BIRC6 modifies risk of invasive bacterial infection in Kenyan children.

James J Gilchrist[1,2,3*], Silvia Kariuki[4], James A Watson[5,6], Gavin Band[3], Sophie Uyoga[4], Carolyne M Ndila[4], Neema Mturi[4], Salim Mwarumba[4], Shebe Mohammed[4], Moses Mosobo[4], Kaur Alasoo[7], Kirk A Rockett[3], Alexander J Mentzer[3], Dominic P Kwiatkowski[3,8], Adrian VS Hill[3,9], Kathryn Maitland[4,10], J Anthony G Scott[4,11], Thomas N Williams[4,12*]

***For correspondence:** james.gilchrist@paediatrics.ox.ac.uk (JJG); twilliams@kemri-wellcome.org (TNW)

[1]Department of Paediatrics, University of Oxford, Oxford, UK; [2]MRC–Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK; [3]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; [4]KEMRI-Wellcome Trust Research Programme, Centre for Geographic Medicine Research-Coast, Kilifi 80108, Kenya; [5]Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK; [6]Mahidol Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand; [7]Institute of Computer Science, University of Tartu, Tartu, Estonia; [8]Wellcome Sanger Institute, Cambridge, UK; [9]The Jenner Institute, University of Oxford, Oxford, UK; [10]Division of Medicine, Imperial College, London, UK; [11]Department of Infectious Disease Epidemiology, London School of Hygiene & Tropical Medicine, London, UK; [12]Institute for Global Health Innovation, Department of Surgery and Cancer, Imperial College, London, UK

Abstract

Invasive bacterial disease is a major cause of morbidity and mortality in African children. Despite being caused by diverse pathogens, children with sepsis are clinically indistinguishable from one another. In spite of this, most genetic susceptibility loci for invasive infection that have been discovered to date are pathogen specific and are not therefore suggestive of a shared genetic architecture of bacterial sepsis. Here we utilise probabilistic diagnostic models to identify children with a high probability of invasive bacterial disease among critically unwell Kenyan children with *P. falciparum* parasitaemia. We construct a joint dataset including 1,445 bacteraemia cases and 1,143 severe malaria cases, and population controls, among critically unwell Kenyan children that have previously been genotyped for human genetic variation. Using these data we perform a cross-trait genome-wide association study of invasive bacterial infection, weighting cases according to their probability of bacterial disease. In doing so we identify and validate a novel risk locus for invasive infection secondary to multiple bacterial pathogens, that has no apparent effect on malaria risk. The locus identified modifies splicing of *BIRC6* in stimulated monocytes, implicating regulation of apoptosis and autophagy in the pathogenesis of sepsis in Kenyan children.

Introduction

Invasive bacterial diseases are a major cause of child morbidity and mortality in Africa (***Berkley et al., 2005***). Although improved control measures, including the rollout of anti-bacterial conjugate vaccines (***Cowgill et al., 2006***; ***Silaba et al., 2019***), have led to recent declines, the burden of conditions such as pneumonia, meningitis and sepsis secondary to bacterial pathogens remains significant (***Vos et al., 2020***). A better understanding of the biology of invasive bacterial infections in African populations might help the development of novel interventions.

Susceptibility to invasive bacterial infections varies widely between individuals. In African children, some of this variability is determined by acquired comorbidities such as HIV, malnutrition and malaria (***Berkley et al., 2005***; ***Church and Maitland, 2014***; ***Scott et al., 2011***). However, the identification of common genetic variants as determinants of bacterial infection suggests that a significant portion of this variability is heritable. Many of these genetic susceptibility loci have pathogen-specific effects (***Davila et al., 2010***; ***Gilchrist et al., 2018***; ***Rautanen et al., 2016***), which is consistent with our understanding of infection susceptibility derived from primary immunodeficiencies. Key examples of pathogen-specificity among primary immunodeficiencies include Mendelian susceptibility to mycobacterial disease and susceptibility to non-tuberculous mycobacteria and nontyphoidal *Salmonella* (***Bustamante et al., 2014***), terminal complement deficiencies and meningococcal disease (***Figueroa et al., 1993***), and IRAK4 deficiency and pneumococcal disease (***Picard et al., 2007***). A major exception to this is the rs334 A>T mutation in *HBB* (sickle haemoglobin), which is associated with raised and lowered risks of infection secondary to a broad range of pathogens among homozygotes (***Williams et al., 2009***) and heterozygotes (***Scott et al., 2011***) respectively. However, the effect sizes associated with sickle haemoglobin are extreme, being maintained in populations by balancing selection, and larger sample sizes will probably be needed for the discovery of new variants with multi-pathogen effects.

Because the clinical features of invasive bacterial infections and severe malaria are broadly similar (***Bejon et al., 2007***), it can be difficult to distinguish between severe illness caused by extensive sequestration of malaria parasites in the microvasculature ('true' severe malaria) and bacterial sepsis in the presence of incidental parasitaemia on the basis of clinical features alone. This is made harder still by the fact that antibiotic pre-treatment and inadequate blood-culture volumes mean that, even in settings with excellent diagnostic facilities, true invasive bacterial infections can often not be confirmed (***Driscoll et al., 2017***). Recently, we illustrated this clinical complexity through a study in which we used probabilistic models based on malaria-specific biomarkers to show that approximately one third of children recruited to studies in Africa with a clinical diagnosis of severe malaria were actually suffering from other conditions (***Watson et al., 2021a***,b).

In the current study, we extend this work to show that invasive bacterial infections are common in children admitted to hospital with a clinical diagnosis of severe malaria, but in whom biomarkers subsequently suggest that malaria was probably not the primary cause for their severe illness. We then construct a dataset of genome-wide genotyped samples from 5,400 Kenyan children, comprising critically unwell Kenyan children with bacteraemia (***Rautanen et al., 2016***) and severe malaria (***Band et al., 2019***), and population controls. Using this dataset we perform a GWAS of invasive bacterial infection in Kenyan children, weighting cases according to the probability that their disease was mediated by a bacterial pathogen. In doing so, we increase our study power and identify *BIRC6* as a novel genetic determinant of invasive bacterial disease in Kenyan children.

Results

Severe malaria probability and risk of bacteraemia

Children admitted to the high dependency ward of Kilifi County Hospital with a clinical diagnosis of severe malaria, defined as a severe febrile illness in the presence of *P. falciparum* parasitaemia (n=2,200), between 11th June 1995 and 12th June 2008 were included in the study. While this definition is sensitive it is not specific, meaning that our study will have included some children

with sepsis accompanied by incidental parasitaemia (***Watson et al., 2021a***). We therefore used two probabilistic models, which included either platelet counts and plasma *Pf*HRP2 concentrations (Model 1, n=1,400) or white blood cell and platelet counts (Model 2, n=2,200), to determine the likelihood of 'true' severe malaria among these children. The estimated probabilities of 'true' severe malaria using each model were well-correlated ($r = 0.64$). Of 1,400 children with a clinical diagnosis of severe malaria with measured plasma *Pf*HRP2 concentrations, 425 (30.4%, Figure 1A and 1B) had a low probability (P(SM|Data)<0.5) of having 'true' severe malaria (941 of 2,220 children using WBC and platelet count data, Figure 1–figure supplement 1A and 1B). That is, while they presented with febrile illness and concomitant malaria parasitaemia, it is unlikely that their illnesses were directly attributable to malaria.

In keeping with the hypothesis that a significant proportion of these critically unwell children represented culture-negative invasive bacterial disease (Figure 1), in-patient mortality was higher among children with a low than a high probability of 'true' severe malaria (Table 1; $OR_{model1} = 1.57$, 95% CI $1.11 - 2.21$, $p = 0.01$; $OR_{model2} = 2.09$, 95% CI $1.60 - 2.72$, $p = 4.91 \times 10^{-8}$). This was also reflected in the rates of concurrent bacteraemia (Table 1; $OR_{model1} = 2.92$, 95% CI $1.66 - 5.13$, $p = 1.07 \times 10^{-4}$; $OR_{model2} = 2.00$, 95% CI $1.27 - 3.17$, $p = 0.003$). Similarly, the constituents of model 1 were each associated with blood culture positivity, both higher platelet counts (OR=2.36, 95% CI 1.19-4.70, $p = 0.014$) and lower *Pf*HRP2 levels (OR=0.52, 95% CI 0.39-0.70, $p = 9.62 \times 10^{-}6$) both being associated with the risk of coincident bacteraemia (Figure 1C and 1D). Conversely, white blood counts in isolation were not associated with risk of concurrent bacteraemia (Figure 1–figure supplement 1). Plasma *Pf*HRP2 is the single best biomarker for severe malaria (***Hendriksen et al., 2012***). In light of this, and given the greater enrichment for concurrent bacteraemia among children with a low probability of 'true' severe malaria as calculated by Model 1 than Model 2, we used Model 1 probabilities in downstream analyses where available (n=1,400) and used Model 2 probabilities for all other cases (n=800).

GWAS of invasive bacterial disease in Kenyan children

Children with a clinical diagnosis of severe malaria but a low probability of having 'true' severe malaria, are thus enriched for invasive bacterial disease. Using genome-wide genotyping data from cases of culture-confirmed bacteraemia and healthy controls, we estimated SNP heritability of bacteraemia in this population to be 19% (95% CI 3-35%, $p = 0.0084$). Despite this, our GWAS of bacteraemia failed to identify genetic correlates of bacteraemia risk beyond the sickle cell locus (***Rautanen et al., 2016***). Motivated by these observations we performed a genome-wide association study of invasive bacterial infection in Kenyan children in which we included both children with culture-confirmed bacteraemia and children with a clinical diagnosis of severe malaria. Children admitted to Kilifi County Hospital between 1st August 1998 and 30th October 2010 with community-acquired bacteraemia were recruited to the study as well as children from the severe malaria study described above. Control children were recruited from the same population between 1st August 2006 and 30th December 2010 as described in detail previously (***Scott et al., 2011***).

Following quality control measures (see Materials and Methods), we included 1,445 cases of culture-confirmed bacteraemia, 1,143 cases of severe malaria and 2,812 control children in our current analysis (Table 2, Figure 2). To account for the varying proportion of invasive bacterial disease among severe malaria cases, we applied weights to our regression analysis to reflect the greater likelihood of invasive bacterial disease among children with a low probability of 'true' severe malaria (sample weight, $w = 1 - P(SM|Data)$). Where *Pf*HRP2 concentrations were available (n=909) we used *Pf*HRP2 and platelet count to determine $P(SM|Data)$ while we used white cell and platelet counts (n=234) in cases where they were not available. Cases with culture-proven bacteraemia and control samples were assigned a sample weight of $w = 1$. Inclusion of the 6 major principal components (PC) of genotyping data and genotyping platform as covariates in the model adequately controlled for confounding variation ($\lambda = 1.0208$, Figure 3–figure supplement 1). In that analysis we found evidence supporting an association between risk of invasive bacterial disease in

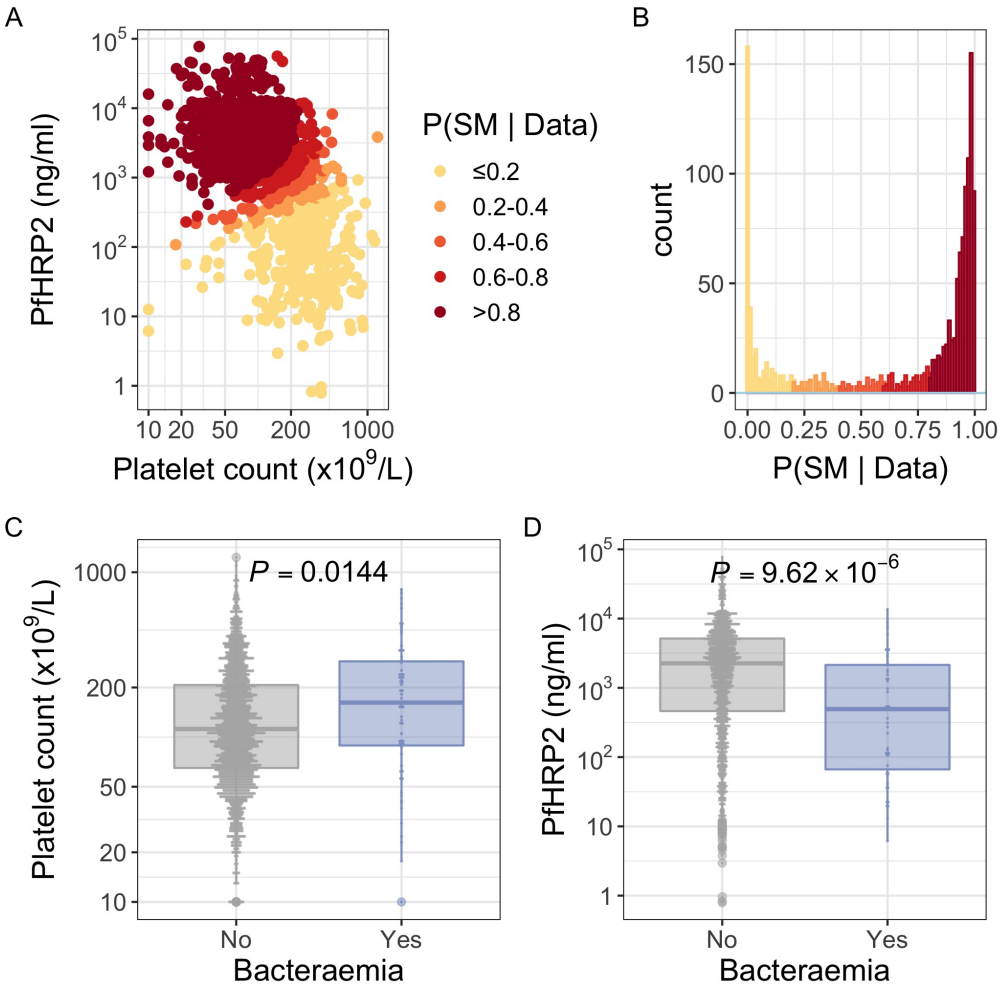


Figure 1. *Pf*HRP2 and platelet count as predictors of severe malaria. (A) Distribution of *Pf*HRP2 concentrations and platelet count among Kenyan children (n=1,400) with a clinical diagnosis of severe malaria. Points are coloured according to the probability of 'true' severe malaria given the data. (B) Distribution of 'true' severe malaria probabilities estimated from platelet count and plasma *Pf*HRP2 concentrations. (C) Platelets counts in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia. (D) *Pf*HRP2 concentrations in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia.

