Genome-wide heritability analysis of severe malaria susceptibility and resistance reveals evidence of polygenic inheritance

Delesa Damena¹ and Emile R. Chimusa^{1*}

¹Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

ABSTRACT

Objective: Estimating SNP-heritability (h^2_g) of severe malaria/resistance and its distribution across the genome might shed new light in to the underlying biology.

Method: We investigated h^2_g of severe malaria susceptibility and resistance from genome-wide association study (GWAS) dataset (sample size =11, 657). We partitioned the h^2_g in to chromosomes, allele frequencies and annotations. We further examined none-cell type specific and cell type specific enrichments from GWAS-summary statistics.

Results: We estimated the h^2_g of severe malaria at 0.21 (se=0.05, p=2.7x10⁻⁵), 0.20 (se=0.05, p=7.5x10⁻⁵) and 0.17 (se=0.05, p= 7.2x10⁻⁴) in Gambian, Kenyan and Malawi populations, respectively. The h^2_g attributed to the GWAS significant SNPs and the well-known sickle cell (*HbS*) variant was approximately 0.07 and 0.03, respectively. We prepared African population reference panel and obtained comparable h^2_g estimate (0.21 (se = 0.02, p< 1x10⁻⁵)) from GWAS-summary statistics meta-analysed across the three populations. Partitioning analysis from raw genotype data showed significant enrichment of h^2_g in protein coding genic SNPs while summary statistics analysis suggests pattern of enrichment in multiple categories.

Conclusion: We report for the first time that the heritability of malaria susceptibility and resistance is largely ascribed by common SNPs and the causal variants are overrepresented in protein coding regions of the genome. Overall, our results suggest that malaria susceptibility and resistance is a polygenic trait. Further studies with larger sample sizes are needed to better understand the underpinning genetics of resistance and susceptibility to severe malaria.

Key words: GWASs, SNP-heritability, severe malaria, resistance, susceptibility

INTRODUCTION

Malaria remains one of the leading public health problems worldwide. The global tally of malaria in 2017 was estimated at 219 million cases and 435 000 deaths worldwide (1). *P. falciparum* malaria results in diverse clinical manifestations ranging from asymptomatic parasitaemia to severe malaria (2,3). Family studies reported that the host genetic factors (heritability) contributes about 25% of the variations observed in clinical severity of malaria (4). Heritability (H^2) is defined as the proportion of the phenotypic variations attributable to genetic differences in a population (5). Narrow-sense heritability (h^2) is a specific term that refers to the phenotypic variations attributable to the additive genetic values. Quantifying the proportion of narrow-sense heritability attributed to common variants from GWAS SNPs without identifying the causal variants, commonly called SNP-heritability (h^2_g), has recently gained a considerable interest in statistical genetics to address the problem of ‘missing’ heritability in GWASs (6).

Several statistical methods now allow accurate estimation of h^2_g from all SNPs of unrelated individuals (7–9). Methods such as the Genome-wide Complex Trait Analysis (GCTA) compute genetic relationships matrix (GRMs) of unrelated subjects by simultaneously fitting all SNPs and estimates the proportion of phenotypic variations explained by genotypic variation using restricted maximum likelihood (GREML) methods (7). Phenotype-Correlation-Genotype-Correlation (PCGC) is a one of recently developed methods specifically designed to estimate h^2_g from case/control (9). PCGC is a Haseman-Elston (H-E) regression model in which normalized phenotype is regressed on the genetic covariances of all unique pair of samples. The slope of the regression is used as an estimator of SNP-heritability. Application of these methods provided new insights in to the genetic architecture of complex disease genetic architecture (10, 11). Alternative contemporary statistical methods that enable estimation of h^2_g from publicly available GWAS-summary statistics without the need of individual genotype data are also widely available (12–14) and gained popularity due to their privacy advantages and computational costs.

Even though a number of severe malaria GWASs have recently been implemented in malaria endemic areas in Africa and reported a number novel variants (15–18), little is known about the proportion of h^2_g and its distribution across chromosomes, functional regions and allele frequency spectrum. We hypothesize that malaria susceptibility and resistance is a complex trait that is mainly under polygenic control. Here we present results from comprehensive SNP-

heritability studies of severe malaria in three African populations including Gambia, Kenya and Malawi. We estimated h^2_g of severe malaria and partitioned in to chromosomes, different minor allele frequency (MAF) bins annotations and cell-types. Overall, our results suggest that malaria susceptibility and resistance is a polygenic trait.

