (25%) and has a mild protective effect against *P. falciparum* growth (-6.5%, p=0.027, Figure 6E). Given its high frequency, as well as its direct association with a number of RBC phenotypes in previous GWAS (Table S4), this synonymous SPTA1 allele could be a promising candidate for investigating anti-malarial selection in Africa. However, the total set of natural RBC variants that limit *P. falciparum* growth *in vitro* do not appear to support the hypothesis of widespread divergent selection related to malaria between Africans and Europeans.

DISCUSSION

Healthy red blood cells (RBCs) harbor extensive phenotypic and genetic variation, both within and between human populations. In this work, we demonstrate that this variation modulates a wide range of RBC susceptibility to *P. falciparum* parasites. Our

452 findings add to a growing understanding of the genetic and phenotypic basis of RBC
453 resistance to *P. falciparum*, especially for RBCs that lack population-specific disease alleles.
454 In addition to suggesting new targets for future malaria interventions, these findings
455 challenge assumptions about the role of malaria selection in shaping human RBC diversity.
456 Since exponential replication of *P. falciparum* is a significant driver of malaria
457 disease progression (Bejon *et al.*, 2007), the ample variation that we observed in this trait
458 *in vitro* could be relevant for clinical outcomes in endemic regions. Growth inhibition from
459 HbAS, for example, reduces the risk of death from malaria by reducing parasite density in
460 the blood (Allison, 1954; Luzzatto, 2012). While HbAS has an extreme effect size, the 3-fold
461 range of parasite replication rates we found among non-carrier RBCs shares substantial
462 overlap with RBCs carrying other protective variants. Although the physiologically complex
463 basis of severe malaria (Okwa, 2012) makes it difficult to estimate the precise contribution
464 of RBC factors to severe malaria risk, the genotypes and phenotypes we have associated
465 with *P. falciparum* fitness may be promising targets for future therapeutic interventions.
466 We have shown here that widespread, 'normal' variation in RBC traits like volume,
467 hydration status, and maximum deformability is associated with *P. falciparum* fitness in
468 non-carrier RBCs. Reassuringly, these phenotypes are present to a stronger degree in RBCs
469 that do carry disease alleles (Clark, Mohandas and Shohet, 1983; Mockenhaupt *et al.*, 2000;
470 Pengon *et al.*, 2018). They are also consistent with published reports of extreme RBC traits
471 that limit parasite replication. For example, (Tiffert *et al.*, 2005) used an experimental
472 gradient of ion concentrations to show that *P. falciparum* growth is strongly reduced in
473 chemically dehydrated RBCs. Here, we found that naturally occurring, mildly dehydrated
474 RBCs also supported less efficient parasite growth and invasion. Similarly, the rare Dantu
475 variant of the glycophorin A/B receptors has been associated with increased RBC
476 membrane tension, decreased *P. falciparum* invasion efficiency, and decreased risk of
477 severe malaria (Field *et al.*, 1994; Leffler *et al.*, 2017; Kariuki *et al.*, 2018). Our data on RBC
478 membrane deformability and fragility support the same link between more flexible RBC
479 membranes and greater *P. falciparum* success. Moreover, they expand upon previous
480 studies by demonstrating for the first time that common, healthy phenotypic variation in
481 RBCs contributes meaningfully to *P. falciparum* fitness.

482 In our linear models of *P. falciparum* growth, phenotypic variation was strongly
483 outperformed by genetic variation in a small number of RBC proteins. This result implies
484 the existence of additional RBC phenotypes that we did not measure (or did not measure
485 with sufficient accuracy), but which have a genetic basis in a small number of genes
486 important to *P. falciparum*. Approximately half of the polymorphisms we identified are
487 non-synonymous, and may therefore exert direct effects on phenotypes like RBC
488 membrane structure or ion transport. The abundances of RBC gene transcripts and
489 proteins are other clear candidates for future phenotype studies, given that the other half
490 of associated polymorphisms were synonymous. A recent metanalysis showed that 'silent'
491 and 'coding' SNPs are equally likely to be associated with human disease, and moreover
492 have similar effect sizes (Chen *et al.*, 2010). Although synonymous SNPs may sometimes be
493 linked to coding variants, some have been shown to directly impact mRNA expression,
494 protein translation speed, and protein folding by altering transcription levels, codon usage,
495 and mRNA stability (Sauna and Kimchi-Sarfaty, 2011). Synonymous SNPs that impact
496 splicing, like rs35356320 in CD44, may also impact protein structure. Other conceivable
497 RBC phenotypes, such as the dynamics of membrane modification during *P. falciparum*
498 development, may only become evident in more detailed time course experiments. The
499 true number of RBC phenotypes important to *P. falciparum* may be effectively infinite
500 (Kinsler, Geiler-Samerotte and Petrov, 2020), making it very useful in practice that static
501 genetic variation is so strongly predictive of parasite growth.
502 Overall, just 31 polymorphisms in 14 RBC genes explained the majority of the
503 variation in *P. falciparum* growth among non-carriers. One reason that our explanatory
504 power is so high is our focus on *P. falciparum* replication in RBCs, which is a relatively
505 simple component of a larger, complex disease. This focus allowed us to use controlled, *in*
506 *vitro* experiments to detect genetic variants with small to moderate effects. Another
507 important explanation is our use of Lasso variable selection on just 106 polymorphisms in
508 genes with strong existing links to malaria. This approach obviated the need to meet an
509 exome-wide significance threshold, while still allowing for the discovery of novel,
510 putatively functional alleles in well-known disease genes. Furthermore, we confirmed that
511 this set of genes was enriched for signal compared to random sets of genes drawn from the
512 exome-wide background. It is important to note that genetic linkage complicates the

513 identification of the exact functional polymorphisms in any population sample (Sohail *et al.*,
514 2019), and we cannot rule out that some variants identified here are merely linked to the
515 true functional variants. Indeed, about half of the variants we identified occur in linkage
516 blocks containing other SNPs associated with RBC traits in GWAS studies. In any case, we
517 present significant evidence that loci containing 14 RBC genes are strongly enriched for
518 polymorphisms with significant impacts on *P. falciparum* growth.

