TITLE

Malaria parasite fitness in healthy human red cells is driven by common host variation

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

[*]corresponding authors

SUMMARY

The replication of *Plasmodium falciparum* malaria parasites within red blood cells (RBCs) causes severe disease in humans, especially in Africa. The influence of host erythrocyte 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 donated by 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. Furthermore, non-pathogenic polymorphisms in 14 RBC proteins including *SPTA1*, *PIEZO1,* and *ATP2B4* explained 45-70% of parasite growth variation. 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.

INTRODUCTION

Malaria caused by the replication of *Plasmodium falciparum* parasites in red blood cells (RBCs) kills hundreds of thousands of children each year[1]. In each 48-hour cycle of blood stage malaria, parasites must deform RBC membranes to invade them[2,3]; consume

32 hemoglobin and tolerate the resulting oxidative stress[4]; multiply to displace half the RBC
33 volume[5]; and remodel the RBC membrane to avoid immune detection[6]. Consequently,
34 genetic disorders that alter aspects of RBC biology are well known to influence malaria
35 susceptibility[7]. For example, sickle cell trait impairs parasite growth by altering
36 hemoglobin polymerization at low oxygen tension[8,9], while deficiency of the G6PD enzyme
37 involved in oxidative stress tolerance is thought to make parasitized RBCs more susceptible
38 to breakdown[10]. Aside from these diseases, however, the genetic basis of malaria
39 susceptibility remains mostly unknown. Several large genome-wide association studies
40 (GWAS) have recently made progress on this question by identifying a few dozen loci that
41 collectively explain up to 11% of the heritability of the risk of severe versus uncomplicated
42 malaria[11–16]. Eleven of these loci are in or near genes expressed predominantly in RBCs,
43 and one new GWAS variant has since been shown to regulate expression of the *ATP2B4*
44 calcium channel[17] and to be associated with RBC dehydration[18]. However, most other RBC
45 loci that have been identified by GWAS are not novel, and a functional link between *ATP2B4*
46 and *P. falciparum* replication has yet to be demonstrated. More discoveries about the
47 influence of RBC variation on malaria susceptibility are not expected from similar
48 GWAS[19,20], in part because severe malaria is a complex phenotype that depends on a
49 combination of factors from RBCs, the vascular endothelium, the immune system, and the
50 environment[21,22]. Alternate approaches are therefore needed to discover more genetic
51 variation that impacts the replication of malaria parasites in human RBCs.
52 RBC phenotypes like mean cell volume (MCV), hemoglobin content (HGB/MCH), and
53 antigenic blood type vary widely within and between human populations[23–25]. 40-90% of
54 the variation in common RBC indices is heritable[26–28], and GWAS have demonstrated that
55 these phenotypes have polygenic bases[29–33]. Average hemoglobin levels, hematocrit, and
56 especially RBC membrane fragility have also been shown to differ between African and
57 European populations, likely for genetic reasons[34–38]. Although African populations are
58 highly genetically diverse[39], some of these differences between Africans and Europeans are
59 explained by RBC disease alleles[24,35,37] that have been widely selected across Africa for
60 their protective effects on malaria. It remains untested whether other extensive phenotypic
61 and genetic diversity in RBCs, aside from these disease alleles, also influences malaria
62 susceptibility and has been similarly shaped by malaria selection.

63 Here, we approach these questions by performing exome sequencing and extensive
64 RBC phenotyping on blood samples from a diverse human cohort of 122 individuals. We
65 show that *P. falciparum* fitness varies widely among donor cells *in vitro*, with the
66 distributions of parasite phenotypes in 'healthy' RBCs overlapping those from RBCs
67 carrying classic disease alleles. We apply Lasso variable selection to identify a small set of
68 alleles and phenotypes that strongly predict parasite fitness outside of the context of RBC
69 disease, highlighting RBC membrane and dehydration genes and phenotypes as key to
70 modulating *P. falciparum* fitness. However, we find no evidence that non-pathogenic alleles
71 or phenotypes that confer parasite protection are associated with African ancestry.
72 Overall, these findings advance our understanding of the origin and function of common
73 RBC variation and suggest new targets for therapeutic intervention for malaria.
74
75 **RESULTS**
76 Many healthy blood donors with African ancestry carry alleles for RBC disease
77 We collected blood samples from 121 donors with no known history of blood
78 disorders, most of whom self-identified as having recent African ancestry (Figure 1A). As a
79 positive control, we also sampled a patient with hereditary elliptocytosis (HE), a polygenic
80 condition characterized by extremely fragile RBC membranes that strongly inhibit *P.*
81 *falciparum* growth[40–43]. We performed whole-exome sequencing, both to check for the
82 presence of known RBC disease alleles and to confirm the population genetic ancestry of
83 our donors. A principal component analysis of more than 35,000 unlinked, exomic SNPs
84 showed that most donors fell along a continuum from African to European ancestry, as
85 defined by data from the 1000 Genomes Project (Figure 1A). We found that 16% of the
86 healthy donors carried pathogenic hemoglobin alleles (Figure 1B), including 5
87 heterozygotes for hemoglobin S (HbAS), 4 heterozygotes for hemoglobin C (HbAC), and 11
88 individuals with one or two copies of an *HBA2* deletion causing α-thalassemia[44]. We also
89 scored eight polymorphisms in *G6PD* that have been functionally associated with various
90 degrees of G6PD deficiency[45,46] and found that 32% of the study population carried at least
91 one, including 12 of the 20 donors with hemoglobinopathies. Among those with wild-type

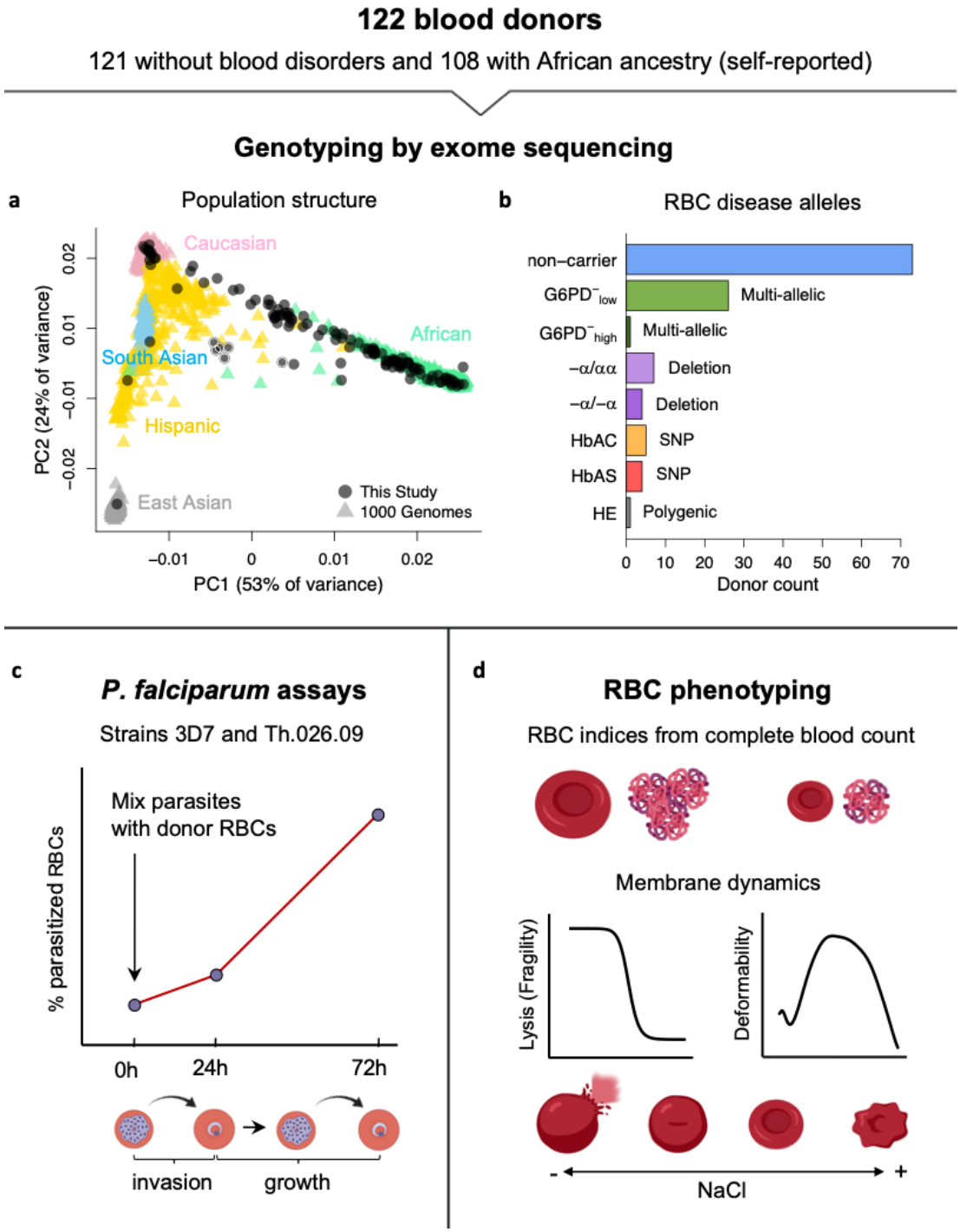


Figure 1. Overview of study design and blood donors. (a) PCA of genetic variation across 35,759 unlinked exome SNPs. Samples collected in this study are plotted on coordinate space derived from 1000 Genomes reference populations. Points with white borders represent the six related individuals in this 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.