RESULTS

SNP-heritability of severe malaria susceptibility and resistance from genotype data

We estimated SNP-heritability at different quality control (QC) levels to determine the appropriate threshold (see Materials and Methods). As expected, the estimates were inflated when less stringent QC thresholds were applied (**Supplementary Figure 1**). However, when more stringent sets of QCs including relatedness threshold (5%), SNP differential missingness proportion ($p \leq 1 \times 10^{-3}$) and SNPs missing proportion ($p > 0.02$) were applied, the estimates became stable in each population (Gambia, Kenya and Malawi). Inclusion of more PCs (15, 20, 50) as a fixed effect didn't further decrease these estimates. At the stringent QC threshold, the SNP-heritability of severe malaria was 0.21 (se=0.05, $p=2.7 \times 10^{-5}$), 0.20 (se=0.05, $p=7.5 \times 10^{-5}$) and 0.17 (se=0.05, $p=7.2 \times 10^{-4}$) in Gambian, Kenyan and Malawi populations, respectively (**Table 1**). The SNP-heritability was approximately similar in major ethnic groups including Mandinka (0.24, se=0.06, $p=5.1 \times 10^{-5}$) in Gambia and Chonye (0.20, se=0.07, $p=5.1 \times 10^{-3}$) and Girimia (0.19, se=0.07, $p=7.3 \times 10^{-3}$) in Kenya. We didn't estimate h^2_g for other ethnic groups because of smaller sample sizes. Furthermore, the PCGC model approximately showed similar results: 0.20, (se=0.06, $p=9.7 \times 10^{-4}$), 0.16 (se=0.06, $p=8 \times 10^{-3}$) and 0.23 (se=0.07, $p=1.3 \times 10^{-3}$) in Gambia, Kenya and Malawi populations, respectively.

Table 1. SNP-heritability of severe malaria susceptibility and resistance determined by GCTA and PCGC methods.

Population	SNPs(n)	Samples (n)	h^2_g GCTA (%)	h^2_g PCGC (%)
Gambia	1486554	4128	0.20(se=0.05)	0.20(se=0.05)
Kenya	1579227	2062	0.20(se=0.05)	0.16 (se=0.05)
Malawi	1502462	2418	0.17(se=0.05)	0.23(se=0.06)
Mandinka	1486554	1281	0.24 (se=0.06)	ne*
Chonye	1579227	637	0.20(se=0.06)	ne*
Girima	1579227	1173	0.19 (se=0.06)	ne*

ne*: Model didn't fit because of small sample size and there was no reliable estimation SNP: Single Nucleotide Polymorphisms, h^2_{gGCTA} :

SNP-heritability estimated using GCTA method, h^2_{gPCGC} : SNP-heritability estimated using PCGC method

Heritability explained by GWAS significant SNPs

To quantify the effects of the known variants, we estimated h^2_g without removing malaria associated loci (see Materials and Methods). This resulted in slight increase of the estimate in Gambian (0.27 (se = 0.05, $p = 1 \times 10^{-5}$)) and Kenyan (0.24 (se = 0.05, $p < 1 \times 10^{-5}$)) populations. Including **rs334** as an additional covariate decreased the estimate to (0.24 (se = 0.05, $p < 1 \times 10^{-5}$)), 0.21 (se = 5%, $p < 1 \times 10^{-5}$) in Gambian and Kenyan populations, respectively. Suggesting that the h^2_g attributed to the significant SNPs and *HbS* locus is about 0.07 and 0.03, respectively across the populations.

Partitioning SNP-heritability by chromosomes and minor allele frequencies

We observed no significant differences in between the h^2_g per chromosome obtained from separate analysis (0.24, se = 0.05, $p < 1 \times 10^{-5}$) and that obtained from the joint analysis (0.20, se = 0.05, $p = 1 \times 10^{-5}$) as shown in **Supplementary Figure 2**, suggesting that the population structure was adequately controlled. Moreover, we observed significant correlations between chromosomal length and h^2_g per chromosome (Adj $r^2 = 0.38$, $p = 0.001$) as shown in **Figure 1**. However, the estimates of three chromosomes (chr 5, 11 and 20) and three other chromosomes (chr 7, 8 and 15) fell above and below the expected h^2_g at 95% CI, respectively. Notably, chr 5 contain a considerable proportion (15%) of the cumulative h^2_g estimate. To determine the relative contribution of variants from various allele frequencies, we partitioned SNPs in to different MAF bins and estimated h^2_g attributed to each bin using joint GRMEL analysis (**Figure 2**). However, we didn't detect significant differences between the proportion of h^2_g attributed to different MAF bins (standard errors overlapped at 95% CI). The total sum of h^2_g across bins (0.27 (se = 0.08, $p = 5.3 \times 10^{-5}$)) was not significantly different from the univariate estimate.