519 One of the unique aspects of our study is the participation of individuals with a
520 range of African ancestry, defined by similarity to donors from five 1000 Genomes
521 reference populations. We found that African ancestry was associated with RBC
522 phenotypes that *improved* parasite fitness, particularly for the clinical Senegalese strain
523 Th.026.09. In the future, it would be very interesting to test for local parasite adaptation to
524 human RBCs using *P. falciparum* strains and RBC samples from around the globe. We also
525 found that the polymorphisms associated with *P. falciparum* growth by Lasso are not
526 enriched in African populations included in the gnomAD database of human variation. This
527 is not consistent with African-specific selection on these smaller-effect variants, despite the
528 famous examples provided by RBC disease alleles. This unexpected result for 'healthy'
529 variation has at least three possible explanations. First, relatively small-effect alleles may
530 not have had sufficient time to increase in frequency since *P. falciparum* began expanding
531 in humans some 5,000-10,000 years ago (Sundararaman *et al.*, 2016; Otto *et al.*, 2018).
532 Second, the complexity of severe malaria could mean that the variants discovered here do
533 not ultimately impact disease outcome, especially relative to known disease variants.
534 Third, human adaptation may be too local to detect with coarse-grain sampling of Sub-
535 Saharan African genetic diversity (e.g (Pankratov *et al.*, 2020)). Our data do suggest,
536 however, that few RBC alleles remain to be discovered that are both widespread in Africa
537 and have large effects on *P. falciparum* proliferation in RBCs (e.g. (Ma *et al.*, 2018)).

538 More broadly, these data show that it is inaccurate to make assumptions about RBC
539 susceptibility to *P. falciparum* based on a person's race or continental ancestry. These kinds
540 of hypotheses (Williams, 2006; Goheen *et al.*, 2016; Kanias *et al.*, 2017; Ma *et al.*, 2018) are
541 based on well-known examples of balanced disease alleles, but our data suggest that these
542 examples are exceptions rather than the rule. It is important to recognize that, at least in
543 biology classes, the use of racially-based genetic examples increases the belief that genes

544 encode absolute, functional differences between races (Donovan, 2017; Sparks, Baldwin
545 and Darner, 2020). This perspective neglects the fact that >90% of genetic differences
546 among humans occur within populations, rather than across them (Rosenberg, 2011). In
547 this study, global population structure captured by PCs showed that non-African variation
548 has its own distinct impacts on *P. falciparum* replication in RBCs. These data are an
549 important reminder that most human genetic diversity is a result of demography, not
550 population-specific selection (Coop *et al.*, 2009; Hofer *et al.*, 2009).
551 In conclusion, this study demonstrates that substantial phenotypic and genetic
552 diversity in healthy human RBCs impacts the replication of malaria parasites. Whether or
553 not this diversity is shaped by malaria selection, a better understanding of how *P.*
554 *falciparum* biology is impacted by natural RBC variation can help lead to new therapies for
555 one of humanity's most important infectious diseases.
556
557 **ACKNOWLEDGMENTS**
558 We gratefully acknowledge the invaluable participation of all volunteer blood donors. Nick
559 Bondy, Sandra Larkin, Brian Fleischer, Ashley Dunn, Talal Seddik, Trung Pham, David Vu,
560 and Spectrum Child Health provided crucial assistance in donor coordination and sample
561 processing. *P. falciparum* strain Th.026.09 was kindly provided by Daouda Ndiaye and
562 Sarah Volkman. For quantitative advice, we thank Grant Kinsler and the Stanford Statistics
563 Consulting Group. This work was primarily supported by a Pilot Early Career award from
564 the Stanford Maternal Child Health Research Institute and a Gabilan Faculty Award from
565 the Stanford University School of Medicine Office of Faculty Development and Diversity.
566 E.R.E. received additional support from the Stanford Center for Computational,
567 Evolutionary, and Human Genomics. D.A.P. was funded through an NIH MIRA award
568 5R35GM118165-05. E.S.E. was funded by the NIH (DP2HL13718601), and is a Tashia and
569 John Morgridge Endowed Faculty Scholar in Pediatric Translational Medicine through the
570 Stanford Maternal Child Health Research Institute. Local blood samples were drawn at the
571 Stanford Clinical and Translational Research Unit, which is supported by CTSA Grant UL1
572 TR001085.
573
574

575 **AUTHOR CONTRIBUTIONS**

576 Conceptualization, E.R.E., D.A.P., and E.S.E; Methodology, E.R.E., D.A.P., F.A.K., and E.S.E.;
577 Investigation, E.R.E., F.A.K., C.L., and E.S.E., Formal Analysis, E.R.E.; Visualization, E.R.E.,
578 Writing- Original Draft, E.R.E.; Writing- Review & Editing, E.R.E., F.A.K., D.A.P., and E.S.E.;
579 Funding Acquisition, E.R.E. and E.S.E.; Resources, F.A.K., D.A.P., and E.S.E; Supervision,
580 D.A.P. and E.S.E.

581

582 **DECLARATION OF INTERESTS**

583 The authors declare no competing interests.

584

585 **METHODS**

586 **Sample collection and preparation**

587 Subjects with no known history of RBC disorders were recruited to donate blood at
588 the Stanford Clinical and Translational Research Unit. Written informed consent was
589 obtained from each subject and/or their parent as part of a protocol approved by the
590 Stanford University Institutional Review Board (#40479). To help control for weekly batch
591 effects, Subject 1111 donated fresh blood for parasite assays each week. 11 other subjects
592 donated blood on at least two different weeks. Whole blood samples from a hereditary
593 elliptocytosis patient were obtained from Dr. Bertil Glader under a separate approved
594 protocol (#14004) that permitted sample sharing among researchers. All samples were de-
595 identified upon collection by labeling with a random four-digit number. Two samples were
596 eventually removed from analysis based on a failed sequencing library (6449KD) and
597 history of stem cell transplant (8715).

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

603

604 **Exome sequencing and genotype calling**

605 Genomic DNA was isolated from frozen buffy coats using a DNeasy Blood and Tissue

606 Kit (Qiagen). Libraries were prepared using a KAPA Hyperplus kit (Roche) and hybridized
607 to human exome probes using the SeqCap EZ Prime Exome kit (Roche). The resulting
608 exome libraries were sequenced with paired-end 150 bp Illumina reads on the HiSeq or
609 NextSeq platforms at Admera Health (South Plainfield, NJ).
610 Reads were aligned to the hg38 human reference genome using bwa mem(Li, 2013),
611 yielding an average coverage of 42X across targeted exome regions (excluding sample
612 6449KD). Variants were called using GATK best practices(Van der Auwera *et al.*, 2013) and
613 hard filtered with the following parameters: QD < 2.0, FS > 60.0, ReadPosRankSum < -2.5,
614 SOR > 2.5, MQ < 55.0, MQRankSum < -1.0, DP < 500. To minimize the effects of sequencing
615 errors, variants not present in 1000 Genomes, dbSNP_138, or the Mills indel
616 collection(Mills *et al.*, 2006) were discarded. Variants were also discarded if their
617 frequencies in our sample did not fall between gnomAD African or European
618 frequencies(Karczewski *et al.*, 2020) and significantly differed from both groups'
619 frequencies by Fisher's Exact Test ($p \leq 0.05$).
620 *PIEZO1* E756del was genotyped via PCR and Sanger sequencing according to a
621 previously published protocol(Nguetse *et al.*, 2020). To call deletion variants that cause α-
622 thalassemia in the paralogous genes HBA2 and HBA1, we extracted reads from each .bam
623 file that lacked any mismatches or soft-clipping and had MAPQ ≥ 13 (i.e., <5% chance of
624 mapping error). Coverage with these well-mapped reads was calculated over the 73 and 81
625 bp of unique sequence in HBA2 and HBA1, respectively, and normalized to each sample's
626 exome-wide coverage. To determine which samples has unusually low coverage, we
627 formed an *ad hoc* reference panel of seven donors who were unlikely to carry deletion
628 alleles based on their normal MCH, MCV, and HGB and >96% exome-wide European
629 ancestry(Weatherall, 2001). We called heterozygous HBA2 deletions when normalized
630 coverage across three unique regions of the HBA2 gene was below the minimum reference
631 value. Similarly, we called homozygous HBA2 deletions when normalized coverage across
632 three unique regions of the HBA2 gene was less than half of the minimum reference value.
633 This approach resulted in an estimated HBA2 copy number of 2.0 in the reference panel,
634 0.95 in eight putative heterozygotes and 0.12 in four putative homozygotes. The same
635 method produced no evidence of HBA1 deletion in any sample.
636