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[47,48]. 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 laboratory adapted strain (3D7) and a recent clinical isolate from Senegal (Th.026.09). 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-hour 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 circles) and other batch-specific factors to correct for variation in parasite growth across multiple experiments (see Methods). Among non-carriers, growth rates ranged from 64-136% for 3D7 (SD = 17.7%) and 76-128% for Th.026.09 (SD = 10.7%) (Figure 2A-B). Per-sample growth rates were strongly correlated between the two strains (Figure 2C, $p < 3x10^{-16}$) and positively correlated when measured in different weeks (Figure S1A, ρ=0.34), demonstrating that these data capture real variation among donor RBCs. Furthermore, as expected[42,49–51], we detected reductions in mean growth rate for both strains in RBCs carrying known disease alleles. These included severe G6PD deficiency (3D7, 60% of the non-carrier average, p<0.014; Th.026.09, 64%, p<0.014), α-thalassemia trait (3D7, 75%, p=0.027; Th.026.09, 81%, p = 0.077), HbAS (3D7, 16%, $p=1.05x10^{-7}$; Th.026.09, 19%, $p =1.2x10^{-4}$), and HE

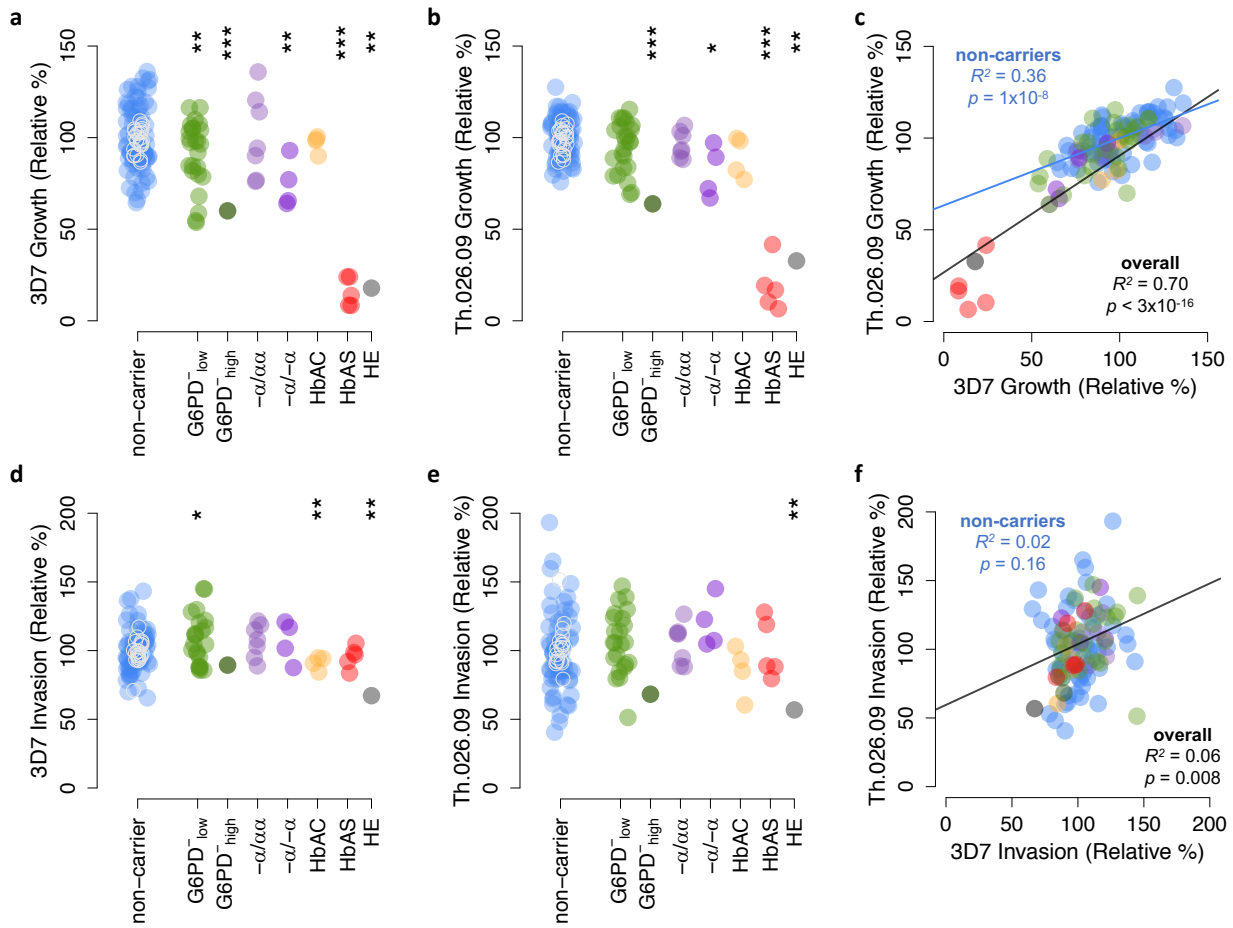


Figure 2. *P. falciparum* replication rate varies widely among donor RBCs. (a-b) Growth of *P. falciparum* strain 3D7 **(a)** or clinical isolate Th026.09 **(b)** over a full 48-hour cycle in donor RBCs (hours 24-72 of assay). Growth is presented relative to the average non-carrier rate after correction 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.

(3D7, 18%, p<0.014; Th.026.09, 33%, p<0.014). 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 effect sizes comparable to the milder known disease alleles.

153 We observed a similarly wide range in the efficiency of *P. falciparum* invasion into
154 donor RBCs (Figure 2D-F). Invasion is defined here as the fold-change in parasitemia over
155 the first 24 hours of the assay, when parasites previously maintained in standard culture
156 conditions egressed and invaded new donor RBCs (Figure 1C). Among non-carriers,
157 invasion rates ranged from 66-173% for 3D7 (SD=15.8%) and 41-193% for Th.026.09
158 (SD=28.9%) (Figure 2D-E). Compared to growth rates, no disease alleles conferred
159 protection against invasion that was extreme enough to fall outside the broad non-carrier
160 range. In 3D7 only, HbAC was associated with a 9% decrease in invasion ($p=0.019$) and
161 mild G6PD deficiency was associated with a 7% increase ($p=0.072$). Only HE had a
162 significant effect on the invasion efficiency of both strains (3D7, 67% of the non-carrier
163 average, $p<0.014$; Th.026.09, 57%, $p<0.014$). There was little correlation between the
164 invasion efficiencies of the two parasite strains in individual donor samples (Figure 2F),
165 potentially reflecting strain-specific differences in the pathways used for invasion[52].
166 However, as we also observed greater variability between repeated samples for invasion
167 than for growth (Figure S1), the invasion results may be influenced by greater
168 experimental noise.
169
170 <u>RBC phenotypes vary widely among non-carriers</u>
171 To assess phenotypic variation across donor RBCs, we measured 19 common
172 indices of RBC size and hemoglobin content from complete blood counts using an Advia
173 hematology analyzer (Figure 3A-D; Figure S2). Mean cellular volume (MCV) and
174 hemoglobin mass (MCH) are closely related traits, which can be represented together as
175 cellular hemoglobin concentration (CHCM) or the fraction of RBCs with 'normal'
176 hemoglobin and volume indices (M5). As expected, each known disease allele was
177 associated with a distinct set of RBC abnormalities. These included elevated CHCM for
178 HbAC ($p=0.033$), consistent with dehydration, and very low MCV ($p=8.7\times10^{-5}$) and MCH
179 ($p=4.6\times10^{-7}$) for α-thalassemia trait (αα/--), consistent with microcytic anemia[44]. RBCs
180 from the HE patient also had very low MCV and MCH (both $p<0.014$), reflecting the
181 membrane breakage and volume loss characteristic of this disease. However, for all of these
182 phenotypic measures, we also observed broad distributions in non-carriers that

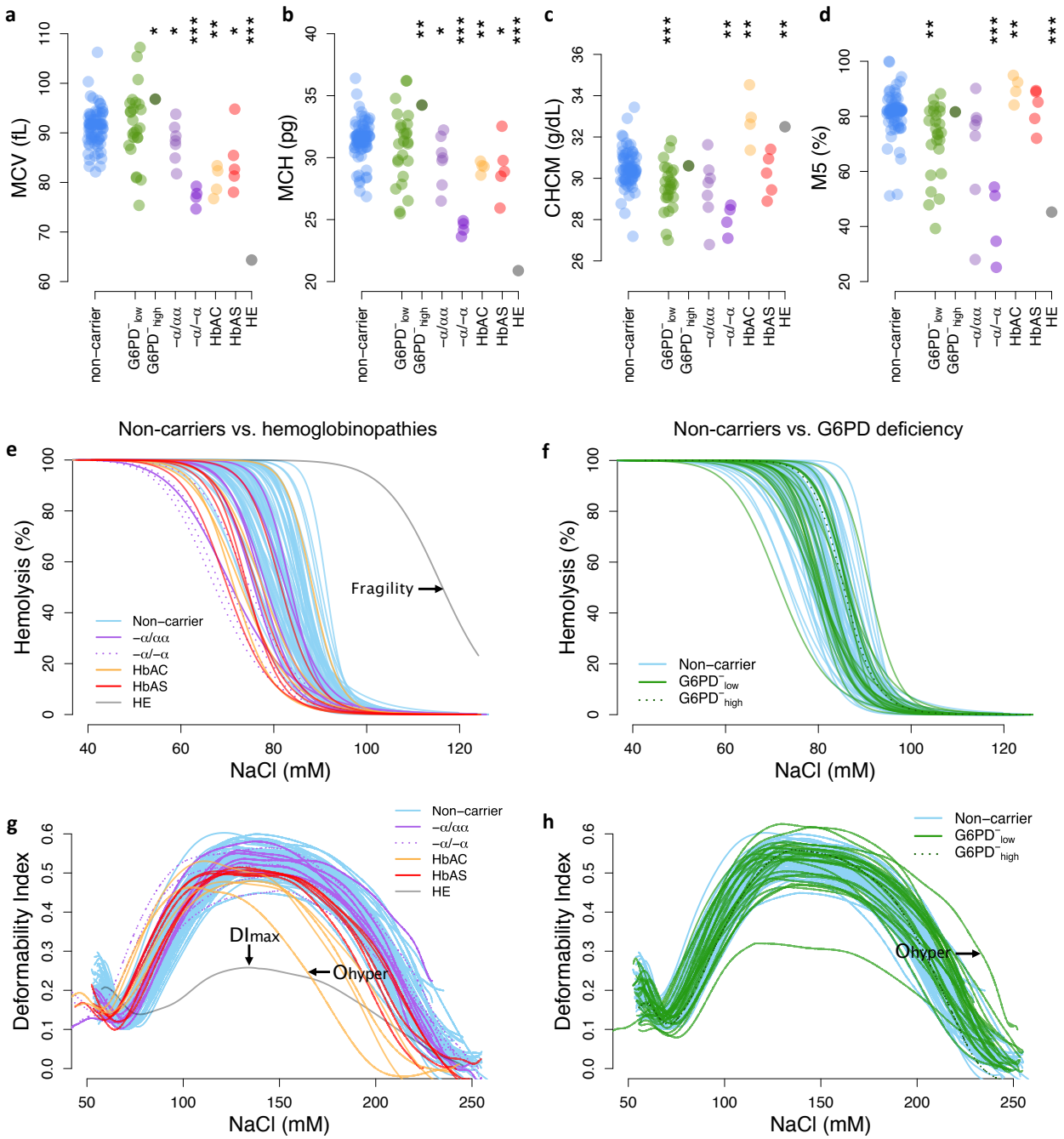


Figure 3. Red cell phenotypes that are abnormal in carriers also vary widely among non-carriers. **MCV:** mean RBC volume. **MCH:** mean cellular hemoglobin. **CHCM:** cellular hemoglobin concentration. **M5:** fraction of RBCs with 'normal' indices of hemoglobin concentration (28-41 g/dL) and volume (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.**