Figure 1. SNP-heritability (h^2_g) per chromosome(y-axis) plotted against chromosome length (x-axis). The blue line represents the h^2_g estimates regressed against chromosome length. The grey shaded areas represent the 95% confidence interval around the slope of the regression model.

Figure 2. SNP-heritability partitioned in to different allele frequency spectrum. We created six MAF-bins and estimated the proportion of h^2_g attributed to each bin. The proportion of h^2_g attributed to each bin is shown in red bar and the proportion of SNPs per bin is shown by blue bar. Error bars represent the 95% CI of the estimate.

Partitioning SNP-heritability by annotation

We estimated the h^2_g explained by genic SNPs and intergenic SNPs at 0.163 (se=0.06) and 0.064 (se=0.05), respectively (**Table 2**). This implies that the proportion of heritability explained by a SNP residing in a gene within 10kb boundary is on average twice as the heritability explained by a SNP residing in an intergenic region. However, the statistical inferences can't be made as standard errors overlapped at 95% CI. We further calculated the h^2_g explained by protein coding genic SNPs and the corresponding intergenic SNPs at 0.165 (se = 0.05) and 0.062 (se = 0.05), respectively. A SNP residing in protein coding genes within 10kb boundary contribute about $2.9 \times h^2_g$ than a SNP residing in an intergenic region and this is statistically significant at 95% confidence interval.

Table 2. SNP-heritability of severe malaria susceptibility and resistance partitioned in to genic and intergenic genomic regions.

SNP location	10 kb boundary		
	SNPs(n)	h^2_g	h^2_g per SNP
Genic	827928	0.163 (se=0.06)	1.9×10^{-5}
Inter-genic	685896	0.064 (se=0.05)	9×10^{-6}
Genic-coding	727996	0.165(se=0.05)	2.3×10^{-5}
Inter-genic	785826	0.062(se=0.05)	7.9×10^{-6}

Functional enrichment from GWAS-summary statistics

After imputation from malaria GWAS-summary statistics and QC filtering, we obtained a total of 20 million high quality SNPs (see Materials and Methods). Using this dataset, we estimated the liability scale SNP-heritability at 0.21 (se = 0.02, $p < 1 \times 10^{-5}$). Partitioning the h^2_g in to 24 main genomic annotations (baseline model) showed patterns of enrichment in multiple categories including 5'UTR (11x), DGF (10x), enhancer (9x), coding (6x), H3K4me1 (4.9x), TSS (5x), TFBS (4x), and FANTOM enhancer (4x) as shown in **Figure 3**. However, none of the enrichments was statistically significant after correction for multiple testing. Further cell-type specific and cell group analysis didn't show significant enrichment.

Figure 3. Enrichment estimates of severe malaria susceptibility and resistance for the 24 main annotations. Error bars represent jackknife standard errors around the estimates of enrichment

DISCUSSIONS

In this study, we estimated the SNP-heritability and functional enrichment of malaria susceptibility and resistance in three African populations and their meta-analysis. We performed GRMEL analysis at different QC levels to determine the appropriate threshold; indeed, the estimates were inflated upward at less stringent QC levels and became stable at more stringent QC levels. At the stringent QC threshold, the h^2_g of severe malaria susceptibility was 0.21 (se=0.05, $p=2.7 \times 10^{-5}$), 0.20 (se =0.05, $p=7.5 \times 10^{-5}$) and 0.17 (se = 0.05, $p = 7.2 \times 10^{-4}$) in Gambian, Kenyan and Malawi populations, respectively. The estimates were approximately similar across major ethnic groups in the study populations. Application of PCGC method in our dataset yielded comparable estimates including 0.20, (se=0.06, $p=9.7 \times 10^{-4}$), 0.16 (se = 0.06, $p = 8 \times 10^{-3}$) and 0.23 (se =0.07, $p=1.3 \times 10^{-3}$) in Gambia, Kenya and Malawi populations, respectively. Despite slight differences, our h^2_g estimate was approximately the same in different populations, arguing that the allele heterogeneity of the malaria susceptibility/resistance and their effect sizes (3) might not substantially affect the overall SNP-heritability across endemic populations. Replication

studies with larger population specific sample sizes might be needed to further investigate this observation.