637 **Variant classification and linkage pruning**

638 Exonic variants in RefSeq genes were identified using ANNOVAR(Wang, Li and
639 Hakonarson, 2010). Variants were classified into three categories: those within 23 malaria-
640 related genes (Table S2); those within 887 other red blood cell proteins (Table S5) derived
641 with a medium-confidence filter from the Red Blood Cell Collection database
642 (rbcc.hegelab.org); and those within any other gene. Singleton variants were removed to
643 avoid excessive false positives.

644 Linkage between all pairs of bi-allelic, exonic variants in our 121 genotyped samples
645 was calculated using the --geno-r2 and --interchrom-geno-r2 functions in vcftools (Danecek
646 *et al.*, 2011). Variants in RBC genes that shared $r^2 > 0.1$ with any variant in the 23-gene set
647 were removed. Within the 23-gene set and RBC-gene set separately, non-carrier variants
648 were ranked by the *p*-values of their OLS regression with all four parasite measures. Then,
649 one variant was removed from each pair with $r^2 > 0.1$, prioritizing retention in the
650 following order: greater significance across models; non-synonymous protein change;
651 higher frequency in our sample; and finally by random sampling. We report results from
652 additive genetic models (genotypes coded 0/1/2), which performed as well or better than
653 recessive (0/0/2) and dominant (0/2/2) models.

654

655 **Population analysis**

656 The population ancestry of our donors was assessed by comparison with African,
657 European, East Asian, and South Asian reference populations from the 1000 Genomes
658 Project(The 1000 Genomes Project Consortium, 2015). Briefly, variants called from an
659 hg38 alignment of the 1000 Genomes data (Lowy-Gallego *et al.*, 2019) were filtered for
660 concordance with the variants genotyped in this study. The --indep-pairwise command in
661 PLINK(Purcell *et al.*, 2007) was used to prune SNPs with $r^2 > 0.1$ with any other SNP in a
662 50-SNP sliding window, producing 35,759 unlinked variants. These variants were analyzed
663 in both PLINK --pca and in ADMIXTURE(Alexander and Lange, 2011) with K=4 for the 121
664 genotyped individuals in this study, alongside 2,458 individuals from 1000 Genomes. Pan-
665 African and pan-European allele frequencies were obtained from gnomAD v3 (Karczewski
666 *et al.*, 2020). F_{ST} for specific alleles was calculated as $(H_T-H_S)/H_T$ and then polarized, such
667 that positive values indicate variants more common in Africa.

668

669 ***P. falciparum* culture and assays**

670 *P. falciparum* strain 3D7 is a laboratory-adapted strain that was obtained from the
671 Walter and Eliza Hall Institute (Melbourne, Australia) and routinely cultured in human
672 erythrocytes obtained from the Stanford Blood Center. Th.026.09 is a clinical strain isolated
673 from a patient in Senegal in 2009 that was then adapted to short-term culture (Park *et al.*,
674 2012) and kindly provided by Daouda Ndiaye and Sarah Volkman. 3D7 was maintained at
675 2% hematocrit in RPMI-1640 supplemented with 25 mM HEPES, 50 mg/L hypoxanthine,
676 2.42 mM sodium bicarbonate and 4.31 mg/ml Albumax (Invitrogen) at 37°C in 5% CO2 and
677 1% O2. Th.026.09 was maintained in the same media, with the exception that half the
678 Albumax was replaced with heat-inactivated human AB serum.

679 Parasite growth and invasion assays were performed using schizont-stage parasites
680 isolated from routine culture using a MACS magnet (Miltenyi). Parasites were added at
681 ~0.5% initial parasitemia to fresh erythrocytes suspended at 1% hematocrit in complete
682 RPMI, as above. Parasites were cultured in each erythrocyte sample for three to five days in
683 triplicate 100 uL wells. Parasitemia was determined on day 0, day 1 (24 hours), day 3 (72
684 hours), and in some cases day 5 (120 hours) by staining with SYBR Green 1 nucleic acid
685 stain (Invitrogen, ThermoFisher Scientific, Eugene, OR, USA) at 1:2000 dilution in
686 PBS/0.3% BSA for 20 minutes, followed by flow cytometry analysis on a MACSQuant flow
687 cytometer (Miltenyi). Raw invasion rate was defined as the day 1 parasitemia divided by
688 the day 0 parasitemia; raw growth rate was defined as the day 3 (or day 5) parasitemia
689 divided by the day 1 (or day 3) parasitemia. Day 0 parasitemia was not measured in weeks
690 1-3, so invasion rate estimates are absent for these samples. The parasite assays failed for
691 both strains in week 9 and for Th.026.09 in week 10, and so were repeated in weeks 10 and
692 11 with RBCs that had been stored for 1 or 2 weeks.

