

1 Title

2 Investigating a causal role for neutrophil count on *P. falciparum* severe malaria: a Mendelian
3 Randomization study

4 Authors

5 Andrei-Emil Constantinescu^{1,2,3}, David A. Hughes^{1,2}, Caroline J. Bull^{1,2,3}, Kathryn Fleming⁴,
6 Ruth E. Mitchell^{1,2}, Jie Zheng^{5,6}, Siddhartha Kar^{1,2}, Nicholas J. Timpson^{1,2}, Borko Amulic^{4*},
7 Emma E. Vincent^{1,2,3*}

8

9 ¹MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK

10 ²Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, UK

11 ³School of Translational Health Sciences, University of Bristol, Bristol, UK

12 ⁴School of Cellular and Molecular Medicine, University of Bristol, Bristol, UK

13 ⁵Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and
14 Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine,
15 Shanghai, China.

16 ⁶Shanghai National Clinical Research Center for Metabolic Diseases, Key Laboratory for
17 Endocrine and Metabolic Diseases of the National Health Commission of the PR China,
18 Shanghai National Center for Translational Medicine, Ruijin Hospital, Shanghai Jiao Tong
19 University School of Medicine, Shanghai, China.

20 * Joint senior authors

21

22 Corresponding authors

23 Correspondence to Borko Amulic (borko.amulic@bristol.ac.uk) and Emma E. Vincent
24 (emma.vincent@bristol.ac.uk).

25

26

27

28 **Abstract**

29 **Background**

30 Malaria caused by *P. falciparum* imposes a tremendous public health burden on people living
31 in sub-Saharan Africa. Severe malaria is associated with high morbidity and mortality and
32 results from complications such as cerebral malaria, severe anaemia or respiratory distress.
33 Individuals living in malaria endemic regions often have a reduced circulating neutrophil
34 count due to a heritable phenomenon called ‘benign ethnic neutropenia’ (BEN). Neutrophils
35 defend against bacterial infections but have been shown to be detrimental in pre-clinical
36 malaria models, raising the possibility that reduced neutrophil counts modulate severity of
37 malaria in susceptible populations. We tested this hypothesis by performing a genome-wide
38 association study (GWAS) of circulating neutrophil count and a Mendelian randomization
39 (MR) analysis of neutrophil counts on severe malaria in individuals of predominantly African
40 ancestry.

41

42 **Results**

43 We carried out a GWAS of neutrophil count in individuals associated to an African
44 continental ancestry group within UK Biobank (N=5,976). We identified previously
45 unknown loci regulating neutrophil count in a non-European population. This was followed
46 by a two-sample bi-directional MR analysis between neutrophil count and severe malaria
47 (MalariaGEN, N=17,056). We identified 73 loci ($r^2=0.1$) associated with neutrophil count,
48 including the well-known rs2814778 variant responsible for BEN. The greatest evidence for
49 an effect was found between neutrophil count and severe anaemia, although the confidence
50 intervals crossed the null. MR analyses failed to suggest evidence for an effect of the
51 combined severe malaria syndromes or individual subtypes (severe malaria anaemia, cerebral
52 malaria, other severe malaria) on neutrophil count.

53

54 **Conclusion**

55 Our GWAS of neutrophil count revealed unique loci present in individuals of African
56 ancestry. We note that a small sample-size reduced our power to identify variants with low
57 allele frequencies and/or low effect sizes in our GWAS. Our work highlights the need for
58 conducting large-scale biobank studies in Africa and for further exploring the link between
59 neutrophils and severe malarial anemia.

60

61 **Keywords**

62 Malaria, Neutrophil count, Mendelian randomization, GWAS, African ancestry

63

64 **Introduction**

65 Malaria is a mosquito-transmitted disease that annually affects approximately 215 million
66 people [1,2]. The disease is caused by protozoan parasites of the *Plasmodium* genus:
67 *Plasmodium falciparum* (*P. falciparum*) causes life-threatening disease in sub-Saharan Africa
68 and accounts for almost all malaria deaths, while *P. vivax* leads to a milder disease that is
69 nonetheless associated with a significant public health burden in diverse geographical regions
70 [2].

71

72 *P. falciparum* malaria causes approximately 400,000-600,000 deaths each year, primarily in
73 African children under the age of five [1]. The majority of *P. falciparum* malaria cases
74 consist of uncomplicated febrile illness, however a portion of nonimmune infected
75 individuals succumb to severe malaria, which can manifest as cerebral malaria, severe
76 anemia, acute respiratory distress or kidney injury [3]. *Plasmodium* resides and proliferates in

77 red blood cells (RBCs) and pathology is triggered by cytoadherence of infected RBCs
78 (iRBCs) to microcapillary endothelia in different organs, which can lead to vascular
79 occlusion and endothelial permeability [3]. Inflammation plays a key role in both facilitating
80 iRBC sequestration [4] and in tissue damage [3,5,6]. In cerebral malaria, the deadliest form
81 of the disease, iRBCs sequester in the neurovasculature, provoking blood brain barrier
82 permeabilization, vascular leak and brain swelling [3].

83

84 Malaria has been the biggest cause of childhood deaths over the past 5000 years [7]. As such,
85 it has exerted the strongest known selective pressure on the human genome and has resulted
86 in the selection of various polymorphisms that confer *Plasmodium* tolerance or resistance.
87 Among the most prominent examples are haemoglobin S (Hbs; sickle cell trait) [8] and
88 alpha-thalassemia variants [9], both of which are common in malaria endemic regions despite
89 causing disease in the homozygous state [7]. The HbS polymorphism in the heterozygous
90 state confers the greatest protection (effect size >80%; [7,10]). The heritability of severe
91 malaria is estimated to be around 30% [11,12] but the cumulative effect of the
92 aforementioned variants is thought to only be 2% [7,11], suggesting that polygenic
93 interactions may account for a large part of the missing heritability of this complex disease.

94

95 Individuals living in malaria-endemic regions, as well as those descended from them, often
96 have reduced numbers of neutrophils in circulation as compared to those living in non-
97 endemic regions. This heritable phenomenon is called ‘benign ethnic neutropenia’ (BEN) and
98 is distinct from life-threatening severe neutropenia. BEN is prominent in South
99 Mediterranean, Middle Eastern, sub-Saharan African and West Indies populations [13]. BEN
100 is estimated to occur in 25-50% of Africans [13–15] and 10.7% of Arabs [16] but in less than
101 1% of people of European ancestry living in the Americas [17]. Neutrophils are essential for

102 immune defense against bacteria and fungi [18], however BEN does not lead to significantly
103 greater susceptibility to infection in the United States [13]. Nevertheless, it remains curious
104 that selection for lower neutrophil counts occurred in sub-Saharan Africa, a region associated
105 with a high infectious disease burden. This observation is partly explained by the finding that
106 in populations of African and Yemenite Jewish ancestry, BEN is strongly associated with a
107 polymorphism in the atypical chemokine receptor 1 (ACKR1/DARC), which encodes the
108 Fy/Duffy antigen, a surface receptor utilized by *P. vivax* to invade RBCs [19]. This variant
109 abolishes expression of ACKR1 on RBCs and is thought to contribute to low prevalence of *P.*
110 *vivax* in sub-Saharan Africa, where the polymorphism is found at levels close to fixation [7].
111 ACKR1, in addition to serving as one of the invasion receptors for *P. vivax*, controls
112 circulating levels of chemokines [20], which also regulate blood neutrophil numbers [20]. It
113 is unclear to what extent other polymorphisms contribute to BEN in individuals living in
114 malaria endemic regions [21].

115

116 Neutrophils have recently been shown to have a detrimental role in malaria, promoting
117 pathogenesis by enhancing sequestration of iRBCs [4] and contributing to inflammatory
118 tissue damage [6,22,23]. Altered neutrophil responses have also been linked to severe
119 malarial anemia in paediatric patients [24]. On the other hand, neutrophils have also been
120 suggested to participate in parasite clearance [25] and in shaping the *Plasmodium* antigenic
121 repertoire [26]. These studies raise the possibility that neutropenia in malaria endemic regions
122 may modulate severity of *P. falciparum* malaria, in addition to conferring resistance to *P.*
123 *vivax*. However, observational studies, such as the ones referenced above, are prone to
124 confounding and reverse causation [27–29]. It is therefore essential to employ additional
125 methods, such as those in population genetics, to study the link between neutrophil count and

126 *P. falciparum* severe malaria, with the overarching aim to improve the health outcomes of the
127 people residing in endemic regions.

128

129 Mendelian randomization (MR) is a method in genetic epidemiology which uses genetic
130 variants as proxies with the aim of providing evidence for causal inference between an
131 exposure and an outcome [27]. As the majority of alleles are assigned randomly at birth, an
132 MR analysis is analogous to that of a randomized control trial (RCT), the most reliable
133 method for evaluating the effectiveness of an intervention [30]. Large-scale studies, such as
134 UK Biobank (UKBB) [31], have increased the potential of MR studies due to the increase in
135 power to detect associations in genome-wide association studies (GWASs) that comes with
136 such a large sample size.

137

138 Recent efforts in genetics have resulted in the generation of hundreds of GWAS using
139 UKBB's non-European participants for many traits in a hypothesis-free manner
140 (<https://pan.ukbb.broadinstitute.org/>). However, the same covariates were used for each trait,
141 and the impact of confounding due to population structure was not studied, this represents a
142 potential limitation for constructing reliable instruments for a MR analysis [32]. A recent
143 study by Chen et al. used individuals of non-European ancestry in UKBB to perform trans-
144 ancestry GWAS of blood cell traits (BCTs) [33]. However, the African continental ancestry
145 groups (CAGs) of UKBB display strong population structure [34]. It therefore remains
146 unclear whether a GWAS of a complex trait, such as neutrophil count, would result in
147 associations that are linked to a biological mechanism, or whether the associations would be a
148 product of confounding due to residual population structure. In order to answer these
149 questions, a more thorough investigation of the sampled dataset is warranted. This becomes

150 even more important when aiming to conduct causal inference analyses in genetic
151 epidemiology, such as two-sample Mendelian randomization [35,36].

152

153 To test the hypothesis that reduced neutrophil counts modulates severity of malaria in
154 susceptible populations, we first performed a GWAS of neutrophil count in individuals
155 associated to the UKBB African continental ancestry group (CAG), described in our previous
156 study [34]. Here, we conducted a series of sensitivity analyses to describe the GWAS results
157 and selection of genetic instruments to proxy for neutrophil count in a MR analysis. We then
158 conducted bi-directional MR to estimate the casual relationship between neutrophil count and
159 SM using data from the MalariaGEN consortium [37].

