Gavin Band[1,2], Ellen Leffler[2], Christina Hubbart[1], Anna Jeffreys[1], Kate Rowlands[1], Thuy Nguyen[2], Jim Stalker[2], Thomas N. Williams[3], Muminatou Jallow[4], Kirk A. Rockett[1], Dominic P. Kwiatkowski[1,2,5] and the Malaria Genomic Epidemiology Network[6]

[1]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; [2]Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK; [3]KEMRI–Wellcome Trust Research Programme, Kilifi, Kenya; [4]Medical Research Council Unit, Fajara, The Gambia; [5]MRC Centre for Genomics and Global Health, Big Data Institute, University of Oxford, Oxford, UK; [6]http://www.malariagen.net

Consortial Project 1

Evidence for interaction between human alleles and *P.falciparum* genetic variation in east and west Africa.

Human genetic variation has been known for over 70 years to affect susceptibility to the severe symptoms of malaria caused by the parasite *Plasmodium falciparum*. The most famous example of natural genetic protection is the the sickle-cell allele (HbS), and despite large-scale GWAS this is still one of only a small number of common variants now known to provide protection (**Figure 1**). However, some of these effects appear heterogeneous for reasons which are currently unexplained (**Figure 2**) .

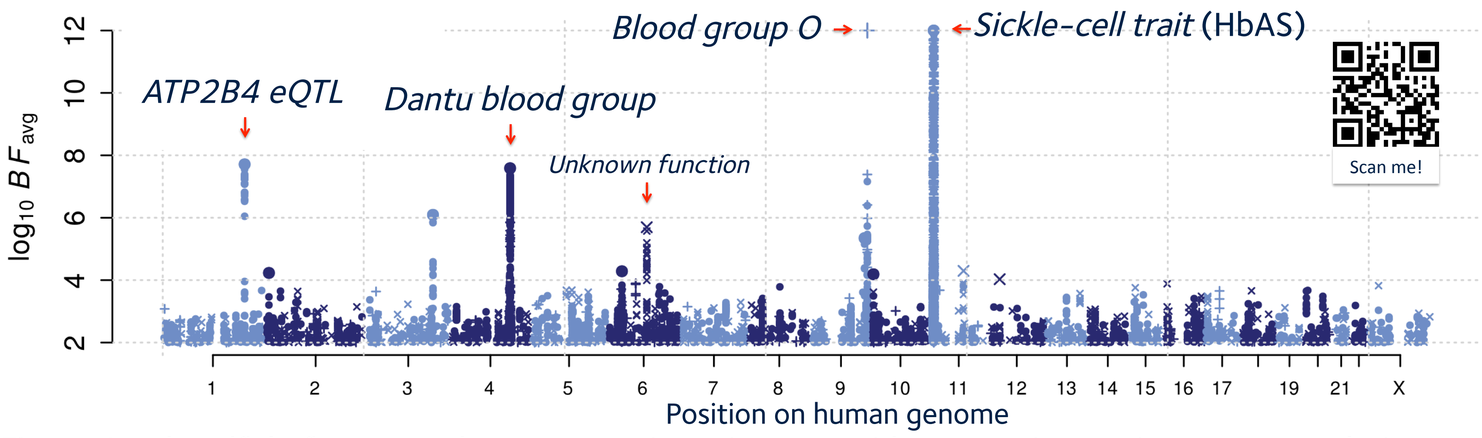

Figure 1: A handful of common human genetic variants provide protection against severe malaria. Plot shows the evidence for association with 18,378,754 genotyped or imputed variants genome-wide. Our study, available in '*New insights into malaria susceptibility from the genomes of 17,000 individuals from Africa, Asia, and Oceania*' on bioRxiv (https://doi.org/10.1101/535898), found five variants with strong evidence that also replicated in an additional 15,000 samples. For all but one of these, a putative functional variant underlying the association is known, and relates to properties of the red cell as noted.