693 To correct for batch effects, such as week-to-week variation in *P. falciparum*
694 replication, we extracted the residuals from a linear regression of the raw parasite values
695 against up to four significantly related batch variables: (1) the raw values for control donor
696 1111 each week; (2) the parasitemia measured at the previous time point; (3) the age in
697 weeks of the RBCs being measured; and (4) the experimenter performing the assays.
698 Notably, there was no additional effect of 'Week' or the length of the experiment (i.e., 3 or 5

699 days) once the previous variables were regressed out. To convert these residuals (mean 0)
700 to relative percents, we first transformed them with a linear model parameterized by data
701 from control donor 1111 and the most extreme donors (HbAS for growth; $G6PD^{-}_{high}$ and HE
702 for invasion), normalized by the 1111 rates from the appropriate week. These values were
703 finally arithmetically adjusted so that the mean invasion and growth values for non-
704 carriers was 100%.
705
706 **Red cell phenotyping and normalization**
707 Complete blood count (CBC) data for RBCs, reticulocytes, and platelets were
708 obtained with an Advia 120 hematology analyzer (Siemens, Laguna Hills, CA) at the Red
709 Cell Laboratory at Children's Hospital Oakland Research Institute. These data were: RBC,
710 HGB, HCT, MCV, MCH, MCHC, CHCM, RDW, HDW, PLT, MPV, Reticulocyte number and
711 percentage, and the fraction of RBCs in each of nine cells of the RBC matrix (M1-M9, with
712 MCH cutoffs of 28 and 41 and MCV cutoffs of 60 and 120). Systematic biases were evident
713 for some measures in certain weeks, but data from control donor 1111 were not available
714 for all weeks. Therefore, CBC data were normalized such that the median value for non-
715 carrier samples was equal across weeks.
716 Osmotic fragility tests were performed by incubating 20 μL of washed erythrocytes
717 for 5 minutes in 500 μL solutions of NaCl in 14 concentrations: 7.17, 6.14, 5.73, 5.32, 4.91,
718 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 minutes at
719 1000g and 100 μL of supernatant was transferred to a 96-well plate in duplicate.
720 Hemoglobin concentration was determined by adding 100 μL of Drabkin's reagent (Ricca
721 Chemical) to each well and measuring absorbance at OD_{540nm} with a Synergy H1 plate
722 reader (Biotek). Relative lysis was determined by normalizing to the maximum hemoglobin
723 concentration in the 14-tube series for each sample. After outlier points were manually
724 removed, sigmoidal osmotic fragility curves were estimated under a self-starting logistic
725 model in the nls package in R. Curves were summarized by the relative tonicity at which
726 50% lysis occurred and normalized within weekly batches, such that this value was equal
727 for control sample 1111 across weeks.

728 Osmotic gradient ektacytometry (Clark, Mohandas and Shohet, 1983; Kuypers *et al.*,
729 1990) was performed at the Red Cell Laboratory at Children's Hospital Oakland Research
730 Institute (CHORI). Red cell deformability estimates across a range of NaCl concentrations
731 were fitted to a 20-parameter polynomial model to generate a smooth curve, which was
732 manually verified to closely fit the data. Each curve was summarized with three standard
733 points (Figure S3A)(Clark, Mohandas and Shohet, 1983), which were normalized such that
734 the median x- and y-values of the three points was equal for non-carrier samples across
735 weeks.
736
737 **Statistical analysis**
738 Student's t-test was used to compare trait values between non-carriers and each
739 other group. Where n=1 (i.e., for $G6PD^{-}_{high}$ and HE), *p*-values were defined as the percentile
740 of the non-carrier distribution. Significance in these tests was defined as $p<0.1$. OLS linear
741 regression was performed with the lm function in R for all comparisons of two continuous
742 variables, unless otherwise specified. Adjusted R^2 values are reported.
743 Lasso regression (Tibshirani and Tibshirani, 1994; Chatterjee, 2013) was performed
744 with the cv.glmnet function in R. To account for moderate uncertainty in the choice of
745 lambda (due to random division of data for cross-validations), each model was run 1000
746 times. For each predictor, we report the fraction of runs in which it was selected (e.g Figure
747 4A; Table S1; Table S3) .The R^2 values reported between actual and predicted parasite
748 measures are medians from 1000 runs. We noted that effect size estimates for each
749 predictor varied considerably across runs; therefore, we used univariate OLS regression to
750 separately estimate the effect size of each predictor selected at least once in at least one
751 model.
752 The significance of each Lasso result was assessed by permutation. Parasite values
753 were randomly shuffled 1000 times, preserving correlations among potential predictors.
754 The median models for 6 cross-validations of each of the 1,000 permuted data sets were
755 compared to the median model from the real data. To compare the 23 malaria-related
756 genes to other sets of genes, a gene-specific R^2 for the target set was calculated using OLS
757 with only specific variants (no phenotypes or PCs) selected in at least 70% of cross-
758 validations. 1000 random sets of 23 genes were then generated from the RBC proteome

759 (Table S5) and analyzed identically. *P*-values were calculated as the percentile of this
760 permuted distribution in which the malaria-related gene set fell.
761
762
763 **LEAD CONTACT:** Further information and requests for resources and reagents should be
764 directed to and will be fulfilled by the Lead Contact, Elizabeth Egan (eegan@stanford.edu).
765
766 **MATERIALS AVAILABILITY**
767 This study did not generate any new unique reagents.
768
769 **DATA and CODE AVAILABILITY**
770 The human sequence datasets generated during this study will be available at dbGAP (in
771 progress). Phenotype data and code generated during this study is available at
772 https://github.com/emily-ebel/RBC.
773
774 **REFERENCES**
775 Alexander, D. H. and Lange, K. (2011) 'Enhancements to the ADMIXTURE algorithm for
776 individual ancestry estimation', *BMC Bioinformatics*. BioMed Central, 12(1), p. 246. doi:
777 10.1186/1471-2105-12-246.
778 Allison, A. C. (1954) 'Protection afforded by sickle-cell trait against subtertian malarial
779 infection', *British Medical Journal*. BMJ Publishing Group, 1(4857), pp. 290–294. doi:
780 10.1136/bmj.1.4857.290.
781 Archer, N. M. *et al.* (2018) 'Resistance to Plasmodium falciparum in sickle cell trait
782 erythrocytes is driven by oxygen-dependent growth inhibition', *Proceedings of the National*
783 *Academy of Sciences of the United States of America*. National Academy of Sciences, 115(28),
784 pp. 7350–7355. doi: 10.1073/pnas.1804388115.
785 Astle, W. J. *et al.* (2016) 'The Allelic Landscape of Human Blood Cell Trait Variation and
786 Links to Common Complex Disease', *Cell*. Cell Press, 167(5), pp. 1415-1429.e19. doi:
787 10.1016/j.cell.2016.10.042.
788 Van der Auwera, G. A. *et al.* (2013) 'From fastQ data to high-confidence variant calls: The
789 genome analysis toolkit best practices pipeline', *Current Protocols in Bioinformatics*. John