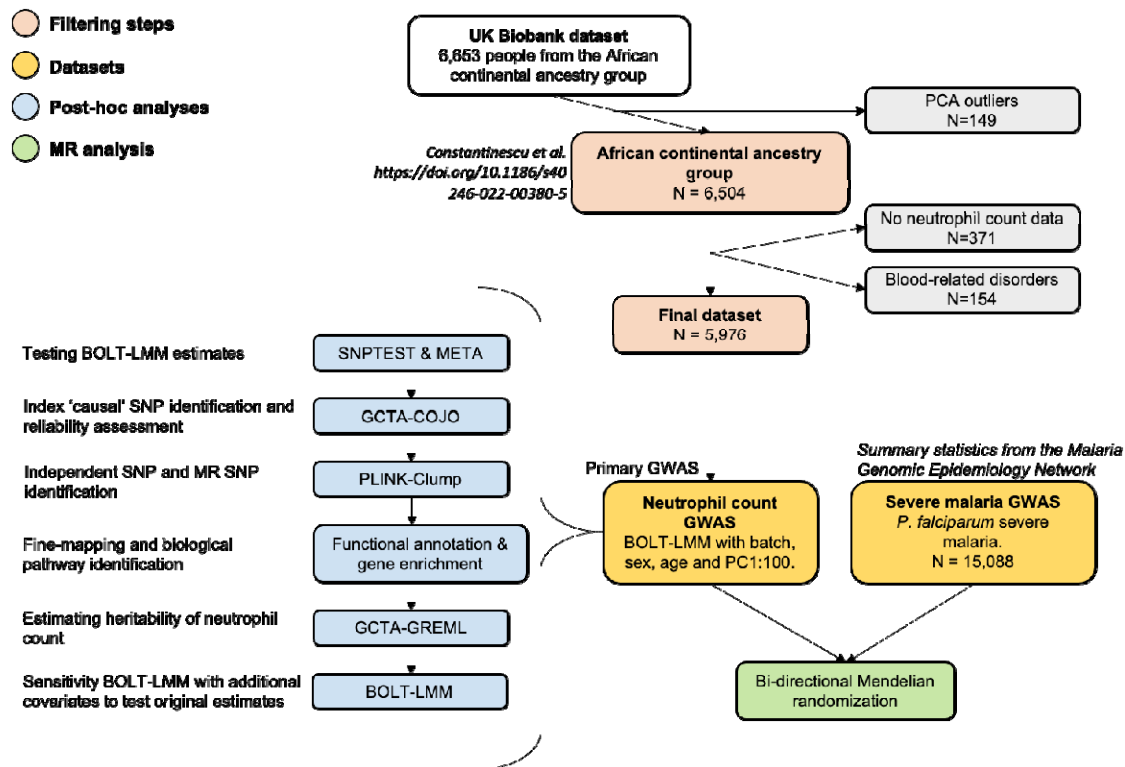
160

161 Materials & Methods

162 Study design

163 6,653 people representing the UKBB African CAG were identified as part of our previous
164 study [34]. After PCA outlier filtering [34], we also excluded those without neutrophil count
165 data and blood-related disorders [38], resulting in a final sample of 5,976. The primary
166 GWAS of neutrophil count used in all other analyses was generated with BOLT-LMM.
167 Several analyses were undertaken afterwards to test the validity of the primary GWAS
168 estimates. Following this, an MR analysis was performed between neutrophil count and
169 severe malaria caused by *P. falciparum* using data from MalariaGEN (**Figure 1**).

170



171

172 **Figure 1.** Study design of the project.

173

174 UK Biobank genetic data

175 UK Biobank's "non-white" British data was studied previously, where 6,653 people
 176 corresponded to the African CAG, of which 6,504 remained (5,989 unrelated; 515 related)
 177 after filtering for principal component analysis (PCA) outliers [34]. These were further
 178 assigned into seven clusters based on a K-means clustering algorithm (K1=527; K2=1,177;
 179 K3=1,176; K4=1,001; K5=1,206; K6=862; K7=184) [34]. This dataset (N=6,504) included
 180 both directly genotyped (N=784,256) and imputed (N=29,363,284) SNPs filtered with a
 181 minor allele count of >20. We filtered out SNPs with an INFO score threshold of 0.3, as it
 182 gives the best balance between data quality and quantity. Another filtering process was a
 183 Hardy-Weinberg equilibrium (HWE) test ($P < 1e-10$), used to identify SNPs of poor
 184 genotyping quality [39]. Finally, related individuals from the dataset were removed, resulting
 185 in 5,509 unrelated people in the filtered African CAG dataset. SNPs with a minor allele count

186 of less than 17 (corresponding to the new sample-size from 20) were removed. 23,530,028
187 SNPs remained after filtering by INFO score, HWE test and minor allele count.

188

189 **UK Biobank phenotypic data**

190 Haematological samples were analysed using four Beckman Coulter LH750 instruments [40].
191 Total white blood cell (WBC) count and neutrophil percentage (%) were measured through
192 the Coulter method, with neutrophil count derived as “neutrophil % / 100 x total WBC” and
193 expressed as 10^9 cells/Litre [40]. Afterwards, the sample collection date was split into year,
194 month, day, and minutes (passed since the start of the day of the appointment visit), while the
195 neutrophil count measurement variable was log-transformed into a variable named “nc_log”,
196 which was used as the default neutrophil count variable throughout the study. Other variables
197 that were used in the main analyses were: age, genetic sex, blood sample device ID, UKBB
198 assessment centre and principal components (PCs) 1 to 100. Filtering was performed based
199 on the selection criteria described by Astle et al. [38] and Chen et al. [33]. Briefly,
200 individuals with disorders/diseases that could affect blood counts (e.g. HIV, leukaemia,
201 congenital anaemias, cirrhosis) were removed, bringing the final sample size to 5,976. This
202 dataset is referred to as “AFR_CAG”.

203

204 **BOLT-LMM GWAS**

205 BOLT-LMM was used as the software to run the primary (main) GWAS. Linkage
206 disequilibrium (LD) scores were generated from the directly genotyped dataset that is
207 required by BOLT-LMM to calibrate the test statistics. After preparing the phenotypic data to
208 match the desired input, BOLT-LMM was run on AFR_CAG adjusting for age, genetic sex,
209 UKBB assessment centre, blood sampling device, sampling year, sampling month, sampling
210 day, minutes passed in sampling day and the first 100 principal components (PCs). Two

211 linear model GWAS in SNPTEST were also completed on each K-means cluster and then
212 meta-analysed: one without accounting for the Duffy SNP rs2814778 called “META-WOD”,
213 and one where the Duffy SNP was included as a covariate, called “META-WD”. Another
214 BOLT-LMM sensitivity run was done with additional covariates to further study the validity
215 of the main GWAS findings (**Supplementary Methods**).

216

217 **Conditional & joint association analysis**

218 We used GCTA-COJO [41,42] to identify independent signals from the BOLT-LMM
219 GWAS, as well as to detect any possible secondary signals arising from a stepwise selection
220 model. SNPs which are close together are usually in LD i.e. their alleles are not random, but
221 correlated [39]. Before running GCTA-COJO, related individuals were filtered out of the
222 dataset. PLINK was then used on this resulting output to perform a greedy filtering of related
223 individuals. Following this step, GCTA-COJO was run on the AFR_CAG filtered dataset to
224 identify conditionally independent SNPs. These were referred to as “index” SNPs in the text.

225

226 **PLINK clumping**

227 After GCTA-COJO, we used PLINK to perform clumping with three different thresholds.
228 The first two represent the thresholds for defining LD independent SNPs for running analyses
229 on the online variant annotation platform Functional Mapping and Annotation (FUMA) [43],
230 while the latter being the clumping conditions used for conducting a Mendelian
231 randomization analysis [44,45].

- 232 1. --clump-p1=5e-8, --clump-r2=0.6, --clump-kb=250
- 233 2. --clump-p1=5e-8, --clump-r2=0.1, --clump-kb=250
- 234 3. --clump-p1=5e-8, --clump-r2=0.001, --clump-kb=10000

235

236 **Heritability analysis**

237 An analysis was conducted with GCTA-GREML to estimate the proportion of variance in
238 neutrophil count explained by all genetic variants present in the filtered AFR_CAG dataset
239 [46], with and without adjusting for the Duffy SNP rs2814778.

240

241 ***P. falciparum* severe malaria genetic data**

242 GWAS summary statistics for *P. falciparum* severe malaria were downloaded from a case-
243 control study that spanned nine African and two Asian countries [37]. In brief, controls
244 samples were gathered from cord blood, and in some cases, from the general population.
245 Cases were diagnosed according to WHO definitions of severe malaria [47] and were
246 categorised according to CM, severe malarial anemia (SMA) and other severe malaria
247 (OTHER) symptoms (**Supplementary Table 1**). The majority of the RSIDs in the
248 MalariaGEN dataset used older identifiers, and some of them had the “kgp” prefix that comes
249 with the Illumina-HumanOmni2.5M array. Ideally, in a two-sample MR setting, the two
250 samples would have a perfect match in the available genetic variants. It is desirable to at least
251 maximise the number of matching variants to test. Therefore, RSID information for the
252 MalariaGEN variants was updated in R by using the filtered AFR_CAG dataset as a
253 reference panel.

254

255 **Meta-analysis of severe malaria African populations**

256 Summary statistics for severe malaria and its sub-phenotypes were generated from a meta-
257 analysis which included individuals from two non-African countries – Vietnam and Papua
258 New Guinea. The inclusion of SNP effect sizes from GWAS conducted in heterogenous
259 population might bias MR estimates [48]. Therefore, per-population summary statistics were

260 downloaded (<https://www.malariagen.net/spl25/>) for each African country in the study and a
261 meta-analysis was conducted on them using METAL [49–51].

262

263 **Mendelian randomization analysis**

264 The “TwoSampleMR” R package [52,53] was used to perform the MR analyses. The two
265 datasets were harmonised i.e. orientated on the same strand and if SNPs were not found in the
266 outcome dataset, we searched for SNP proxies. We then conducted a bi-directional MR
267 analysis, where the effect of neutrophil count on overall severe malaria, along with the three
268 sub-phenotypes was estimated and vice-versa. The main analysis was conducted using an
269 IVW model [54]. Additionally, we ran a sensitivity MR analysis to outline the effect
270 estimates of each SNP on the desired outcome, with IVW and MR-Egger [55,56] estimates
271 where the number of instruments was larger than two and three, respectively.

272

273 **Results**

274 **Analysis of study sample**

275 5,976 out of 6,504 individuals in AFR_CAG remained after filtering for missing data and
276 traits affecting blood cells. The mean value for neutrophil count was 2.9×10^9 cells/litre, as
277 expected this was lower than a European sample (4.21×10^9 cells/Litre) [33,38]. The GWAS
278 sample had a larger proportion of females (57%), was of a higher mean age (39 vs. 58.1
279 years) [57] and slightly higher body mass index (BMI) (27.6 vs. 29.8 kg/m²) [58] than the
280 general UK population (**Table 1**).

281

282 **Table 1.** Description of GWAS sample.

Characteristic	N = 5,9761
----------------	------------

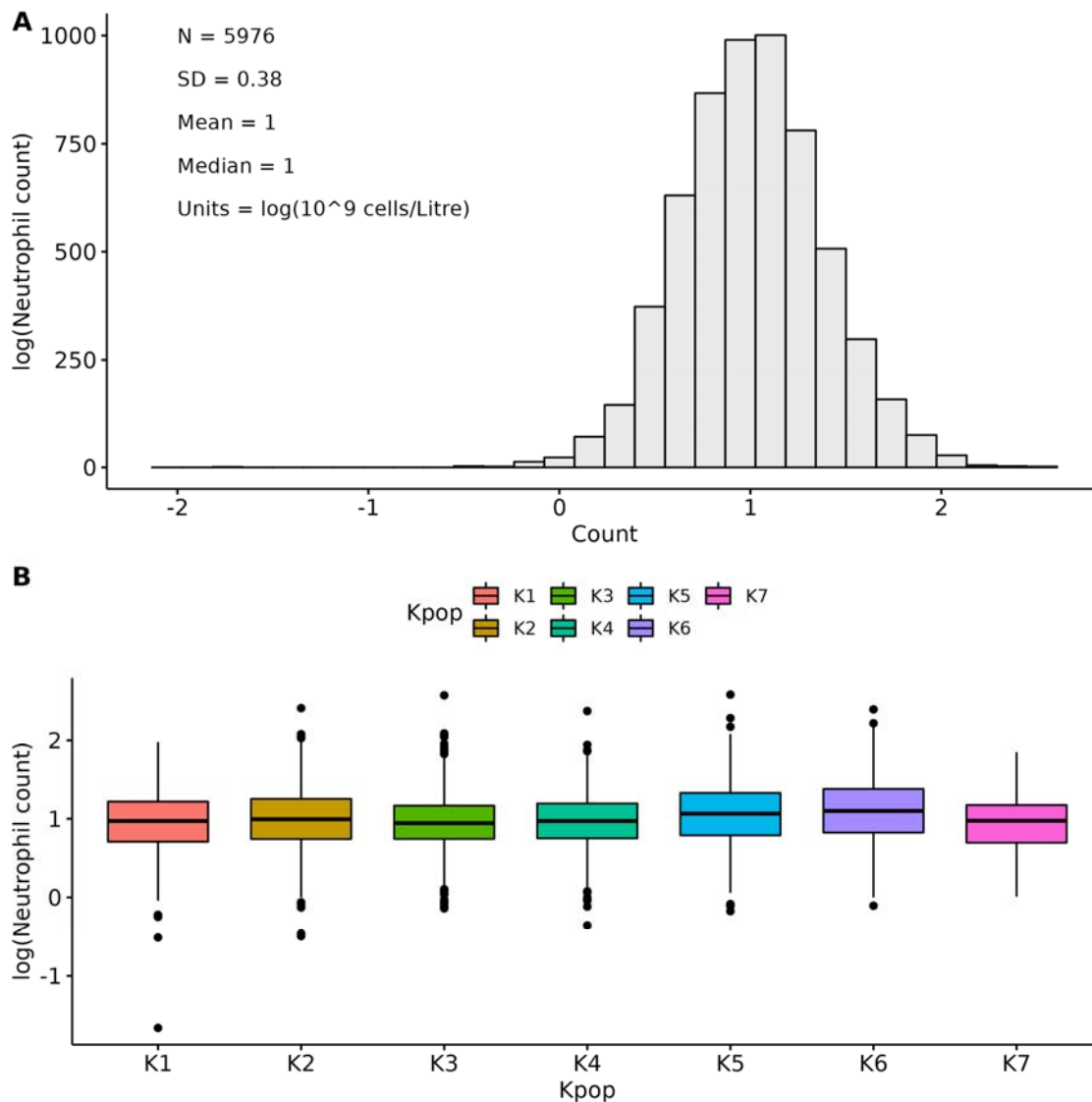
Neutrophil count (10 ⁹ cells/Litre)	2.9 (1.2)
Genetic sex	
Female	3,399 / 5,976 (57%)
Male	2,577 / 5,976 (43%)
Age (years)	51.8 (8.1)
BMI (kg/m ²)	29.8 (5.3)
Missing	100
K-means cluster	
K1	519 / 5,976 (8.7%)
K2	1,142 / 5,976 (19%)
K3	1,150 / 5,976 (19%)
K4	967 / 5,976 (16%)
K5	1,174 / 5,976 (20%)
K6	841 / 5,976 (14%)
K7	183 / 5,976 (3.1%)
<hr/>	
1Mean (SD); n / N (%)	