Figure 1–Figure supplement 1. White blood cell and platelet count as predictors of severe malaria. (A) Distribution of white blood cell and platelet count among Kenyan children (n=2,200) with a clinical diagnosis of severe malaria. Points are coloured according to the probability of 'true' severe malaria given the data. (B) Distribution of 'true' severe malaria probabilities estimated from platelet count and white blood cell count. (C) Platelets counts in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia. (D) White blood cell counts in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia.

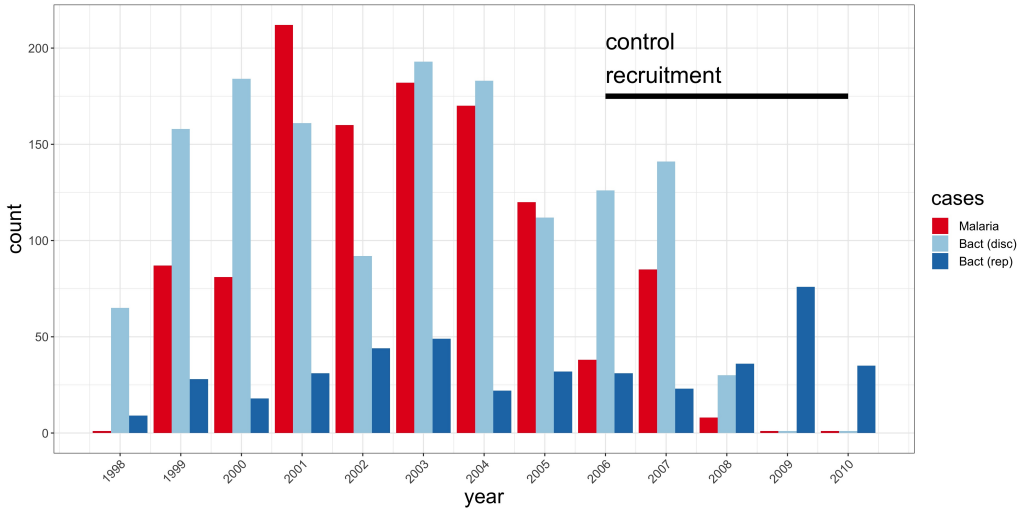


Figure 2. Study sample recruitment. Recruitment of severe malaria cases (n=1,143), discovery analysis bacteraemia cases (n=1,445) and replication analysis bacteraemia cases (n=434) between 1998 and 2010. The time period during which control samples (n=2,812) were recruited is also highlighted (black bar). Sample numbers represent children with genome-wide genotype data, who passed quality control filters and were included in this study.

Kenyan children and 7 SNPs at a single locus on chromosome 2 (peak SNP: rs183868412:T, OR=2.13, 95% CI 1.65-2.74, $p = 4.64 \times 10^{-9}$) (Figure 3, Table 3). Fine-mapping of this association identified a credible set of 7 SNPs with a 95% probability of containing the causal variant (Table 3), spanning a 212kb region: chr2:32,402,640-32,614,746.

To address the possibility that the observed association at this locus is driven by confounding secondary to population structure, we used ABERRANT (***Bellenguez et al., 2012***) to define a set of outlier samples on the first two PCs of genotyping data (n=22, Figure 3–figure supplement 2). These individuals are all genotyped on the Illumina Omni 2.5M array and are both cases and controls (15 and 7 respectively). While they are predominantly individuals with less common self-reported ethnicities in our study population (19 of 22 are not Giriama, Chonyi or Kauma), they are not representative of a single self-reported ethnicity (the most common single ethnicity in this group is Digo, n=7). Excluding these samples from the association analysis at rs183868412:T did not significantly alter the association with invasive bacterial infection ($p = 2.38 \times 10^{-8}$, OR=2.05, 95% CI 1.59-2.64). We further estimated the effect of rs183868412:T on invasive bacterial disease risk in four subpopulations defined by self-reported ethnicity (Giriama, $n = 2,501$; Chonyi, $n = 1,560$; Kauma, $n = 472$; Other, $n = 384$). Within each subpopulation we tested for association between genotype and case status in a weighted logistic regression model, including platform as a categorial covariate (Table 4). The minor allele frequency at rs183868412 ranged from 0.016 (Giriama) to 0.037 (Kauma), with no evidence of differentiation between subpopulations ($F_{ST} = 0.001$). We observed consistent effect sizes in both of the major study subpopulations; Giriama (OR=1.97, 95% CI 1.30-3.01, $p = 0.0015$) and Chonyi (OR=2.18, 95% CI 1.34-3.54, $p = 0.0017$) samples (Table 4), which together make up 83% of the study samples. Genotype at rs183868412 was also associated with less common self-reported ethnicities grouped together (OR=2.46, 95% CI 1.01-5.96, $p = 0.047$). Genotype was not associated with invasive bacterial disease risk in the Kauma subpopulation, however the sample size in the stratum is very limited (154 cases, 318 controls) and may simply reflect insufficient power to detect an association.

To assess whether our analysis could be affected by our choice of model to define severe malaria case weights, we restricted our analysis to samples with data available to calculate estimates for $P(SM|Data)$ using both Model 1 and Model 2 (n=909 severe malaria cases). We recalculated effect estimates for the rs183868412 association with invasive bacterial disease using each

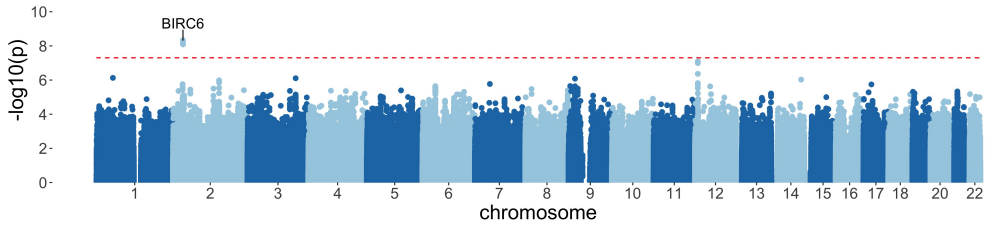


Figure 3. Manhattan plot of invasive bacterial infection in Kenyan children. Evidence for association with invasive bacterial disease at genotyped and imputed autosomal SNPs and indels (n=14,010,600) in Kenyan children (bacteraemia cases = 1,445, severe malaria cases = 1,143, controls = 2,812). Association statistics represent additive association. The red, dashed line denotes $p = 5 \times 10^{-8}$.

Figure 3–Figure supplement 1. Genome-wide association analysis quality control. (A) Principal components of genome-wide genotyping data in discovery samples. Individuals are color-coded according to genotyping platform; Affymetrix SNP 6.0 in purple, Illumina Omni 2.5M in orange. (B) Principal components of genome-wide genotyping data in discovery samples. Individuals are color-coded according to self-reported ethnicity; Chonyi in red, Giriama in blue, Kauma in green and other in grey. (C) Quantile-quantile plot of invasive bacterial infection in Kenyan children. QQ plot of weighted logistic regression GWAS of invasive bacterial disease in Kenyan children (2,588 cases, 2,812 controls).

Figure 3–Figure supplement 2. Principal components of genome-wide genotyping data in discovery samples. Outlier samples identified by ABERRANT (n=22) are highlighted (orange).

Figure 3–Figure supplement 3. Sensitivity analysis of case weights. Comparison of effect sizes of the rs183868412:T association with invasive bacterial infection in the combined analysis (bacteraemia cases = 1,445, severe malaria cases = 1,143, controls = 2,812), and restricted to cases weights calculated with *Pf*HRP2 plasma concentration and platelet count (Model 1; bacteraemia cases = 1,445, severe malaria cases = 909, controls = 2,812) or white cell count and platelet count (Model 2; bacteraemia cases = 1,445, severe malaria cases = 909, controls = 2,812).

Figure 3–Figure supplement 4. Genotyping concordance between Illumina and Affymetrix platforms. Pairwise genotyping concordance between samples genotyped on both Affymetrix SNP 6.0 and Illumina Omni 2.5M platforms.

model alone. The association with invasive bacterial disease at rs183868412:T is robust to the choice of the model for case weights, with effect estimates derived using Model 1 alone (OR=2.13, 95% CI 1.65-2.75, $p = 6.92 \times 10^{-9}$) and Model 2 alone (OR=2.05, 95% CI 1.59-2.65, $p = 2.63 \times 10^{-8}$) being entirely consistent with those seen in the main analysis (Figure 3–figure supplement 3). Moreover, restricting our analysis to cases of culture-confirmed bacteraemia, the effect estimate for bacteraemia risk observed in the discovery analysis (1,445 cases, 2,812 controls; OR=2.12, 95% CI 1.60-2.82, $p = 1.97 \times 10^{-7}$) is consistent with that seen in the main model.

We sought to replicate evidence of association in our discovery analysis through use of an independent case-control collection of Kenyan children with bacteraemia (n=434) and healthy controls (n=1,258) conducted in the same population. The peak trait-associated variants in the discovery analysis were well-imputed in the replication data (rs183868412 imputation info score $= 0.84$). In that analysis, we found evidence supporting the association at chromosome 2 with invasive bacterial disease (Figure 4, Table 5: rs183868412:T, OR=2.85, 95% CI 1.54-5.28, $p = 8.90 \times 10^{-4}$). In a fixed effects meta-analysis, rs183868412:T was strongly associated with risk of invasive bacterial disease in Kenyan children: OR=2.22, 95% CI 1.76-2.80, $p = 2.35 \times 10^{-11}$. That association was driven by children with culture-confirmed bacteraemia and critically unwell children with malaria parasites, but a low probability of 'true' severe malaria. In a stratified analysis (Figure 4, Table 5), rs183868412 was associated with culture-confirmed bacteraemia (OR=2.12, 95% CI 1.60-2.82, $p = 1.94 \times 10^{-7}$) and critical illness with parasitaemia and with a low probability of 'true' severe malaria (P(SM|Data)<0.5: OR=2.37, 95% CI 1.27-4.43, $p = 6.82 \times 10^{-3}$), but was not associated with risk of critical illness with a

high probability of 'true' severe malaria (P(SM | Data)>0.5: $p = 0.823$).

rs183868412 is associated with risk of invasive bacterial disease secondary to diverse pathogens and is independent of malaria.

Previous data describing the genetic risk of invasive bacterial disease in this population have identified pathogen-specific effects. To better-understand the range of pathogens to which genetic variation at *BIRC6* modifies risk we estimated the effect of rs183868412 on the risk of bacteraemia caused by the seven most common causative pathogens within this population (Figure 5A). In that analysis, the data best-supported a model in which genotype increases risk of bacteraemia caused by a broad range of pathogens, including bacteraemia secondary to pneumococcus, nontyphoidal *Salmonellae*, *E. coli*, β-haemolytic *Streptococci*, *S. aureus* and other less common pathogens grouped as a single stratum (log10 Bayes factor= 4.72). Genotype at rs183868412 similarly modified risk of bacteraemia in the neonatal period and in older children (log10 Bayes factor= 2.70, Figure 5B).

Malaria infection results in increased risk of invasive bacterial disease secondary to a broad range of pathogens (***Scott et al., 2011***), and genetic determinants of malaria risk (e.g. sickle cell trait) modify risk of bacterial infection in malaria-endemic settings (***Scott et al., 2011***). The observation that, among children with a clinical diagnosis of severe malaria, risk of disease is only modified by rs183868412 in children with a low probability that their disease represents 'true' severe malaria (Figure 5) suggests that the effect of genetic variation at *BIRC6* on invasive bacterial disease risk operates independently of malaria. In keeping with this, the data best-supports an effect of rs183868412 genotype on bacteraemia risk in children both with and without concomitant parasitaemia (log10 Bayes factor= 2.73, Figure 5C). In addition, unlike sickle cell trait (***Scott et al., 2011***), the increased risk of invasive bacterial infection conferred by rs183868412:T carriage in the study setting is stable across a period of declining malaria prevalence (Bayes factor= 8.7, Figure 5D).

Evidence of selection pressure and pleiotropy at rs183868412

Common genetic variation associated with a two-fold increased risk of invasive bacterial infection in children, in particular across a broad range of pathogens, will be subject to considerable negative selection pressure. The derived allele rs183868412:T, associated with increased risk of invasive bacterial disease in Kenyan children, is absent in non-African populations (https://gnomad.broadinstitute.org). Within Africa, rs183868412:T is present in all 9 African populations included in the MalariaGEN consortium project (***Band et al., 2019***) (Table 6), minor allele frequencies ranging from 0.011 in The Gambia to 0.034 in Malawi. There is no evidence for within-Africa differentiation at rs183868412 ($p = 0.601$) providing no support for a selective sweep at the locus. We further evaluated evidence for recent directional selection pressure, examining integrate haplotype scores (iHS) (***Voight et al., 2006***) in 7 African populations included in the 1000 Genomes Project. In those data, there is no evidence to support recent selection at the locus (minimum rank $p = 0.07$, maximum iHS = 1.3). Finally, to understand whether variation at this locus may be maintained in populations through beneficial effects on other traits, we examined evidence for pleiotropy at rs183868412 in a meta-analysis of GWAS data (***Gurdasani et al., 2019***) of 33 cardio-metabolic traits from Uganda, Ghana, Nigeria, South Africa and Kenya (sample size ranging from 2,671 to 14,126 individuals). In keeping with the MalariaGEN consortium project data, minor allele frequencies at rs183868412 range from 0.015 to 0.028. In these data there is no evidence for pleiotropy at the locus, with no evidence for an effect of rs183868412 on any of the 33 traits tested (minimum meta-analysis $p = 0.078$). The explanation for the persistence of this polymorphism, therefore, remains an open question.

rs183868412 is associated with alternative splicing of *BIRC6* in stimulated monocytes.