790 Wiley and Sons Inc., 11(SUPL.43), p. 11.10.1. doi: 10.1002/0471250953.bi1110s43.
791 Band, G. *et al.* (2015) 'A novel locus of resistance to severe malaria in a region of ancient
792 balancing selection', *Nature*. Nature Publishing Group, 526(7572), pp. 253–257. doi:
793 10.1038/nature15390.
794 Bejon, P. *et al.* (2007) 'Defining Childhood Severe Falciparum Malaria for Intervention
795 Studies', *PLoS Medicine*. Edited by G. Pasvol. Public Library of Science, 4(8), p. e251. doi:
796 10.1371/journal.pmed.0040251.
797 Beutler, E. and West, C. (2005) 'Hematologic differences between African-Americans and
798 whites: The roles of iron deficiency and α-thalassemia on hemoglobin levels and mean
799 corpuscular volume', *Blood*. American Society of Hematology, 106(2), pp. 740–745. doi:
800 10.1182/blood-2005-02-0713.
801 Boyle, E. A., Li, Y. I. and Pritchard, J. K. (2017) 'An Expanded View of Complex Traits: From
802 Polygenic to Omnigenic', *Cell*. Cell Press, pp. 1177–1186. doi: 10.1016/j.cell.2017.05.038.
803 Campbell, M. C. and Tishkoff, S. A. (2008) 'African Genetic Diversity: Implications for
804 Human Demographic History, Modern Human Origins, and Complex Disease Mapping',
805 *Annual Review of Genomics and Human Genetics*. Annual Reviews , 9(1), pp. 403–433. doi:
806 10.1146/annurev.genom.9.081307.164258.
807 Canela-Xandri, O., Rawlik, K. and Tenesa, A. (2018) 'An atlas of genetic associations in UK
808 Biobank', *Nature Genetics*. Nature Publishing Group, 50(11), pp. 1593–1599. doi:
809 10.1038/s41588-018-0248-z.
810 Cao, A. and Galanello, R. (2010) 'Beta-thalassemia', *Genetics in Medicine*. Nature Publishing
811 Group, pp. 61–76. doi: 10.1097/GIM.0b013e3181cd68ed.
812 Chami, N. *et al.* (2016) 'Exome Genotyping Identifies Pleiotropic Variants Associated with
813 Red Blood Cell Traits', *American Journal of Human Genetics*. Cell Press, 99(1), pp. 8–21. doi:
814 10.1016/j.ajhg.2016.05.007.
815 Chatterjee, S. (2013) 'Assumptionless consistency of the Lasso', *arXiv*. Available at:
816 http://arxiv.org/abs/1303.5817 (Accessed: 10 March 2020).
817 Chen, M.-H. *et al.* (2020) 'Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667
818 Individuals from 5 Global Populations', *Cell*. Cell Press, 182(5), pp. 1198-1213.e14. doi:
819 10.1016/j.cell.2020.06.045.
820 Chen, R. *et al.* (2010) 'Non-synonymous and synonymous coding SNPS show similar

821 likelihood and effect size of human disease association', *PLoS ONE*. PLoS One, 5(10). doi:
822 10.1371/journal.pone.0013574.
823 Clark, M., Mohandas, N. and Shohet, S. (1983) 'Osmotic gradient ektacytometry:
824 comprehensive characterization of red cell volume and surface maintenance', *Blood*, 61(5),
825 pp. 899–910. doi: 10.1182/blood.V61.5.899.899.
826 Clarke, G. M. *et al.* (2017) 'Characterisation of the opposing effects of G6PD deficiency on
827 cerebral malaria and severe malarial anaemia', *eLife*. eLife Sciences Publications Ltd, 6. doi:
828 10.7554/eLife.15085.
829 Cooling, L. (2015) 'Blood groups in infection and host susceptibility', *Clinical Microbiology*
830 *Reviews*. American Society for Microbiology, 28(3), pp. 801–870. doi: 10.1128/CMR.00109-
831 14.
832 Coop, G. *et al.* (2009) 'The Role of Geography in Human Adaptation', *PLoS Genetics*. Edited
833 by M. H. Schierup. Public Library of Science, 5(6), p. e1000500. doi:
834 10.1371/journal.pgen.1000500.
835 Crosnier, C. *et al.* (2011) 'Basigin is a receptor essential for erythrocyte invasion by
836 Plasmodium falciparum', *Nature*. NIH Public Access, 480(7378), pp. 534–537. doi:
837 10.1038/nature10606.
838 Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp.
839 2156–2158. doi: 10.1093/bioinformatics/btr330.
840 Dhermy, D., Schrével, J. and Lecomte, M.-C. (2007) 'Spectrin-based skeleton in red blood
841 cells and malaria', *Current Opinion in Hematology*, 14(3), pp. 198–202. doi:
842 10.1097/MOH.0b013e3280d21afd.
843 Donovan, B. M. (2017) 'Learned inequality: Racial labels in the biology curriculum can
844 affect the development of racial prejudice', *Journal of Research in Science Teaching*. John
845 Wiley and Sons Inc., 54(3), pp. 379–411. doi: 10.1002/tea.21370.
846 Egan, E. S. *et al.* (2015) 'A forward genetic screen identifies erythrocyte CD55 as essential
847 for Plasmodium falciparum invasion', *Science*. American Association for the Advancement
848 of Science, 348(6235), pp. 711–714. doi: 10.1126/science.aaa3526.
849 Egan, E. S. *et al.* (2018) 'Erythrocytes lacking the Langereis blood group protein ABCB6 are
850 resistant to the malaria parasite Plasmodium falciparum', *Communications Biology*. Nature
851 Research, 1(1), pp. 1–9. doi: 10.1038/s42003-018-0046-2.