283

284

285 We used the natural log-transformation, `nc_log`, in the GWAS. There was some variation in
286 `nc_log` between each K-means cluster (Kpop) (**Figure 2B**), although this was low, with the
287 median hovering around 1 (**Figure 2A**).

288



289

290 **Figure 2.** Neutrophil count variation in the GWAS sample. A histogram outlining the
291 distribution of neutrophil levels is shown in the whole AFR_CAG population (A), along with
292 representative boxplots describing neutrophil count variation by K-means cluster sample (B).

293

294 GWAS are usually performed using individuals of a similar genetic background to avoid
295 SNP-trait associations that are biased or are false-positives due to a confounding effect by
296 ancestry [59]. However, even in white British individuals from UKBB, latent population
297 structure can affect SNP effect sizes, which may not be completely removed by adjusting for

298 PCs [60]. Nevertheless, we investigated the number of PCs that should be added into the
299 GWAS to control for population structure. The Tracy-Widom statistic from the EIGENSOFT
300 package [61] indicated over 100 significant PCs. However, there is no exact way to establish
301 how many PCs should be added into a GWAS, although an excessive number of PCs can lead
302 to a reduction in power, while too few might bias GWAS effect sizes due to residual
303 population structure [48]. Previous studies have added 40 to 100 PCs in their UKBB GWA
304 analyses [60,62,63], hence our inclusion of the first 100 PCs as covariates.

305

306 Next, we conducted a power calculation supposing a linear, additive, GWA model. Sample-
307 size is an important factor in GWAS [64,65], and the higher the sample-size, the higher the
308 power to detect SNPs which explain a smaller proportion of the variance (heritability) in a
309 particular trait [66]. The power to detect an association was >80% when the proportion of
310 variance explained by SNPs was higher than 0.75% (**Supplementary Figure 1**).

311

312 **Genome-wide association study**

313 We used BOLT-LMM for the main GWAS, which employs a linear-mixed model algorithm
314 for conducting association testing [67]. It is unknown how well linear mixed model using
315 PCs and kinship matrixes performs in highly stratified population samples with complex
316 demographic histories and unique allele frequencies and linkage disequilibrium [68]. To
317 ensure that results derived by a linear mixed model as implemented by BOLT-LMM were
318 reliable, we also aimed to conduct additional GWAS using a standard linear model on less
319 stratified sub-samples of our sample population – as identified using an unsupervised
320 machine learning methodology (**Supplementary Methods**).

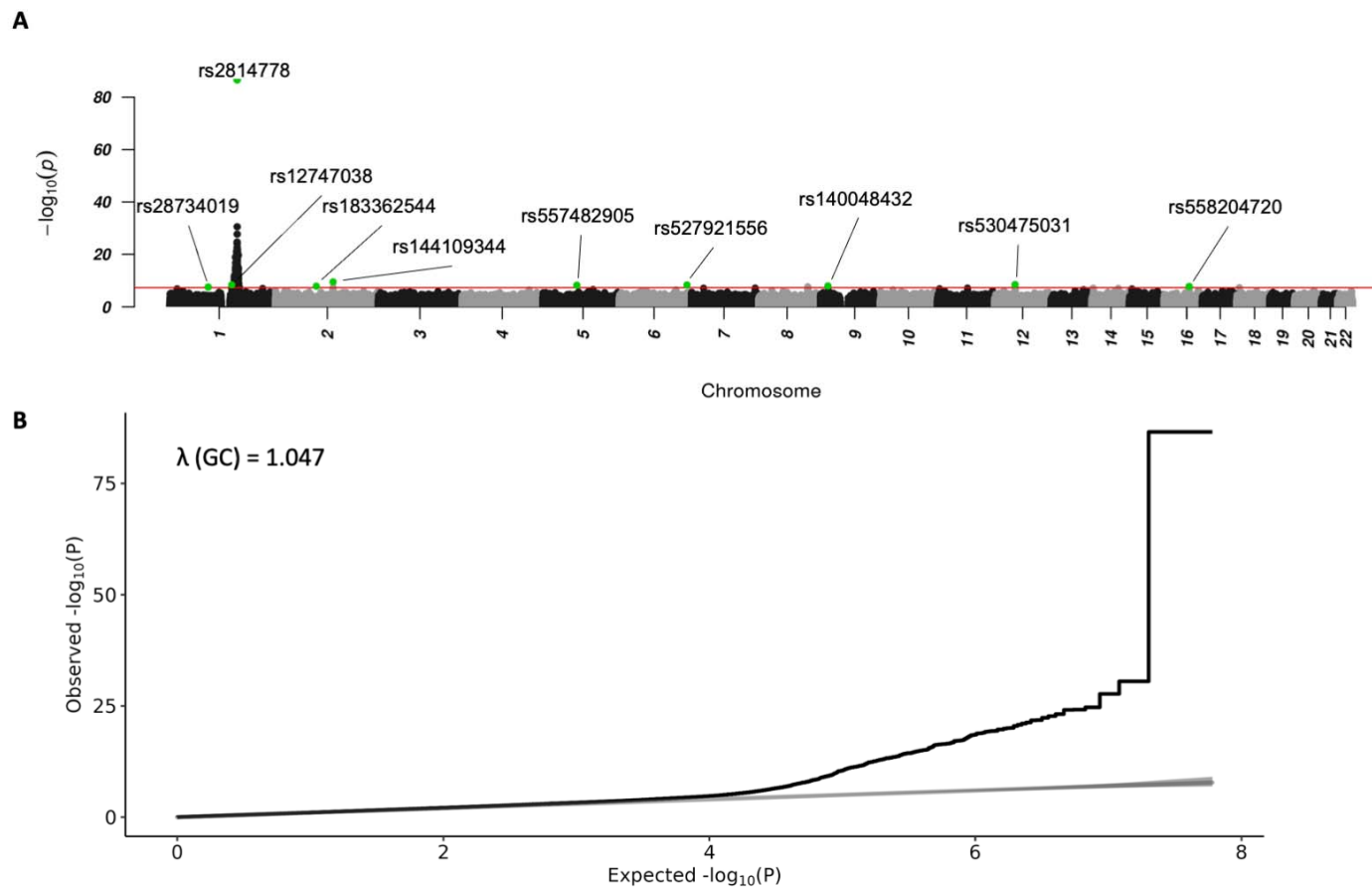
321

322 This AFR_CAG filtered sample was taken forward for further analyses. 704 genetic variants
323 passed the GWAS significance threshold of $P < 5e-8$ in the primary GWAS. Most of these
324 signals were in chromosome 1, in the proximity of the ACKR1-associated rs2814778, which
325 had the lowest P-value across the genome ($2.7E-87$) (**Figure 3A**). The META-WOD GWAS
326 had 373 variants passing the threshold, while the META-WD (with Duffy adjustment)
327 GWAS had 31 significant SNPs, evidencing that most of the identified top signals in META-
328 WOD were likely in LD with rs2814778. The QQ-plot of the BOLT-LMM GWAS did not
329 display an early deviation from the expected P-value, indicating low likelihood of systemic
330 bias in association statistics [69] (**Figure 3B**).

331

332 Next, we aimed to identify which SNPs might causally associate with neutrophil count. To do
333 this, we used a conservative GCTA-COJO approach [42], which yielded 10 index SNPs
334 (**Figure 3**, **Table 2**). Genomic location context of each index SNP is available in
335 **Supplementary Figures 2-4**.

336



337

338 **Figure 3.** Manhattan plot of neutrophil count GWAS. The x-axis is the base-pair position inside each chromosome, while the y-axis is the $-\log$ of
339 the association P-value. A GWAS significance line is drawn to correspond to $P=5 \times 10^{-8}$ on the $-\log(P)$ axis (A). Index SNPs from the GCTA-COJO

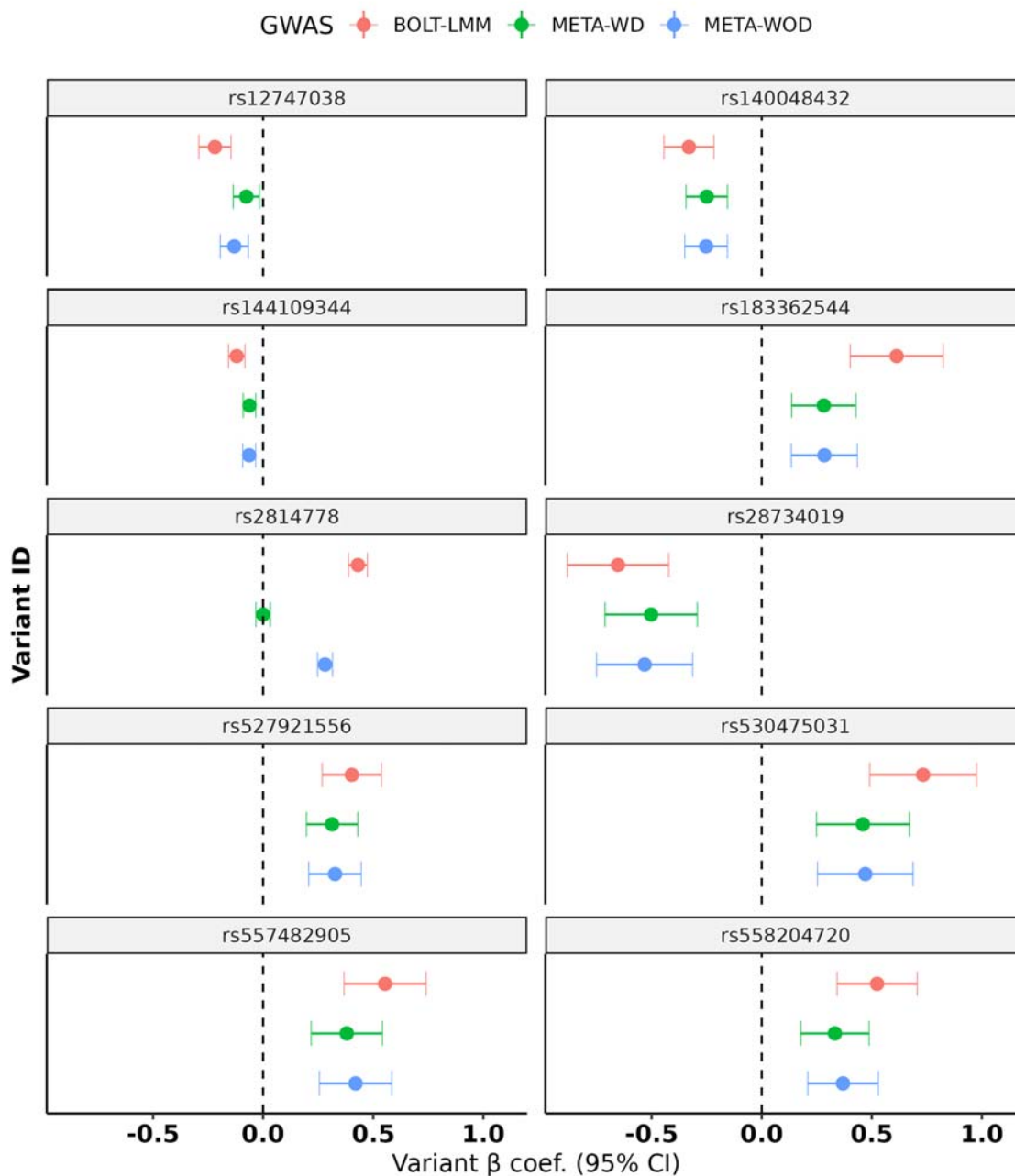
340 run are highlighted in green. QQ-Plot of observed vs. expected P-values for each SNP, along with the genomic inflation factor on the top-left
341 **(B)**.