Our h^2_g estimates was roughly close to a previous report from a family-based study in which the narrow-sense heritability estimates of severe malaria was estimated to be 0.25 (4). The same study estimated the contribution of the well-known *HbS* variant at 0.025 which is also consistent with our estimate (0.03). Suggesting that the malaria susceptibility/resistance is influenced by several variants each individually having a small effect sizes in population and thus, it is a polygenic trait. This is consistent with the hypothesis that the vast proportion of heritability of complex traits is attributed to SNPs with effect sizes too small to attain the stringent genome wide significance threshold ($p < 5 \times 10^{-8}$) at the current sample sizes (19). To gain better insights in to the genetic architecture of severe malaria, we partitioned the h^2_g in to different chromosomes, allele frequency spectrum and annotations using joint analysis implemented in GCTA software (7). We found no significant difference between the separate analysis (0.25, se = 0.05, $p < 1 \times 10^{-5}$) and joint analysis (0.24, se=0.05, $p = 1 \times 10^{-5}$), suggesting that the population structure is adequately controlled. The rationale is that separate analysis in which one chromosome is fitted to the model at a time, captures variances due to correlated SNPs on other chromosomes and thus, yield inflated estimates in the presence of population structure. However, the joint analysis in which GRMs of all chromosomes are simultaneously fitted in to a single GRMEL, effectively controls upward biases that can be created from correlated SNPs on different chromosomes and thus, provide more accurate estimates (29).

Supporting the polygenic view of genetic architecture, we found a correlation between h^2_g per chromosome and chromosomal lengths (Adj $r^2 = 0.39$, $p = 0.001$). However, the estimates of three chromosomes (chr 5, 11 and 20) and three other chromosomes (chr 7, 8 and 15) fell below above and below the expected h^2_g at 95% CI. Notably, chr 5 contain a considerable proportion 15% (0.036) of the cumulative h^2_g estimate, suggesting that this chromosome might contain loci with larger effects against the polygenic background. Previous family-based studies reported the association of 5q31-q33 with mild malaria susceptibility and resistance (20, 21).

MAF-stratified analysis didn't reveal significant differences between the proportion of h^2_g attributed to different MAF bins (standard errors overlapped at 95% CI). This might assert that h^2_g of severe malaria is broadly uniform across the allele frequency spectrum and is not over-represented by rare alleles. Moreover, the total sum of h^2_g in MAF-stratified analysis (0.27 (se = 0.08, $p = 5.3 \times 10^{-5}$)) was not significantly different from the cumulative univariate

h^2_g estimate, further suggesting that allelic heterogeneity of malaria protective variants in different populations might not have substantial effects on the overall SNP-heritability. However, more powered studies might be needed to replicate this observation.

Partitioning by annotation revealed that the h^2_g of severe malaria susceptibility and resistance is concentrated in genic regions of the genome. The proportion of heritability explained by a SNP in genes within 10kb boundary is on average twice as the heritability explained by a SNP in intergenic regions. On average, a SNP in protein coding genes within 10kb boundary contribute about $2.9 \times h^2_g$ than a SNP in intergenic regions. The later was statistically significant at 95% CI. This finding is arguing against the notion that SNP-heritability of complex traits is largely attributed to SNPs residing in none coding regions of the genome.

In addition to the direct estimation from the raw genotype, functional enrichment analysis from GWAS-summary statistics using stratified LDSC (12) method has recently been shedding new lights in to complex disease genetic architectures (22). However, this approach require population specific reference panel to produce reliable results making it difficult to apply in African populations which has been poorly represented in public reference domain (12). In effort to address this challenge, we prepared African population specific reference panel for the first time which can be obtained from <http://web.cbio.uct.ac.za/~emile/software.html>. Our liability scale SNP-heritability estimate from GWAS-summary statistics meta analysed across the study populations (0.21 ($se = 0.02$, $p < 1 \times 10^{-5}$)) was comparable to the direct estimates from raw genotype data. However, our full base line model and cell specific didn't reveal significant enrichments. Of note, the coding and regions surround coding genes were among the top categories in our base line model, further implying the importance of genic regions of the genome in influencing the malaria susceptibility and resistance trait. One of the downsides of stratified LDSC analysis is that it requires very large sample sizes to detect significant enrichments. We strongly believe that replication studies with large sample sizes can shed new lights in to the genetic architecture of severe malaria susceptibility and resistance.

CONCLUSIONS

In conclusion, our study showed for the first time that heritability of malaria susceptibility and resistance is largely explained by common SNPs and is disproportionately enriched in protein coding regions of the genome. Consistent with polygenic genetic architecture, we observed broadly even distribution of SNP-heritability across chromosomes and different allele

frequency spectrum, implying that very large sample sizes will be needed to identify novel variants. We prepared African population specific reference panel and showed that LDSC analysis can provide reliable SNP-heritability estimates in African populations. Further studies with larger sample sizes are needed to understand the unpinning genetic architecture of severe malaria susceptibility/resistance phenotype.