852 Evans, D. M., Frazer, I. H. and Martin, N. G. (1999) 'Genetic and environmental causes of
853 variation in basal levels of blood cells', *Twin Research (1999)*. Twin Res, 2(4), pp. 250–257.
854 doi: 10.1375/136905299320565735.
855 Facer, C. A. (1995) *Erythrocytes carrying mutations in spectrin and protein 4.1 show differing*
856 *sensitivities to invasion by Plasmodium falciparum*, *Parasitol Res*. Springer Verlag.
857 Field, S. P. *et al.* (1994) 'Glycophorin variants and Plasmodium falciparum: protective effect
858 of the Dantu phenotype in vitro', *Human Genetics*. Springer-Verlag, 93(2), pp. 148–150. doi:
859 10.1007/BF00210600.
860 Flynn, C. J., Hurvich, C. M. and Simonoff, J. S. (2017) 'On the sensitivity of the Lasso to the
861 number of predictor variables', *Statistical Science*. Institute of Mathematical Statistics,
862 32(1), pp. 88–105. doi: 10.1214/16-STS586.
863 Francis, S. E., Sullivan, D. J. and Goldberg, and D. E. (1997) 'HEMOGLOBIN METABOLISM IN
864 THE MALARIA PARASITE *PLASMODIUM FALCIPARUM*', *Annual Review of Microbiology*.
865 Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA ,
866 51(1), pp. 97–123. doi: 10.1146/annurev.micro.51.1.97.
867 Friedman, M. J. (1978) 'Erythrocytic mechanism of sickle cell resistance to malaria',
868 *Proceedings of the National Academy of Sciences of the United States of America*. National
869 Academy of Sciences, 75(4), pp. 1994–1997. doi: 10.1073/pnas.75.4.1994.
870 Galanello, R. and Cao, A. (2011) 'Alpha-thalassemia', *Genetics in Medicine*. Genet Med, pp.
871 83–88. doi: 10.1097/GIM.0b013e3181fcb468.
872 Gallagher, P. G. (2013) 'Abnormalities of the erythrocyte membrane', *Pediatric Clinics of*
873 *North America*. NIH Public Access, pp. 1349–1362. doi: 10.1016/j.pcl.2013.09.001.
874 Garn, S. M. (1981) 'Lower Hematocrit Levels in Blacks Are Not Due to Diet or
875 Socioeconomic Factors', *Pediatrics*, 67(4).
876 Glogowska, E. *et al.* (2017) 'Novel mechanisms of PIEZO1 dysfunction in hereditary
877 xerocytosis', *Blood*. American Society of Hematology, 130(16), pp. 1845–1856. doi:
878 10.1182/blood-2017-05-786004.
879 Goheen, M. M. *et al.* (2016) 'Anemia Offers Stronger Protection Than Sickle Cell Trait
880 Against the Erythrocytic Stage of Falciparum Malaria and This Protection Is Reversed by
881 Iron Supplementation', *EBioMedicine*. Elsevier B.V., 14, pp. 123–130. doi:
882 10.1016/j.ebiom.2016.11.011.

883 Greene, L. S. (1993) 'G6PD deficiency as protection againstfalciparum malaria: An
884 epidemiologic critique of population and experimental studies', *American Journal of*
885 *Physical Anthropology*, 36(S17), pp. 153–178. doi: 10.1002/ajpa.1330360609.
886 GTEx Consortium (2017) 'Genetic effects on gene expression across human tissues', *Nature*.
887 Nature Publishing Group, 550(7675), pp. 204–213. doi: 10.1038/nature24277.
888 Hanssen, E. *et al.* (2012) 'Soft X-ray microscopy analysis of cell volume and hemoglobin
889 content in erythrocytes infected with asexual and sexual stages of Plasmodium falciparum',
890 *Journal of Structural Biology*. Academic Press, 177(2), pp. 224–232. doi:
891 10.1016/j.jsb.2011.09.003.
892 van der Harst, P. *et al.* (2012) 'Seventy-five genetic loci influencing the human red blood
893 cell.', *Nature*. NIH Public Access, 492(7429), pp. 369–75. doi: 10.1038/nature11677.
894 Hofer, T. *et al.* (2009) 'Large allele frequency differences between human continental
895 groups are more likely to have occurred by drift during range expansions than by
896 selection', *Annals of Human Genetics*. Blackwell Publishing Ltd, 73(1), pp. 95–108. doi:
897 10.1111/j.1469-1809.2008.00489.x.
898 Ifediba, T. *et al.* (1985) 'Plasmodium falciparum in vitro: diminished growth in hemoglobin
899 H disease erythrocytes', *Blood*. American Society of Hematology, 65(2), pp. 452–455. doi:
900 10.1182/blood.V65.2.452.452.
901 Ilboudo, Y. *et al.* (2018) 'A common functional PIEZO1 deletion allele associates with red
902 blood cell density in sickle cell disease patients', *American Journal of Hematology*. Wiley-
903 Liss Inc., pp. E362–E365. doi: 10.1002/ajh.25245.
904 Ioannidis, J. P. A. (2005) 'Why Most Published Research Findings Are False', *PLoS Medicine*.
905 Springer International Publishing, 2(8), p. e124. doi: 10.1371/journal.pmed.0020124.
906 Kanias, T. *et al.* (2017) 'Ethnicity, sex, and age are determinants of red blood cell storage
907 and stress hemolysis: Results of the REDS-III RBC-Omics study', *Blood Advances*. American
908 Society of Hematology, 1(15), pp. 1132–1141. doi: 10.1182/bloodadvances.2017004820.
909 Karczewski, K. J. *et al.* (2020) 'The mutational constraint spectrum quantified from
910 variation in 141,456 humans', *Nature*. Nature Research, 581(7809), pp. 434–443. doi:
911 10.1038/s41586-020-2308-7.
912 Kariuki, S. N. *et al.* (2018) 'Red blood cell tension controls Plasmodium falciparum invasion
913 and protects against severe malaria in the Dantu blood group', *bioRxiv*. Cold Spring Harbor

914 Laboratory, (November), p. 475442. doi: 10.1101/475442.
915 Kariuki, S. N. *et al.* (2020) 'Red blood cell tension protects against severe malaria in the
916 Dantu blood group', *Nature*. Nature Publishing Group, pp. 1–5. doi: 10.1038/s41586-020-
917 2726-6.
918 Kinsler, G., Geiler-Samerotte, K. and Petrov, D. (2020) 'A genotype-phenotype-fitness map
919 reveals local modularity and global pleiotropy of adaptation', *bioRxiv*. Cold Spring Harbor
920 Laboratory, p. 2020.06.25.172197. doi: 10.1101/2020.06.25.172197.
921 Koch, M. *et al.* (2017) 'Plasmodium falciparum erythrocyte-binding antigen 175 triggers a
922 biophysical change in the red blood cell that facilitates invasion', *Proceedings of the*
923 *National Academy of Sciences of the United States of America*. National Academy of Sciences,
924 114(16), pp. 4225–4230. doi: 10.1073/pnas.1620843114.
925 Kuypers, F. A. *et al.* (1990) 'Use of ektacytometry to determine red cell susceptibility to
926 oxidative stress', *The Journal of Laboratory and Clinical Medicine*, 116(4), pp. 527–534.
927 Kwiatkowski, D. P. (2005) 'How malaria has affected the human genome and what human
928 genetics can teach us about malaria', *American Journal of Human Genetics*. University of
929 Chicago Press, pp. 171–192. doi: 10.1086/432519.
930 Leffler, E. M. *et al.* (2017) 'Resistance to malaria through structural variation of red blood
931 cell invasion receptors', *Science*. American Association for the Advancement of Science,
932 356(6343), pp. 1140–1152. doi: 10.1126/science.aam6393.
933 Lessard, S. *et al.* (2017) 'An erythroid-specific ATP2B4 enhancer mediates red blood cell
934 hydration and malaria susceptibility', *Journal of Clinical Investigation*. American Society for
935 Clinical Investigation, 127(8), pp. 3065–3074. doi: 10.1172/JCI94378.
936 Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-
937 MEM', *arXiv*. Available at: http://arxiv.org/abs/1303.3997 (Accessed: 22 February 2020).
938 Li, J. *et al.* (2013) 'GWAS of blood cell traits identifies novel associated loci and epistatic
939 interactions in Caucasian and African-American children', *Human Molecular Genetics*. doi:
940 10.1093/hmg/dds534.
941 Lin, S.-H., Brown, D. and Machiela, M. (2020) 'LDtrait: an online tool for identifying
942 published phenotype associations in linkage disequilibrium', *Cancer Research*. American
943 Association for Cancer Research, p. canres.0985.2020. doi: 10.1158/0008-5472.CAN-20-
944 0985.