343 **Table 2.** GCTA-COJO index SNPs. BETA, SE and P BOLT are the regression statistics of the BOLT-LMM neutrophil count GWAS. BETA,
 344 SE and P for META-WOD and META-WD are the regression statistics of the meta-analysed GWAS done on each Kpop, without and with
 345 adjustment for the Duffy SNP, respectively, META-N and META-N-Studies indicate the number of meta-analysed individuals in each Kpop,
 346 and the number of Kpops included in the meta-analysis.

SNP	BETA-BOLT	SE-BOLT	P-BOLT	BETA-META-WOD	BETA-META-WD	META-N	META-N-Studies	CHR	BP (GRCh37)	EA	NEA	EAF
rs2814778	0.43	0.02	2.66E-87	0.28	0	5,793	6	1	159174683	T	C	0.036
rs144109344	-0.12	0.02	3.12E-10	-0.06	-0.06	5,976	7	2	136787730	C	T	0.964
rs530475031	0.73	0.12	3.16E-09	0.47	0.46	4,952	5	12	48810860	G	T	0.998
rs12747038	-0.22	0.04	3.89E-09	-0.13	-0.08	5,976	7	1	146651428	T	G	0.990
rs527921556	0.4	0.07	4.48E-09	0.33	0.31	5,793	6	6	160605701	T	C	0.996
rs557482905	0.55	0.1	5.79E-09	0.42	0.38	3,778	4	5	80629499	C	T	0.998
rs140048432	-0.33	0.06	1.11E-08	-0.25	-0.25	5,976	7	9	17700893	T	C	0.996
rs183362544	0.61	0.11	1.27E-08	0.28	0.28	2,717	4	2	97045902	C	T	0.998
rs558204720	0.52	0.09	1.67E-08	0.37	0.33	1,486	2	16	59472815	T	C	0.998
rs28734019	-0.65	0.12	2.89E-08	-0.53	-0.5	4,124	4	1	90800573	C	T	0.998

347

348 The effect sizes of the primary GWAS index SNPs were compared with those from the
349 SNPTEST/META GWAS. The direction was consistent and effect sizes were similar
350 between the three GWAS, with those generated from the BOLT-LMM run (primary GWAS)
351 being slightly larger, most likely due to the improved sample size (minor allele count) and
352 power of the linear-mixed model (**Figure 4**). As expected, the META-WD effect size for the
353 rs2814778 SNP was zero.
354



355

356 **Figure 4.** Effect estimates of the index SNPs. The beta coefficient for each index SNP is
357 displayed along with 95% CIs. These are displayed for the BOLT-LMM, META-WOD and
358 META-WD GWAS.

359

360 We next investigated the association statistics of the index SNPs in each Kpop. This was
361 done to detect discrepancies in directionality and effect sizes, which could indicate residual

362 population structure or a SNP association with a specific Kpop. Overall, there was agreement
363 in direction, and some variation in effect sizes was detected across Kpops (**Supplementary**
364 **Figure 5**).

365

366 The GCTA-COJO analysis was also run on the two SNPTEST/META GWASs. The META-
367 WOD analysis identified rs2814778, rs138163369 and rs570518709 as index SNPs.
368 Similarly, the META-WD analysis identified rs138163369 and rs570518709. These two
369 latter SNPs were not identified as index SNPs in the BOLT-LMM analysis, but their P-values
370 were similar (rs138163369 – 4.90E-08, 2.28E-08, 1.22E-08; rs570518709 – 8.10E-08, 1.07E-
371 09, 3.03E-09) (**Supplementary Table 2**). As another sensitivity analysis to test the reliability
372 of the BOLT-LMM results, the effect sizes of all GCTA-COJO SNPs were compared in a
373 pair-wise manner across the three GWAS. A regression line was fit through the scatter plots,
374 showing a large degree of correlation between the BOLT-LMM effect sizes and the
375 SNPTEST/META runs (META-WOD $R^2 = 0.91$, META-WD $R^2 = 0.93$) (**Supplementary**
376 **Figure 6**).

377

378 Two PLINK clumping analyses were performed on the filtered AFR_CAG summary
379 statistics using the same clumping parameters on the well-known FUMA platform [43]. Here,
380 193 SNPs were identified as loci at the relaxed threshold of $r^2=0.6$ and 73 independent loci at
381 the stringent threshold of $r^2=0.1$. Finally, 12 top loci were identified at $r^2=0.001$ and a 10Mb
382 window, which are the very conservative MR clumping parameters [44,45]. Furthermore, a
383 FUMA analysis was run on the filtered AFR_CAG dataset for the top loci ($r^2=0.1$). This was
384 done to visualise which genomic locations are affecting neutrophil count and if they are more
385 likely to have a particular genetic function compared to the whole genome i.e. functional
386 variants [70]. Seventeen genomic risk loci were identified (**Supplementary Figure 7A**). The

387 ANNOVAR analysis [71] showed evidence for changes in genetic function enrichment
388 relative to all SNPs in the reference panel. In brief, seven genomic regions were enriched, all
389 indicating an enrichment in genic rather than intergenic spaces (**Supplementary Figure 7B**).

390

391 Next, we investigated the independent SNPs in the GWAS Catalog [72], as we aimed to see
392 if they have been previously associated with WBC count or immunity. Here, SNPs
393 predominantly showed associations with white blood cell count variation, further improving
394 the reliability of the GWAS (**Supplementary Table 3**). We compared the AFR_CAG GWAS
395 with a neutrophil count GWAS meta-analysis of Africans from UKBB and additional studies
396 from Chen et al. [33], and found that 81.71% of the GWAS significant SNPs from Chen et al.
397 were replicated (using the same covariates) in the AFR_CAG dataset ($P < 0.05$)
398 (**Supplementary Table 4**). The Manhattan plots also visually showed a good degree of
399 overlap (**Supplementary Figure 8**), in contrast with a GWAS of neutrophil count in
400 Europeans (**Supplementary Figure 9**) [38]. Finally, SNPs that were top loci at $r^2 = 0.1$ were
401 investigated in the Astle et al. [38] and Chen et al. [33] summary statistics, as well as in the
402 GWAS Catalog [72]. Nineteen genetic variants were not present in these three datasets, 7 of
403 which were index SNPs (**Table 3**).

404

405 **Table 3.** Top loci not found in other studies. Only independent SNPs clumped at $r = 0.1$ are shown.

SNP	CHR	BP (GRCh37)	r0.001 lead?	cojo_index	Novel?	Nearest gene	Type
rs28734019	1	90800573	Yes	Yes	Yes	RNU6-695P	Intergenic
rs61823703	1	159542164	No	No	Yes	OR10AE1P	Intergenic
rs539456851	1	158731459	No	No	Yes	OR6N1	Intergenic
rs371178711	1	158186653	No	No	Yes	RP11-404O13.5	Intergenic
rs146677619	1	158995984	No	No	Yes	IFI16	Intronic
rs11576058	1	161111446	No	No	Yes	UFC1	Intergenic
1:158777618_CT_C	1	158777618	No	No	Yes	OR10AA1P	Downstream
rs183362544	2	97045902	Yes	Yes	Yes	NCAPH	Intergenic
rs11422063	1	159799599	No	No	Yes	SLAMF8	Intronic
rs112483667	1	151651180	No	No	Yes	SNX27	Intronic
rs12406899	1	157540651	No	No	Yes	FCRL4	Intergenic
rs1103805	1	158924741	No	No	Yes	PYHIN1	Intronic
rs557482905	5	80629499	Yes	Yes	Yes	ACOT12	Intronic
rs527921556	6	160605701	Yes	Yes	Yes	SLC22A2	Intronic
rs10096834	8	116281087	Yes	No	Yes	TRPS1	Intergenic
rs140048432	9	17700893	Yes	Yes	Yes	SH3GL2	Intronic
rs530475031	12	48810860	Yes	Yes	Yes	C12orf54	Intronic
rs558204720	16	59472815	Yes	Yes	Yes	LOC105371298	Intronic
rs138163369	18	6492075	Yes	No	Yes	CTD-2124B20.2	Intergenic

406

407 **Heritability analysis**

408 Without adjusting for rs2814778, the genetic variance was estimated at 0.101 (10.1%) (SE =
409 0.018), and the phenotypic variance at 0.133 (13.3%) (SE = 0.003) with an analysis P-value
410 of 2.29e-09. When adjusting for the ACKR1/Duffy SNP, the genetic variance was estimated
411 at 0.050 (5%) (SE = 0.017), twice as low as in the previous analysis, and the phenotypic
412 variance was estimated at 0.123 (12.3%) (SE = 0.002), with the analysis P-value of 1.36E-03
413 (**Supplementary Table 5**).

414

415 **Descriptive analyses of neutrophil count**

416 Next, we aimed to assess if the index SNPs remained associated with neutrophil count when
417 adjusting for variables such as BMI and smoking status. This was done to investigate the
418 reliability of the index SNPs in the context of their relationship with neutrophil count. First,
419 the descriptive statistics of the AFR_CAG dataset were studied with these additional
420 variables (**Supplementary Table 6**). Several variables had missing data or had values
421 assigned as “prefer not to answer” / “not sure” in the case of self-reported traits. There was
422 no evidence of a difference in neutrophil count between these data types and those that were
423 kept in the dataset (**Supplementary Table 7**). 5,310 individuals remained in the dataset after
424 filtering out these data types. The proportion of variance explained for the additional
425 covariates was explored with an ANOVA analysis (**Supplementary Figure 10**).

426

427 **BOLT-LMM with additional covariates**

428 A sensitivity BOLT-LMM GWAS was conducted with six additional covariates on 5,310
429 individuals: UN region of birth, K-means cluster, smoking status, alcohol drinker status,
430 menstrual status and BMI. The association statistics of this sensitivity run and the main
431 BOLT-LMM GWAS run were compared, showing very similar results (**Supplementary**

432 **Table 8**). This provides evidence that the effect of these additional variables on the main
433 GWAS were modest, and that the PCs and kinship matrix derived by BOLT-LMM appears to
434 have accounted for any population stratification.