METHODS

Description of the study datasets

GWAS data of three African populations including Gambia, Kenya and Malawi were obtained from European Phenome Genome Archive (EGA) through the MalariaGen consortium standard data access protocols (23, 24). The dataset contains information about a total of 11, 657 samples including 4921 samples from Gambia (2491 cases and 2430 controls), 3752 samples from Malawi (3752 cases and 13220 controls), 2984 samples from Kenya (1506 cases and 1478 controls). Cases were obtained from children who were admitted to Hospitals and fulfilled WHO case definition for severe malaria (24) and controls were obtained from the general population (15–18). All the samples were genotyped on Illumina Omni 2.5M array and were provided in VCF format. Information about phenotypes, imputation and quality control (QC) was provided with the data.

Quality control

The basic QC analyses including plate effects, sample relatedness, Hardy-Weinberg equilibrium, heterozygosity, missingness and differential missingness were conducted for the previously published GWASs as described elsewhere (15, 25). Taking in to consideration that small artifacts can have substantial cumulative effects in heritability analysis (26), we applied further stringent QC protocols. Briefly, we aligned the quality filtered VCF files to the forward strand of the human reference sequence (GRCh37), using the Illumina-supplied files (www.well.ox.ac.uk/~wrayner/strand) and removed all SNPs with position and strand mismatches. We further removed SNPs with minor allele frequency (MAF) below 0.01, deviate from Hardy-Weinberg at p-value below 0.01 using PLINK software (27). We then implemented step-wise QC control based on SNPs missingness proportion, differential missingness and sample relatedness.

Estimating heritability from genotype data

We applied GCTA (7) to estimate the SNP-heritability of severe malaria susceptibility and resistance from raw genotype datasets. Briefly, we excluded the region of extended inversion (723,8552-12,442,658) on chromosome 8p23 (28), the Major Histocompatibility Complex (MCH) region (25,000,000-40,000,000) on chr6 and the known malaria susceptibility associated loci including the *ATP2B4* region on chr1:203,154,024-204,154,024, cluster of *glycophorin* (*GYP A/B/E*) region on chr4:143,000,000-146,000,000, *ABO* blood group region on chr9:135,630,000-136,630,000, and the sickle cell (*HbS*) region on chr11:2,500,000-6,500,000 to avoid potential biases.

We constructed Genomic Relatedness Matrix (GRMs) from pruned high quality independent autosomal SNPs (independent pairwise 50 10.2) using GCTA software (7) and obtained list of samples with relatedness threshold above 5%. We then computed GRMs using all SNPs for each cohort and excluded one of any pair of samples with relatedness threshold above 5% as recommended elsewhere (8). The final sample of unrelated individuals was 4128, 2062 and 2418 for Gambia, Kenya and Malawi, respectively. The distribution of off-diagonal element of the GRMs for each population is shown in **supplementary Figure 3**.

We estimated SNP-heritability (h^2_g) using GCTA software (7). We performed principal components analysis (PCA) and included the top 10 PCs as fixed effects to account for population structure. We used population prevalence of 1% of severe malaria as previously described in (24). We transformed the estimates to liability scale as outlined in Lee *et al.* (29). Using the same GRMs, we estimated the h^2_g using PCGC model as outlined in Golan *et al.* (12). We also computed separate GRMs for ethnic groups including Mandinka of Gambia and, Girmia and Chonye of Kenya and estimated the h^2_g attributed to each ethnic group. To quantify the effects of significant SNPs, we created GRMs without excluding the association loci and performed GRMEL analysis. We repeated the analysis by including **rs334** as additional covariate to estimate the h^2_g attributed to *HbS*.

Partitioning SNP-heritability from genotype data

Using Gambian dataset (largest sample size), we partitioned h^2_g by chromosomes, MAF bins and annotations. To investigate the biases that might be created by population structure, we computed GRMs for individual chromosome and estimated h^2_g attributed to each chromosome by separate GREML in which one chromosome is fitted to the model at a time. We then performed a joint analysis in which GRMs of all chromosomes are simultaneously fitted in to

a single GREML analysis and compared the results obtained from both analyses. We further created five MAF bins including $>0.01-0.05$, $>0.05-0.1$, $>0.1-0.2$, $>0.2-0.3$, $>0.3-0.4$, $>0.4-0.5$, constructed separate GRMs for each bin and estimated h^2_g attributed to each category using the joint analysis implemented in GCTA software (7). Finally, we partitioned the SNP-heritability in to genic and intergenic regions of the genome. Briefly, we mapped all the autosomal SNPs to the reference genome using QCTOOLV2 (<https://www.well.ox.ac.uk/~gav/qctool>). We labelled all SNPs that are mapped to genes as genic and the remaining SNPs with no annotation as intergenic. We defined gene boundaries as regions within 10 kilobases (kb) regions upstream and downstream of a gene based on UCSC Genome Browser hg19 assembly annotation databases (<http://genome.ucsc.edu>). We estimated the h^2_g explained by the respective category using joint analysis in GCTA software (7).