193 overlapped the distributions of most carriers (Figure 3A-D; Figure S2). Notably, the
194 breadth of the non-carrier distribution for each phenotype was large (e.g., 24 fL range for
195 MCV) compared to the average difference between Africans and Caucasians (e.g., 3-5
196 fL[24,35]). This wide diversity and substantial overlap between non-carrier and carrier traits
197 suggest that 'healthy' RBCs exist on the same phenotypic continuum as RBCs carrying
198 known disease alleles.
199 We observed similar patterns of variation in RBC membrane fragility (Figure 3E-F)
200 and membrane deformability (Figure 3G-H), as measured with osmotic fragility tests and
201 osmotic gradient ektacytometry (see Methods). Both sets of curves represent RBC
202 tolerance to osmotic stress, which can result in swelling and lysis (fragility, Figure 3E-F) or
203 dehydration and decreased deformability (O_{hyper}, Figure 3G-H). Specific
204 hemoglobinopathies were associated with moderate to strong reductions in fragility (αα/α-
205 p=0.039; αα/-- p=0.003; HbAS p=0.014), deformability (DI_{max}; HbAC p=0.007; HbAS
206 p=4.4x10^{-6}), and/or resistance to loss of deformability when dehydrated (O_{hyper}; HbAC p =
207 0.019; HbAS p=0.041). HE cells were both extremely fragile (p<0.014) and extremely non-
208 deformable (p<0.014). In non-carriers, the distributions for all membrane measures were
209 wide, continuous, and overlapped the distribution of most carriers (Figure 3E-H; Figure
210 S3). Overall, these data demonstrate that multiple phenotypic alterations associated with
211 RBC disease alleles are also present in non-carrier RBCs.
212
213 Non-carrier variation in RBC phenotypes predicts *P. falciparum* replication rate
214 To identify sets of phenotypes associated with *P. falciparum* replication in non-
215 carrier RBCs, we used a machine learning method called Lasso (least absolute shrinkage
216 and selection operator) that performs regularization and variable selection[53] (see
217 Methods). Briefly, Lasso shrinks the regression coefficients for some possible predictors (in
218 this case, phenotypes) to zero, in order to obtain a subset of predictors that minimizes
219 prediction error. This method is well-suited for datasets in which possible predictors are
220 correlated, as are standard measurements of RBC size, hemoglobin, and membrane
221 dynamics. For each trait in each predictive subset selected by Lasso, we then applied
222 univariate OLS regression to more accurately estimate the sign of its effect on all measured
223 components of parasite fitness. The highest-confidence results are summarized in

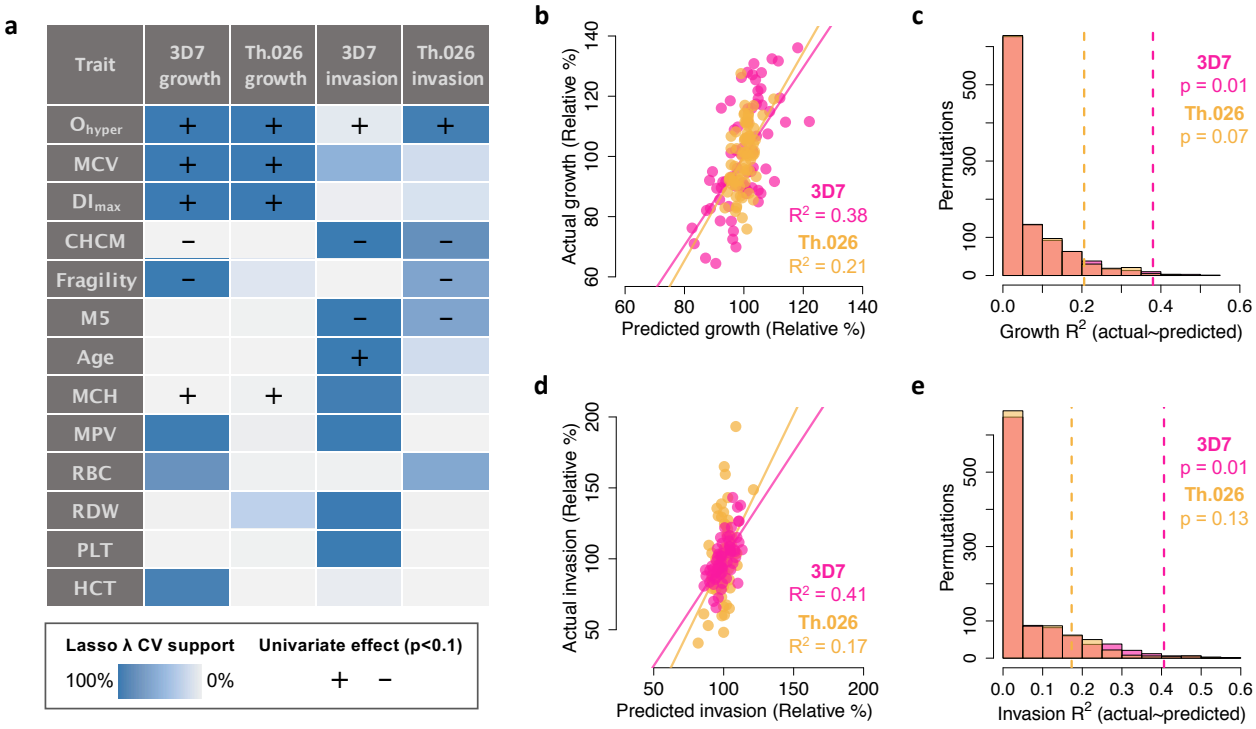


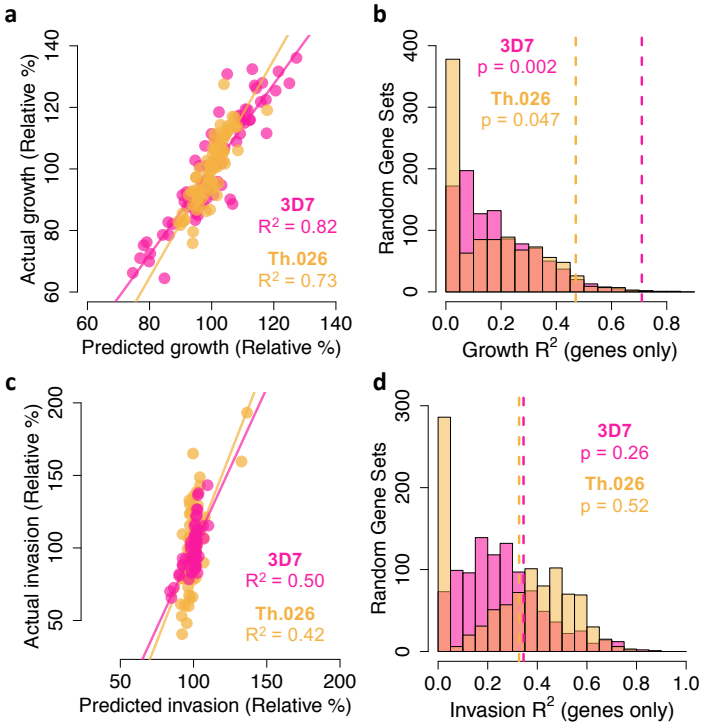
Figure 4. RBC phenotypes predict *P. falciparum* fitness in non-carriers. (a) Phenotypes shown were selected by Lasso in at least one of four models (columns) with at least 60% cross-validation (CV) support, as indicated by blue shading. **(+/-)** shows the direction of effect if the phenotype was significantly 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 (%).

Figure 4A, with complete details provided in Table S2.

P. falciparum fitness in non-carrier RBCs was strongly predicted by variation in several RBC traits related to deformability, dehydration, volume, and hemoglobin content. Among 24 tested phenotypes, the most important trait was the ektacytometry parameter O_{hyper}, which represents a cell's tendency to retain deformability in the face of dehydration (Figure 3G). In univariate models, non-carrier RBCs with the largest O_{hyper} values supported

251 1.2-1.5X faster parasite growth (3D7 1.47X, p=0.003; Th.026.09 1.24X, p=0.005) and 1.4-
252 2.0X more effective invasion (3D7 1.36X, p=0.018; Th.026.09 1.94X, p=0.006) than RBCs
253 with the smallest O_{hyper} values (Table S2; Figure S5A). Consistent with this result, *P.*
254 *falciparum* replication was relatively inhibited in RBCs that were more dehydrated at
255 baseline, as indicated by higher CHCM (3D7 invasion p=0.007; 3D7 growth p=0.042;
256 Th.026.09 invasion p=0.006; Th.026.09 growth p=0.36). Parasites also grew faster in RBCs
257 with larger mean volume, or MCV (3D7 p=0.001; Th.026.09, p=0.001); a greater mass of
258 hemoglobin per cell, or MCH (3D7 p=0.073; Th.026.09 p=0.019); and more deformable
259 membranes, measured by DI_{max} (3D7 p=0.0006; Th.026.09 p=0.006). Some measures of
260 parasite fitness were also reduced in RBCs with more fragile membranes (3D7 invasion
261 p=0.65; 3D7 growth p=0.003; Th.026.09 invasion p=0.067; Th.026.09 growth p=0.14).
262 Additional phenotypes related to platelets and RBC density were selected for some models,
263 but the direction of their effects was unclear since they were not significant when
264 considered individually. These results indicate that common, non-pathogenic variation in
265 RBC size, membrane dynamics, and other correlated traits have meaningful effects on *P.*
266 *falciparum* replication rate in RBCs.
267 Together, 13 phenotype measurements explained approximately 40% of the
268 variation in 3D7 growth and invasion in non-carriers (Figure 4B,D). To test for overfitting,
269 we applied the same model-fitting procedure many times to randomly shuffled parasite
270 data, which usually failed to produce predictive models (i.e., R^2=0, Figure 4C,E). These
271 permutation tests found that real data explained, on average, 6.5X the variance in 3D7
272 growth and 6.3X the variance in 3D7 invasion than expected by chance (both p=0.01,
273 Figure 4C,E). For Th.026.09, the phenotypes selected by Lasso explained 17-21% of the
274 total variation in parasite fitness (Figure 4B), which was marginally better than expected
275 by chance for growth (3.7X, p=0.07, Figure 4C) and marginally worse than expected for
276 invasion (3.0X, p=0.13, Figure 4E). The weaker result for Th.026.09 may be explained by its
277 smaller dynamic growth range (Figure 2A-B), perhaps because clinical isolates are
278 relatively poorly adapted to laboratory conditions. Overall, these results demonstrate that
279 multiple, variable phenotypes impact *P. falciparum* susceptibility in healthy RBCs. Non-
280 carrier cells that are less hospitable to parasites share specific traits with RBCs that carry