435

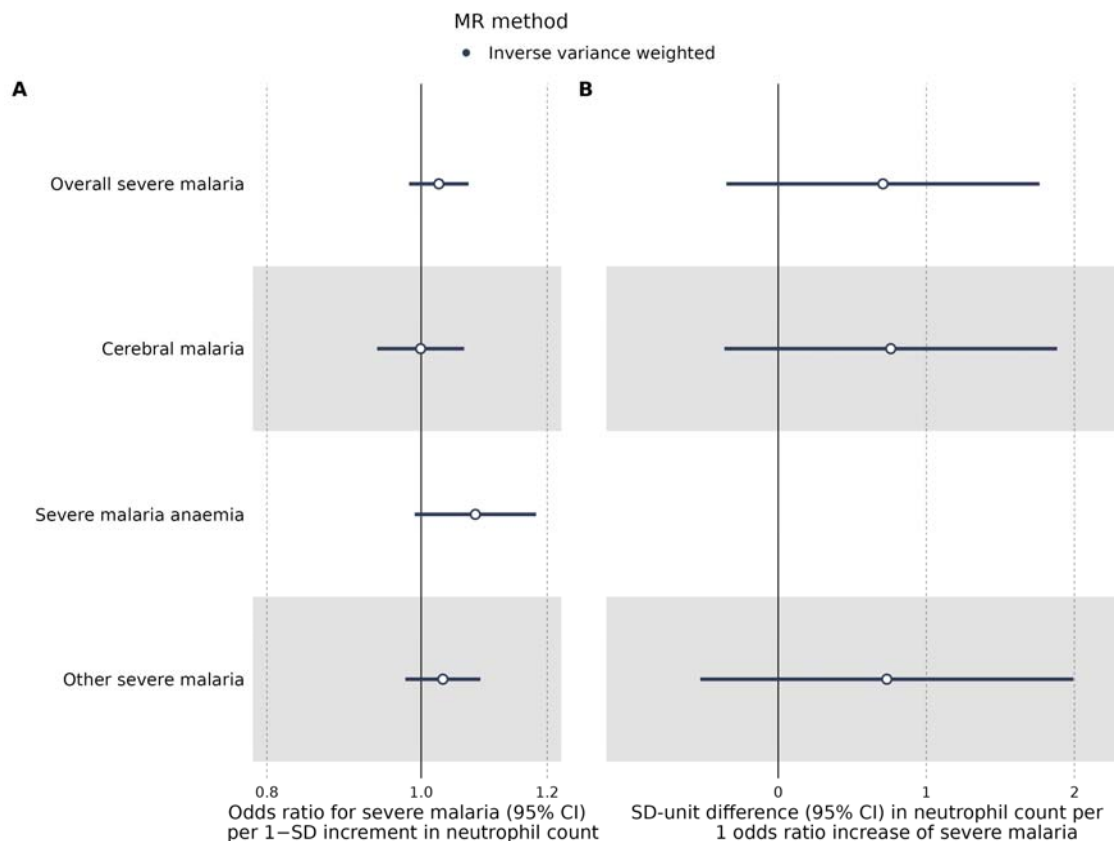
436 **Mendelian randomization**

437 Finally, a bi-directional MR was performed between neutrophil count and severe malaria. For
438 the latter, we used summary statistics from the MalariaGEN study [37]. Only 3 SNPs were
439 available to proxy for neutrophil count after data harmonization with the malaria dataset. For
440 severe malaria as an exposure, 7 SNPs were available for overall severe malaria, 2 for CM
441 and 3 for OTHER.

442

443 The MR analysis did not suggest an effect of increasing neutrophil count on CM risk (IVW
444 OR: 1.00, 95% CI: 0.94 to 1.06; P = 0.98. There was limited evidence of an effect of
445 neutrophil count on overall severe malaria (IVW OR: 1.03, 95% CI: 0.98 to 1.07; P = 0.24),
446 OTHER (IVW OR: 1.03, 95% CI: 0.98 to 1.09; P = 0.26) and SMA (IVW OR: 1.08, 95% CI:
447 0.99 to 1.18; P = 0.08), although the effect estimates were trending towards an increased risk
448 of severity, particularly for SMA (**Figure 5A**, **Supplementary Table 9**). When running the
449 MR analysis in the other direction, there was little evidence of an effect of overall severe
450 malaria (IVW OR: 2.03, 95% CI: 0.70 to 5.84; P = 0.19), CM (IVW OR: 2.14, 95% CI: 0.70
451 to 6.57; P = 0.18) and OTHER (IVW OR: 2.08, 95% CI: 0.59 to 7.34; P = 0.25) on neutrophil
452 count. However, there was a directional agreement in effect estimates towards an increase in
453 neutrophil count (**Figure 5B**, **Supplementary Table 9**). No SNPs passed the GWAS
454 significance threshold for SMA, meaning this analysis could not be conducted.

455



456

457 **Figure 5.** Bi-directional Mendelian randomization. Forest plot of the IVW MR analysis with
458 neutrophil count as an exposure (**A**) and severe malaria as an exposure (**B**). Overall severe
459 malaria and its sub-phenotypes are listed on the y-axis, with the effect estimates on the x-
460 axis. In the first instance, the MR results are interpreted as an OR increase severe malaria per
461 1-SD increase in neutrophil count, while in the latter as a 1-SD unit difference in neutrophil
462 count per 1-OR

463

464 A single-SNP MR analysis was performed to study the effect of each genetic variant on the
465 outcome. For neutrophil count as the exposure, SNPs rs2325919 (proxy for rs2814778),
466 rs7460611 (proxy for rs10096834), and rs144109344 were used. There was little evidence of
467 an effect by any single SNP, although the general direction was towards an increased risk of
468 severe malaria (**Supplementary Table 10**). The estimated conditional F-statistic for SNPs
469 rs2325919, rs7460611 and rs144109344 were 182, 16 and 36 respectively. For severe malaria

470 as an exposure, SNPs rs113892119, rs116423146, rs1419114, rs553707144, rs557568961,
471 rs57032711, rs8176751 were used to proxy for overall severe malaria, rs113892119 and
472 rs543034558 for CM, and rs113892119, rs116423146, rs557568961 for OTHER
473 (**Supplementary Table 11**). The estimated conditional F-statistic for SNPs rs113892119,
474 rs116423146, rs1419114, rs553707144, rs557568961, rs57032711 and rs8176751 were 96,
475 32, 30, 38, 119, 32 and 44 respectively.

476

477 Discussion

478 Here, we conducted a GWAS of neutrophil count in individuals from the AFR CAG in
479 UKBB. Seventy-three independent loci were identified, of which nineteen were novel and
480 rare. Ten index SNPs were found using the conservative GCTA-COJO approach, and another
481 two through MR clumping. Moreover, BOLT-LMM was found to be reliable in conducting
482 GWAS on UKBB participants of African ancestry. Ultimately, this allowed us to run an MR
483 analysis between neutrophil count and *P. falciparum* severe malaria.

484

485 An aim of our study was to assess whether BOLT-LMM could provide reliable results when
486 performing a GWAS in people of non-European ancestry, such as those in the UKBB AFR
487 CAG. In their meta-analysis of BCT in non-European datasets, Chen et al. used a linear
488 model in PLINK to run their GWAS, restricting BOLT-LMM only to the European dataset
489 [33]. Compared to our META-WD and META-WOD GWAS, the BOLT-LMM approach
490 was more similar to that of Chen et al conducted with a larger sample-size (N=15,171). These
491 findings indicate that a linear mixed model framework using a kinship matrix might reliably
492 account for extensive population structure in a complex data set such as that seen in the
493 African CAG used here. If this observation holds true this would be advantageous in

494 identifying more causal ancestry-specific SNPs in future studies, as the power of BOLT-
495 LMM scales with increasing GWAS sample-size [67].

496

497 Next, we found a marked difference between the genetic architecture of neutrophil count in
498 people of African vs. European ancestry [38]. Interestingly, tissue expression for BCTs has
499 been found to vary between ancestries as well [73], further showing the importance of
500 conducting GWAS in diverse populations to improve the understanding of BCT biology. We
501 investigated some of the GCTA-COJO index SNPs in relation to a biological mechanism that
502 could explain how allele variation might affect neutrophil count levels in people of African
503 ancestry.

504

505 One such SNP is rs12747038, an index SNP located on chromosome 1 (1q21.1), was also
506 identified by Chen et al. and Hu et al. to be associated with neutrophil count and they found a
507 similar effect size to us (AFR_CAG: BETA = -0.22, P-value = 3.90e-09; Chen et al: BETA =
508 -0.31, P-value = 3e-20; Hu et al: BETA = -0.21, P-value = 8e-36) [33,74]. Interestingly,
509 rs12747038 has a role as a splicing QTL (sQTL) i.e. affecting alternative splicing to make
510 different protein isoforms [75], which can be more relevant mechanistically to a phenotype
511 compared to expression data [76]. The strongest association as an sQTL was with *NBPF12*
512 gene (NES = 0.49, P-value = 2.9e-9) in the thyroid. McCartney et al. had found that
513 rs11239931, an sQTL for *NBPF12*, was also associated with a decrease in granulocyte count
514 (BETA = -0.23, P-value = 4e-12) in people of African ancestry (N=6,152) [77]. *NBPF12* is
515 part of the neuroblastoma breakpoint family, which has been associated with an array of
516 traits, such as autism, psoriasis and various cancers [78].

517

518 The rs2814778 (chromosome 1q23.2) index SNP has been the most replicated genetic variant
519 in people of African ancestry known to affect neutrophil count [33,79–84], with the CC
520 genotype (most common in Africans) associated with decreased neutrophil count [20]. The
521 exact location of rs2814778 is inside a promoter upstream of the *ACKRI/DARC* (Atypical
522 Chemokine Receptor 1/Duffy Antigen Receptor for Chemokines) gene [13]. The CC
523 genotype inhibits the binding of the GATA transcription factor and therefore *ACKRI*
524 expression in erythrocytes, preventing the production of a glycosylated transmembrane
525 receptor [20]. This receptor is heavily involved in chemokine signalling, such as CXCL8 and
526 CCL5 [13].

527

528 rs144109344 is an index variant on chromosome 2 (2q21.3), and its association was similar
529 to that in the studies of Chen et al. and Soremekun et al. (N=17,802 Africans): AFR_CAG
530 BETA = -0.12, P-value = 3.10e-10; Chen BETA = -0.27, P-value = 3.39e-14; Soremekun
531 BETA = -0.21, P-value = 2e-13) [33,82]. Similarly, other SNPs mapping to the
532 *DARS/CXCR4* (Aspartyl-TRNA Synthetase 1/C-X-C Motif Chemokine Receptor 4) genes
533 have been associated with neutrophil and monocyte count [33,38,85–88]. CXCR4 is a
534 chemokine receptor which binds to CXCL12 [89], and is known to regulate the release of
535 neutrophils from the bone marrow during both homeostasis and infections [90]. Interestingly,
536 CXCR4 has been implicated in *P. falciparum* pathogenesis. Macrophage migration inhibitory
537 factor (MIF) can interact with CXCR4 to recruit neutrophils [91], and *P. falciparum* is known
538 to also produce MIF (PfMIF) [92]. A previous laboratory study using both murine (*P*
539 *berghei*) and human (*P falciparum*) models found impairment of the parasite liver-cycle in
540 both genetically deficient and drug-targeted CXCR4 [93].

541

542 We note that the process of mapping SNPs to a biological function is a difficult process, and
543 the brief discussion above only serves as an inquiry into a possible explanation for the
544 primary GWAS results.

545

546 Finally, in the MR analysis, there was limited evidence for an effect of increased circulating
547 neutrophil on the risk of SM. The strongest effect was observed for the SMA sub-phenotype,
548 however, this did not reach statistical significance. Interestingly, a recent report demonstrated
549 an association between circulating neutrophil transcriptional activity and levels of anaemia in
550 children with malaria [24], highlighting the need for further pathophysiological studies. We
551 also observed little evidence for an effect of SM on neutrophil count. Previously, Band et al.
552 performed a MR analysis between neutrophil count and *P. falciparum* SM [37], however,
553 they used SNPs for neutrophil count generated from a GWAS in Europeans from UKBB
554 [38], where they found no evidence of an effect on SM (AFR_CAG BETA = 0.03, P-value =
555 0.24; Band BETA = 0.00, P-value = 0.87) [37].