Trait associated genetic variation identified by GWAS is highly enriched for regulatory variation. The African specificity of the trait-associated variation identified here makes annotation with publicly-

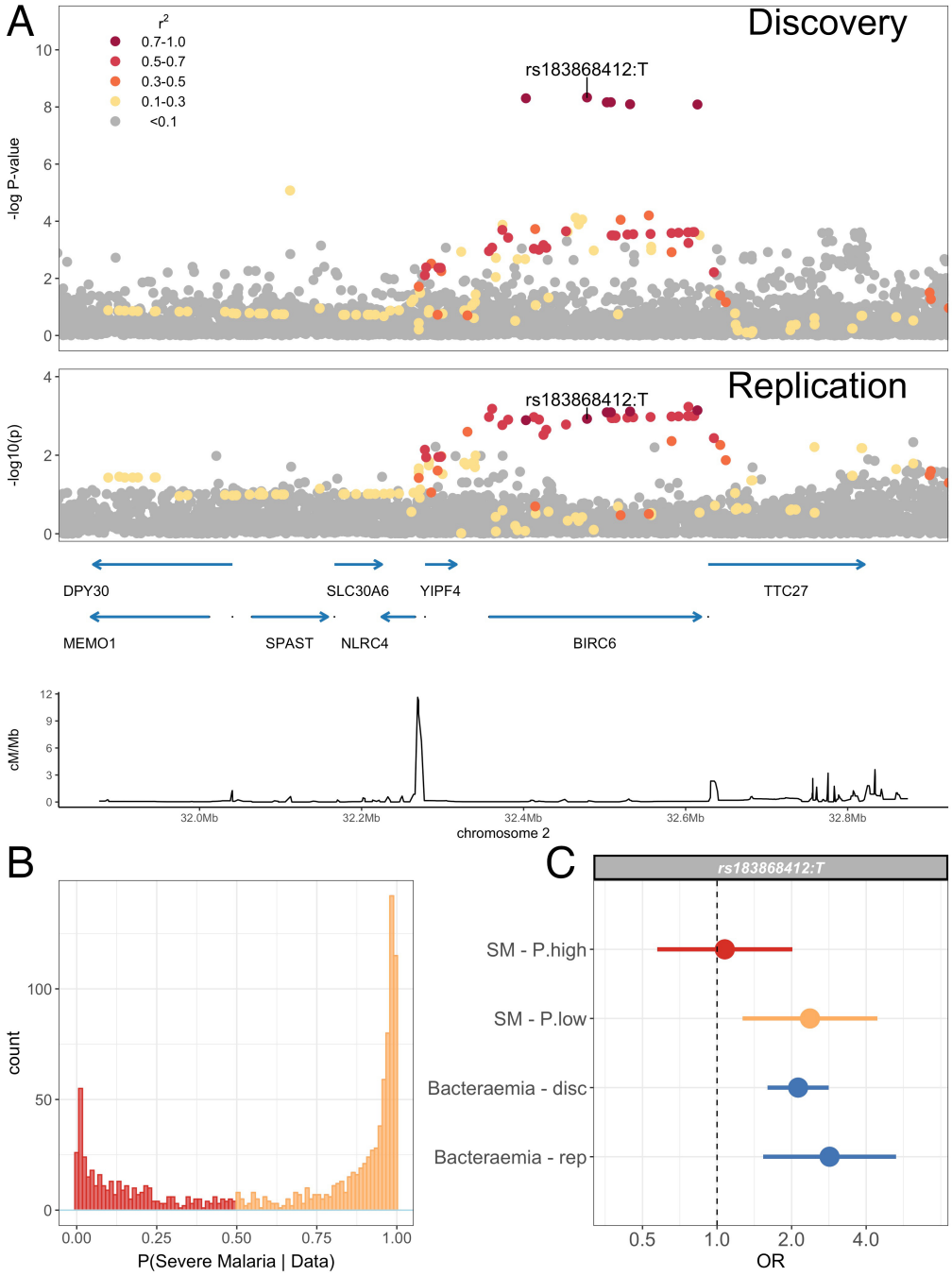


Figure 4. Association with invasive bacterial disease at the *BIRC6* locus. (A) Regional association plot of invasive bacterial disease association at the *BIRC6* locus in the discovery and replication analyses. SNPs are coloured according to linkage disequilibrium to rs183868412. Genomic coordinates are GRCh38. (B) Distribution of 'true' severe malaria probabilities among malaria cases estimated from plasma *Pf* HRP2 concentration and platelet count (n=909) and white blood cell count and platelet count (n=234). (C) Odds ratios and 95% confidence intervals of rs183868412 association with disease stratified by malaria cases with high (P>0.5, red) and low (P<0.5, orange) probabilities of 'true' severe malaria and culture-proven invasive bacterial disease (blue). P(SM | Data) represents the probability of 'true' severe malaria estimated from plasma PfHRP2 concentration and platelet count (n=909) or white blood cell count and platelet count (n=234).

Figure 4–Figure supplement 1. Principal components of genome-wide genotyping data in replication samples. Individuals are color-coded according to self-reported ethnicity; Chonyi in red, Giriama in blue, Kauma in green and other in grey.

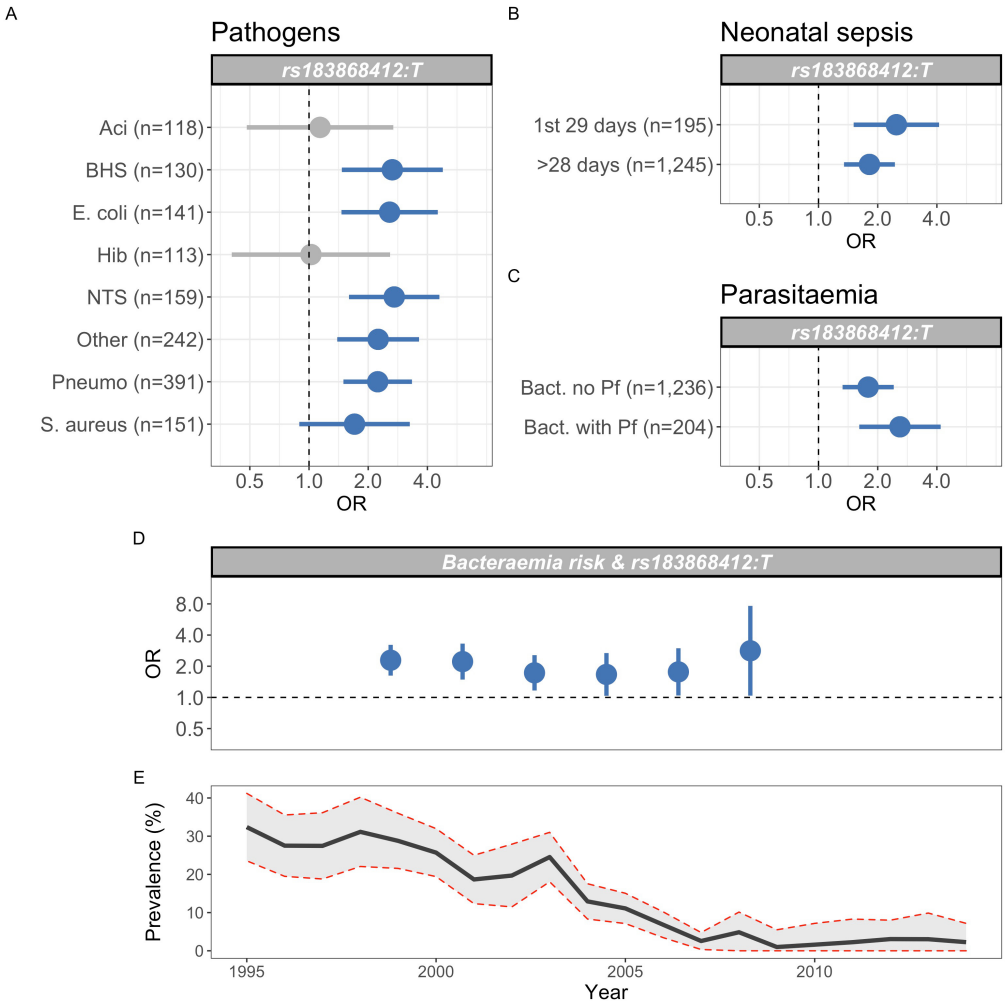


Figure 5. Genetic variation at *BIRC6* confers broad susceptibility to invasive bacterial disease. Odds ratios and 95% confidence intervals of rs183868412 association with invasive bacterial disease stratified by pathogen (A), neonatal and non-neonatal sepsis (B) and bacteraemia with and without malaria parasitaemia (C). Odds ratios and 95% confidence intervals of rs183868412 association with invasive bacterial disease stratified by year (D), compared to age-standardized, annual malaria parasite prevalence in Kilifi, Kenya, as estimated from parasite prevalence among trauma admissions (E). We compared models of association across strata using a Bayesian approach (see Methods). Strata associated with rs183868412 genotype in the most likely model in each analysis are highlighted in blue.

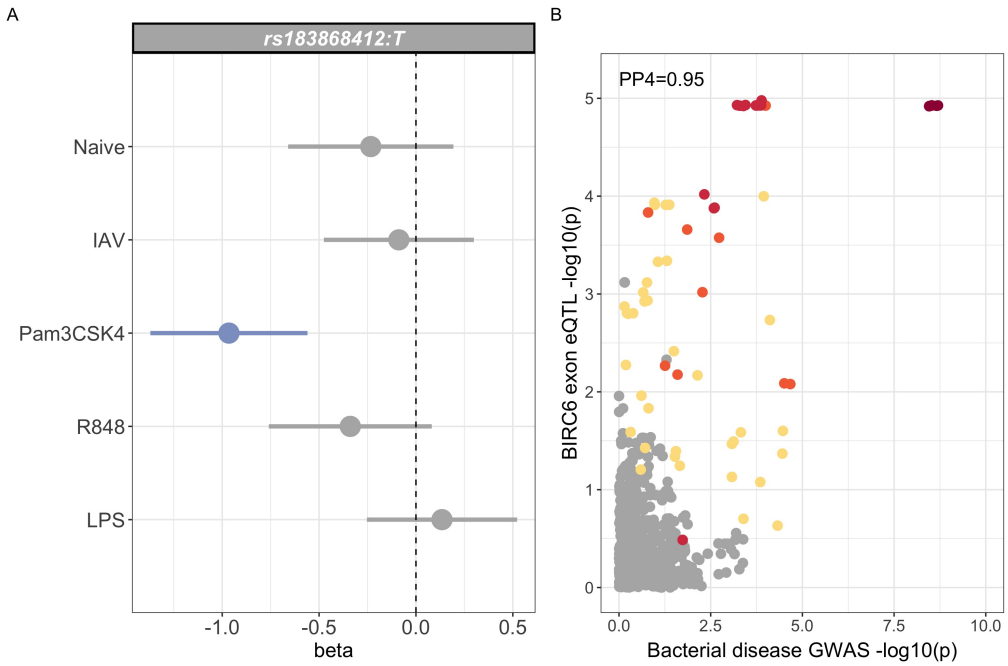


Figure 6. Regulatory function of rs183868412 monocytes. (A) Betas and 95% confidence intervals of rs183868412 association with expression of a 12bp *BIRC6* exonic sequence (chr2:32,453,943-32,453,954) in monocytes. Monocytes are naive or stimulated with; LPS (lipopolysaccharide), IAV (influenza A virus), Pam3CSK4 and R848. (B) Colocalisation of the *BIRC6* exon eQTL in Pam3CSK4-stimulated monocytes colocalizes with the risk locus for invasive bacterial disease ($PP4 = 0.951$). SNPs are coloured according to linkage disequilibrium to rs183868412.

available regulatory mapping data challenging. To investigate the regulatory function of rs183868412 in immune cells in African populations we used eQTL Catalogue (https://www.ebi.ac.uk/eqtl) mapping data (***Kerimov et al., 2021***) from 100 African ancestry individuals in primary monocytes with and without innate stimulation (***Quach et al., 2016***); influenza A virus, LPS, Pam3CSK4 (synthetic lipoprotein and TLR1/2 agonist) and R848 (a TLR7/8 agonist). In this dataset, rs183868412 is well-imputed ($r^2 = 0.998$), with a minor allele frequency of 0.05 (10 heterozygous individuals). We found no evidence for a regulatory effect of rs183868412 at the gene level in monocytes regardless of stimulation. We did, however, observe an effect of rs183868412 genotype on expression of a 12bp *BIRC6* exonic sequence (chr2:32,453,943-32,453,954, $p = 1.18 \times 10^{-5}$), with evidence for co-localisation of this eQTL with our GWAS signal (posterior probability of colocalisation, PP4 = 0.951, Figure 6). This effect was only observed following stimulation with Pam3CSK4 (Figure 6), with the bacteraemia risk allele, rs183868412:T, being associated with reduced expression of this sequence. That 12bp sequence is associated with an alternative splicing event that results in extension of a *BIRC6* exon. The 23rd exon (ENSE00001189810, chr2:32,453,808-32,453,942) of the canonical *BIRC6* transcript, ENST00000421745.6, is 135bp long and terminates immediately before the 12bp sequence associated with rs183868412:T genotype. The 22nd exon (ENSE00003835010, chr2:32,453,808-32,453,942) of an alternative *BIRC6* transcript, ENST00000648282.1, is 147bp long, having the same start site but including the 12bp sequence at its 3' end. Thus, increased risk of invasive bacterial disease may be associated with decreased expression of an alternative *BIRC6* transcript in TLR1/2 stimulated monocytes.