281 disease alleles, including smaller size, decreased deformability, and an increased tendency
282 to lose deformability when dehydrating.
283
284 A small set of RBC genotypes predicts *P. falciparum* replication in non-carriers
285 Next, we tested whether non-carrier genotypes derived from exome sequencing
286 could improve our predictions of *P. falciparum* replication rate. To limit false positives and
287 increase statistical power[54,55], we focused on 23 candidate RBC proteins that have been
288 previously associated with *P. falciparum* resistance in the literature (Table S3). These
289 proteins contained 106 unlinked genetic variants (pairwise $r^2 < 0.1$) among the 73 non-
290 carriers. Along with these polymorphisms and phenotypes described above, we also
291 treated the top 10 principal components (PCs) of 1000 Genomes population structure as
292 possible predictors of *P. falciparum* fitness.
293 The addition of genetic factors substantially improved all four models of parasite
294 fitness. Across strains, 73-82% of the parasite variation in non-carriers was explained by a
295 total of 47 genetic and phenotypic variables selected by Lasso (Figure 5A; Table S4). For
296 3D7 and Th.026.09 respectively, polymorphisms in 14 of 23 malaria-related genes
297 explained 71% and 41% of the growth variation in non-carriers (Figure 5B). These models
298 were significant by permutation (both p=0.011; Figure S6AB), indicating that the results
299 were unlikely to be due to overfitting. Nonetheless, to confirm that the genetic signal was
300 specific to our set of 23 malaria-related genes, we used polymorphisms from 1000 sets of
301 random genes from the RBC proteome (Table S6) to build 1000 alternative models (see
302 Methods). On average, variants from random RBC genes explained only 17.9% of the
303 observed variation in *P. falciparum* growth for 3D7 (p=0.002) and 16.4% for Th.026.09
304 (p=0.047) (Figure 5B). In contrast to growth, the genotype/phenotype models of *P.*
305 *falciparum* invasion (Figure 5C) were only marginally significant by permutation (3D7
306 p=0.09, Th.026.09 p=0.11, Figure S6C-D) and the genetic signal was not specific to the set of
307 23 malaria-related genes (3D7 p=0.26, Th.026.09 p=0.52, Figure 5D). Nonetheless, 4 of the
308 8 variants selected by Lasso for 3D7 invasion were also selected for 3D7 growth. Together,
309 these results indicate that *P. falciparum* growth rate in RBCs is strongly determined by host
310 genetic variation concentrated in a small number of genes.



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.

Of the top 31 variants associated by Lasso with *P. falciparum* growth, seventeen (55%) were synonymous (Table S4). This fraction did not significantly differ from the input set of 106 variants (p=0.64, 2-tailed z-score). Although 15/17 of these synonymous variants have no known, direct function, 7/15 are in linkage disequilibrium ($r^2 > 0.1$) with other variants that have been associated with RBC phenotypes by GWAS (Table S4). Similarly, 6 of the top 14 non-synonymous polymorphisms associated with *P. falciparum* growth have been previously linked to RBC or other traits (Table S4). This overlap with previous studies suggests that the variants associated with *P. falciparum* growth in this study are likely to tag functional genetic loci.

336 Nearly all of the top 31 growth-associated polymorphisms occurred in either (1) ion
337 channel proteins, which regulate RBC hydration; (2) components of the flexible RBC
338 membrane backbone; or (3) red cell plasma membrane proteins, including known invasion
339 receptors (Figure 5E). In the first category, the highly polymorphic ion channel *PIEZO1*
340 contained six polymorphisms associated with small (-3%) to moderate (-32%) reductions
341 of *P. falciparum* growth rate (Table S4). The microsatellite variant *PIEZO1*-E756del, which
342 has been a focus of several recent studies[56–59], predicted a moderate reduction in Th.026.09
343 growth (-8.2%, p=0.011) but was not related to RBC dehydration in these data (Table S5).
344 We also detected one growth-associated variant in *ATP2B4,* which encodes the primary
345 RBC calcium channel PMCA4b (-13% 3D7, p=0.006). This variant tags an *ATP2B4* haplotype
346 implicated by GWAS in protection from severe malaria[16,17,60] and many RBC
347 phenotypes[18,31,60,61] including RDW, which was replicated in our data set (Table S4). *SPTA1*
348 and *SPTB,* which encode the flexible spectrin backbone of RBCs, also contained several
349 variants associated with the growth of at least one *P. falciparum* strain, as did the structural
350 linker proteins ANK1, SLC4A1, and EPB42 (Figure 5E; Table S4). Finally, we identified one
351 polymorphism in ABCB6, two in CD44, and two in basigin (BSG) that were strongly
352 associated with *P. falciparum* growth (Figure 5E; Table S4). These plasma membrane
353 proteins have previously been implicated in *P. falciparum* invasion by *in vitro* genetic
354 deficiency studies[62–64]. Notably, two of the natural polymorphisms identified here are
355 synonymous quantitative trait loci (QTL) for CD44 splicing (rs35356320) and BSG
356 expression (rs4682)[65]. We also detected one variant in the RBC iron exporter *FPN* and one
357 in *G6PD* with small effects on Th.026.09 growth (2.4%, p=0.015 and 3.4%, p=0.010
358 respectively; Table S4). No associated variants were detected in the other nine tested genes
359 (Table S2), including three hemoglobin proteins, three glycophorins, *CD55*, *EPB41*, and *HP*.
360 Together, these data demonstrate that *P. falciparum* growth rate in non-carrier RBCs is
361 strongly determined by dozens of host genetic variants, which together shape the
362 phenotypic distribution of red cell susceptibility.
363 In addition to these specific genetic variants, PCs of population structure explained
364 22-26% of the variation in non-carrier invasion and 10-12% of the variation in growth.
365 Interestingly, neither PC1 nor PC2—which distinguish Africans from other populations
366 (Figure 1A)—were related to parasite fitness. Instead, this signal of correlated variation

367 was partially driven by a six-member family with unique ancestry in our sample (Figure
368 1A, white borders) and extreme parasite fitness values (Figure S6). Although the present
369 study is limited by sample size, the association between global genetic PCs and *P.*
370 *falciparum* fitness suggests that additional functional variants remain to be discovered in
371 many populations.
372
373 African ancestry does not predict *P. falciparum* resistance in red cells
374 Based on evidence from balanced disease alleles like HbAS, it has been suggested
375 that anti-malarial selection has shaped polygenic red cell phenotypes in African
376 populations as a whole[37,56,66]. We tested this hypothesis by examining the correlation
377 between African ancestry and *P. falciparum* replication rate in RBCs from non-carriers
378 (Figure 6A-D). We found no evidence that these traits were related, apart from an weakly
379 positive relationship between African ancestry and the invasion rate of Th.026.09, a clinical
380 Senegalese strain ($p=0.01$, $R^2=0.09$, Figure 6D). To understand this result, we next
381 examined how key RBC phenotypes identified in this study (Figure 4A) vary with African
382 ancestry. We found that greater African ancestry predicts reduced osmotic fragility ($p=$
383 9.03×10^{-6}), increased hypotonic deformability (DI_{min}, $p=0.001$), lower cellular hemoglobin
384 concentration (CHCM, reflecting reduced dehydration, $p=0.007$), and a reduced fraction of
385 RBCs with normal volume and hemoglobin content (M5, $p=0.033$). All of these traits
386 actually predict *greater* red cell susceptibility to *P. falciparum* (Figure 4A), although
387 together they explain less than 9% of the non-carrier variation in parasite fitness. The
388 remaining key phenotypes (Figure 4A) do not vary with African ancestry, which may
389 explain why African ancestry itself is mostly unassociated with *P. falciparum* fitness in non-
390 carrier RBCs (Fig 6A-D).
391 Next, we used allele frequency data from the global gnomAD collection[67] to test
392 whether polymorphisms associated with *P. falciparum* fitness in RBCs (Figure 5E) vary
393 between African and European populations. Geographical differences in malaria selection
394 are hypothesized to have increased the frequency of hundreds or thousands of
395 undiscovered anti-malarial alleles in Africa[22,68], as has been shown for several variants
396 causing common RBC disorders (Figure 1B)[7]. Here, aside from well-known disease
397 variants, we find no evidence that African populations as a whole are enriched for RBC

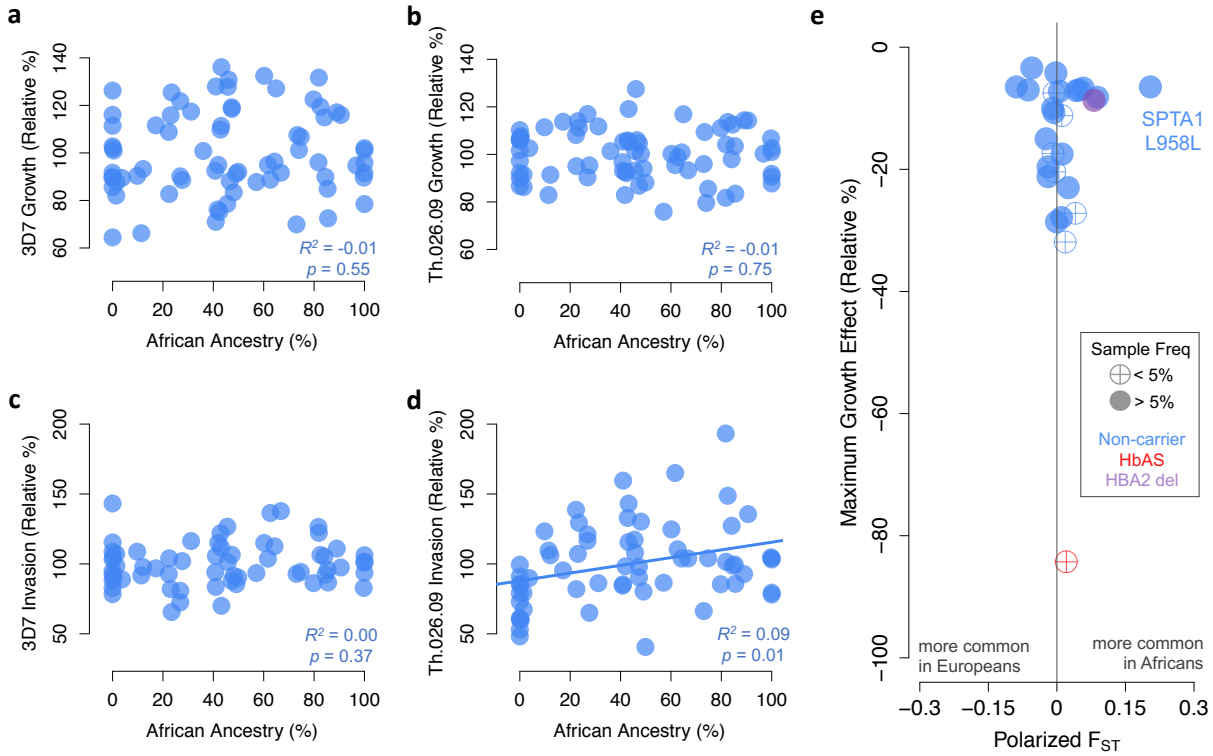


Figure 6. No evidence of widespread selection in Africa for slower *P. falciparum* replication or protective alleles in non-carriers. (a-d) Parasite growth and invasion measurements plotted against the exome-wide fraction of African ancestry, determined by comparison to 1000 Genomes reference populations. 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).