556

557 Our study has certain limitations. Firstly, the novel genetic variants identified here may be a
558 result of Winner's curse [94] - SNPs can pass the "significance" threshold (commonly set at
559 $5e-8$ [95,96]) in GWAS by chance in the first discovery study, which is then not replicated in
560 subsequent studies [97,98].

561

562 Secondly, only a limited number of instruments were available to proxy for neutrophil count
563 in the MR analysis. Seven index SNPs had a very high effect allele count, which might have
564 been fixed in the MalariaGEN study population and so could not be used in the MR analysis.
565 The rs2814778 SNP (associated with the *ACKRI* gene) most likely had a very small allele
566 frequency and might have been eliminated, although we were able to use another SNP in LD

567 with it as a proxy. While LD proxies are useful, they can also come with the caveat of not
568 precisely instrumenting the trait [36].

569

570 Finally, the most impactful limitation in this study is the small sample-size and hence
571 statistical power. As mentioned previously, we have chosen to use BOLT-LMM here to best
572 address the issues of a small sample-size and the presence of population structure. Current
573 studies performed on people living in sub-Saharan Africa have been small [33,80–82]
574 compared to those currently being carried out in Europe, East Asia and the US [31,88,99].
575 Having a large-scale study akin to UKBB in sub-Saharan African would allow for finding
576 common SNPs with smaller effect sizes that could be used reliably for polygenic risk score
577 generation or MR analyses for complex traits such as neutrophil count.

578

579 In conclusion, our GWAS of neutrophil count in people from the UKBB African CAG
580 identified several SNPs associated with neutrophil count. Additionally, our analyses would
581 support a conclusion that linear mixed model frameworks can properly account for possible
582 confounding due to population stratification in complex highly stratified sample populations.
583 Finally, while the MR results were largely inconclusive, this only demonstrates the
584 importance of conducting large-scale biobank studies in Africa.

585

586 Availability of data and materials

587 Genetic data from UK Biobank were made available as part of project code 15825. Analytical
588 code is available on GitHub at <https://github.com/andrewcon/AFR-GWAS-neutrophil>.

589

590 Contributions

591 AC, BA, DA, EEV, CJB and REM conceived the study. AC conducted the analysis. All
592 authors contributed to the interpretation of the findings. AC, EEV, CJB, DH and BA wrote
593 the manuscript. All authors critically revised the paper for intellectual content and approved
594 the final version of the manuscript.

595

596 Acknowledgements

597 We are grateful to the UK Biobank study and its participants. This research has been
598 conducted using the UK Biobank resource under Application 15825. We thank the Malaria
599 GEN Network for their study and their participants.

600

601 Funding

602 AC acknowledges funding from grant MR/N0137941/1 for the GW4 BIOMED MRC DTP,
603 awarded to the Universities of Bath, Bristol, Cardiff and Exeter from the Medical Research
604 Council (MRC)/UKRI. NJT and REM acknowledge funding from the MRC
605 (MC_UU_00011/1). NJT is the PI of the Avon Longitudinal Study of Parents and Children
606 (MRC & Wellcome Trust 217065/Z/19/Z) and is supported by the University of Bristol
607 NIHR Biomedical Research Centre (BRC-1215-2001). NJT and DH acknowledge funding
608 from the Wellcome Trust (202802/Z/16/Z). EEV, CJB, and NJT also acknowledge funding
609 by the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019). EEV and
610 CJB are supported by Diabetes UK (17/0005587) and the World Cancer Research Fund
611 (WCRF UK), as part of the World Cancer Research Fund International grant program
612 (IIG_2019_2009). S.K. is supported by a United Kingdom Research and Innovation Future

613 Leaders Fellowship (MR/T043202/1). JZ is supported by Shanghai Thousand Talents
614 Program and the National Health Commission of the PR China. BA acknowledges funding
615 from the Medical Research Council (MR/R02149x/1). The funders of the study had no role in
616 the study design, data collection, data analysis, data interpretation, or writing of the report.

617

618 Ethics declarations

619 Ethics approval and consent to participate

620 UK Biobank received ethical approval from the NHS National Research Ethics Service North
621 West (11/NW/0382; 16/NW/0274) and was conducted in accordance with the Declaration of
622 Helsinki. All participants provided written informed consent before enrolment in the study.

623

624 Consent for publication

625 All authors consented to the publication of this work.

626

627 Competing interests

628 The authors declare no competing interests.

629

630 Supplementary information

631 Additional file 1

632 **Supplementary Figures**

633 Additional file 2

634 **Supplementary Tables**

635 Additional file 3

636 **Supplementary Methods**

637 Additional file 4

638 **STROBE-MR checklist**

639

640 References

- 641 [1] WHO Africa. World malaria report 2019. 2019.
- 642 [2] Price RN, Commons RJ, Battle KE, Thriemer K, Mendis K. Plasmodium vivax in the
643 Era of the Shrinking P. falciparum Map. Trends Parasitol 2020;36:560–70.
644 <https://doi.org/10.1016/j.pt.2020.03.009>.
- 645 [3] Moxon CA, Gibbins MP, McGuinness D, Milner DA, Marti M. New Insights into
646 Malaria Pathogenesis. Annual Review of Pathology: Mechanisms of Disease
647 2020;15:315–43. <https://doi.org/10.1146/annurev-pathmechdis-012419-032640>.
- 648 [4] Knackstedt SL, Georgiadou A, Apel F, Abu-Abed U, Moxon CA, Cunnington AJ, et
649 al. Neutrophil extracellular traps drive inflammatory pathogenesis in malaria
650 2019;4:336. <https://doi.org/10.1126/SCIIMMUNOL.AAW0336>.
- 651 [5] Sierra F, Grau GER. The ins and outs of cerebral malaria pathogenesis:
652 Immunopathology, extracellular vesicles, immunometabolism, and trained immunity.
653 Front Immunol 2019;10:830. <https://doi.org/10.3389/fimmu.2019.00830>.
- 654 [6] Cela D, Knackstedt SL, Groves S, Rice CM, Kwon JTW, Mordmüller B, et al. PAD4
655 controls chemoattractant production and neutrophil trafficking in malaria. J Leukoc
656 Biol 2021. <https://doi.org/10.1002/JLB.4AB1120-780R>.
- 657 [7] Kariuki SN, Williams TN. Human genetics and malaria resistance. vol. 139. Springer;
658 2020. <https://doi.org/10.1007/S00439-020-02142-6>.
- 659 [8] Allison AC. Protection afforded by sickle-cell trait against subtertian malarial
660 infection. Br Med J 1954;1:290–4. <https://doi.org/10.1136/bmj.1.4857.290>.
- 661 [9] Kwiatkowski DP. How Malaria Has Affected the Human Genome and What Human
662 Genetics Can Teach Us about Malaria. The American Journal of Human Genetics
663 2005;77:171–92. <https://doi.org/10.1086/432519>.
- 664 [10] Ndila CM, Uyoga S, Macharia AW, Nyutu G, Peshu N, Ojal J, et al. Human candidate
665 gene polymorphisms and risk of severe malaria in children in Kilifi, Kenya: a case-
666 control association study. Lancet Haematol 2018;5:e333–45.
667 [https://doi.org/10.1016/S2352-3026\(18\)30107-8](https://doi.org/10.1016/S2352-3026(18)30107-8).
- 668 [11] Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of
669 Malaria in Africa. PLoS Med 2005;2:e340.
670 <https://doi.org/10.1371/journal.pmed.0020340>.
- 671 [12] Sakuntabhai A, Ndiaye R, Casadémont I, Peerapittayamonkol C, Rogier C, Tortevoye
672 P, et al. Genetic Determination and Linkage Mapping of Plasmodium falciparum
673 Malaria Related Traits in Senegal. PLoS One 2008;3:e2000.
674 <https://doi.org/10.1371/journal.pone.0002000>.
- 675 [13] Atallah-Yunes SA, Ready A, Newburger PE. Benign ethnic neutropenia. Blood Rev
676 2019;37:100586. <https://doi.org/10.1016/j.blre.2019.06.003>.

- 677 [14] Shoenfeld Y, Alkan ML, Asaly A, Carmeli Y, Katz M. Benign familial leukopenia and
678 neutropenia in different ethnic groups. *Eur J Haematol* 1988;41:273–7.
679 <https://doi.org/https://doi.org/10.1111/j.1600-0609.1988.tb01192.x>.
- 680 [15] Rippey JJ. LEUCOPENIA IN WEST INDIANS AND AFRICANS. *The Lancet*
681 1967;290:44. [https://doi.org/https://doi.org/10.1016/S0140-6736\(67\)90086-4](https://doi.org/https://doi.org/10.1016/S0140-6736(67)90086-4).
- 682 [16] Denic S, Showqi S, Klein C, Takala M, Nagelkerke N, Agarwal MM. Prevalence,
683 phenotype and inheritance of benign neutropenia in Arabs. *BMC Blood Disord*
684 2009;9:3. <https://doi.org/10.1186/1471-2326-9-3>.
- 685 [17] Hsieh MM, Everhart JE, Byrd-Holt DD, Tisdale JF, Rodgers GP. Prevalence of
686 Neutropenia in the U.S. Population: Age, Sex, Smoking Status, and Ethnic
687 Differences. *Ann Intern Med* 2007;146:486. [https://doi.org/10.7326/0003-4819-146-7-](https://doi.org/10.7326/0003-4819-146-7-200704030-00004)
688 [200704030-00004](https://doi.org/10.7326/0003-4819-146-7-200704030-00004).
- 689 [18] Amulic B, Cazalet C, Hayes GL, Metzler KD, Zychlinsky A. Neutrophil Function:
690 From Mechanisms to Disease. [https://doi.org/10.1146/Annurev-Immunol-020711-](https://doi.org/10.1146/Annurev-Immunol-020711-074942)
691 [074942](https://doi.org/10.1146/Annurev-Immunol-020711-074942) 2012;30:459–89. [https://doi.org/10.1146/Annurev-Immunol-020711-](https://doi.org/10.1146/Annurev-Immunol-020711-074942)
692 [074942](https://doi.org/10.1146/Annurev-Immunol-020711-074942).
- 693 [19] Reich D, Nalls MA, Kao WHL, Akylbekova EL, Tandon A, Patterson N, et al.
694 Reduced neutrophil count in people of African descent is due to a regulatory variant in
695 the Duffy antigen receptor for chemokines gene. *PLoS Genet* 2009;5.
696 <https://doi.org/10.1371/journal.pgen.1000360>.
- 697 [20] Rappoport N, Simon AJ, Amariglio N, Rechavi G. The Duffy antigen receptor for
698 chemokines, ACKR1,– ‘Jeanne DARC’ of benign neutropenia. *Br J Haematol*
699 2019;184:497–507. <https://doi.org/10.1111/bjh.15730>.
- 700 [21] Palmblad J, Höglund P. Ethnic benign neutropenia: A phenomenon finds an
701 explanation. *Pediatr Blood Cancer* 2018;65:e27361. <https://doi.org/10.1002/xbc.27361>.
- 702 [22] Amulic B, Moxon C, Cunningham A. A more granular view of neutrophils in malaria 1
703 2. n.d.
- 704 [23] Aitken EH, Alemu A, Rogerson SJ. Neutrophils and Malaria. *Front Immunol*
705 2018;9:3005. <https://doi.org/10.3389/fimmu.2018.03005>.
- 706 [24] Anyona S, Cheng Q, Guo Y, Seidenberg P, Schneider K, Lambert C, et al. Entire
707 Expressed Peripheral Blood Transcriptome in Pediatric Severe Malarial Anemia 2023.
708 <https://doi.org/10.21203/RS.3.RS-3150748/V1>.
- 709 [25] Garcia-Senosain A, Kana IH, Singh S, Das MK, Dziegiel MH, Hertegonne S, et al.
710 Neutrophils dominate in opsonic phagocytosis of *P. falciparum* blood-stage merozoites
711 and protect against febrile malaria. *Commun Biol* 2021;4.
712 <https://doi.org/10.1038/S42003-021-02511-5>.
- 713 [26] Zelter T, Strahilevitz J, Simantov K, Yajuk O, Jensen AR, Dzikowski R, et al.
714 Neutrophils impose strong selective pressure against PfEMP1 variants implicated in
715 cerebral malaria. *BioRxiv* 2021:2021.05.09.443317.
716 <https://doi.org/10.1101/2021.05.09.443317>.
- 717 [27] Smith GD, Ebrahim S. “Mendelian randomization”: Can genetic epidemiology
718 contribute to understanding environmental determinants of disease? *Int J Epidemiol*
719 2003;32:1–22. <https://doi.org/10.1093/ije/dyg070>.
- 720 [28] Smith GD, Ebrahim S. Mendelian randomization: prospects, potentials, and
721 limitations. *Int J Epidemiol* 2004;33:30–42. <https://doi.org/10.1093/ije/dyh132>.
- 722 [29] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal
723 inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–98.
724 <https://doi.org/10.1093/hmg/ddu328>.