Functional enrichment analysis of SNP-heritability from GWAS-summary statistics

Preparation of African-specific reference panel

Partitioning h^2_g in to cell-types and functional categories using stratified Linkage Disequilibrium score regression (LDSC) approach has recently been shedding new lights in to the genetic architecture of several complex diseases (12, 22, 30). The method is based on the fact that a given category of SNPs is enriched for h^2_g if SNPs with high LD to that category have higher χ^2 statistics than SNPs with low LD to that category (22, 30). However, the LDSC analysis require population specific reference panel and very large sample sizes to produce reliable results (12). Consequently, the current European 1000G haplotype reference panel that is used as a default in LDSC software(12, 30) might not well-represent our study populations. To address this challenge, we first merged African population datasets obtained from 1000 Genomes Project and African Genome Variation Project (31) based on overlapped variants and removed structural variants and ambiguous SNPs using plink tool (27). This resulted in a merged dataset of sample size ($n = 4975$). From this dataset, we excluded admixed populations including Americans of African Ancestry and African Caribbean. By taking in to consideration the high genetic diversity of African populations, we clustered the dataset into sub-regions comprised of eastern and western Africans using smart pca software (32) as shown in **Supplementary Figure 4**. We performed QC filtering based on $MAF < 1\%$, missingness test ($GENO > 0.05$) and HWE test in controls (alpha level 0.0001). We retained a total of 22,473,268 SNPs and 18,919,068 SNPs in eastern (Sample size = 2,112) and western (Sample size = 380) African sub-region datasets. We eventually calculated the MAF of the

panel for later partitioning analysis. Owing to the fact that our study populations are comprised of both east African (Malawi and Kenya) and west Africa (Gambian) populations, we utilized our entire dataset as a reference panel.

Baseline model and functional annotations

We created baseline model and cell type specific annotations for our reference panel as described in (12). The baseline-LD model included 24 main annotations that are not cell-type specific including coding, UTRs (3'UTR and 5'UTR), promoter and intronic regions obtained from UCSC genome browser and processed by Gusev *et al.* (10), the histone marks (H3) such as: acetylation of histone at lysine 9 (H3K9ac), monomethylation (H3K4me1) and trimethylation (H3K4me3) of H3 at lysine 4 obtained from Trynka *et al.* (33), acetylation of H3 at lysine 27(H3K27ac) version one processed by Hnisz *et al.* (34) and version two Psychiatric Genomics Consortium (PGC), combined chromHMM and Segway predictions obtained from Hoffman *et al.* (35), regions that are conserved in mammals (36, 37), super enhancers (34), FANTOM5 enhancers (38), transcription factor binding sites (TFBS) and digital genomic footprint (DGF) post-processed by Gusev *et al.* (10). Around each category, we added 500bp windows as separate categories to prevent biases that might arise from adjacent categories. The 24 main annotations together with the additional windows and a category containing all SNPs yielded 53 overlapping baseline model. Next, we created 220 cell-type-specific annotations for the four histone marks: H3K4me1, H3K4me3, H3K9ac, and H3K27ac (12) using our reference panel and computed LD score for each annotation. We then combined the 120 cell specific annotations in to 10 cell groups including adrenal and pancreas, central nervous system (CNS), cardiovascular, connective and bone, gastrointestinal, immune and hematopoietic, kidney, liver, skeletal muscle, and other as described in (12). For each of the 10 categories, we computed the corresponding LD scores.

Application LDSC analysis to malaria susceptibility GWAS-summary statistics

We obtained meta-analysed GWAS-summary statistics of the three populations (n =15122094 SNPs) from the previous GWAS (15). We performed imputation on this dataset using ImpG software (39). Briefly, we removed SNPs that mismatch with 1000G phase three markers, computed z-score from the association statistics and performed the imputation using ImGv.1.1 under default settings. We used all 661 individuals labelled as “AFRICAN” haplotypes in phase 1 of 1000 Genome Project version-3 calls (40). We removed all imputed SNPs with a predicated accuracy less than 0.9 and SNPs with MAF < 0.01. After QC filtering, we performed

stratified LD score regression analysis using our reference panels as described in (12). Briefly, we converted the summary statistics to LDSC format, filtered SNPs with imputation accuracy greater than nine and MAF greater than 1%, removed structural variants, ambiguous SNPs, the MHC region and significant SNPs. We then performed none-cell type and cell type specific analyses as described in (12).