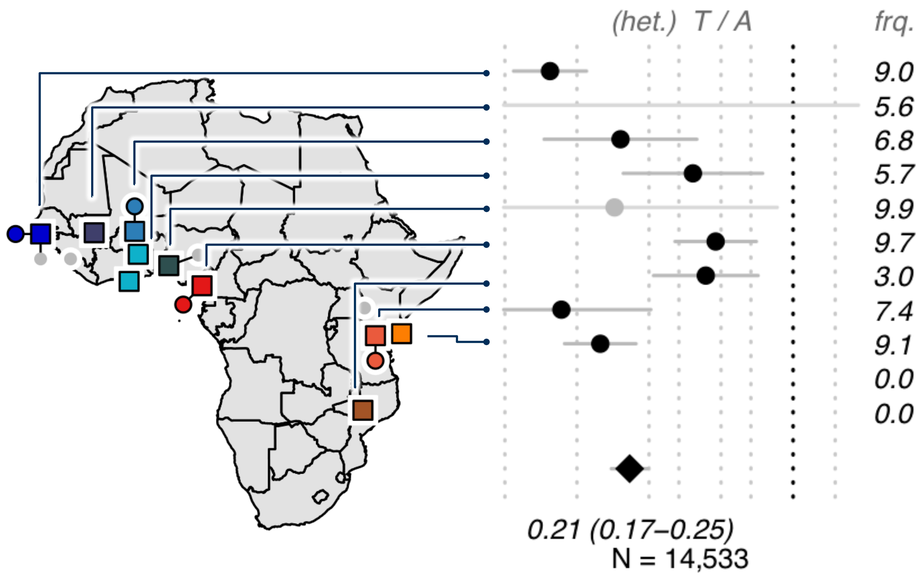

Figure 2: Variation in population effect-size estimates remains unexplained. Plot shows per-population estimates of the protective effect of sickle-cell trait (HbAS) under a heterozygote model of association. The large differences observed between populations could result from several factors, including differences in phenotyping between sites or, potentially, variation in parasites in these populations.

Malaria parasites are themselves diverse and parasite variation might be one factor that contributes to heterogeneity in susceptibility. To investigate this, we sequenced parasite DNA in cases ascertained with severe malaria from the above study from **Banjul, The Gambia** and **Kilifi, Kenya** using either **whole-genome sequencing** (WGS; **Figure 3**) or **selective whole-genome amplification** (sWGA; **Figure 4**). In total we sequenced **5,303** samples, of which a subset of **3,346** passed our QC processes and had genome-wide human data.

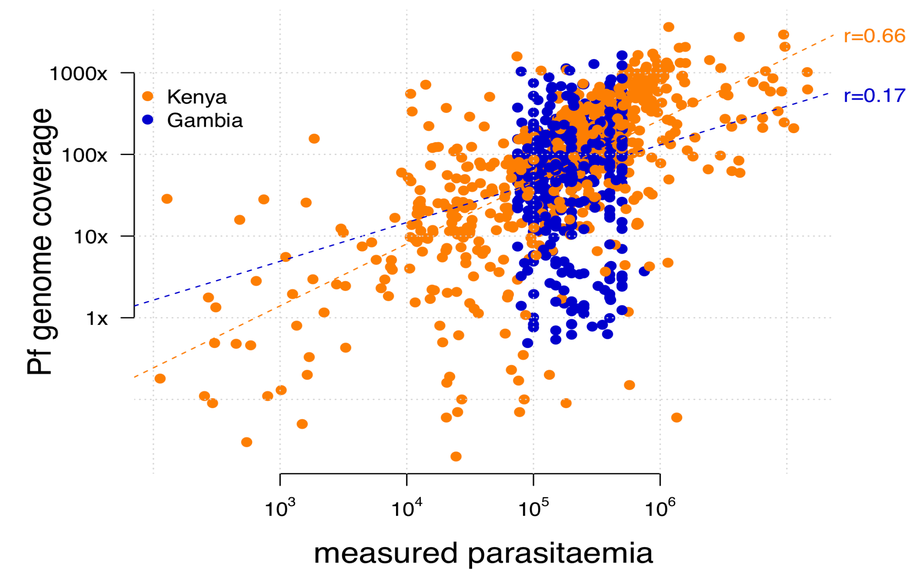

Figure 3: Whole-genome sequencing recovers parasite genomes in individuals with high-parasitaemia infections. Plot shows *P.falciparum* parasitaemia (parasites per microlitre of blood) measured by blood slide (X axis), versus mean sequencing read depth of the *Plasmodium falciparum* genome (Y axis) in samples from The Gambia (blue) and Kenya (Orange). Samples were selected for sequencing based on measured parasitaemia. Variation in blood slide measurements arises from a number of sources.