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

420 direct association with a number of RBC phenotypes in previous GWAS (Table S4), this
421 synonymous SPTA1 allele could be a promising candidate for investigating anti-malarial
422 selection in Africa. However, the total set of natural RBC variants that limit *P. falciparum*
423 growth *in vitro* do not appear to support that general hypothesis.
424
425 **DISCUSSION**
426 Healthy red blood cells (RBCs) harbor extensive phenotypic and genetic variation,
427 both within and between human populations. In this work, we demonstrate that this
428 variation modulates a wide range of RBC susceptibility to *P. falciparum* parasites. Our
429 findings add to a growing understanding of the genetic and phenotypic basis of RBC
430 resistance to *P. falciparum*, especially for RBCs that lack population-specific disease alleles.
431 In addition to suggesting new targets for future malaria interventions, these findings
432 challenge assumptions about the role of malaria selection in shaping human RBC diversity.
433 Since exponential replication of *P. falciparum* is a significant driver of malaria
434 disease progression[69], the ample variation in this trai*t* that we observed *in vitro* could be
435 relevant for clinical outcomes in endemic regions. Growth inhibition from HbAS, for
436 example, reduces the risk of death from malaria by reducing parasite density in the
437 blood[70,71]. While HbAS has an extreme effect size, the 3-fold range of parasite replication
438 rates we found among non-carrier RBCs shares substantial overlap with RBCs carrying
439 other protective variants. Although the physiologically complex basis of severe malaria[72]
440 makes it difficult to estimate the precise contribution of RBC factors to severe malaria risk,
441 the genotypes and phenotypes we have associated with *P. falciparum* fitness may be
442 promising targets for future therapeutic interventions.
443 We have shown here that widespread, 'normal' variation in RBC traits like volume,
444 hydration status, and maximum deformability is associated with *P. falciparum* fitness in
445 non-carrier RBCs. Reassuringly, these phenotypes are present to a stronger degree in RBCs
446 that do carry disease alleles[73–75]. They are also consistent with published reports of
447 extreme RBC traits that limit parasite replication. For example, Tiffert *et al.*[76] used an
448 experimental gradient of ion concentrations to show that *P. falciparum* growth is strongly
449 reduced in chemically dehydrated RBCs. Here, we found that naturally occurring, mildly
450 dehydrated RBCs also supported less efficient parasite growth and invasion. Similarly, the

451 rare Dantu variant of the glycophorin A/B receptors has been associated with increased
452 RBC membrane tension, decreased *P. falciparum* invasion efficiency, and decreased risk of
453 severe malaria[11,77,78]. Our data on RBC membrane deformability and fragility support the
454 same link between more flexible RBC membranes and greater *P. falciparum* success.
455 Moreover, they expand upon previous studies by demonstrating for the first time that
456 common, healthy phenotypic variation in RBCs contributes meaningfully to *P. falciparum*
457 fitness.

458 In our linear models of *P. falciparum* growth, phenotypic variation was strongly
459 outperformed by genetic variation in a small number of RBC proteins. This result implies
460 the existence of additional RBC phenotypes that we did not measure (or did not measure
461 with sufficient accuracy), but which have a genetic basis in a small number of genes
462 important to *P. falciparum*. Approximately half of the polymorphisms we identified are
463 non-synonymous, and may therefore exert direct effects on phenotypes like RBC
464 membrane structure or ion transport. The abundances of RBC gene transcripts and
465 proteins are other clear candidates for future phenotype studies, given that the other half
466 of associated polymorphisms were synonymous. A recent metanalysis showed that 'silent'
467 and 'coding' SNPs are equally likely to be associated with human disease, and moreover
468 have similar effect sizes[83]. Although synonymous SNPs may sometimes be linked to coding
469 variants, some have been shown to directly impact mRNA expression, protein translation
470 speed, and protein folding by altering transcription levels, codon usage, and mRNA
471 stability[84]. Synonymous SNPs that impact splicing, like rs35356320 in CD44, may also
472 impact protein structure. Other conceivable RBC phenotypes, such as the dynamics of
473 membrane modification during *P. falciparum* development, may only become evident in
474 more detailed time course experiments. The true number of RBC phenotypes important to
475 *P. falciparum* may be effectively infinite[85], making it very useful in practice that static
476 genetic variation is so strongly predictive of parasite growth.

477 Overall, just 31 polymorphisms in 14 RBC genes explained the majority of the
478 variation in *P. falciparum* growth among non-carriers. One reason that our explanatory
479 power is so high is our focus on *P. falciparum* replication in RBCs, which is a relatively
480 simple component of a larger, complex disease. This focus allowed us to use controlled, *in*
481 *vitro* experiments to detect genetic variants with small to moderate effects. Another

482 important explanation is our use of Lasso variable selection on just 106 polymorphisms in
483 genes with strong existing links to malaria. This approach obviated the need to meet an
484 exome-wide significance threshold, while still allowing for the discovery of novel,
485 putatively functional alleles in well-known disease genes. Furthermore, we confirmed that
486 this set of genes was enriched for signal compared to random sets of genes drawn from the
487 exome-wide background. It is important to note that genetic linkage complicates the
488 identification of the exact functional polymorphisms in any population sample[87], and we
489 cannot rule out that some variants identified here are merely linked to the true functional
490 variants. Indeed, about half of the variants we identified occur in linkage blocks containing
491 other SNPs associated with RBC traits in GWAS studies. In any case, we present significant
492 evidence that loci containing 14 RBC genes are strongly enriched for polymorphisms with
493 significant impacts on *P. falciparum* growth.
494 One of the unique aspects of our study is the participation of individuals with a
495 range of African ancestry, defined by similarity to donors from five 1000 Genomes
496 reference populations. We found that African ancestry was associated with RBC
497 phenotypes that *improved* parasite fitness, particularly for the clinical Senegalese strain
498 Th.026.09. In the future, it would be very interesting to test for local parasite adaptation to
499 human RBCs using *P. falciparum* strains and RBC samples from around the globe. We also
500 found that the polymorphisms associated with *P. falciparum* growth by Lasso are not
501 enriched in African populations included in the gnomAD database of human variation. This
502 is not consistent with African-specific selection on these smaller-effect variants, despite the
503 famous examples provided by RBC disease alleles. This unexpected result for 'healthy'
504 variation has at least three possible explanations. First, relatively small-effect alleles may
505 not have had sufficient time to increase in frequency since *P. falciparum* began expanding
506 in humans some 5,000-10,000 years ago[88,89]. Second, the complexity of severe malaria
507 could mean that the variants discovered here do not ultimately impact disease outcome,
508 especially relative to known disease variants. Third, human adaptation may be too local to
509 detect with coarse-grain sampling of Sub-Saharan African genetic diversity (e.g [90]). Our
510 data do suggest, however, that few RBC alleles remain to be discovered that are both
511 widespread in Africa and have large effects on *P. falciparum* proliferation in RBCs (e.g. [56]).

512 More broadly, these data show that it is inaccurate to make assumptions about RBC
513 susceptibility to *P. falciparum* based on a person's race or continental ancestry. These kinds
514 of hypotheses[37,56,66,68] are based on well-known examples of balanced disease alleles, but
515 our data suggest that these examples are exceptions rather than the rule. It is important to
516 recognize that, at least in biology classes, the use of racially-based genetic examples
517 increases the belief that genes encode absolute, functional differences between races[91,92].
518 This perspective neglects the fact that >90% of genetic differences among humans occur
519 within populations, rather than across them[93]. In this study, global population structure
520 captured by PCs showed that non-African variation has its own distinct impacts on *P.*
521 *falciparum* replication in RBCs. These data are an important reminder that most human
522 genetic diversity is a result of demography, not population-specific selection[94,95].
523 In conclusion, this study demonstrates that substantial phenotypic and genetic
524 diversity in healthy human RBCs impacts the replication of malaria parasites. Whether or
525 not this diversity is shaped by malaria selection, a better understanding of how *P.*
526 *falciparum* biology is impacted by natural RBC variation can help lead to new therapies for
527 one of humanity's most important infectious diseases.
528
529
530 **METHODS**
531 Sample collection and preparation
532 Subjects with no known history of RBC disorders were recruited to donate blood at
533 the Stanford Clinical and Translational Research Unit. Written informed consent was
534 obtained from each subject and/or their parent as part of a protocol approved by the
535 Stanford University Institutional Review Board (#40479). To help control for weekly batch
536 effects, Subject 1111 donated fresh blood for parasite assays each week. 11 other subjects
537 donated blood on at least two different weeks. Whole blood samples from a hereditary
538 elliptocytosis patient were obtained from Dr. Bertil Glader under a separate approved
539 protocol (#14004) that permitted sample sharing among researchers. All samples were de-
540 identified upon collection by labeling with a random four-digit number. Two samples were
541 eventually removed from analysis based on a failed sequencing library (6449KD) and
542 history of stem cell transplant (8715).