Discussion

In this study, we have leveraged the close relationship between *P. falciparum* infection and bacteraemia in African children (***Scott et al., 2011***) to perform a GWAS of invasive bacterial infection in

5,400 Kenyan children. We approached this by defining the probability with which each critically unwell child with a clinical diagnosis of severe malaria has a disease process directly mediated by malaria, that is 'true' severe malaria. We hypothesised that critically unwell children, with a low probability of having 'true' severe malaria, are enriched for invasive bacterial infections. We explored the validity of this approach, demonstrating that children with a low probability of 'true' severe malaria were indeed enriched for culture-proven bacteraemia and were at a higher risk of death than children with a higher probability. We therefore performed a GWAS weighting cases according to their likelihood of invasive bacterial disease. In doing so, we have identified and validated *BIRC6* as a novel genetic susceptibility locus for all-cause invasive bacterial disease in Kenyan children.

The disease-associated locus modifies risk of invasive bacterial disease caused by a broad range pathogens, including β-haemolytic *Streptococci*, *E. coli*, nontyphoidal *Salmonella*, *S. pneumoniae* and *S. aureus*. Moreover it modifies risk of invasive infection in both the neonatal period and in older children. Furthermore, in contrast to the rs334 *HBB* A>T mutation (***Scott et al., 2011***), rs183868412 modifies risk of invasive bacterial disease in a manner independent of malaria, with rs183868412:T carriage increasing risk of disease across a period of falling malaria transmission and in children with and without concurrent parasitaemia.

We further demonstrate that rs183868412 mediates risk of invasive bacterial disease through the modification of *BIRC6* splicing in Pam3CSK4-stimulated monocytes. *BIRC6* (Baculovirus inhibitor of apoptosis protein repeat containing 6), also known as *BRUCE* (BIR repeat containing ubiquitin-conjugating enzyme), encodes a large member of the inhibitor of apoptosis protein (IAP) family (***Hauser et al., 1998***). Members of the IAP family bind to cognate caspases, inhibiting their activity, and thereby cell death, through occlusion of their active site (***Verhagen et al., 2001***). A proportion of IAPs also contain an E3 ubiquitin ligase allowing both direct inhibition of caspases and targeting of caspases for proteasomal degradation (***Verhagen et al., 2001***). *BIRC6* contains both inhibitor of apoptosis domains and an E2/E3 ubiquitin ligase, which function to inhibit apoptosis in response to a variety of stimuli, both by interaction with and degradation of caspase-9, but also through the ubiquitination and degradation of SMAC, an IAP antagonist (***Hao et al., 2004***; ***Bartke et al., 2004***). *BIRC6* also regulates autophagosome-lysosome fusion (***Ebner et al., 2018***), and ubiquitinates (and targets for degradation) LC3, a key effector of autophagosome formation. Thus, *BIRC6* also acts as a negative regulator of autophagy (***Jia and Bonifacino, 2019***).

It is highly plausible that *BIRC6* could determine susceptibility to invasive infection through either its regulation of apoptosis or autophagy (or both). Sepsis induces marked changes in apoptosis across a range of immune cells (***Hotchkiss and Nicholson, 2006***). There is markedly enhanced apoptosis in both dendritic cells (***Hotchkiss et al., 2002***) and in lymphocytes. Enhanced lymphocyte apoptosis is most striking in B cells and $CD4^+$ T cells (***Hotchkiss et al., 2001***) which, at least in part, is mediated by caspase-9. The consequent lymphopoenia is correlated with both severity of sepsis and outcome (***Le Tulzo et al., 2002***). In addition to the direct effects of immune cell loss on the innate and adaptive immune responses to invasive infection, sepsis-induced apoptosis induces immune cell dysfunction, phagocytosis of apoptotic cells resulting in reduction in pro-inflammatory cytokine production and cross-presentation of antigen from apoptotic cells to adaptive immune cells (***Albert, 2004***). In keeping with a role for regulators of apoptosis in the pathogenesis of sepsis, members of the IAP family, including *NAIP/BIRC1* and *BIRC3*, are downregulated in immune cells in patients with sepsis, as is the *BIRC6* ubiquitination target *SMAC* (***Hoogerwerf et al., 2010***). Autophagy contributes to the direct removal of intracellular pathogens and, through the degradation of invading organisms in autophagosomes, directs antigen presentation and pro-inflammatory cytokine secretion (***Deretic et al., 2013***). As above, *BIRC6* regulates autophagy through its interaction with LC3, and overexpression of *LC3B* limits inflammation and tissue injury in a mouse model of sepsis (***Lo et al., 2013***).

In keeping with a role for *BIRC6* in autophagy and apoptosis in sepsis, our data identify a role for genetic variation at *BIRC6* in determining risk of invasive infection secondary to a broad range

of bacteria. This is in contrast to previously-published data describing susceptibility to invasive bacterial infection, which has highlighted a prominent role for genetic risk factors that are specific to single pathogens (**Davila et al., 2010**; **Gilchrist et al., 2018**; **Rautanen et al., 2016**). In this study, the derived allele (T) at rs183868412 was associated with increased risk of bacteraemia secondary to gram positive (β-haemolytic *Streptococci*, *S. pneumoniae* and *S. aureus*) and gram negative (*E. coli*, nontyphoidal *Salmonella*) pathogens, including intracellular and extracellular bacteria, and enteric and respiratory pathogens. Moreover, rs183868412 modified risk of bacteraemia in both the neonatal period, when infection is likely to be maternally derived, and in older children, when sources of community acquired infection will be more diverse. This modulation of invasive bacterial disease risk, despite diverse sources and routes of infection and diverse mechanisms of invasion, suggests a mechanism in which genetic variation at *BIRC6* modifies risk of invasive infection downstream of initial mechanisms of infection and invasion. In an interesting parallel to this, common genetic variation at another ubiquitin-conjugating enzyme, *UBE2U*, has been shown to modify outcome in meningitis secondary to diverse pathogens in individuals of European ancestry (**Lees et al., 2019**).

Our study has some important limitations. The African-specific nature of the trait-associated variation identified here limits our ability to comprehensively interrogate the effect of that variation in immune cells. The eQTL mapping data that we utilise here (**Quach et al., 2016**) is limited in that it allows us only to consider a single immune cell type. It is also important to note that there are relatively few African-ancestry individuals in the eQTL mapping data we utilise here, and the splicing signal at *BIRC6* is based on only 10 individuals heterogygous for rs183868412:T. A more complete understanding of the role played by genetic variation at *BIRC6* plays in the pathogenesis of sepsis in African children will require more detailed expression and functional studies in African populations. In addition, there is a paucity of large-scale genetic association studies performed in African populations. The African-specific nature of the variation identified in our study therefore limits our ability to explore pleiotropic effects at this locus. Allied to this, the low minor allele frequency of rs183868412 will result in very limited power to detect selection events at the locus. Larger and richer datasets detailing genetic variation in African populations will be required to explore the wider phenotypic consequences of variation at *BIRC6*. Our study uses plasma *Pf*HRP2 concentrations to help identify children at low risk of 'true' severe malaria. Given the increasing frequency of *pfhrp2* and *pfhrp2* deletions in many settings, including in Africa (**Agaba et al., 2019**; **Feleke et al., 2021**; **Gamboa et al., 2010**), an understanding of the local prevalence of *pfhrp2/3* deletions will be important in considering how to translate this model to other settings.

Taken together, our data identify a role for *BIRC6* in the pathogenesis of invasive bacterial infections in Kenyan children. By maximising our available sample size to include children with a high likelihood of invasive bacterial infection, but without culture-confirmed infection, we facilitate novel variant discovery and reveal a common genetic architecture of invasive bacterial disease secondary to diverse pathogens. In doing so we expand our understanding of the biology of invasive infection in African children. In particular, these data inform our understanding of the biology shared by diverse bacterial infections causing a common clinical syndrome: sepsis.

Materials and Methods

Study samples

Recruitment of the severe malaria cases, bacteraemia cases and healthy controls have been described in detail elsewhere (**Ndila et al., 2018**; **Rautanen et al., 2016**). In brief, children under 14 years admitted to the high dependency ward of Kilifi County Hospital with a clinical diagnosis of severe malaria, defined as *P. falciparum* parasites on blood film and at least one of; reduced Blantyre Coma Score, severe anaemia (Hb<50g/L), evidence of respiratory distress, hypoglycaemia or hyperparasitaemia were eligible for recruitment as cases of severe malaria. During the study period, all children admitted to hospital, with the exception of elective surgical admissions and minor

trauma, had a blood sample taken for bacterial culture (BACTEC 9050 instrument, Becton Dickinson, USA). Children under 14 years, in whom a pathogenic organism was identified in blood were eligible for study inclusion (*Bacillus* species, coryneform bacteria, coagulase-negative *Staphylococci*, *Staphylococcus saprophyticus* and Viridans group *Streptococci* were considered contaminants). Control children were recruited between 3 and 12 months of age from consecutive live births from the population which Kilifi County Hospital serves. Control children have been subject to longitudinal follow-up. Following explanation of the study, written informed consent was obtained from the parent or guardian of each child included in the study. Ethical approval was obtained from the Kenya Medical Research Institute (KEMRI) National Scientific Steering and Research Committees (approval numbers; SCC1192 and SCC967) and the Oxford Tropical Research Ethics Committee (OxTREC, approval numbers; 020-06 and 014-01).

Models to define the probability of 'true' severe malaria.

Among cases of severe malaria recruited to the study, we used probabilistic models to assign a probability of that child's clinical presentation being mediated by parasite sequestration, that is 'true' severe malaria, as described previously (***Watson et al., 2021a***,b). Where available (n=1,400), we used platelet counts and *Pf*HRP2 concentrations to derive the probabilistic model. In cases where *Pf*HRP2 concentration was not measured (n=800), we used white blood cell counts and platelet counts as input data to the model.

For both models (Model 1: *Pf*HRP2/platelet counts; Model 2: platelet counts/white blood cell counts), the probabilities of 'true' severe malaria were derived by fitting parametric latent class models. These assumed that each patient had a binary latent state (true severe malaria versus not severe malaria). For the Model 1, we assumed that in each latent state the data were distributed as a single bivariate normal distribution. For the model 2, the data did not fit well to a two-component bivariate normal (white blood cell counts have much weaker diagnostic value) so we assumed that the data had bivariate student-t distribution for the severe malaria state, and a flexible mixture of bivariates for the not severe malaria state.

Genotyping & imputation

For the discovery analysis, we utilised genotypes generated as part of genome-wide association studies of severe malaria (***Band et al., 2019***) and bacteraemia (***Rautanen et al., 2016***) previously reported in this population. Bacteraemia cases and controls were genotyped using the Affymetrix SNP 6.0 array and the severe malaria samples using the Illumina Omni 2.5 M platform. SNP and sample quality control for both datasets are highly analogous, and have been described previously (***Band et al., 2019***; ***Rautanen et al., 2016***). In brief, MalariaGEN SNP QC excluded poorly genotyped SNPs using the following metrics; SNP missingness >2.5%, minor allele frequency (MAF) <1%, Hardy-Weinberg equilibrium (HWE) $p < 1 \times 10^{-20}$, plate effect $p < 1 \times 10^{-3}$ and a recall test quantifying changes in genotype following a re-clustering process $p < 1 \times 10^{-6}$. For Affymetrix-genotyped samples, SNP QC excluded poorly performing SNPs using the following metrics; SNP missingness >2%, MAF <1%, genotype probability (info) <0.975, plate effect $p < 1 \times 10^{-6}$, and HWE $p < 1 \times 10^{-20}$. Sample QC on both platforms excluded sample outliers with respect to channel intensity, missingness, heterozygosity, population outliers and duplicated samples (relatedness coefficient > 0.75). In addition, for Affymetrix-genotyped samples, samples were excluded in cases of discordant sex as recorded in the clinical record and imputed from mean intensities from X and Y chromosome markers.

To facilitate combining datasets we applied an additional set of cross-platform QC procedures. We defined a shared subset of SNPs genotyped and passing SNP QC on both platforms (n=167,108), observing high levels of genotype concordance (median concordance 0.993, Figure 3–figure supplement 4) among the subset of samples genotyped on both platforms (n=1,365). We used this shared SNP set to compute relatedness estimates and PCs in PLINK v1.90 (***Chang et al., 2015***). The major six PCs of shared genotypes differentiate self-reported ethnicity (Figure 3–figure supple-

ment 1) and are non-differential with respect to genotyping platform (Figure 3–figure supplement 1). To harmonise QC across both platforms we excluded MalariaGEN samples with discordant clinical and genetic sex (n=136). We further excluded one of duplicate or related sample pairs (relatedness coefficient > 0.2) across platforms, retaining case samples where possible and excluding equal numbers of control samples genotyped on each platform (n=1,973). Following QC, genotypes were phased using SHAPEIT2 (***Delaneau et al., 2012***), and untyped genotypes imputed genome-wide using IMPUTE2 (v2.3.2) (***Howie et al., 2011***, ***2009***) with 1000 Genomes Phase III as a reference panel. Following imputation, we excluded SNPs with imputation info scores <0.5, MAF <1% and HWE $p < 1 \times 10^{-5}$, applying each threshold for each platform and overall. Following SNP and sample QC, 14,010,600 autosomal SNPs and indels from 5,400 samples (1,445 bacteraemia cases, 1,143 severe malaria cases and 2,812 control samples: 917 Illumina genotyped and 1,895 Affymetrix genotyped) were taken forward as a combined discovery dataset for association analysis. Following QC and association analysis, we identified a further set of population outliers using ABERRANT (***Bellenguez et al., 2012***) for downstream sensitivity analysis.

Estimation of SNP heritability

To estimate SNP heritability of bacteraemia in this population, we used genome-wide genotyping data from culture-confirmed bacteraemia cases and healthy controls genotyped with the Affymetrix SNP 6.0 array. For this analysis we used directly typed markers passing quality control as described above ($n = 783,094$). In addition to the sample quality control applied above, we additionally excluded one individual from each sample pair with relatedness coefficient > 0.05, leaving a final sample size of 2,559 (1,042 cases, 1,517 controls). We estimated SNP heritability using GCTA-GREML (***Yang et al., 2011***). For transformation of the heritability estimate to the liability scale we assumed a population prevalence for bacteraemia of 2% in this setting.