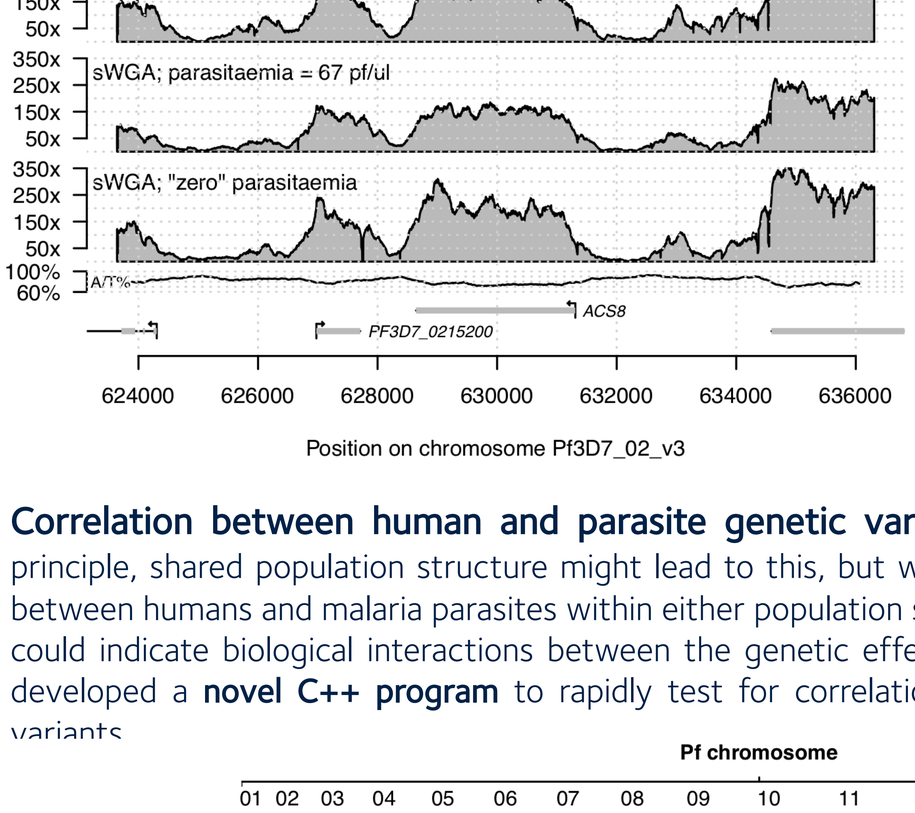

Figure 4: Selective whole-genome amplification (sWGA) recovers parasite genomes from low-parasitaemia infections. Plot shows an illustration of coverage of the *Plasmodium falciparum* 3D7 reference genome sequence, in samples sequenced from whole DNA or using sWGA to amplify *Pf* DNA. Panels show a high-parasitaemia sample (1st panel, whole-genome sequencing; 2nd panel, sWGA-based sequencing), a low-parasitaemia sample (3rd panel) and a sample recorded as having zero parasitaemia (4th panel). Sequencing performance for both methods is correlated with A/T coverage (5th panel) which also varies with genic content (6th panel).

Correlation between human and parasite genetic variation could arise in several ways. In principle, shared population structure might lead to this, but we found little evidence of co-structure between humans and malaria parasites within either population studied. Most compellingly, correlations could indicate biological interactions between the genetic effects of host and parasite variants. We developed a **novel C++ program** to rapidly test for correlation between pairs of host and parasite variants.