543 Whole blood was drawn into CPDA tubes and spun down within 36 hours to
544 separate serum, buffy coat, and red blood cells. Red blood cells were washed and stored in
545 RPMI-1640 medium (Sigma) supplemented with 25 mM HEPES, 50 mg/L hypoxanthine,
546 and 2.42 mM sodium bicarbonate at 4°C. Buffy coat was transferred directly to cryotubes
547 and stored at -80°C.
548
549 Exome sequencing and genotype calling
550 Genomic DNA was isolated from frozen buffy coats using a DNeasy Blood and Tissue
551 Kit (Qiagen). Libraries were prepared using a KAPA Hyperplus kit (Roche) and hybridized
552 to human exome probes using the SeqCap EZ Prime Exome kit (Roche). The resulting
553 exome libraries were sequenced with paired-end 150 bp Illumina reads on the HiSeq or
554 NextSeq platforms at Admera Health (South Plainfield, NJ).
555 Reads were aligned to the hg38 human reference genome using bwa mem[96],
556 yielding an average coverage of 42X across targeted exome regions (excluding sample
557 6449KD). Variants were called using GATK best practices[97] and hard filtered with the
558 following parameters: QD < 2.0, FS > 60.0, ReadPosRankSum < -2.5, SOR > 2.5, MQ < 55.0,
559 MQRankSum < -1.0, DP < 500. To minimize the effects of sequencing errors, variants not
560 present in 1000 Genomes, dbSNP_138, or the Mills indel collection[98] were discarded.
561 Variants were also discarded if their frequencies in our sample did not fall between
562 gnomAD African or European frequencies[67] and significantly differed from both groups'
563 frequencies by Fisher's Exact Test ($p \leq 0.05$).
564 *PIEZO1* E756del was genotyped via PCR and Sanger sequencing according to a
565 previously published protocol[57]. To call deletion variants that cause α-thalassemia in the
566 paralogous genes HBA2 and HBA1, we extracted reads from each .bam file that lacked any
567 mismatches or soft-clipping and had MAPQ ≥ 13 (i.e., <5% chance of mapping error).
568 Coverage with these well-mapped reads was calculated over the 73 and 81 bp of unique
569 sequence in HBA2 and HBA1, respectively, and normalized to each sample's exome-wide
570 coverage. To determine which samples has unusually low coverage, we formed an *ad hoc*
571 reference panel of seven donors who were unlikely to carry deletion alleles based on their
572 normal MCH, MCV, and HGB and >96% exome-wide European ancestry[99]. We called
573 heterozygous HBA2 deletions when normalized coverage across three unique regions of

574 the HBA2 gene was below the minimum reference value. Similarly, we called homozygous
575 HBA2 deletions when normalized coverage across three unique regions of the HBA2 gene
576 was less than half of the minimum reference value. This approach resulted in an estimated
577 HBA2 copy number of 2.0 in the reference panel, 0.95 in eight putative heterozygotes and
578 0.12 in four putative homozygotes. The same method produced no evidence of HBA1
579 deletion in any sample.
580
581 Variant classification and linkage pruning
582 Exonic variants in RefSeq genes were identified using ANNOVAR[100]. Variants were
583 classified into three categories: those within 23 malaria-related genes (Table S2); those
584 within 887 other red blood cell proteins (Table S5) derived with a medium-confidence
585 filter from the Red Blood Cell Collection database (rbcc.hegelab.org); and those within any
586 other gene. Singleton variants were removed to avoid excessive false positives.
587 Linkage between all pairs of bi-allelic, exonic variants in our 121 genotyped samples
588 was calculated using the --geno-r2 and --interchrom-geno-r2 functions in vcftools [101].
589 Variants in RBC genes that shared $r^2 > 0.1$ with any variant in the 23-gene set were
590 removed. Within the 23-gene set and RBC-gene set separately, non-carrier variants were
591 ranked by the *p*-values of their OLS regression with all four parasite measures. Then, one
592 variant was removed from each pair with $r^2 > 0.1$, prioritizing retention in the following
593 order: greater significance across models; non-synonymous protein change; higher
594 frequency in our sample; and finally by random sampling. We report results from additive
595 genetic models (genotypes coded 0/1/2), which performed as well or better than recessive
596 (0/0/2) and dominant (0/2/2) models.
597
598 Population analysis
599 The population ancestry of our donors was assessed by comparison with African,
600 European, East Asian, and South Asian reference populations from the 1000 Genomes
601 Project[102]. Briefly, variants called from an hg38 alignment of the 1000 Genomes data [103]
602 were filtered for concordance with the variants genotyped in this study. The --indep-
603 pairwise command in PLINK[104] was used to prune SNPs with $r^2 > 0.1$ with any other SNP in
604 a 50-SNP sliding window, producing 35,759 unlinked variants. These variants were

605 analyzed in both PLINK --pca and in ADMIXTURE[105] with K=4 for the 121 genotyped
606 individuals in this study, alongside 2,458 individuals from 1000 Genomes. Pan-African and
607 pan-European allele frequencies were obtained from gnomAD v3[67]. F_{ST} for specific alleles
608 was calculated as $(H_T-H_S)/H_T$ and then polarized, such that positive values indicate variants
609 more common in Africa.
610
611 *P. falciparum* culture and assays
612 *P. falciparum* strain 3D7 is a laboratory-adapted strain that was obtained from the
613 Walter and Eliza Hall Institute (Melbourne, Australia) and routinely cultured in human
614 erythrocytes obtained from the Stanford Blood Center. Th.026.09 is a clinical strain isolated
615 from a patient in Senegal in 2009 that was then adapted to short-term culture[106] and kindly
616 provided by Daouda Ndiaye and Sarah Volkman. 3D7 was maintained at 2% hematocrit in
617 RPMI-1640 supplemented with 25 mM HEPES, 50 mg/L hypoxanthine, 2.42 mM sodium
618 bicarbonate and 4.31 mg/ml Albumax (Invitrogen) at 37°C in 5% CO2 and 1% O2.
619 Th.026.09 was maintained in the same media, with the exception that half the Albumax was
620 replaced with heat-inactivated human AB serum.
621 Parasite growth and invasion assays were performed using schizont-stage parasites
622 isolated from routine culture using a MACS magnet (Miltenyi). Parasites were added at
623 ~0.5% initial parasitemia to fresh erythrocytes suspended at 1% hematocrit in complete
624 RPMI, as above. Parasites were cultured in each erythrocyte sample for three to five days in
625 triplicate 100 uL wells. Parasitemia was determined on day 0, day 1 (24 hours), day 3 (72
626 hours), and in some cases day 5 (120 hours) by staining with SYBR Green 1 nucleic acid
627 stain (Invitrogen, ThermoFisher Scientific, Eugene, OR, USA) at 1:2000 dilution in
628 PBS/0.3% BSA for 20 minutes, followed by flow cytometry analysis on a MACSQuant flow
629 cytometer (Miltenyi). Raw invasion rate was defined as the day 1 parasitemia divided by
630 the day 0 parasitemia; raw growth rate was defined as the day 3 (or day 5) parasitemia
631 divided by the day 1 (or day 3) parasitemia. Day 0 parasitemia was not measured in weeks
632 1-3, so invasion rate estimates are absent for these samples. The parasite assays failed for
633 both strains in week 9 and for Th.026.09 in week 10, and so were repeated in weeks 10 and
634 11 with RBCs that had been stored for 1 or 2 weeks.

635 To correct for batch effects, such as week-to-week variation in *P. falciparum*
636 replication, we extracted the residuals from a linear regression of the raw parasite values
637 against up to four significantly related batch variables: (1) the raw values for control donor
638 1111 each week; (2) the parasitemia measured at the previous time point; (3) the age in
639 weeks of the RBCs being measured; and (4) the experimenter performing the assays.
640 Notably, there was no additional effect of 'Week' or the length of the experiment (i.e., 3 or 5
641 days) once the previous variables were regressed out. To convert these residuals (mean 0)
642 to relative percents, we first transformed them with a linear model parameterized by data
643 from control donor 1111 and the most extreme donors (HbAS for growth; $G6PD^{-}_{high}$ and HE
644 for invasion), normalized by the 1111 rates from the appropriate week. These values were
645 finally arithmetically adjusted so that the mean invasion and growth values for non-
646 carriers was 100%.
647
648 Red cell phenotyping and normalization
649 Complete blood count (CBC) data for RBCs, reticulocytes, and platelets were
650 obtained with an Advia 120 hematology analyzer (Siemens, Laguna Hills, CA) at the Red
651 Cell Laboratory at Children's Hospital Oakland Research Institute. These data were: RBC,
652 HGB, HCT, MCV, MCH, MCHC, CHCM, RDW, HDW, PLT, MPV, Reticulocyte number and
653 percentage, and the fraction of RBCs in each of nine cells of the RBC matrix (M1-M9, with
654 MCH cutoffs of 28 and 41 and MCV cutoffs of 60 and 120). Systematic biases were evident
655 for some measures in certain weeks, but data from control donor 1111 were not available
656 for all weeks. Therefore, CBC data were normalized such that the median value for non-
657 carrier samples was equal across weeks.
658 Osmotic fragility tests were performed by incubating 20 μL of washed erythrocytes
659 for 5 minutes in 500 μL solutions of NaCl in 14 concentrations: 7.17, 6.14, 5.73, 5.32, 4.91,
660 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
661 1000g and 100 μL of supernatant was transferred to a 96-well plate in duplicate.
662 Hemoglobin concentration was determined by adding 100 μL of Drabkin's reagent (Ricca
663 Chemical) to each well and measuring absorbance at OD_{540nm} with a Synergy H1 plate
664 reader (Biotek). Relative lysis was determined by normalizing to the maximum hemoglobin

665 concentration in the 14-tube series for each sample. After outlier points were manually
666 removed, sigmoidal osmotic fragility curves were estimated under a self-starting logistic
667 model in the nls package in R. Curves were summarized by the relative tonicity at which
668 50% lysis occurred and normalized within weekly batches, such that this value was equal
669 for control sample 1111 across weeks.
670 Osmotic gradient ektacytometry[74,107] was performed at the Red Cell Laboratory at
671 Children's Hospital Oakland Research Institute (CHORI). Red cell deformability estimates
672 across a range of NaCl concentrations were fitted to a 20-parameter polynomial model to
673 generate a smooth curve, which was manually verified to closely fit the data. Each curve
674 was summarized with three standard points (Figure S3A)[74], which were normalized such
675 that the median x- and y-values of the three points was equal for non-carrier samples
676 across weeks.
677
678 Statistical analysis
679 Student's t-test was used to compare trait values between non-carriers and each
680 other group. Where n=1 (i.e., for $G6PD^{-}_{high}$ and HE), *p*-values were defined as the percentile
681 of the non-carrier distribution. Significance in these tests was defined as $p<0.1$. OLS linear
682 regression was performed with the lm function in R for all comparisons of two continuous
683 variables, unless otherwise specified. Adjusted R^2 values are reported.
684 Lasso regression[53,108] was performed with the cv.glmnet function in R. To account
685 for moderate uncertainty in the choice of lambda (due to random division of data for cross-
686 validations), each model was run 1000 times. For each predictor, we report the fraction of
687 runs in which it was selected (e.g Figure 4A; Table S1; Table S3) .The R^2 values reported
688 between actual and predicted parasite measures are medians from 1000 runs. We noted
689 that effect size estimates for each predictor varied considerably across runs; therefore, we
690 used univariate OLS regression to separately estimate the effect size of each predictor
691 selected at least once in at least one model.
692 The significance of each Lasso result was assessed by permutation. Parasite values
693 were randomly shuffled 1000 times, preserving correlations among potential predictors.
694 The median models for 6 cross-validations of each of the 1,000 permuted data sets were
695 compared to the median model from the real data. To compare the 23 malaria-related

genes to other sets of genes, a gene-specific R^2 for the target set was calculated using OLS with only specific variants (no phenotypes or PCs) selected in at least 70% of cross-validations. 1000 random sets of 23 genes were then generated from the RBC proteome (Table S5) and analyzed identically. *P*-values were calculated as the percentile of this permuted distribution in which the malaria-related gene set fell.