Association analysis & fine mapping

In the discovery analysis, we tested for association between genotype at each variant passing QC and invasive bacterial disease by logistic regression in an additive linear model. We used weighted logistic regression to reflect the probability of each case sample being a 'true' case of invasive bacterial infection. Cases of culture-confirmed bacteraemia we assigned a weight of 1, whereas cases of severe malaria were assigned weights 1-P(SM|Data), re-weighting the contribution of a case to the log-likelihood according to its probability of representing invasive bacterial infection. Control samples were assigned a weight of 1. Our regression thus assumes that the lower the probability of 'true' severe malaria, the greater the probability that a case represents culture-negative invasive bacterial disease. To control for confounding variation, we included the 6 major PCs of genotyping data and genotyping platform as covariates in the model. Weighted logistic regression was implemented using the *glm* function in R. As described previously (***Watson et al., 2021a***), standard errors were transformed to reflect the reduced effective sample size resulting from inclusion of sample weights in the model. We considered $p < 5 \times 10^{-8}$ to be significant.

We used a Bayesian approach to identify a set of SNPs with 95% probability of containing the causal variant at the trait-associated locus. Approximate Bayes' factors (***Wakefield, 2009***) were calculated for each SNP in the region (a 200kb surrounding rs183868412) with a prior distribution of $N(0, 0.2^2)$. All SNPs were considered equally likely to be the causal variant a priori. A set of SNPs with 95% probability of containing the causal SNP was defined as the smallest number of SNPs for which the summed posterior probabilities exceed 0.95.

Replication samples and analysis

To replicate our findings from the discovery analysis we used a second sample set, recruited from the same population as the discovery samples. Replication case samples were cases of bacteraemia only, and did not include cases of severe malaria without culture-confirmed bacterial infection. Case samples were recruited between 1st August 1998 and 30th October 2010. As for the

discovery case samples, children under 14 years with a bacterial pathogen isolated from blood on admission to hospital were eligible for recruitment to the study. As above, control samples were recruited as part of a birth cohort from the same population, with children recruited between the ages of 3 and 12 months. Genotyping and QC procedures for these samples have been described previously. In brief, study samples were genotyped using the Immunochip Consortium (***Cortes and Brown, 2011***) array (Illumina). Sample QC was performed as for the discovery samples (above), with duplicate control samples (samples common to MalariaGEN and Immunochip controls, n=78) being removed from the replication set. As above, relatedness estimates and PCs were computed in PLINK v1.90 (***Chang et al., 2015***) (Figure 4-figure supplement 1). SNP QC excluded the following variants; SNP missingness >1%, MAF <1% and HWE $p < 1 \times 10^{-10}$. Following QC, 143,000 genotyped variants in 434 cases and 1,258 control samples were taken forward for imputation. As above, imputation was performed with SHAPEIT2 (***Delaneau et al., 2012***) and IMPUTE2 (v2.3.2) (***Howie et al., 2011***, ***2009***) with 1000 Genomes Phase III as a reference panel.

Following imputation, we further excluded poorly-imputed (imputation info score<0.5) and rare (MAF<1%) variants and variants with HWE $p < 1 \times 10^{-10}$. At variants associated with invasive bacterial disease ($p < 5 \times 10^{-8}$) in the discovery analysis, we tested for association with bacteraemia case-control status using logistic regression in an additive model in SNPTEST v2.5.6 (***Marchini et al., 2007***). To exclude confounding variation, we included the major six PCs of genotyping data in the model. We considered evidence of association with bacteraemia in the replication samples with $p < 0.05$ with the same direction of effect as in the discovery analysis to be significant.

Bayesian comparison of models of association

At the locus of interest, we used multinomial logistic regression, implemented in SNPTEST v2.5.6 (***Marchini et al., 2007***) to estimate the additive effect of genotype on risk of bacteraemia stratified by pathogen, bacteraemia in the neonatal and non-neonatal periods, bacteraemia with and without *P. falciparum* parasitaemia, and bacteraemia presenting at different time periods across a period of declining malaria transmission intensity. For these analyses we used only samples with culture-confirmed bacteraemia. In each case we used control status as the baseline stratum, and included the six major principal components of genotyping data to control for confounding variation as above.

For the pathogen-stratified analysis, we defined eight case strata among the discovery cases, one for each of the seven most commonly isolated organisms (*Acinetobacter*, n=118; β-haemolytic *Streptococci*, n=130; *E. coli*, n=141; *H. influenzae* type b, n=113; nontyphoidal *Salmonella*, n=159; pneumococci, n=390; *S. aureus*, n=152) and one stratum for the remaining other organisms (n=242). For the neonatal/non-neonatal disease analysis we stratified cases as presenting in the first 28 days of life (n=195) or beyond that (n=1,245). For the analysis comparing bacteraemia with and without malaria we stratified cases with (n=204) and without (n=1,236) *P. falciparum* on their admission blood film. For each of these analysis case strata were compared to Affymetrix-genotyped discovery control samples (n=1,895) as a baseline stratum.

For the analysis stratified across year of admission, we defined case strata by grouping into six time periods according to their date of admission; 1998-2000 (n=498), 2001-2002 (n=349), 2003-2004 (n=467), 2005-2006 (n=310), 2007-2008 (n=237), 2009-2010 (n=111). For this analysis we used both discovery (Affymetrix) and replication (Immunochip) case and control samples. This allowed better coverage of the years later in the study, which were underrepresented in the discovery samples (the discovery median admission year is 2003, c.f. 2005 for the replication samples). In that analysis we used multinomial logistic regression in each cohort to estimate stratum-specific effects, combining these results in a fixed effects meta-analysis using BINGWA (***Band et al., 2015***).

We then compared models of association using a Bayesian approach (***Rautanen et al., 2016***), considering the following models:

"Null": effect size = i.e. no association with bacteraemia.

"Same": effect size $N(0, 0.2^2)$ and fixed across strata.

Additional models consider each possible combination of a fixed effect size for associated strata and no association at other strata. For each model we calculated approximate Bayes factors (***Wakefield, 2009***) and posterior probabilities, assuming each model to be equally likely a priori. Statistical analysis was performed in R.

eQTL mapping and colocalisation analysis

We used the eQTL catalogue (***Kerimov et al., 2021***) mapping pipeline (https://github.com/eQTL-Catalogue/) to map eQTL in naive and stimulated monocytes (***Quach et al., 2016***). These data include bulk RNA-Seq and genome-wide genotyping data from naive and stimulated primary monocytes isolated from individuals of European (n=100) and African (n=100) ancestry (***Quach et al., 2016***). Given the African-specific nature of variation at rs183868412, we performed eQTL mapping in this dataset restricting our analysis to samples of African ancestry. The eQTL Catalogue mapping pipeline has been described previously (***Kerimov et al., 2021***). In brief, sample genotypes (Illumina HumanOmni5-Quad genotyped) were pre-phased with Eagle v.2.4.1 (***Loh et al., 2016***) and genotypes imputed with Minimac4 v.1.0.2 (***Das et al., 2016***) using 1000 Genomes phase III as a reference panel. Gene expression, exon expression, transcript usage and transcriptional event usage were quantified from RNA-Seq data using HISAT2 v.2.1.0 (***Kim et al., 2019***), DEXSeq v.1.18.4 (***Anders et al., 2012***) and Salmon v.0.13.1 (***Patro et al., 2017***). Nominal mapping in *cis* was performed for each phenotype for variants within a 1Mb window of the start of each gene using FastQTL ***Ongen et al.*** (***2016***), including 6 principal components of genotyping and phenotype data as covariates in the model.

We then used the R package coloc v5.1.0 (***Giambartolomei et al., 2014***) to identify evidence of causal variants shared by our bacterial disease-associated locus of interest and regulatory genetic variation identified in our eQTL mapping data. Coloc adopts a Bayesian approach to compare evidence for independent or shared association signals for two traits at a given genetic locus. We used the coloc.susie() command to allow colocalisation of multiple independent signals at a single locus for each trait. We considered evidence for colocalisation for each gene and exon within a 250kb window of the peak association (rs183868412). We considered a posterior probability > 0.8 supporting a shared causal locus to be significant.

Acknowledgments

This publication uses genotyping data from the MalariaGEN consortial project, as described in Malaria Genomic Epidemiology Network, et al. Nature Communications, 2019 (https://doi.org/10.1038/s414
019-13480-z). This study makes use of data generated by the Wellcome Trust Case Control Consortium 2 project (Grant Reference 085475/B/08/Z). JJG and AJM are funded by National Institute for Health Research (NIHR) Clinical Lectureships. TNW and JAGS are supported by Senior Research Fellowships from the Wellcome Trust (202800 and 098532 respectively). JAW is a Sir Henry Dale Fellow funded by the Wellcome Trust (223253/Z/21/Z). SMAART (Severe Malaria Africa – A consortium for Research and Trials) is funded by a Wellcome Collaborative Award in Science grant (209265/Z/17/Z) held in part by KM and TNW. During this work AVSH was supported by a Wellcome Trust Senior Investigator Award (HCUZZ0) and by a European Research Council advanced grant (294557). The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This research was funded by The Wellcome Trust. A CC BY or equivalent licence is applied to the author accepted manuscript arising from this submission, in accordance with the grant's open access conditions. This paper is published with the permission of the Director of KEMRI.

References

Agaba BB, Yeka A, Nsobya S, Arinaitwe E, Nankabirwa J, Opigo J, Mbaka P, Lim CS, Kalyango JN, Karamagi C, Kamya MR. Systematic review of the status of pfhrp2 and pfhrp3 gene deletion, approaches and methods

used for its estimation and reporting in Plasmodium falciparum populations in Africa: review of published studies 2010-2019. Malar J. 2019 Nov; 18(1):355. doi: 10.1186/s12936-019-2987-4.

Albert ML. Death-defying immunity: do apoptotic cells influence antigen processing and presentation? Nature Reviews Immunology. 2004; 4(3):223–231. https://doi.org/10.1038/nri11308, doi: 10.1038/nri11308.

Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. Genome Res. 2012 Oct; 22(10):2008–2017. doi: 10.1101/gr.133744.111.

Band G, Le QS, Clarke GM, Kivinen K, Hubbart C, Jeffreys AE, Rowlands K, Leffler EM, Jallow M, Conway DJ, Sisay-Joof F, Sirugo G, d'Alessandro U, Toure OB, Thera MA, Konate S, Sissoko S, Mangano VD, Bougouma EC, Sirima SB, et al. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. Nature Communications. 2019; 10(1):5732. https://doi.org/10.1038/s41467-019-13480-z, doi: 10.1038/s41467-019-13480-z.

Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Si Le Q, Clarke GM, Kivinen K, Leffler EM, Rockett KA, Kwiatkowski DP, Spencer CCA, Rockett KA, Spencer CCA, Cornelius V, Conway DJ, Williams TN, Taylor T, Kwiatkowski DP, Conway DJ, Bojang KA, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature. 2015; 526(7572):253–257. https://doi.org/10.1038/nature15390, doi: 10.1038/nature15390.

Bartke T, Pohl C, Pyrowolakis G, Jentsch S. Dual Role of BRUCE as an Antiapoptotic IAP and a Chimeric E2/E3 Ubiquitin Ligase. Molecular Cell. 2004; 14(6):801–811. https://www.sciencedirect.com/science/article/pii/S1097276504003041, doi: https://doi.org/10.1016/j.molcel.2004.05.018.

Bejon P, Berkley JA, Mwangi T, Ogada E, Mwangi I, Maitland K, Williams T, Scott JAG, English M, Lowe BS, Peshu N, Newton CRJC, Marsh K. Defining childhood severe falciparum malaria for intervention studies. PLoS Med. 2007 Aug; 4(8):e251. doi: 10.1371/journal.pmed.0040251.

Bellenguez C, Strange A, Freeman C, Consortium WTCC, Donnelly P, Spencer CCA. A robust clustering algorithm for identifying problematic samples in genome-wide association studies. Bioinformatics (Oxford, England). 2012 01; 28(1):134–135. https://pubmed.ncbi.nlm.nih.gov/22057162, doi: 10.1093/bioinformatics/btr599.

Berkley JA, Lowe BS, Mwangi I, Williams T, Bauni E, Mwarumba S, Ngetsa C, Slack MPE, Njenga S, Hart CA, Maitland K, English M, Marsh K, Scott JAG. Bacteremia among children admitted to a rural hospital in Kenya. N Engl J Med. 2005 Jan; 352(1):39–47. doi: 10.1056/NEJMoa040275.

Bustamante J, Boisson-Dupuis S, Abel L, Casanova JL. Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN- immunity. Semin Immunol. 2014 Dec; 26(6):454–470. doi: 10.1016/j.smim.2014.09.008.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7. doi: 10.1186/s13742-015-0047-8.

Church J, Maitland K. Invasive bacterial co-infection in African children with Plasmodium falciparum malaria: a systematic review. BMC Med. 2014 Feb; 12:31. doi: 10.1186/1741-7015-12-31.

Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011 Feb; 13(1):101. doi: 10.1186/ar3204.

Cowgill KD, Ndiritu M, Nyiro J, Slack MPE, Chiphatsi S, Ismail A, Kamau T, Mwangi I, English M, Newton CRJC, Feikin DR, Scott JAG. Effectiveness of Haemophilus influenzae Type b Conjugate Vaccine Introduction Into Routine Childhood Immunization in Kenya. JAMA. 2006 11/3/2021; 296(6):671–678. https://doi.org/10.1001/jama.296.6.671, doi: 10.1001/jama.296.6.671.

Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY, Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A, Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, et al. Next-generation genotype imputation service and methods. Nat Genet. 2016 Oct; 48(10):1284–1287. doi: 10.1038/ng.3656.

Davila S, Wright VJ, Khor CC, Sim KS, Binder A, Breunis WB, Inwald D, Nadel S, Betts H, Carrol ED, de Groot R, Hermans PWM, Hazelzet J, Emonts M, Lim CC, Kuijpers TW, Martinon-Torres F, Salas A, Zenz W, Levin M, et al. Genome-wide association study identifies variants in the CFH region associated with host susceptibility to meningococcal disease. Nature Genetics. 2010; 42(9):772–776. https://doi.org/10.1038/ng.640, doi: 10.1038/ng.640.

Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nature Methods. 2012; 9(2):179–181. https://doi.org/10.1038/nmeth.1785, doi: 10.1038/nmeth.1785.

Deretic V, Saitoh T, Akira S. Autophagy in infection, inflammation and immunity. Nature Reviews Immunology. 2013; 13(10):722–737. https://doi.org/10.1038/nri3532, doi: 10.1038/nri3532.

Driscoll AJ, Deloria Knoll M, Hammitt LL, Baggett HC, Brooks WA, Feikin DR, Kotloff KL, Levine OS, Madhi SA, O'Brien KL, Scott JAG, Thea DM, Howie SRC, Adrian PV, Ahmed D, DeLuca AN, Ebruke BE, Gitahi C, Higdon MM, Kaewpan A, et al. The Effect of Antibiotic Exposure and Specimen Volume on the Detection of Bacterial Pathogens in Children With Pneumonia. Clinical Infectious Diseases. 2017 11/16/2021; 64(suppl_3):S368–S377. https://doi.org/10.1093/cid/cix101, doi: 10.1093/cid/cix101.

Ebner P, Poetsch I, Deszcz L, Hoffmann T, Zuber J, Ikeda F. The IAP family member BRUCE regulates autophagosome-lysosome fusion. Nat Commun. 2018 Feb; 9(1):599. doi: 10.1038/s41467-018-02823-x.

Feleke SM, Reichert EN, Mohammed H, Brhane BG, Mekete K, Mamo H, Petros B, Solomon H, Abate E, Hennelly C, Denton M, Keeler C, Hathaway NJ, Juliano JJ, Bailey JA, Rogier E, Cunningham J, Aydemir O, Parr JB. Plasmodium falciparum is evolving to escape malaria rapid diagnostic tests in Ethiopia. Nat Microbiol. 2021 Oct; 6(10):1289–1299. doi: 10.1038/s41564-021-00962-4.

Figueroa J, Andreoni J, Densen P. Complement deficiency states and meningococcal disease. Immunol Res. 1993; 12(3):295–311. doi: 10.1007/BF02918259.

Gamboa D, Ho MF, Bendezu J, Torres K, Chiodini PL, Barnwell JW, Incardona S, Perkins M, Bell D, McCarthy J, Cheng Q. A large proportion of P. falciparum isolates in the Amazon region of Peru lack pfhrp2 and pfhrp3: implications for malaria rapid diagnostic tests. PLoS One. 2010 Jan; 5(1):e8091. doi: 10.1371/journal.pone.0008091.

Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014 May; 10(5):e1004383. doi: 10.1371/journal.pgen.1004383.

Gilchrist JJ, Rautanen A, Fairfax BP, Mills TC, Naranbhai V, Trochet H, Pirinen M, Muthumbi E, Mwarumba S, Njuguna P, Mturi N, Msefula CL, Gondwe EN, MacLennan JM, Chapman SJ, Molyneux ME, Knight JC, Spencer CCA, Williams TN, MacLennan CA, et al. Risk of nontyphoidal Salmonella bacteraemia in African children is modified by STAT4. Nature Communications. 2018; 9(1):1014. https://doi.org/10.1038/s41467-017-02398-z, doi: 10.1038/s41467-017-02398-z.

Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, Bouman H, Abascal F, Haber M, Tachmazidou I, Mathieson I, Ekoru K, DeGorter MK, Nsubuga RN, Finan C, Wheeler E, Chen L, Cooper DN, Schiffels S, Chen Y, et al. Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. Cell. 2019 Oct; 179(4):984–1002. doi: 10.1016/j.cell.2019.10.004.

Hao Y, Sekine K, Kawabata A, Nakamura H, Ishioka T, Ohata H, Katayama R, Hashimoto C, Zhang X, Noda T, Tsuruo T, Naito M. Apollon ubiquitinates SMAC and caspase-9, and has an essential cytoprotection function. Nature Cell Biology. 2004; 6(9):849–860. https://doi.org/10.1038/ncb1159, doi: 10.1038/ncb1159.

Hauser HP, Bardroff M, Pyrowolakis G, Jentsch S. A giant ubiquitin-conjugating enzyme related to IAP apoptosis inhibitors. J Cell Biol. 1998 Jun; 141(6):1415–1422. doi: 10.1083/jcb.141.6.1415.

Hendriksen ICE, Mwanga-Amumpaire J, von Seidlein L, Mtove G, White LJ, Olaosebikan R, Lee SJ, Tshefu AK, Woodrow C, Amos B, Karema C, Saiwaew S, Maitland K, Gomes E, Pan-Ngum W, Gesase S, Silamut K, Reyburn H, Joseph S, Chotivanich K, et al. Diagnosing severe falciparum malaria in parasitaemic African children: a prospective evaluation of plasma PfHRP2 measurement. PLoS Med. 2012; 9(8):e1001297. doi: 10.1371/journal.pmed.1001297.

Hoogerwerf JJ, van Zoelen MA, Wiersinga WJ, van 't Veer C, de Vos AF, de Boer A, Schultz MJ, Hooibrink B, de Jonge E, van der Poll T. Gene expression profiling of apoptosis regulators in patients with sepsis. J Innate Immun. 2010; 2(5):461–468. doi: 10.1159/000317035.

Hotchkiss RS, Tinsley KW, Swanson PE, Schmieg REJ, Hui JJ, Chang KC, Osborne DF, Freeman BD, Cobb JP, Buchman TG, Karl IE. Sepsis-induced apoptosis causes progressive profound depletion of B and CD4+ T lymphocytes in humans. J Immunol. 2001 Jun; 166(11):6952–6963. doi: 10.4049/jimmunol.166.11.6952.

Hotchkiss RS, Nicholson DW. Apoptosis and caspases regulate death and inflammation in sepsis. Nat Rev Immunol. 2006 Nov; 6(11):813–822. doi: 10.1038/nri1943.

Hotchkiss RS, Tinsley KW, Swanson PE, Grayson MH, Osborne DF, Wagner TH, Cobb JP, Coopersmith C, Karl IE. Depletion of dendritic cells, but not macrophages, in patients with sepsis. J Immunol. 2002 Mar; 168(5):2493–2500. doi: 10.4049/jimmunol.168.5.2493.

Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. G3 (Bethesda). 2011 Nov; 1(6):457–470. doi: 10.1534/g3.111.001198.

Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009 Jun; 5(6):e1000529. doi: 10.1371/journal.pgen.1000529.

Jia R, Bonifacino JS. Negative regulation of autophagy by UBA6-BIRC6-mediated ubiquitination of LC3. Elife. 2019 Nov; 8. doi: 10.7554/eLife.50034.

Kerimov N, Hayhurst JD, Peikova K, Manning JR, Walter P, Kolberg L, Samoviča M, Sakthivel MP, Kuzmin I, Trevanion SJ, Burdett T, Jupp S, Parkinson H, Papatheodorou I, Yates AD, Zerbino DR, Alasoo K. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nature Genetics. 2021; 53(9):1290–1299. https://doi.org/10.1038/s41588-021-00924-w, doi: 10.1038/s41588-021-00924-w.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019 Aug; 37(8):907–915. doi: 10.1038/s41587-019-0201-4.

Le Tulzo Y, Pangault C, Gacouin A, Guilloux V, Tribut O, Amiot L, Tattevin P, Thomas R, Fauchet R, Drénou B. Early circulating lymphocyte apoptosis in human septic shock is associated with poor outcome. Shock. 2002 Dec; 18(6):487–494. doi: 10.1097/00024382-200212000-00001.

Lees JA, Ferwerda B, Kremer PHC, Wheeler NE, Serón MV, Croucher NJ, Gladstone RA, Bootsma HJ, Rots NY, Wijmega-Monsuur AJ, Sanders EAM, Trzciński K, Wyllie AL, Zwinderman AH, van den Berg LH, van Rheenen W, Veldink JH, Harboe ZB, Lundbo LF, de Groot LCPGM, et al. Joint sequencing of human and pathogen genomes reveals the genetics of pneumococcal meningitis. Nat Commun. 2019 May; 10(1):2176. doi: 10.1038/s41467-019-09976-3.

Lo S, Yuan SSF, Hsu C, Cheng YJ, Chang YF, Hsueh HW, Lee PH, Hsieh YC. Lc3 Over-Expression Improves Survival and Attenuates Lung Injury Through Increasing Autophagosomal Clearance in Septic Mice. Annals of Surgery. 2013; 257(2). https://journals.lww.com/annalsofsurgery/Fulltext/2013/02000/Lc3_Over_Expression_Improves_Survival_and.26.aspx.

Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, Schoenherr S, Forer L, McCarthy S, Abecasis GR, Durbin R, L Price A. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016 Nov; 48(11):1443–1448. doi: 10.1038/ng.3679.

Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nature Genetics. 2007; 39(7):906–913. https://doi.org/10.1038/ng2088, doi: 10.1038/ng2088.

Ndila CM, Uyoga S, Macharia AW, Nyutu G, Peshu N, Ojal J, Shebe M, Awuondo KO, Mturi N, Tsofa B, Sepúlveda N, Clark TG, Band G, Clarke G, Rowlands K, Hubbart C, Jeffreys A, Kariuki S, Marsh K, Mackinnon M, et al. Human candidate gene polymorphisms and risk of severe malaria in children in Kilifi, Kenya: a case-control association study. Lancet Haematol. 2018 Aug; 5(8):e333–e345. doi: 10.1016/S2352-3026(18)30107-8.

Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 2016 May; 32(10):1479–1485. doi: 10.1093/bioinformatics/btv722.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017 Apr; 14(4):417–419. doi: 10.1038/nmeth.4197.

Picard C, von Bernuth H, Ku CL, Yang K, Puel A, Casanova JL. Inherited human IRAK-4 deficiency: an update. Immunol Res. 2007; 38(1-3):347–352. doi: 10.1007/s12026-007-0006-2.

Quach H, Rotival M, Pothlichet J, Loh YHE, Dannemann M, Zidane N, Laval G, Patin E, Harmant C, Lopez M, Deschamps M, Naffakh N, Duffy D, Coen A, Leroux-Roels G, Clément F, Boland A, Deleuze JF, Kelso J, Albert ML, et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell. 2016 Oct; 167(3):643–656. doi: 10.1016/j.cell.2016.09.024.

Rautanen A, Pirinen M, Mills TC, Rockett KA, Strange A, Ndungu AW, Naranbhai V, Gilchrist JJ, Bellenguez C, Freeman C, Band G, Bumpstead SJ, Edkins S, Giannoulatou E, Gray E, Dronov S, Hunt SE, Langford C, Pearson RD, Su Z, et al. Polymorphism in a lincRNA Associates with a Doubled Risk of Pneumococcal Bacteremia in Kenyan Children. Am J Hum Genet. 2016 Jun; 98(6):1092–1100. doi: 10.1016/j.ajhg.2016.03.025.

Scott JAG, Berkley JA, Mwangi I, Ochola L, Uyoga S, Macharia A, Ndila C, Lowe BS, Mwarumba S, Bauni E, Marsh K, Williams TN. Relation between falciparum malaria and bacteraemia in Kenyan children: a population-based, case-control study and a longitudinal study. Lancet. 2011 Oct; 378(9799):1316–1323. doi: 10.1016/S0140-6736(11)60888-X.

Silaba M, Ooko M, Bottomley C, Sande J, Benamore R, Park K, Ignas J, Maitland K, Mturi N, Makumi A, Otiende M, Kagwanja S, Safari S, Ochola V, Bwanaali T, Bauni E, Gleeson F, Deloria Knoll M, Adetifa I, Marsh K, et al. Effect of 10-valent pneumococcal conjugate vaccine on the incidence of radiologically-confirmed pneumonia and clinically-defined pneumonia in Kenyan children: an interrupted time-series analysis. The Lancet Global Health. 2019 2021/11/03; 7(3):e337–e346. https://doi.org/10.1016/S2214-109X(18)30491-1, doi: 10.1016/S2214-109X(18)30491-1.

Verhagen AM, Coulson EJ, Vaux DL. Inhibitor of apoptosis proteins and their relatives: IAPs and other BIRPs. Genome Biology. 2001; 2(7):reviews3009.1. https://doi.org/10.1186/gb-2001-2-7-reviews3009, doi: 10.1186/gb-2001-2-7-reviews3009.

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006 Mar; 4(3):e72. doi: 10.1371/journal.pbio.0040072.

Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, Abbasi-Kangevari M, Abbastabar H, Abd-Allah F, Abdelalim A, Abdollahi M, Abdollahpour I, Abolhassani H, Aboyans V, Abrams EM, Abreu LG, Abrigo MRM, Abu-Raddad LJ, Abushouk AI, Acebedo A, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020 2021/11/03; 396(10258):1204–1222. https://doi.org/10.1016/S0140-6736(20)30925-9, doi: 10.1016/S0140-6736(20)30925-9.

Wakefield J. Bayes factors for genome-wide association studies: comparison with P-values. Genet Epidemiol. 2009 Jan; 33(1):79–86. doi: 10.1002/gepi.20359.

Watson JA, Ndila CM, Uyoga S, Macharia A, Nyutu G, Mohammed S, Ngetsa C, Mturi N, Peshu N, Tsofa B, Rockett K, Leopold S, Kingston H, George EC, Maitland K, Day NP, Dondorp AM, Bejon P, Williams T, Holmes CC, et al. Improving statistical power in severe malaria genetic association studies by augmenting phenotypic precision. Elife. 2021 Jul; 10. doi: 10.7554/eLife.69698.

Watson JA, Uyoga S, Wanjiku P, Makale J, Nyutu GM, Mturi N, George EC, Woodrow CJ, Day NP, Bejon P, Opoka RO, Dondorp AM, John CC, Maitland K, Williams TN, White NJ. Improving the diagnosis of severe malaria in African children using platelet counts and plasma Pf HRP2 concentrations. medRxiv. 2021; https://www.medrxiv.org/content/early/2021/10/30/2021.10.27.21265557, doi: 10.1101/2021.10.27.21265557.