- 725 [30] Zheng J, Baird D, Borges M-C, Bowden J, Hemani G, Haycock P, et al. Recent
726 Developments in Mendelian Randomization Studies. *Curr Epidemiol Rep* 2017;4:330–
727 45. <https://doi.org/10.1007/s40471-017-0128-6>.
- 728 [31] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK
729 Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.
730 <https://doi.org/10.1038/s41586-018-0579-z>.
- 731 [32] Skrivankova VW, Richmond RC, Woolf BAR, Davies NM, Swanson SA,
732 Vanderweele TJ, et al. Strengthening the reporting of observational studies in
733 epidemiology using mendelian randomisation (STROBE-MR): explanation and
734 elaboration. *BMJ* 2021;375. <https://doi.org/10.1136/BMJ.N2233>.
- 735 [33] Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-
736 ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals from 5
737 Global Populations. *Cell* 2020;182:1198-1213.e14.
738 <https://doi.org/10.1016/j.cell.2020.06.045>.
- 739 [34] Constantinescu A-E, Mitchell RE, Zheng J, Bull CJ, Timpson NJ, Amulic B, et al. A
740 framework for research into continental ancestry groups of the UK Biobank. *Human*
741 *Genomics* 2022 16:1 2022;16:1–14. <https://doi.org/10.1186/S40246-022-00380-5>.
- 742 [35] Davey Smith G, Ebrahim S, Smith GD, Ebrahim S. ‘Mendelian randomization’: can
743 genetic epidemiology contribute to understanding environmental determinants of
744 disease? *Int J Epidemiol* 2003;32:1–22. <https://doi.org/10.1093/ije/dyg070>.
- 745 [36] Hartwig FP, Davies NM, Hemani G, Smith GD. Two-sample Mendelian
746 randomization: avoiding the downsides of a powerful, widely applicable but
747 potentially fallible technique. *Int J Epidemiol* 2016;45:1717–26.
- 748 [37] Malaria Genomic Epidemiology Network. Insights into malaria susceptibility using
749 genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nat Commun*
750 2019;10:5732. <https://doi.org/10.1038/s41467-019-13480-z>.
- 751 [38] Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The Allelic
752 Landscape of Human Blood Cell Trait Variation and Links to Common Complex
753 Disease. *Cell* 2016;167:1415-1429.e19. <https://doi.org/10.1016/j.cell.2016.10.042>.
- 754 [39] Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A
755 tutorial on conducting genome-wide association studies: Quality control and statistical
756 analysis. *Int J Methods Psychiatr Res* 2018;27:1–10. <https://doi.org/10.1002/mpr.1608>.
- 757 [40] Sheard S, Nicholls R, Froggatt J. UK Biobank Haematology Data Companion
758 Document n.d.
- 759 [41] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex
760 trait analysis. *Am J Hum Genet* 2011;88:76–82.
761 <https://doi.org/10.1016/j.ajhg.2010.11.011>.
- 762 [42] Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of AnTC,
763 Consortium DiaGRAM (DIAGRAM), et al. Conditional and joint multiple-SNP
764 analysis of GWAS summary statistics identifies additional variants influencing
765 complex traits. *Nat Genet* 2012;44:369-S3. <https://doi.org/10.1038/ng.2213>.
- 766 [43] Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and
767 annotation of genetic associations with FUMA. *Nature Communications* 2017 8:1
768 2017;8:1–11. <https://doi.org/10.1038/s41467-017-01261-5>.
- 769 [44] Choi KW, Stein MB, Nishimi KM, Ge T, Coleman JRI, Chen CY, et al. An exposure-
770 wide and mendelian randomization approach to identifying modifiable factors for the
771 prevention of depression. *American Journal of Psychiatry* 2020;177:944–54.
772 <https://doi.org/10.1176/APPI.AJP.2020.19111158/ASSET/IMAGES/LARGE/APPI.AJ>
773 [P.2020.19111158F4.JPEG](https://doi.org/10.1176/APPI.AJP.2020.19111158F4.JPEG).

- 774 [45] Noyce AJ, Bandres-Ciga S, Kim J, Heilbron K, Kia D, Hemani G, et al. The
775 Parkinson's Disease Mendelian Randomization Research Portal. *Movement Disorders*
776 2019;34:1864. <https://doi.org/10.1002/MDS.27873>.
- 777 [46] Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al.
778 Common SNPs explain a large proportion of heritability for human height. *Nat Genet*
779 2010;42:565. <https://doi.org/10.1038/NG.608>.
- 780 [47] WHO Africa. Severe Malaria 2014. <https://doi.org/10.1111/tmi.12313>.
- 781 [48] Lawson DJ, Davies NM, Haworth S, Ashraf B, Howe L, Crawford A, et al. Is
782 population structure in the genetic biobank era irrelevant, a challenge, or an
783 opportunity? *Hum Genet* 2020;139:23–41. [https://doi.org/10.1007/s00439-019-02014-](https://doi.org/10.1007/s00439-019-02014-8)
784 8.
- 785 [49] Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of
786 genomewide association scans. *Bioinformatics* 2010;26:2190–1.
787 <https://doi.org/10.1093/BIOINFORMATICS/BTQ340>.
- 788 [50] Sanna S, Jackson AU, Nagaraja R, Willer CJ, Chen WM, Bonnycastle LL, et al.
789 Common variants in the GDF5-UQCC region are associated with variation in human
790 height. *Nat Genet* 2008;40:198–203. <https://doi.org/10.1038/NG.74>.
- 791 [51] Willer CJ, Sanna S, Jackson AU, Scuteri A, Bonnycastle LL, Clarke R, et al. Newly
792 identified loci that influence lipid concentrations and risk of coronary artery disease.
793 *Nat Genet* 2008;40:161–9. <https://doi.org/10.1038/NG.76>.
- 794 [52] Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between
795 imprecisely measured traits using GWAS summary data. *PLoS Genet* 2017;13.
796 <https://doi.org/10.1371/JOURNAL.PGEN.1007081>.
- 797 [53] Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base
798 platform supports systematic causal inference across the human phenome. *Elife*
799 2018;7.
- 800 [54] Burgess S, Dudbridge F, Thompson SG. Combining information on multiple
801 instrumental variables in Mendelian randomization: comparison of allele score and
802 summarized data methods. *Stat Med* 2016;35:1880–906.
803 <https://doi.org/https://doi.org/10.1002/sim.6835>.
- 804 [55] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid
805 instruments: effect estimation and bias detection through Egger regression. *Int J*
806 *Epidemiol* 2015;44:512–25. <https://doi.org/10.1093/ije/dyv080>.
- 807 [56] Bowden J, Hemani G, Davey Smith G. Invited Commentary: Detecting Individual and
808 Global Horizontal Pleiotropy in Mendelian Randomization—A Job for the Humble
809 Heterogeneity Statistic? *Am J Epidemiol* 2018;187:2681–5.
810 <https://doi.org/10.1093/AJE/KWY185>.
- 811 [57] Age groups - GOV.UK Ethnicity facts and figures n.d. [https://www.ethnicity-facts-](https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest)
812 [figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest](https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest)
813 (accessed August 17, 2022).
- 814 [58] Health Survey for England: Weight n.d. [http://healthsurvey.hscic.gov.uk/data-](http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/weight.aspx)
815 [visualisation/data-visualisation/explore-the-trends/weight.aspx](http://healthsurvey.hscic.gov.uk/data-visualisation/data-visualisation/explore-the-trends/weight.aspx) (accessed August 17,
816 2022).
- 817 [59] Sul JH, Martin LS, Eskin E. Population structure in genetic studies: Confounding
818 factors and mixed models. *PLoS Genet* 2018;14.
819 <https://doi.org/10.1371/JOURNAL.PGEN.1007309>.
- 820 [60] Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, et al.
821 Apparent latent structure within the UK Biobank sample has implications for
822 epidemiological analysis. *Nat Commun* 2019;10. [https://doi.org/10.1038/S41467-018-](https://doi.org/10.1038/S41467-018-08219-1)
823 08219-1.

- 824 [61] Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*
825 2006;2:2074–93. <https://doi.org/10.1371/journal.pgen.0020190>.
- 826 [62] Zaidi AA, Mathieson I. Demographic history mediates the effect of stratification on
827 polygenic scores. *Elife* 2020;9:1–30. <https://doi.org/10.7554/ELIFE.61548>.
- 828 [63] Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, et al.
829 Genetic correlates of social stratification in Great Britain. *Nature Human Behaviour*
830 2019 3:12 2019;3:1332–42. <https://doi.org/10.1038/s41562-019-0757-5>.
- 831 [64] Klein RJ. Power analysis for genome-wide association studies. *BMC Genet* 2007;8:58.
832 <https://doi.org/10.1186/1471-2156-8-58>.
- 833 [65] Pierce BL, Burgess S. Efficient Design for Mendelian Randomization Studies:
834 Subsample and 2-Sample Instrumental Variable Estimators. *Am J Epidemiol*
835 2013;178:1177. <https://doi.org/10.1093/AJE/KWT084>.
- 836 [66] Visscher PM, Hemani G, Vinkhuyzen AAE, Chen GB, Lee SH, Wray NR, et al.
837 Statistical Power to Detect Genetic (Co)Variance of Complex Traits Using SNP Data
838 in Unrelated Samples. *PLoS Genet* 2014;10:e1004269.
839 <https://doi.org/10.1371/JOURNAL.PGEN.1004269>.
- 840 [67] Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et
841 al. Efficient Bayesian mixed-model analysis increases association power in large
842 cohorts. *Nat Genet* 2015;47:284–90. <https://doi.org/10.1038/ng.3190>.
- 843 [68] Weissbrod O, Kanai M, Shi H, Gazal S, Peyrot WJ, Khera A V., et al. Leveraging
844 fine-mapping and multi-population training data to improve cross-population
845 polygenic risk scores. *Nat Genet* 2022;54:450. <https://doi.org/10.1038/S41588-022-01036-9>.
- 846 [69] Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance
847 component model to account for sample structure in genome-wide association studies.
848 *Nat Genet* 2010;42:348. <https://doi.org/10.1038/NG.548>.
- 849 [70] Lichou F, Trynka G. Functional studies of GWAS variants are gaining momentum.
850 *Nature Communications* 2020 11:1 2020;11:1–4. <https://doi.org/10.1038/s41467-020-20188-y>.
- 851 [71] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants
852 from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164–e164.
853 <https://doi.org/10.1093/NAR/GKQ603>.
- 854 [72] Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al.
855 The NHGRI-EBI GWAS Catalog of published genome-wide association studies,
856 targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47.
857 <https://doi.org/10.1093/NAR/GKY1120>.
- 858 [73] Wen J, Xie M, Rowland B, Rosen JD, Sun Q, Chen J, et al. Transcriptome-wide
859 association study of blood cell traits in african ancestry and hispanic/latino
860 populations. *Genes (Basel)* 2021;12:1049. <https://doi.org/10.3390/genes12071049>.
- 861 [74] Hu Y, Bien SA, Nishimura KK, Haessler J, Hodonsky CJ, Baldassari AR, et al. Multi-
862 ethnic genome-wide association analyses of white blood cell and platelet traits in the
863 Population Architecture using Genomics and Epidemiology (PAGE) study. *BMC*
864 *Genomics* 2021;22:1–11. <https://doi.org/10.1186/S12864-021-07745-5>.
- 865 [75] Nilsen TW, Graveley BR. Expansion of the eukaryotic proteome by alternative
866 splicing. *Nature* 2010;463:457. <https://doi.org/10.1038/NATURE08909>.
- 867 [76] Garrido-Martín D, Borsari B, Calvo M, Reverter F, Guigó R. Identification and
868 analysis of splicing quantitative trait loci across multiple tissues in the human genome.
869 *Nature Communications* 2021 12:1 2021;12:1–16. <https://doi.org/10.1038/s41467-020-20578-2>.
- 870
871
872