ACKNOWLEDGEMENTS

We gratefully acknowledge the invaluable participation of all 123 volunteer blood donors. Nick Bondy, 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. For quantitative advice, we thank Grant Kinsler and the Stanford Statistics Consulting Group. This work 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. E.R.E. received additional support from the Stanford Center for Computational, Evolutionary, and Human Genomics. D.A.P. was funded through an NIH MIRA award 5R35GM118165-05. E.S.E. was funded by the Doris Duke Charitable Foundation (no. 2016098) and the NIH (DP2HL13718601). Local blood samples were drawn at the Stanford Clinical and Translational Research Unit, which is supported by CTSA Grant UL1 TR001085.

REFERENCES

1. WHO. *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 (2019).
2. Kariuki, S. N. *et al.* Red blood cell tension protects against severe malaria in the Dantu blood group. *Nature* 1–5 (2020) doi:10.1038/s41586-020-2726-6.
3. Koch, M. *et al.* Plasmodium falciparum erythrocyte-binding antigen 175 triggers a biophysical change in the red blood cell that facilitates invasion. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4225–4230 (2017).
4. Francis, S. E., Sullivan, D. J. & Goldberg, and D. E. HEMOGLOBIN METABOLISM IN

727 THE MALARIA PARASITE *PLASMODIUM FALCIPARUM*. *Annu. Rev. Microbiol.* **51**, 97–
728 123 (1997).
729 5. Hanssen, E. *et al.* Soft X-ray microscopy analysis of cell volume and hemoglobin
730 content in erythrocytes infected with asexual and sexual stages of Plasmodium
731 falciparum. *J. Struct. Biol.* **177**, 224–232 (2012).
732 6. Zhang, Y. *et al.* Multiple stiffening effects of nanoscale knobs on human red blood
733 cells infected with Plasmodium falciparum malaria parasite. *Proc. Natl. Acad. Sci. U. S.*
734 *A.* **112**, 6068–6073 (2015).
735 7. Kwiatkowski, D. P. How malaria has affected the human genome and what human
736 genetics can teach us about malaria. *American Journal of Human Genetics* vol. 77 171–
737 192 (2005).
738 8. Pasvol, G., Weatherall, D. J. & Wilson, R. J. M. Cellular mechanism for the protective
739 effect of haemoglobin S against P. falciparum malaria [22]. *Nature* vol. 274 701–703
740 (1978).
741 9. Archer, N. M. *et al.* Resistance to Plasmodium falciparum in sickle cell trait
742 erythrocytes is driven by oxygen-dependent growth inhibition. *Proc. Natl. Acad. Sci.*
743 *U. S. A.* **115**, 7350–7355 (2018).
744 10. Ruwende, C. & Hill, A. Glucose-6-phosphate dehydrogenase deficiency and malaria.
745 *Journal of Molecular Medicine* vol. 76 581–588 (1998).
746 11. Leffler, E. M. *et al.* Resistance to malaria through structural variation of red blood cell
747 invasion receptors. *Science (80-.).* **356**, 1140–1152 (2017).
748 12. MGE Network. Insights into malaria susceptibility using genome-wide data on 17,000
749 individuals from Africa, Asia and Oceania. *Nat. Commun.* **10**, 5732 (2019).
750 13. Ndila, C. M. *et al.* Human candidate gene polymorphisms and risk of severe malaria in
751 children in Kilifi, Kenya: a case-control association study. *Lancet Haematol.* **5**, e333–
752 e345 (2018).
753 14. Band, G. *et al.* A novel locus of resistance to severe malaria in a region of ancient
754 balancing selection. *Nature* **526**, 253–257 (2015).
755 15. Rockett, K. A. *et al.* Reappraisal of known malaria resistance loci in a large
756 multicenter study. *Nat. Genet.* **46**, 1197–1204 (2014).
757 16. Timmann, C. *et al.* Genome-wide association study indicates two novel resistance loci

758 for severe malaria. *Nature* **489**, 443–446 (2012).
759 17. Zámbó, B. *et al.* Decreased calcium pump expression in human erythrocytes is
760 connected to a minor haplotype in the ATP2B4 gene. *Cell Calcium* **65**, 73–79 (2017).
761 18. Li, J. *et al.* GWAS of blood cell traits identifies novel associated loci and epistatic
762 interactions in Caucasian and African-American children. *Hum. Mol. Genet.* (2013)
763 doi:10.1093/hmg/dds534.
764 19. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From
765 Polygenic to Omnigenic. *Cell* vol. 169 1177–1186 (2017).
766 20. MGE Network. Insights into malaria susceptibility using genome-wide data on 17,000
767 individuals from Africa, Asia and Oceania. *Nat. Commun.* **10**, 5732 (2019).
768 21. De Mendonça, V. R. R., Goncalves, M. S. & Barral-Netto, M. The Host Genetic Diversity
769 in Malaria Infection. *J. Trop. Med.* **2012**, 17 (2012).
770 22. Mackinnon, M. J., Mwangi, T. W., Snow, R. W., Marsh, K. & Williams, T. N. Heritability
771 of malaria in Africa. *PLoS Med.* **2**, 1253–1259 (2005).
772 23. Canela-Xandri, O., Rawlik, K. & Tenesa, A. An atlas of genetic associations in UK
773 Biobank. *Nat. Genet.* **50**, 1593–1599 (2018).
774 24. Lo, K. S. *et al.* Genetic association analysis highlights new loci that modulate
775 hematological trait variation in Caucasians and African Americans. *Hum. Genet.* **129**,
776 307–317 (2011).
777 25. Cooling, L. Blood groups in infection and host susceptibility. *Clin. Microbiol. Rev.* **28**,
778 801–870 (2015).
779 26. Whitfield, J. B., Martin, N. G. & Rao, D. C. Genetic and environmental influences on the
780 size and number of cells in the blood. *Genet. Epidemiol.* **2**, 133–144 (1985).
781 27. Evans, D. M., Frazer, I. H. & Martin, N. G. Genetic and environmental causes of
782 variation in basal levels of blood cells. *Twin Res.* **2**, 250–257 (1999).
783 28. Pilia, G. *et al.* Heritability of Cardiovascular and Personality Traits in 6,148
784 Sardinians. *PLoS Genet.* **2**, e132 (2006).
785 29. Chen, M.-H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667
786 Individuals from 5 Global Populations. *Cell* **182**, 1198-1213.e14 (2020).
787 30. Vuckovic, D. *et al.* The Polygenic and Monogenic Basis of Blood Traits and Diseases.
788 *Cell* **182**, 1214-1231.e11 (2020).

789 31. van der Harst, P. *et al.* Seventy-five genetic loci influencing the human red blood cell.
790 *Nature* **492**, 369–75 (2012).
791 32. Chami, N. *et al.* Exome Genotyping Identifies Pleiotropic Variants Associated with
792 Red Blood Cell Traits. *Am. J. Hum. Genet.* **99**, 8–21 (2016).
793 33. Astle, W. J. *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links
794 to Common Complex Disease. *Cell* **167**, 1415-1429.e19 (2016).
795 34. Perry, G. S., Byers, T., Yip, R. & Margen, S. Iron nutrition does not account for the
796 hemoglobin differences between blacks and whites. *J. Nutr.* **122**, 1417–1424 (1992).
797 35. Beutler, E. & West, C. Hematologic differences between African-Americans and
798 whites: The roles of iron deficiency and α-thalassemia on hemoglobin levels and
799 mean corpuscular volume. *Blood* **106**, 740–745 (2005).
800 36. Garn, S. M. Lower Hematocrit Levels in Blacks Are Not Due to Diet or Socioeconomic
801 Factors. *Pediatrics* **67**, (1981).
802 37. Kanias, T. *et al.* Ethnicity, sex, and age are determinants of red blood cell storage and
803 stress hemolysis: Results of the REDS-III RBC-Omics study. *Blood Adv.* **1**, 1132–1141
804 (2017).
805 38. Page, G. *et al.* GWAS of Osmotic Hemolysis in 12,352 Healthy Blood Donors Identifies
806 Red Cell Genetic Variants Associated with Steady State Hemolysis in Patients with
807 Sickle Cell Disease. *Blood* **130**, 1117–1117 (2017).
808 39. Campbell, M. C. & Tishkoff, S. A. African Genetic Diversity: Implications for Human
809 Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annu.*
810 *Rev. Genomics Hum. Genet.* **9**, 403–433 (2008).
811 40. Gallagher, P. G. Abnormalities of the erythrocyte membrane. *Pediatric Clinics of North*
812 *America* vol. 60 1349–1362 (2013).
813 41. Schulman, S. *et al.* Growth of Plasmodium falciparum in human erythrocytes
814 containing abnormal membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 7339–7343
815 (1990).
816 42. Facer, C. A. *Erythrocytes carrying mutations in spectrin and protein 4.1 show differing*
817 *sensitivities to invasion by Plasmodium falciparum*. *Parasitol Res* vol. 81 (1995).
818 43. Dhermy, D., Schrével, J. & Lecomte, M.-C. Spectrin-based skeleton in red blood cells
819 and malaria. *Curr. Opin. Hematol.* **14**, 198–202 (2007).