ACKNOWLEDGEMENTS

We thank Kwiatkowski's group from University of Oxford for their constructive comments and assistance. We thank Gavin Band for his supervision and guidance in heritability analysis from genotype data. This work was supported through the DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

FUNDING

DD is a PhD student funded by The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) program (grant #PD00217ML). EC is funded by NIH projects

DECLARATION OF INTEREST

The authors declare that they have no competing interests.

AUTHORS CONTRIBUTIONS

DD designed, performed the data analysis and drafted the manuscript, EC contributed in designing, data-analysis and revision of the manuscript and supervised the work.

REFERENCES

1. World Health Organization (2018) World malaria report
2. World Health Organization (2014) Severe Malaria. *Trop. Med & Int Health.*, **19**, 7–131.
3. Kwiatkowski,D.P. (2005) How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. *Am. J. Hum. Genet.*, **77**, 171–192.
4. Mackinnon,M.J., Mwangi,T.W., Snow,R.W., Marsh,K. and Williams,T.N. (2005) Heritability of malaria in Africa. *PLoS Med.*,**12**, e340 doi:10.1371/journal.pmed.0020340.
5. Tenesa,A. and Haley,C.S.(2013) The heritability of human disease : estimation , uses and abuses. *Nat. Genet.*,**14**, 139-149, doi:10.1038/nrg3377.
6. Manolio,T.A., Collins,F.S., Cox,N.J., Goldstein,D.B., Hindorff,L.A., Hunter,D.J., McCarthy,M.I., Ramos,E.M., Cardon,L.R., Chakravarti,A., *et al.* (2009) Finding the missing heritability of complex diseases. *Nature.*, **461**, 747–753.
7. Yang,J., Lee,S.H., Goddard,M.E. and Visscher,P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
8. Speed,D., Cai,N., Johnson,M.R., Nejentsev,S. and Balding,D.J. (2017) Reevaluation of SNP heritability in complex human traits. *Nat. Genet.*, **49**, 986–992.
9. Golan,D., Lander,E.S. and Rosset,S. (2014) Measuring missing heritability: Inferring the contribution of common variants. *Proc. Natl. Acad. Sci.*, **111**, 5272–5281.
10. Gusev,A., Lee,S.H., Trynka,G., Finucane,H., Vilhja,B.J., Xu,H., Zang,C., Ripke,S., Bulik-sullivan,B., Stahl,E., *et al.* (2014) Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *Am. J. Hum. Genet.*, **95**, 535–552 doi:10.1016/j.ajhg.2014.10.004.
11. The Brian Consortium. (2018) Analysis of shared heritability in common disorders of the brain. *Sci.*,**360**, 8757- 8769
12. Finucane,H.K., Bulik-sullivan,B., Gusev,A. and Trynka,G. (2015) Partitioning heritability by functional category using GWAS summary statistics. *Nat Genet.*, **47**, 1228–1235. doi:10.1038/ng.3404.
13. Weissbrod,O., Flint,J. and Rosset,S. (2018) Estimating Heritability and GeneticCorrelation in Case Control Studies Directly and with Summary Statistics. *Am. J. Hum. Genet.*, **103**,89-99. doi: 10.1016/j.ajhg.2018.06.002.
14. Speed,D. and Balding,D.J.(2019) SumHer better estimates the SNP heritability of complex traits from summary statistics complex traits from summary statistics. *Nat. Genet.*,**51**, 277–284.
15. Band,G., Rockett,K.A., Spencer,C.C.A., Kwiatkowski,D.P., Band,G., Si Le,Q., Clarke,G.M., Kivinen,K., Leffler,E.M., Rockett,K.A., *et al.* (2015) A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature.*, **526**, 253–257.
16. Timmann,C., Thye,T., Vens,M., Evans,J., May,J., Ehmen,C., Sievertsen,J., Muntau,B., Ruge,G., Loag,W., *et al.* (2012) Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature.*, **489**, 443–446.
17. Jallow,M., Teo,Y.Y., Small,K.S., Rockett,K.A., Deloukas,P., Clark,T.G., Kivinen,K., Bojang,K.A., Conway,D.J., Pinder,M., *et al.* (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet.*, **41**, 657–665. doi:10.1038/ng.
18. Ravenhall,M., Campino,S., Sepu,N., Nadjm,B., Mtove,G., Wangai,H., Maxwell,C., Olomi,R., Reyburn,H., Drakeley,C.J., *et al.* (2018) Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genet.*, **14**: e1007172 <https://doi.org/10.1371/journal.pgen.1007172>
19. Visscher,P.M., Brown,M.A., McCarthy,M.I. and Yang,J. (2012) Five Years of GWAS