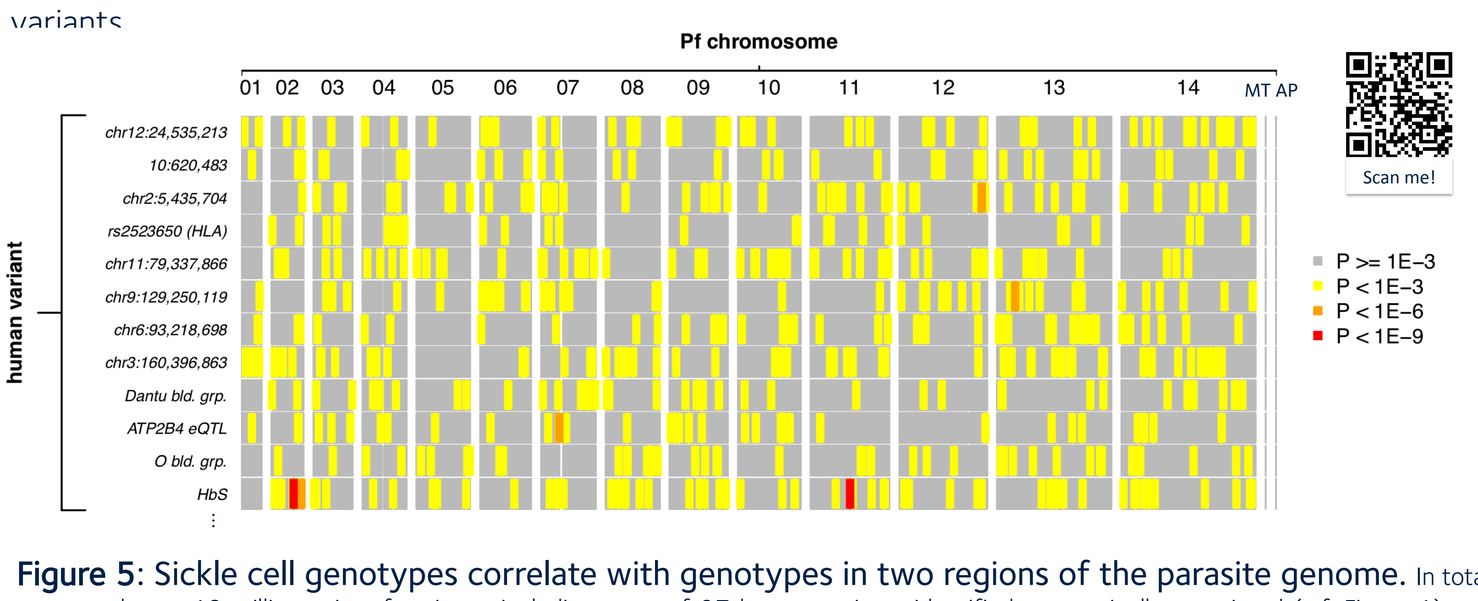

Figure 5: Sickle cell genotypes correlate with genotypes in two regions of the parasite genome. In total we tested over 12 million pairs of variants, including a set of 97 human variants identified as marginally associated (c.f. Figure 1) and ~4,000 variants within blood group genes, and a set of 51,553 biallelic SNPs called in the parasite genome using a **GATK pipeline**. We found evidence that the HbS genotype correlates with genotypes at variants in ***PfACS8*** on chromosome 2 (c.f. **Figure 4**), and a second region on chromosome 11 whose function is not currently annotated. ACS8 encodes a member of the **acyl-CoA synthetase family** which is involved in scavenging fatty acids for parasite use within erythrocytes. This family appears to have expanded in *P.falciparum*[1] and has been identified as a target of natural selection in *Plasmodium* populations[2,3]. See right for further interpretation of these results.

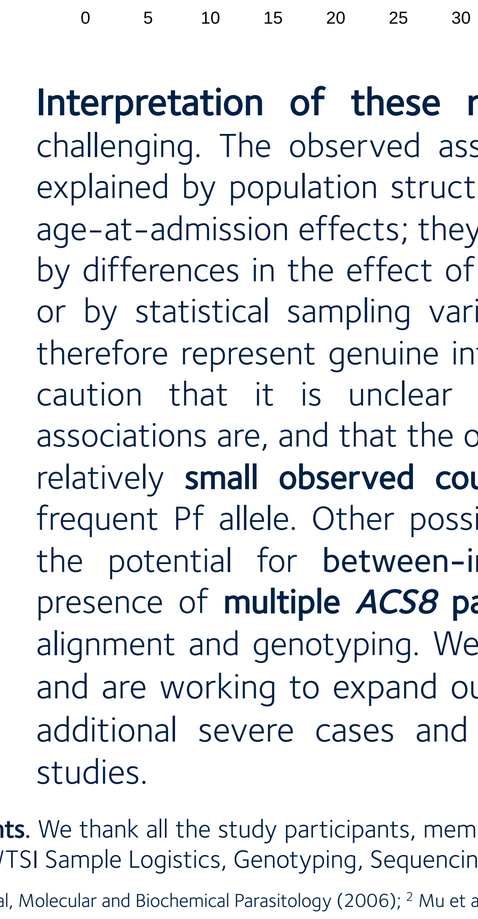

Figure 6: Little evidence of confounding by population structure in our data. Quantile-quantile plot showing observed chi-squared statistics (y axis) versus the expectation under the null. For the host-parasite correlation tests shown we conditioned on 10 host and 10 parasite principal components in each population, but this did not have a strong effect on the results.

Interpretation of these results is currently somewhat challenging. The observed associations do not appear to be explained by population structure, or by date-of-admission or age-at-admission effects; they are also not obviously explained by differences in the effect of HbS between malaria subtypes, or by statistical sampling variation (**Figure 6**). They might therefore represent genuine interaction effects. However, we caution that it is unclear how a priori plausible these associations are, and that the observed signals currently rely on relatively **small observed counts** of the HbS and the less frequent Pf allele. Other possible confounding factors include the potential for **between-infection relatedness**, and the presence of **multiple *ACS8* paralogs** which could affect read alignment and genotyping. We are investigating these issues, and are working to expand our sample set by sequencing of additional severe cases and further samples from cohort studies.

Acknowledgements. We thank all the study participants, members of the MalariaGEN consortial projects, and staff of the WTSI Sample Logistics, Genotyping, Sequencing and Informatics facilities.

References. [1]Bethke et al, Molecular and Biochemical Parasitology (2006); [2] Mu et al Nat Genet (2010); [3]van Tyne et al PLoS Genet (2011)