820 44. Galanello, R. & Cao, A. Alpha-thalassemia. *Genetics in Medicine* vol. 13 83–88 (2011).
821 45. Yoshida, A., Beutler, E. & Motulsky, A. G. Human glucose-6-phosphate dehydrogenase
822 variants. *Bull. World Health Organ.* **45**, 243–253 (1971).
823 46. Clarke, G. M. *et al.* Characterisation of the opposing effects of G6PD deficiency on
824 cerebral malaria and severe malarial anaemia. *Elife* **6**, (2017).
825 47. Cao, A. & Galanello, R. Beta-thalassemia. *Genetics in Medicine* vol. 12 61–76 (2010).
826 48. Glogowska, E. *et al.* Novel mechanisms of PIEZO1 dysfunction in hereditary
827 xerocytosis. *Blood* **130**, 1845–1856 (2017).
828 49. Greene, L. S. G6PD deficiency as protection againstfalciparum malaria: An
829 epidemiologic critique of population and experimental studies. *Am. J. Phys. Anthropol.*
830 **36**, 153–178 (1993).
831 50. Ifediba, T., Stern, A., Ibrahim, A. & Rieder, R. Plasmodium falciparum in vitro:
832 diminished growth in hemoglobin H disease erythrocytes. *Blood* **65**, 452–455 (1985).
833 51. Friedman, M. J. Erythrocytic mechanism of sickle cell resistance to malaria. *Proc. Natl.*
834 *Acad. Sci. U. S. A.* **75**, 1994–1997 (1978).
835 52. Wright, G. J. & Rayner, J. C. Plasmodium falciparum Erythrocyte Invasion: Combining
836 Function with Immune Evasion. *PLoS Pathog.* **10**, (2014).
837 53. Tibshirani, R. & Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R.*
838 *Stat. Soc. Ser. B* **58**, 267--288 (1994).
839 54. Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLoS Med.* **2**, e124
840 (2005).
841 55. Flynn, C. J., Hurvich, C. M. & Simonoff, J. S. On the sensitivity of the Lasso to the
842 number of predictor variables. *Stat. Sci.* **32**, 88–105 (2017).
843 56. Ma, S. *et al.* Common PIEZO1 Allele in African Populations Causes RBC Dehydration
844 and Attenuates Plasmodium Infection. *Cell* **173**, 443-455.e12 (2018).
845 57. Nguetse, C. N. *et al.* A common polymorphism in the mechanosensitive ion channel
846 PIEZO1 is associated with protection from severe malaria in humans. *Proc. Natl.*
847 *Acad. Sci.* (2020) doi:10.1073/pnas.1919843117/-/DCSupplemental.
848 58. Rooks, H. *et al.* A gain of function variant in PIEZO1 (E756del) and sickle cell disease.
849 *Haematologica* vol. 104 e91–e93 (2019).
850 59. Ilboudo, Y. *et al.* A common functional PIEZO1 deletion allele associates with red

851 blood cell density in sickle cell disease patients. *American Journal of Hematology* vol.
852 93 E362–E365 (2018).
853 60. Lessard, S. *et al.* An erythroid-specific ATP2B4 enhancer mediates red blood cell
854 hydration and malaria susceptibility. *J. Clin. Invest.* **127**, 3065–3074 (2017).
855 61. Lin, S.-H., Brown, D. & Machiela, M. LDtrait: an online tool for identifying published
856 phenotype associations in linkage disequilibrium. *Cancer Res.* canres.0985.2020
857 (2020) doi:10.1158/0008-5472.CAN-20-0985.
858 62. Crosnier, C. *et al.* Basigin is a receptor essential for erythrocyte invasion by
859 Plasmodium falciparum. *Nature* **480**, 534–537 (2011).
860 63. Egan, E. S. *et al.* Erythrocytes lacking the Langereis blood group protein ABCB6 are
861 resistant to the malaria parasite Plasmodium falciparum. *Commun. Biol.* **1**, 1–9
862 (2018).
863 64. Egan, E. S. *et al.* A forward genetic screen identifies erythrocyte CD55 as essential for
864 Plasmodium falciparum invasion. *Science (80-.).* **348**, 711–714 (2015).
865 65. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*
866 **550**, 204–213 (2017).
867 66. Goheen, M. M. *et al.* Anemia Offers Stronger Protection Than Sickle Cell Trait Against
868 the Erythrocytic Stage of Falciparum Malaria and This Protection Is Reversed by Iron
869 Supplementation. *EBioMedicine* **14**, 123–130 (2016).
870 67. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation
871 in 141,456 humans. *Nature* **581**, 434–443 (2020).
872 68. Williams, T. N. Human red blood cell polymorphisms and malaria. *Current Opinion in*
873 *Microbiology* vol. 9 388–394 (2006).
874 69. Bejon, P. *et al.* Defining Childhood Severe Falciparum Malaria for Intervention
875 Studies. *PLoS Med.* **4**, e251 (2007).
876 70. Allison, A. C. Protection afforded by sickle-cell trait against subtertian malarial
877 infection. *Br. Med. J.* **1**, 290–294 (1954).
878 71. Luzzatto, L. Sickle cell anaemia and malaria. *Mediterranean Journal of Hematology*
879 *and Infectious Diseases* vol. 4 2012065 (2012).
880 72. Okwa, O. O. *Malaria Parasites*. *Malaria Parasites* (InTech, 2012). doi:10.5772/1477.
881 73. Pengon, J., Svasti, S., Kamchonwongpaisan, S. & Vattanaviboon, P. Hematological

882 parameters and red blood cell morphological abnormality of Glucose-6-Phosphate
883 dehydrogenase deficiency co-inherited with thalassemia. *Hematol. Oncol. Stem Cell*
884 *Ther.* **11**, 18–24 (2018).
885 74. Clark, M., Mohandas, N. & Shohet, S. Osmotic gradient ektacytometry: comprehensive
886 characterization of red cell volume and surface maintenance. *Blood* **61**, 899–910
887 (1983).
888 75. Mockenhaupt, F. P. *et al.* Anaemia in pregnant Ghanaian women: importance of
889 malaria, iron deficiency, and haemoglobinopathies. *Trans. R. Soc. Trop. Med. Hyg.* **94**,
890 477–483 (2000).
891 76. Tiffert, T. *et al.* The hydration state of human red blood cells and their susceptibility
892 to invasion by Plasmodium falciparum. *Blood* **105**, 4853–4860 (2005).
893 77. Kariuki, S. N. *et al.* Red blood cell tension controls Plasmodium falciparum invasion
894 and protects against severe malaria in the Dantu blood group. *bioRxiv* 475442 (2018)
895 doi:10.1101/475442.
896 78. Field, S. P., Hempelmann, E., Mendelow, B. V. & Fleming, A. F. Glycophorin variants
897 and Plasmodium falciparum: protective effect of the Dantu phenotype in vitro. *Hum.*
898 *Genet.* **93**, 148–150 (1994).
899 79. Bei, A. K. *et al.* A flow cytometry-based assay for measuring invasion of red blood
900 cells by *Plasmodium falciparum*. *Am. J. Hematol.* **85**, 234–237 (2010).
901 80. Awandare, G. A., Nyarko, P. B., Aniweh, Y., Ayivor-Djanie, R. & Stoute, J. A.
902 Plasmodium falciparum strains spontaneously switch invasion phenotype in
903 suspension culture. *Sci. Rep.* **8**, 1–10 (2018).
904 81. Dolan, S. A., Miller, L. H. & Wellems, T. E. Evidence for a switching mechanism in the
905 invasion of erythrocytes by Plasmodium falciparum. *J. Clin. Invest.* **86**, 618–624
906 (1990).
907 82. Boyle, M. J. *et al.* Isolation of viable Plasmodium falciparum merozoites to define
908 erythrocyte invasion events and advance vaccine and drug development. *Proc. Natl.*
909 *Acad. Sci. U. S. A.* **107**, 14378–14383 (2010).
910 83. Chen, R., Davydov, E. V., Sirsota, M. & Butte, A. J. Non-synonymous and synonymous
911 coding SNPS show similar likelihood and effect size of human disease association.
912 *PLoS One* **5**, (2010).

913 84. Sauna, Z. E. & Kimchi-Sarfaty, C. Understanding the contribution of synonymous
914 mutations to human disease. *Nature Reviews Genetics* vol. 12 683–691 (2011).
915 85. Kinsler, G., Geiler-Samerotte, K. & Petrov, D. A genotype-phenotype-fitness map
916 reveals local modularity and global pleiotropy of adaptation. *bioRxiv*
917 2020.06.25.172197 (2020) doi:10.1101/2020.06.25.172197.
918 86. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants
919 across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
920 87. Sohail, M. *et al.* Polygenic adaptation on height is overestimated due to uncorrected
921 stratification in genome-wide association studies. *Elife* **8**, (2019).
922 88. Otto, T. D. *et al.* Genomes of all known members of a Plasmodium subgenus reveal
923 paths to virulent human malaria. *Nat. Microbiol.* **3**, 687–697 (2018).
924 89. Sundararaman, S. A. *et al.* Genomes of cryptic chimpanzee Plasmodium species reveal
925 key evolutionary events leading to human malaria. *Nat. Commun.* **7**, (2016).
926 90. Pankratov, V. *et al.* Differences in local population history at the finest level: the case
927 of the Estonian population. *Eur. J. Hum. Genet.* 1–12 (2020) doi:10.1038/s41431-020-
928 0699-4.
929 91. Donovan, B. M. Learned inequality: Racial labels in the biology curriculum can affect
930 the development of racial prejudice. *J. Res. Sci. Teach.* **54**, 379–411 (2017).
931 92. Sparks, R. A., Baldwin, K. E. & Darner, R. Using Culturally Relevant Pedagogy to
932 Reconsider the Genetics Canon †. *J. Microbiol. Biol. Educ.* **21**, (2020).
933 93. Rosenberg, N. A. A Population-Genetic Perspective on the Similarities and Differences
934 among Worldwide Human Populations. *Hum. Biol.* **83**, 659–684 (2011).
935 94. Coop, G. *et al.* The Role of Geography in Human Adaptation. *PLoS Genet.* **5**, e1000500
936 (2009).
937 95. Hofer, T., Ray, N., Wegmann, D. & Excoffier, L. Large allele frequency differences
938 between human continental groups are more likely to have occurred by drift during
939 range expansions than by selection. *Ann. Hum. Genet.* **73**, 95–108 (2009).
940 96. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
941 MEM. *arXiv* (2013).
942 97. Van der Auwera, G. A. *et al.* From fastQ data to high-confidence variant calls: The
943 genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinforma.* **11**, 11.10.1

944 (2013).
945 98. Mills, R. E. *et al.* An initial map of insertion and deletion (INDEL) variation in the
946 human genome. *Genome Res.* **16**, 1182–1190 (2006).
947 99. Weatherall, D. J. Phenotype-genotype relationships in monogenic disease: Lessons
948 from the thalassaemias. *Nature Reviews Genetics* vol. 2 245–255 (2001).
949 100. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: Functional annotation of genetic
950 variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, (2010).
951 101. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158
952 (2011).
953 102. The 1000 Genomes Project Consortium. A global reference for human genetic
954 variation. *Nature* vol. 526 68–74 (2015).
955 103. Lowy-Gallego, E. *et al.* Variant calling on the GRCh38 assembly with the data from
956 phase three of the 1000 Genomes Project. *Wellcome Open Res.* **4**, 50 (2019).
957 104. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-
958 based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
959 105. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for
960 individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).
961 106. Park, D. J. *et al.* Sequence-based association and selection scans identify drug
962 resistance loci in the Plasmodium falciparum malaria parasite. *Proc. Natl. Acad. Sci. U.*
963 *S. A.* **109**, 13052–13057 (2012).
964 107. Kuypers, F. A., Scott, M. D., Schott, M. A., Lubin, B. & Chiu, D. T. Y. Use of ektacytometry
965 to determine red cell susceptibility to oxidative stress. *J. Lab. Clin. Med.* **116**, 527–534
966 (1990).
967 108. Chatterjee, S. Assumptionless consistency of the Lasso. *arXiv* (2013).
968