945 Lo, K. S. *et al.* (2011) 'Genetic association analysis highlights new loci that modulate
946 hematological trait variation in Caucasians and African Americans', *Human Genetics*.
947 Springer, 129(3), pp. 307–317. doi: 10.1007/s00439-010-0925-1.
948 Lowy-Gallego, E. *et al.* (2019) 'Variant calling on the GRCh38 assembly with the data from
949 phase three of the 1000 Genomes Project', *Wellcome Open Research*. F1000 Research Ltd, 4,
950 p. 50. doi: 10.12688/wellcomeopenres.15126.2.
951 Luzzatto, L. (2012) 'Sickle cell anaemia and malaria', *Mediterranean Journal of Hematology*
952 *and Infectious Diseases*. Catholic University in Rome, p. 2012065. doi:
953 10.4084/MJHID.2012.065.
954 Ma, S. *et al.* (2018) 'Common PIEZO1 Allele in African Populations Causes RBC Dehydration
955 and Attenuates Plasmodium Infection', *Cell*. Cell Press, 173(2), pp. 443-455.e12. doi:
956 10.1016/j.cell.2018.02.047.
957 Mackinnon, M. J. *et al.* (2005) 'Heritability of malaria in Africa', *PLoS Medicine*. Public
958 Library of Science, 2(12), pp. 1253–1259. doi: 10.1371/journal.pmed.0020340.
959 De Mendonça, V. R. R., Goncalves, M. S. and Barral-Netto, M. (2012) 'The Host Genetic
960 Diversity in Malaria Infection', *Journal of Tropical Medicine*. Hindawi Publishing
961 Corporation, 2012, p. 17. doi: 10.1155/2012/940616.
962 MGE Network (2019a) 'Insights into malaria susceptibility using genome-wide data on
963 17,000 individuals from Africa, Asia and Oceania', *Nature Communications*. NLM (Medline),
964 10(1), p. 5732. doi: 10.1038/s41467-019-13480-z.
965 MGE Network (2019b) 'Insights into malaria susceptibility using genome-wide data on
966 17,000 individuals from Africa, Asia and Oceania', *Nature communications*. NLM (Medline),
967 10(1), p. 5732. doi: 10.1038/s41467-019-13480-z.
968 Mills, R. E. *et al.* (2006) 'An initial map of insertion and deletion (INDEL) variation in the
969 human genome', *Genome Research*. Cold Spring Harbor Laboratory Press, 16(9), pp. 1182–
970 1190. doi: 10.1101/gr.4565806.
971 Mockenhaupt, F. P. *et al.* (2000) 'Anaemia in pregnant Ghanaian women: importance of
972 malaria, iron deficiency, and haemoglobinopathies', *Transactions of the Royal Society of*
973 *Tropical Medicine and Hygiene*. Royal Society of Tropical Medicine and Hygiene, 94(5), pp.
974 477–483. doi: 10.1016/S0035-9203(00)90057-9.
975 Ndila, C. M. *et al.* (2018) 'Human candidate gene polymorphisms and risk of severe malaria

976 in children in Kilifi, Kenya: a case-control association study', *The Lancet Haematology*.
977 Elsevier Ltd, 5(8), pp. e333–e345. doi: 10.1016/S2352-3026(18)30107-8.
978 Nguetse, C. N. *et al.* (2020) 'A common polymorphism in the mechanosensitive ion channel
979 PIEZO1 is associated with protection from severe malaria in humans', *Proceedings of the*
980 *National Academy of Sciences*. doi: 10.1073/pnas.1919843117/-/DCSupplemental.
981 Okwa, O. O. (2012) *Malaria Parasites*, *Malaria Parasites*. InTech. doi: 10.5772/1477.
982 Otto, T. D. *et al.* (2018) 'Genomes of all known members of a Plasmodium subgenus reveal
983 paths to virulent human malaria', *Nature Microbiology*. Nature Publishing Group, 3(6), pp.
984 687–697. doi: 10.1038/s41564-018-0162-2.
985 Page, G. *et al.* (2017) 'GWAS of Osmotic Hemolysis in 12,352 Healthy Blood Donors
986 Identifies Red Cell Genetic Variants Associated with Steady State Hemolysis in Patients
987 with Sickle Cell Disease', *Blood*. American Society of Hematology, 130(Supplement 1), pp.
988 1117–1117. doi: 10.1182/BLOOD.V130.SUPPL_1.1117.1117.
989 Pankratov, V. *et al.* (2020) 'Differences in local population history at the finest level: the
990 case of the Estonian population', *European Journal of Human Genetics*. Springer Science and
991 Business Media LLC, pp. 1–12. doi: 10.1038/s41431-020-0699-4.
992 Park, D. J. *et al.* (2012) 'Sequence-based association and selection scans identify drug
993 resistance loci in the Plasmodium falciparum malaria parasite', *Proceedings of the National*
994 *Academy of Sciences of the United States of America*. National Academy of Sciences, 109(32),
995 pp. 13052–13057. doi: 10.1073/pnas.1210585109.
996 Pasvol, G., Weatherall, D. J. and Wilson, R. J. M. (1978) 'Cellular mechanism for the
997 protective effect of haemoglobin S against P. falciparum malaria [22]', *Nature*. Nature, pp.
998 701–703. doi: 10.1038/274701a0.
999 Pengon, J. *et al.* (2018) 'Hematological parameters and red blood cell morphological
1000 abnormality of Glucose-6-Phosphate dehydrogenase deficiency co-inherited with
1001 thalassemia', *Hematology/ Oncology and Stem Cell Therapy*. King Faisal Specialist Hospital
1002 and Research Centre, 11(1), pp. 18–24. doi: 10.1016/j.hemonc.2017.05.029.
1003 Perry, G. S. *et al.* (1992) 'Iron nutrition does not account for the hemoglobin differences
1004 between blacks and whites', *Journal of Nutrition*, 122(7), pp. 1417–1424. doi:
1005 10.1093/jn/122.7.1417.
1006 Pilia, G. *et al.* (2006) 'Heritability of Cardiovascular and Personality Traits in 6,148