Williams TN, Uyoga S, Macharia A, Ndila C, McAuley CF, Opi DH, Mwarumba S, Makani J, Komba A, Ndiritu MN, Sharif SK, Marsh K, Berkley JA, Scott JAG. Bacteraemia in Kenyan children with sickle-cell anaemia: a retrospective cohort and case-control study. Lancet. 2009 Oct; 374(9698):1364–1370. doi: 10.1016/S0140-6736(09)61374-X.

Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011 Jan; 88(1):76–82. doi: 10.1016/j.ajhg.2010.11.011.

Tables

Table 1. Demographics & clinical characteristics of Kenyan children with severe malaria.

Model	Numbers	Sex (female)	Age (months)	Bacteraemia	Mortality
*Pf*HRP2/Plt	Total (n=1,400)	695 (49.6%)	29 (17-44)	51 (3.6%)	155 (11.1%)
	P(SM|Data)>0.5 (n=975)	497 (51.0%)	29 (18-45)	23 (2.4%)	94 (7.4%)
	P(SM|Data)<0.5 (n=425)	198 (46.6%)	28 (16-43)	28 (6.6%)	61 (14.4%)
WBC/Plt	Total (n=2,220)	1,074 (48.4%)	28 (15-43)	78 (3.5%)	256 (11.6%)
	P(SM|Data)>0.5 (n=1,279)	623 (48.7%)	29 (17-44)	32 (2.5%)	106 (8.4%)
	P(SM|Data)<0.5 (n=941)	451 (47.9%)	25 (13-40)	46 (4.9%)	150 (15.9%)

Mortality reflects in-patient deaths. Figures are absolute numbers with percentages or interquartile ranges in parentheses. P(SM|Data) reflects the probability of 'true' severe malaria estimated from each model (*Pf*HRP2/Platelet count, White blood cell count/Platelet count).

Table 2. Demographics & clinical characteristics of GWAS study samples.

				Severe malaria subtypes		Concurrent infection		
	Numbers	Sex (female)	Age (months)	SMA	CM	Malaria	Bacteraemia	Mortality
Bacteraemia (overall)	1,445	614 (43%)	14 (5-34)			94 (12%)		358 (26%)
Acinetobacter	118	45 (38%)	13 (3-28)			11 (13%)		12 (10%)
β-haemolytic *Strep.*	130	60 (46%)	5 (1-20)			6 (8%)		37 (30%)
E. coli	141	58 (41%)	11 (6-25)			12 (15%)		45 (34%)
Hib	113	53 (47%)	12 (5-25)			3 (8%)		29 (26%)
NTS	159	75 (47%)	15 (9-26)			15 (25%)		31 (20%)
S. pneumoniae	390	151 (39%)	23 (9-61)			20 (9%)		86 (23%)
S. aureus	152	64 (42.1%)	26 (9-88)			15 (15%)		22 (15%)
Other	242	110 (46%)	10 (1-28)			10 (8%)		96 (41%)
Malaria (overall)	1,143	559 (49%)	27 (16-41)	298 (26%)	697 (61%)		40 (4%)	140 (12%)
P(SM|Data)<0.5	375	176 (47%)	28 (17-42)	62 (17%)	262 (70%)		23 (6%)	62 (17%)
P(SM|Data)>0.5	768	383 (50%)	26 (15-40)	236 (31%)	435 (57%)		17 (2%)	17 (2%)
Controls (SNP 6.0)	1,895	955 (50%)	*					
Controls (Omni 2.5M)	917	451 (49.2%)						

P(SM|Data) reflects the probability of 'true' severe malaria estimated from *Pf*HRP2 concentration and platelet count or white blood cell count and platelet count. Blood cultures were taken from all children severe malaria at admission.

*Control children were recruited between 3 and 12 months of age and have been subject to longitudinal follow-up.

SMA, severe malarial anaemia; CM, cerebral malaria.

Table 3. 95% credible SNP set of invasive bacterial disease association.

SNP	Effect allele	Chr	BP	MAF	Info score	OR (95% CI)	p-value
rs183868412	T	2	32,478,169	0.021	0.956	2.13 (1.65-2.74)	4.64×10^{-9}
rs139827594	G	2	32,402,640	0.020	0.966	2.12 (1.65-2.73)	4.96×10^{-9}
rs144257579	G	2	32,507,619	0.021	0.954	2.11 (1.64-2.72)	6.82×10^{-9}
rs145056232	C	2	32,503,024	0.021	0.955	2.11 (1.64-2.72)	6.86×10^{-9}
rs145315025	G	2	32,502,654	0.021	0.955	2.11 (1.64-2.72)	6.87×10^{-9}
rs143909151	T	2	32,531,452	0.021	0.962	2.11 (1.64-2.71)	8.01×10^{-9}
rs150430979	T	2	32,614,746	0.021	0.955	2.11 (1.64-2.72)	8.18×10^{-9}

MAF, minor allele frequency. CI, confidence interval. Genomic coordinates are GRCh38.

Table 4. Effect of rs183868412 genotype on risk of invasive bacterial disease stratifies by self-reported ethnicity.

Discovery population			Numbers	Genotypes	MAF	OR (95% CI)	p-value
Giriama	Cases	Overall	1,232	0/56/1,176	0.023	1.97 (1.30-3.01)	$p = 0.0015$
		Bacteraemia	558	0/38/520	0.034		
		SM - P(SM|Data)<0.5	199	0/12/187	0.030		
		SM - P(SM|Data)>0.5	475	0/6/469	0.006		
	Controls		1,269	0/41/1,228	0.016		
Chonyi	Cases	Overall	503	0/38/465	0.038	2.18 (1.34-3.54)	$p = 0.0017$
		Bacteraemia	238	0/27/211	0.057		
		SM - P(SM|Data)<0.5	105	0/4/101	0.019		
		SM - P(SM|Data)>0.5	160	0/7/153	0.022		
	Controls		1,057	0/43/1,014	0.020		
Kauma	Cases	Overall	154	0/8/146	0.026	1.20 (0.50-2.85)	$p = 0.686$
		Bacteraemia	70	0/6/64	0.043		
		SM - P(SM|Data)<0.5	25	0/1/24	0.020		
		SM - P(SM|Data)>0.5	59	0/1/58	0.008		
	Controls		318	0/20/298	0.031		
Other	Cases	Overall	219	1/16/202	0.041	2.46 (1.01-5.96)	$p = 0.047$
		Bacteraemia	101	1/8/92	0.050		
		SM - P(SM|Data)<0.5	38	0/5/33	0.066		
		SM - P(SM|Data)>0.5	80	0/3/77	0.019		
	Controls		165	0/7/158	0.021		
Total	Cases	Overall	2,588	3/164/2,421	0.033	2.13 (1.65-2.74)	$p = 4.64 \times 10^{-9}$
		Bacteraemia	1,445	3/125/1,317	0.045		
		SM - P(SM|Data)<0.5	375	0/20/355	0.027		
		SM - P(SM|Data)>0.5	768	0/19/749	0.012		
	Controls		2,812	0/111/2,701	0.020		

Self-reported ethnicity data is missing in 482 samples (480 of which are cases). Effect estimates derived with weighted logistic regression. P(SM|Data) represent the probability of 'true' severe malaria estimated from plasma PfHRP2 concentration and platelet count (n=909) or white blood cell count and platelet count (n=234). OR, odds ratio. MAF, minor allele frequency. CI, confidence interval.

Table 5. Effect of rs183868412 genotype on risk of invasive bacterial disease in Kenyan children.

			Numbers	Genotypes	MAF	OR (95% CI)	p-value
Discovery	Cases	Overall	2,588	3/164/2,421	0.033	2.13 (1.65-2.74)	$p = 4.64 \times 10^{-9}$
		Bacteraemia*	1,445	3/125/1,317	0.045	2.12 (1.60-2.82)	$p = 1.94 \times 10^{-7}$
		SM - P(SM|Data)<0.5*	375	0/20/355	0.027	2.37 (1.27-4.43)	$p = 6.82 \times 10^{-3}$
		SM - P(SM|Data)>0.5*	768	0/19/749	0.012	1.07 (0.57-2.01)	$p = 0.823$
	Controls		2,812	0/111/2,701	0.020		
Replication	Cases		434	0/24/410	0.028	2.85 (1.54-5.28)	$p = 8.90 \times 10^{-4}$
	Controls		1,258	0/28/1,230	0.011		
Meta-analysis	Cases		3,022	3/188/2,831	0.032	2.23 (1.76-2.80)	$p = 2.35 \times 10^{-11}$
	Controls		4,070	0/139/3,931	0.017		

*estimates derived from multinomial logistic regression model. P(SM|Data) represent the probability of 'true' severe malaria estimated from plasma *Pf*HRP2 concentration and platelet count (n=909) or white blood cell count and platelet count (n=234). SM, severe malaria. MAF, minor allele frequency. CI, confidence interval.

Table 6. rs183868412 frequencies in Africa.

Population	Number	MAF
Gambia	2,605	0.011
Mali	183	0.021
Burkina Faso	596	0.009
Ghana	320	0.014
Nigeria	22	0.024
Cameroon	685	0.031
Malawi	1,317	0.034
Tanzania	402	0.028
Kenya	1,681	0.017

Numbers reflect healthy control samples. MAF, minor allele frequency.

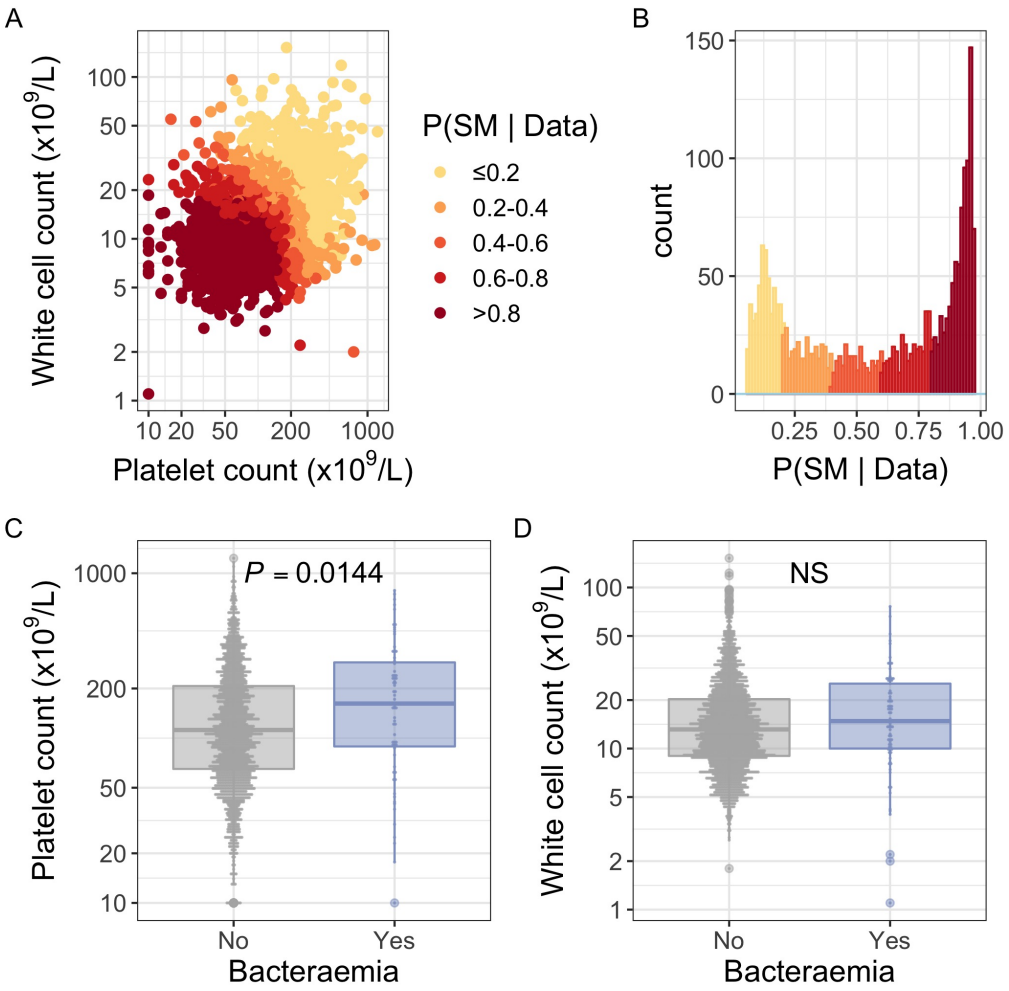


Figure 1–Figure supplement 1. White blood cell and platelet count as predictors of severe malaria. (A) Distribution of white blood cell and platelet count among Kenyan children (n=2,200) with a clinical diagnosis of severe malaria. Points are coloured according to the probability of 'true' severe malaria given the data. (B) Distribution of 'true' severe malaria probabilities estimated from platelet count and white blood cell count. (C) Platelets counts in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia. (D) White blood cell counts in children with a clinical diagnosis of severe malaria with and without concomitant bacteraemia.