- Discovery. *Am. J. Hum. Genet.*, **90**, 7–24.
20. Flori, L., Sawadogo, S., Esnault, C., Fre, N., Fumoux, F. and Rihet, P. (2003) Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. *Hum. Mol. Genet.*, **12**, 375–378.
21. Brisebarre, A., Kumulungui, B., Sawadogo, S., Atkinson, A., Garnier, S., Fumoux, F. and Rihet, P. (2014) A genome scan for *Plasmodium falciparum* malaria identifies quantitative trait loci on chromosomes 5q31, 6p21.3, 17p12, and 19p13. *Malar. J.*, **13**, 1–7.
22. Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Llesneur, C., Kuchenbaecker, K.B., Dennis, J., *et al.* (2019) Shared heritability and functional enrichment across six solid cancers. *Nat. Commun.*, **10**, 431–454. doi: 10.1038/s41467-018-08054-4.
23. Parker, M., Bull, S.J., Vries, J. De, Agbenyega, T., Doumbo, O.K. and Dominic, P. (2009) Ethical Data Release in Genome-Wide Association Studies in Developing Countries. *PLoS Med*, **6**, 11e1000143. doi:10.1371/journal.pmed.1000143
24. Achidi, E.A., Agbenyega, T., Allen, S., Amodu, O., Bojang, K., Conway, D., Corran, P., Deloukas, P., Djimde, A., Dolo, A., *et al.* (2008) A global network for investigating the genomic epidemiology of malaria. *Nature*, **456**, 732–737.
25. Band, G., Le, Q.S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G., *et al.* (2013) Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. *PLoS Genet.*, **9**, e1003509.
26. Speed, D., Hemani, G., Johnson, M.R. and Balding, D.J. (2012) Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.*, **91**, 1011–1021.
27. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., *et al.* (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
28. Antonacci, F., Kidd, J.M., Marques-bonet, T., Ventura, M., Siswara, P., Jiang, Z. and Eichler, E.E. (2009) Characterization of six human disease-associated inversion polymorphisms. *Hum. Mol. Genet.*, **18**, 2555–2566 doi:10.1093/hmg/ddp187.
29. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.
30. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P. and Palamara, P.F. (2017) Linkage disequilibrium dependent architecture of human complex traits shows action of negative selection. *Nat Genet.* **49**, 1421–1427.
31. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A., *et al.* (2015) The African Genome Variation Project shapes medical genetics in Africa. *Nature*, **517**, 327–332.
32. Patterson, N., Price, A.L. and Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genet.*, **10**.1371/journal.pgen.0020190.
33. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. and Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.* **45**:124–130. doi: 10.1038/ng.2504.
34. Sigova, A.A., Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-andre, V., Hoke, H.A. and Young, R.A. (2013) Resource Super-Enhancers in the Control of Cell Identity and Disease. *Cell*, **155**, 934–947.
35. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E., *et al.* (2013) Integrative annotation of chromatin elements from ENCODE data. *Nuc. Ac. Res.*, **41**, 827–841
36. Lindblad-toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P.,

- Ernst,J., Jordan,G., Mauceli,E., *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature.*, **478**, 476–481.
37. Ward,L.D. and Kellis,M. (2012) Evidence of Abundant Purifying Selection in Humans for Recently Acquired Regulatory Functions. *Sci.***337**,1675-1683.
38. Andersson,R., Gebhard,C., Miguel-escalada,I., Hoof,I., Bornholdt,J., Boyd,M., Chen,Y., Zhao,X., Schmidl,C., Suzuki,T., *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature.*,**507**,455-461. doi: 10.1038/nature12787.
39. Pasaniuc,B., Zaitlen,N., Shi,H., Bhatia,G., Gusev,A., Pickrell,J., Hirschhorn,J., Strachan,D.P., Patterson,N. and Price,A.L. (2014) Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics.* **30**, 2906–2914. doi: 10.1093/bioinformatics/btu416
40. The 1000 Genomes Project Consortium (2011) A map of human genome variation from population scale sequencing. *Nature*, **467**, 1061–1073.