1007 Sardinians', *PLoS Genetics*. Edited by W. N. Frankel. Public Library of Science, 2(8), p. e132.
1008 doi: 10.1371/journal.pgen.0020132.
1009 Purcell, S. *et al.* (2007) 'PLINK: A tool set for whole-genome association and population-
1010 based linkage analyses', *American Journal of Human Genetics*. Cell Press, 81(3), pp. 559–
1011 575. doi: 10.1086/519795.
1012 Rockett, K. A. *et al.* (2014) 'Reappraisal of known malaria resistance loci in a large
1013 multicenter study', *Nature Genetics*. Nature Publishing Group, 46(11), pp. 1197–1204. doi:
1014 10.1038/ng.3107.
1015 Rooks, H. *et al.* (2019) 'A gain of function variant in PIEZO1 (E756del) and sickle cell
1016 disease', *Haematologica*. Ferrata Storti Foundation, pp. e91–e93. doi:
1017 10.3324/haematol.2018.202697.
1018 Rosenberg, N. A. (2011) 'A Population-Genetic Perspective on the Similarities and
1019 Differences among Worldwide Human Populations', *Human Biology*. Wayne State
1020 University Press, 83(6), pp. 659–684. doi: 10.3378/027.083.0601.
1021 Ruwende, C. and Hill, A. (1998) 'Glucose-6-phosphate dehydrogenase deficiency and
1022 malaria', *Journal of Molecular Medicine*. J Mol Med (Berl), pp. 581–588. doi:
1023 10.1007/s001090050253.
1024 Sauna, Z. E. and Kimchi-Sarfaty, C. (2011) 'Understanding the contribution of synonymous
1025 mutations to human disease', *Nature Reviews Genetics*. Nature Publishing Group, pp. 683–
1026 691. doi: 10.1038/nrg3051.
1027 Schulman, S. *et al.* (1990) 'Growth of Plasmodium falciparum in human erythrocytes
1028 containing abnormal membrane proteins', *Proceedings of the National Academy of Sciences*
1029 *of the United States of America*. National Academy of Sciences, 87(18), pp. 7339–7343. doi:
1030 10.1073/pnas.87.18.7339.
1031 Sohail, M. *et al.* (2019) 'Polygenic adaptation on height is overestimated due to uncorrected
1032 stratification in genome-wide association studies', *eLife*. eLife Sciences Publications Ltd, 8.
1033 doi: 10.7554/eLife.39702.
1034 Sparks, R. A., Baldwin, K. E. and Darner, R. (2020) 'Using Culturally Relevant Pedagogy to
1035 Reconsider the Genetics Canon †', *Journal of Microbiology & Biology Education*. American
1036 Society for Microbiology, 21(1). doi: 10.1128/jmbe.v21i1.1901.
1037 Sundararaman, S. A. *et al.* (2016) 'Genomes of cryptic chimpanzee Plasmodium species

1038 reveal key evolutionary events leading to human malaria', *Nature Communications*. Nature
1039 Publishing Group, 7. doi: 10.1038/ncomms11078.
1040 The 1000 Genomes Project Consortium (2015) 'A global reference for human genetic
1041 variation', *Nature*. Nature Publishing Group, pp. 68–74. doi: 10.1038/nature15393.
1042 Tibshirani, R. and Tibshirani, R. (1994) 'Regression Shrinkage and Selection Via the Lasso',
1043 *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 58, pp. 267--288. Available at:
1044 http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7574 (Accessed: 22 February
1045 2020).
1046 Tiffert, T. *et al.* (2005) 'The hydration state of human red blood cells and their susceptibility
1047 to invasion by Plasmodium falciparum', *Blood*. The American Society of Hematology,
1048 105(12), pp. 4853–4860. doi: 10.1182/blood-2004-12-4948.
1049 Timmann, C. *et al.* (2012) 'Genome-wide association study indicates two novel resistance
1050 loci for severe malaria', *Nature*, 489(7416), pp. 443–446. doi: 10.1038/nature11334.
1051 Vuckovic, D. *et al.* (2020) 'The Polygenic and Monogenic Basis of Blood Traits and Diseases',
1052 *Cell*. Cell Press, 182(5), pp. 1214-1231.e11. doi: 10.1016/j.cell.2020.08.008.
1053 Wang, K., Li, M. and Hakonarson, H. (2010) 'ANNOVAR: Functional annotation of genetic
1054 variants from high-throughput sequencing data', *Nucleic Acids Research*, 38(16). doi:
1055 10.1093/nar/gkq603.
1056 Weatherall, D. J. (2001) 'Phenotype-genotype relationships in monogenic disease: Lessons
1057 from the thalassaemias', *Nature Reviews Genetics*. Nature Publishing Group, pp. 245–255.
1058 doi: 10.1038/35066048.
1059 Whitfield, J. B., Martin, N. G. and Rao, D. C. (1985) 'Genetic and environmental influences on
1060 the size and number of cells in the blood', *Genetic Epidemiology*, 2(2), pp. 133–144. doi:
1061 10.1002/gepi.1370020204.
1062 WHO (2019) *Malaria eradication: benefits, future scenarios and feasibility. Executive*
1063 *summary, WHO Strategic Advisory Group on Malaria Eradication*. Available at:
1064 https://www.who.int/publications/i/item/who-cds-gmp-2019-10 (Accessed: 7 August
1065 2020).
1066 Williams, T. N. (2006) 'Human red blood cell polymorphisms and malaria', *Current Opinion*
1067 *in Microbiology*. Elsevier Current Trends, pp. 388–394. doi: 10.1016/j.mib.2006.06.009.
1068 Wright, G. J. and Rayner, J. C. (2014) 'Plasmodium falciparum Erythrocyte Invasion:

1069 Combining Function with Immune Evasion', *PLoS Pathogens*. Public Library of Science,
1070 10(3). doi: 10.1371/journal.ppat.1003943.
1071 Yoshida, A., Beutler, E. and Motulsky, A. G. (1971) 'Human glucose-6-phosphate
1072 dehydrogenase variants.', *Bulletin of the World Health Organization*. World Health
1073 Organization, 45(2), pp. 243–253.
1074 Zámbó, B. *et al.* (2017) 'Decreased calcium pump expression in human erythrocytes is
1075 connected to a minor haplotype in the ATP2B4 gene', *Cell Calcium*. Elsevier Ltd, 65, pp. 73–
1076 79. doi: 10.1016/j.ceca.2017.02.001.
1077 Zhang, Y. *et al.* (2015) 'Multiple stiffening effects of nanoscale knobs on human red blood
1078 cells infected with Plasmodium falciparum malaria parasite', *Proceedings of the National*
1079 *Academy of Sciences of the United States of America*. National Academy of Sciences, 112(19),
1080 pp. 6068–6073. doi: 10.1073/pnas.1505584112.
1081