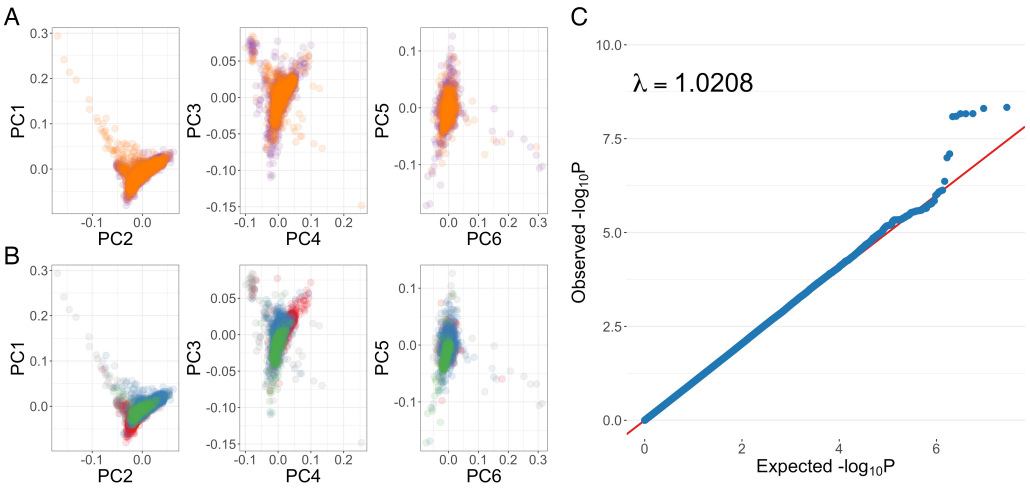


Figure 3–Figure supplement 1. Genome-wide association analysis quality control. (A) Principal components of genome-wide genotyping data in discovery samples. Individuals are color-coded according to genotyping platform; Affymetrix SNP 6.0 in purple, Illumina Omni 2.5M in orange. (B) Principal components of genome-wide genotyping data in discovery samples. Individuals are color-coded according to self-reported ethnicity; Chonyi in red, Giriama in blue, Kauma in green and other in grey. (C) Quantile-quantile plot of invasive bacterial infection in Kenyan children. QQ plot of weighted logistic regression GWAS of invasive bacterial disease in Kenyan children (2,588 cases, 2,812 controls).

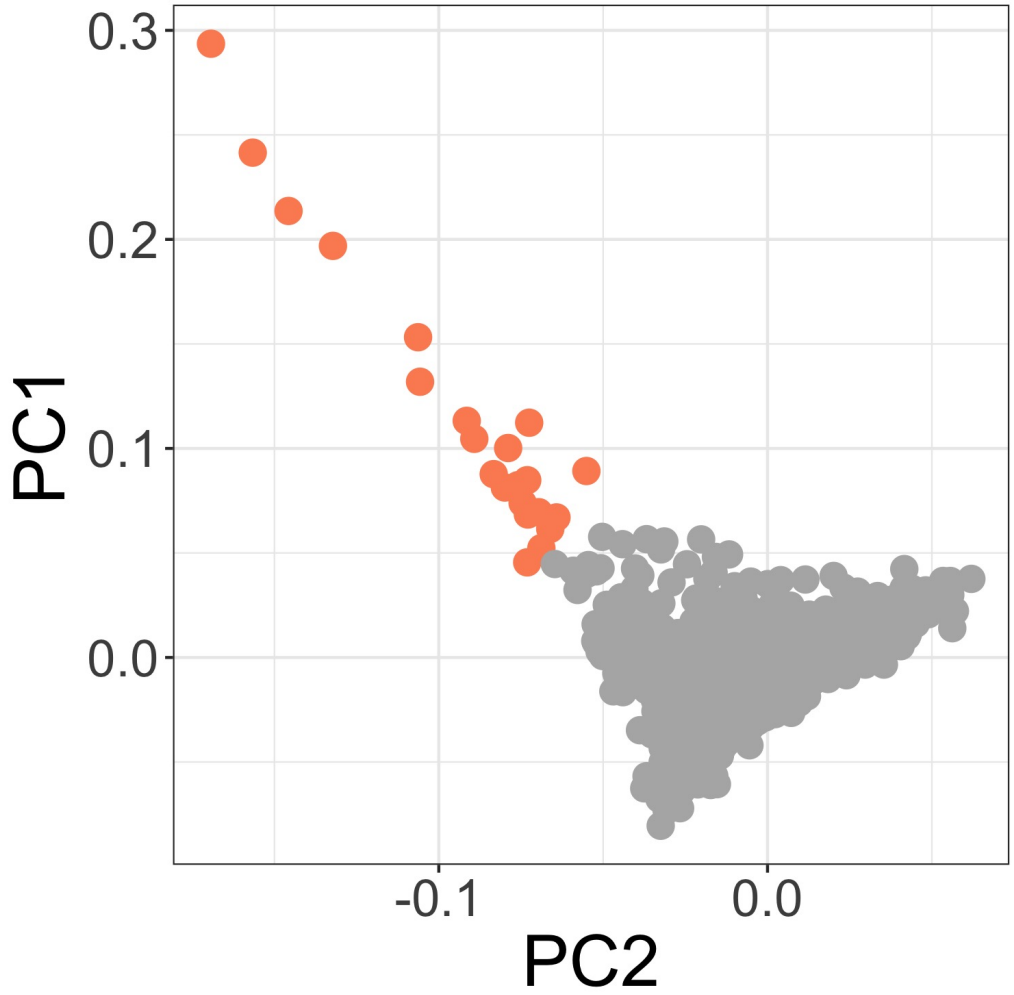
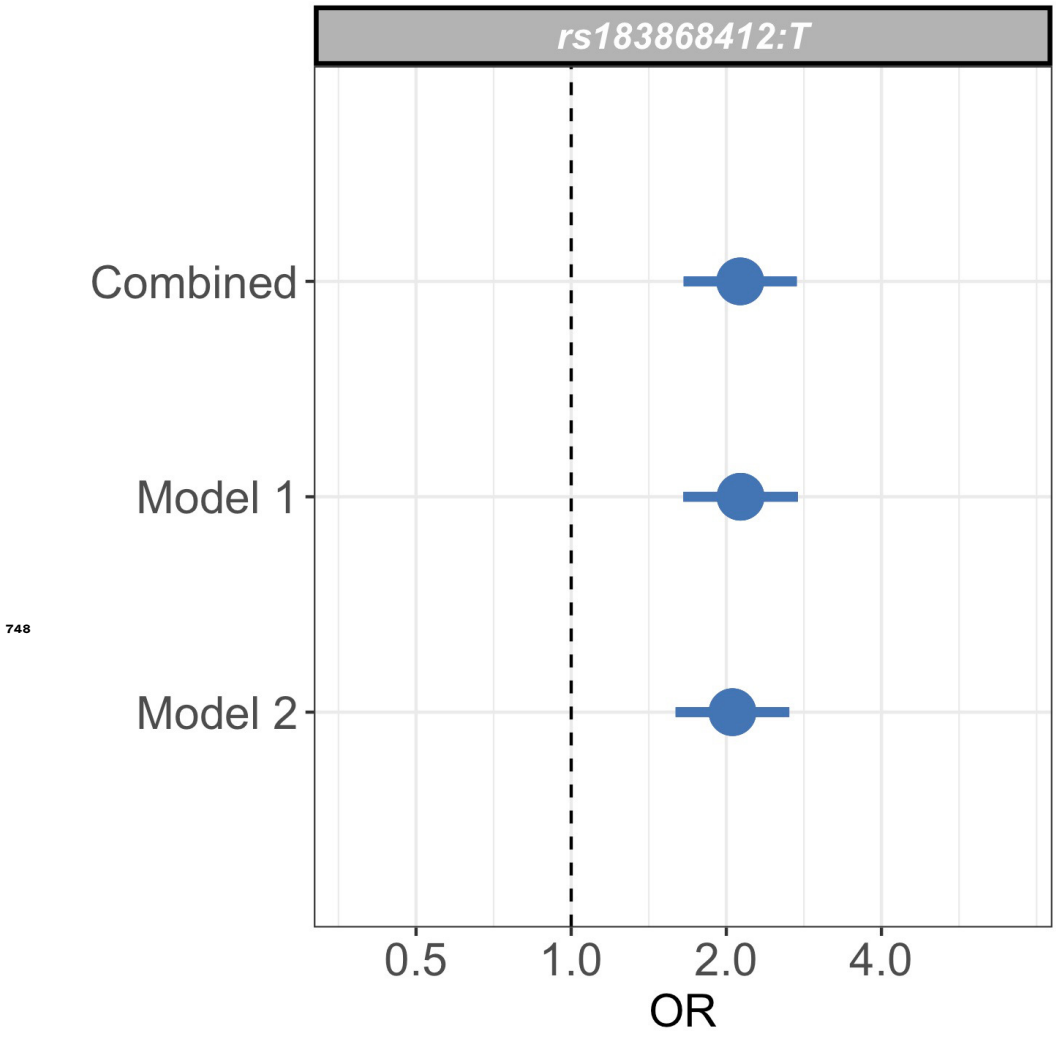


Figure 3–Figure supplement 2. Principal components of genome-wide genotyping data in discovery samples. Outlier samples identified by ABERRANT (n=22) are highlighted (orange).



748

Figure 3–Figure supplement 3. Sensitivity analysis of case weights. Comparison of effect sizes of the rs183868412:T association with invasive bacterial infection in the combined analysis (bacteraemia cases = 1,445, severe malaria cases = 1,143, controls = 2,812), and restricted to cases weights calculated with *Pf*HRP2 plasma concentration and platelet count (Model 1; bacteraemia cases = 1,445, severe malaria cases = 909, controls = 2,812) or white cell count and platelet count (Model 2; bacteraemia cases = 1,445, severe malaria cases = 909, controls = 2,812).

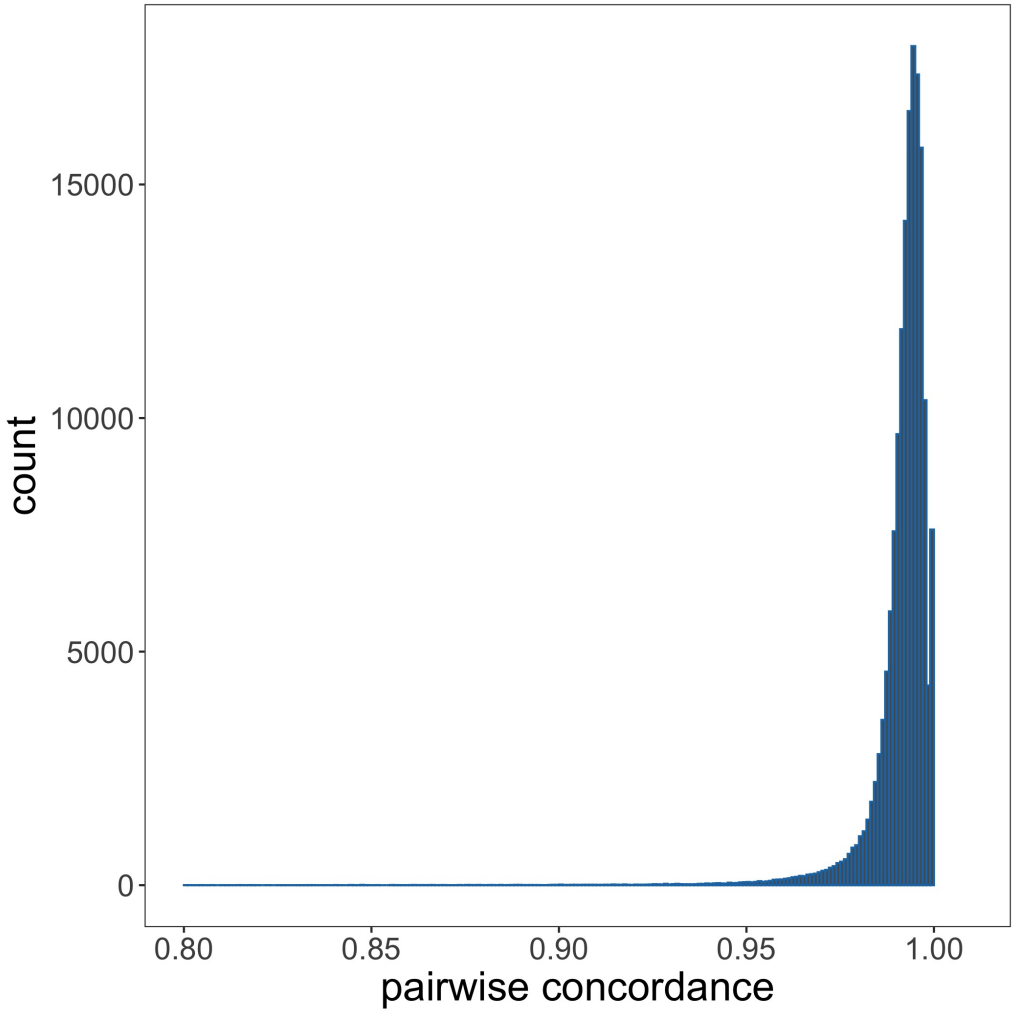


Figure 3–Figure supplement 4. Genotyping concordance between Illumina and Affymetrix platforms. Pairwise genotyping concordance between samples genotyped on both Affymetrix SNP 6.0 and Illumina Omni 2.5M platforms.

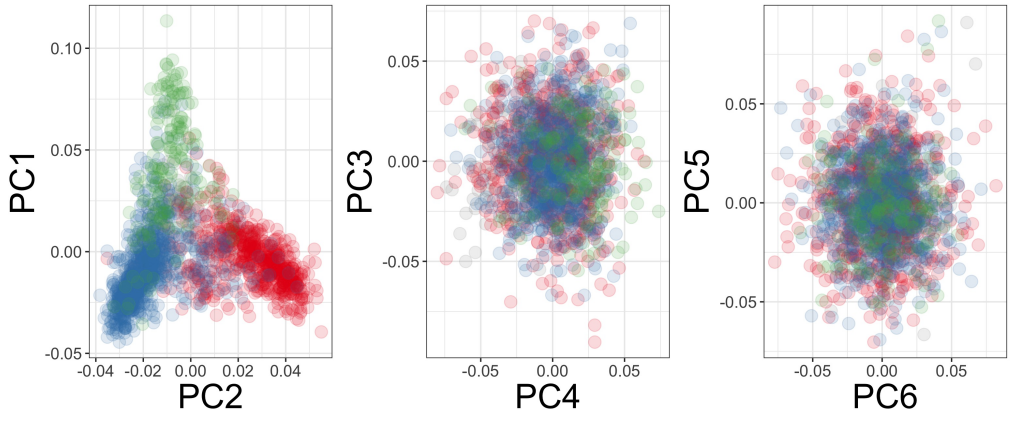


Figure 4–Figure supplement 1. Principal components of genome-wide genotyping data in replication samples. Individuals are color-coded according to self-reported ethnicity; Chonyi in red, Giriama in blue, Kauma in green and other in grey.