- 873 [77] McCartney DL, Min JL, Richmond RC, Lu AT, Sobczyk MK, Davies G, et al.
874 Genome-wide association studies identify 137 genetic loci for DNA methylation
875 biomarkers of aging. *Genome Biol* 2021;22:25. [https://doi.org/10.1186/S13059-021-](https://doi.org/10.1186/S13059-021-02398-9)
876 [02398-9](https://doi.org/10.1186/S13059-021-02398-9).
- 877 [78] Zhou F, Xing Y, Xu X, Yang Y, Zhang J, Ma Z, et al. NBPF is a potential DNA-
878 binding transcription factor that is directly regulated by NF- κ B. *Int J Biochem Cell*
879 *Biol* 2013;45:2479–90. <https://doi.org/10.1016/J.BIOCEL.2013.07.022>.
- 880 [79] Moore CB, Verma A, Pendergrass S, Verma SS, Johnson DH, Daar ES, et al.
881 Phenome-wide Association Study Relating Pretreatment Laboratory Parameters With
882 Human Genetic Variants in AIDS Clinical Trials Group Protocols. *Open Forum Infect*
883 *Dis* 2015;2. <https://doi.org/10.1093/OFID/OFU113>.
- 884 [80] Gurdasani D, Carstensen T, Fatumo S, Chen G, Franklin CS, Prado-Martinez J, et al.
885 Uganda Genome Resource Enables Insights into Population History and Genomic
886 Discovery in Africa. *Cell* 2019;179. <https://doi.org/10.1016/J.CELL.2019.10.004>.
- 887 [81] Reiner AP, Lettre G, Nalls MA, Ganesh SK, Mathias R, Austin MA, et al. Genome-
888 Wide Association Study of White Blood Cell Count in 16,388 African Americans: the
889 Continental Origins and Genetic Epidemiology Network (COGENT). *PLoS Genet*
890 2011;7:e1002108. <https://doi.org/10.1371/journal.pgen.1002108>.
- 891 [82] Soremekun O, Soremekun C, Machipisa T, Soliman M, Nashiru O, Chikowore T, et al.
892 Genome-Wide Association and Mendelian Randomization Analysis Reveal the Causal
893 Relationship Between White Blood Cell Subtypes and Asthma in Africans. *Front*
894 *Genet* 2021;12:749415. <https://doi.org/10.3389/FGENE.2021.749415/FULL>.
- 895 [83] Jain D, Hodonsky CJ, Schick UM, Morrison J V., Minnerath S, Brown L, et al.
896 Genome-wide association of white blood cell counts in Hispanic/Latino Americans:
897 the Hispanic Community Health Study/Study of Latinos. *Hum Mol Genet* 2017;26.
898 <https://doi.org/10.1093/HMG/DDX024>.
- 899 [84] Legge SE, Pardiñas AF, Helthuis M, Jansen JA, Jollie K, Knapper S, et al. A genome-
900 wide association study in individuals of African ancestry reveals the importance of the
901 Duffy-null genotype in the assessment of clozapine-related neutropenia. *Molecular*
902 *Psychiatry* 2019 24:3 2019;24:328–37. <https://doi.org/10.1038/s41380-018-0335-7>.
- 903 [85] Kachuri L, Jeon S, DeWan AT, Metayer C, Ma X, Witte JS, et al. Genetic
904 determinants of blood-cell traits influence susceptibility to childhood acute
905 lymphoblastic leukemia. *Am J Hum Genet* 2021;108:1823–35.
906 <https://doi.org/10.1016/J.AJHG.2021.08.004>.
- 907 [86] Vuckovic D, Bao EL, Akbari P, Lareau CA, Mousas A, Jiang T, et al. The Polygenic
908 and Monogenic Basis of Blood Traits and Diseases. *Cell* 2020;182:1214-1231.e11.
909 [https://doi.org/10.1016/J.CELL.2020.08.008/ATTACHMENT/347CE04A-7337-4664-](https://doi.org/10.1016/J.CELL.2020.08.008/ATTACHMENT/347CE04A-7337-4664-BB5B-5ED6234B8F9E/MMC11.DOCX)
910 [BB5B-5ED6234B8F9E/MMC11.DOCX](https://doi.org/10.1016/J.CELL.2020.08.008/ATTACHMENT/347CE04A-7337-4664-BB5B-5ED6234B8F9E/MMC11.DOCX).
- 911 [87] Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, et al. A cross-
912 population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*
913 2021 53:10 2021;53:1415–24. <https://doi.org/10.1038/s41588-021-00931-x>.
- 914 [88] Kanai M, Akiyama M, Takahashi A, Matoba N, Momozawa Y, Ikeda M, et al. Genetic
915 analysis of quantitative traits in the Japanese population links cell types to complex
916 human diseases. *Nature Genetics* 2018 50:3 2018;50:390–400.
917 <https://doi.org/10.1038/s41588-018-0047-6>.
- 918 [89] De Filippo K, Rankin SM. CXCR4, the master regulator of neutrophil trafficking in
919 homeostasis and disease. *Eur J Clin Invest* 2018;48.
920 <https://doi.org/10.1111/ECI.12949>.

- 921 [90] Eash KJ, Means JM, White DW, Link DC. CXCR4 is a key regulator of neutrophil
922 release from the bone marrow under basal and stress granulopoiesis conditions. *Blood*
923 2009;113:4711. <https://doi.org/10.1182/BLOOD-2008-09-177287>.
- 924 [91] Weber C, Kraemer S, Drechsler M, Lue H, Koenen RR, Kapurniotu A, et al. Structural
925 determinants of MIF functions in CXCR2-mediated inflammatory and atherogenic
926 leukocyte recruitment. *Proc Natl Acad Sci U S A* 2008;105:16278.
927 <https://doi.org/10.1073/PNAS.0804017105>.
- 928 [92] Ghosh S, Jiang N, Farr L, Ngoben R, Moonah S. Parasite-produced MIF cytokine:
929 Role in immune evasion, invasion, and pathogenesis. *Front Immunol* 2019;10:1995.
930 <https://doi.org/10.3389/FIMMU.2019.01995/BIBTEX>.
- 931 [93] Bando H, Pradipta A, Iwanaga S, Okamoto T, Okuzaki D, Tanaka S, et al. CXCR4
932 regulates Plasmodium development in mouse and human hepatocytes. *J Exp Med*
933 2019;216:1733.
- 934 [94] Thaler RH. Anomalies: The Winner's Curse. *Journal of Economic Perspectives*
935 1988;2:191–202. <https://doi.org/10.1257/JEP.2.1.191>.
- 936 [95] Panagiotou OA, Ioannidis JPA, Hirschhorn JN, Abecasis GR, Frayling TM, McCarthy
937 MI, et al. What should the genome-wide significance threshold be? Empirical
938 replication of borderline genetic associations. *Int J Epidemiol* 2012;41:273–86.
939 <https://doi.org/10.1093/IJE/DYR178>.
- 940 [96] Chen Z, Boehnke M, Wen X, Mukherjee B. Revisiting the genome-wide significance
941 threshold for common variant GWAS. *G3 Genes|Genomes|Genetics* 2021;11.
942 <https://doi.org/10.1093/G3JOURNAL/JKAA056>.
- 943 [97] Kraft P. Curses - Winner's and otherwise - In genetic epidemiology. *Epidemiology*
944 2008;19:649–51. <https://doi.org/10.1097/EDE.0B013E318181B865>.
- 945 [98] Ioannidis JPA. Why most discovered true associations are inflated. *Epidemiology*
946 2008;19:640–8. <https://doi.org/10.1097/EDE.0B013E31818131E7>.
- 947 [99] Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million
948 Veteran Program: A mega-biobank to study genetic influences on health and disease. *J*
949 *Clin Epidemiol* 2016;70:214–23. <https://doi.org/10.1016/J.JCLINEPI.2015.09.016>.
- 950
951

SUPPLEMENTARY FIGURES

Figure S1: Power calculation.

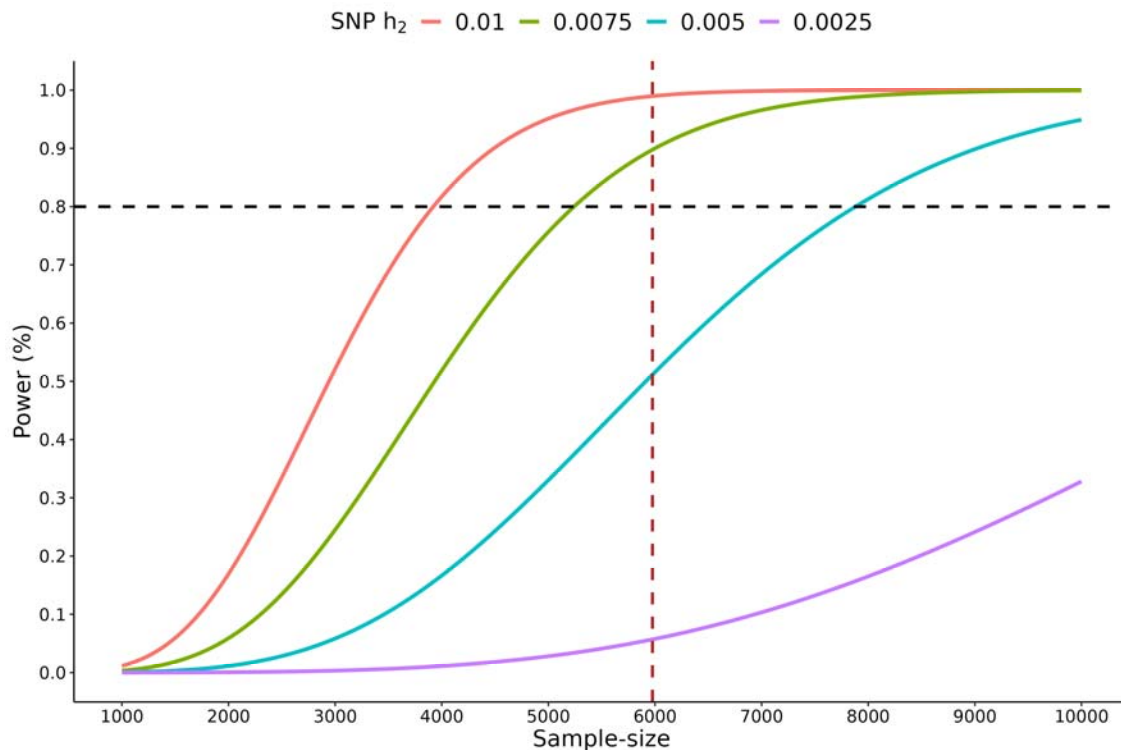


Figure S1. Power calculation of a GWAS AFR_CAG sample. The x-axis indicates the sample-size, while the y-axis is the statistical power of an association test. Each curved line shows how power varies by sample-size at different degrees of the variance explained by all the SNPs on neutrophil count (1%, 0.75%, 0.5%, 0.25%). A black horizontal line is fixed at Power=80%, and a red vertical line is drawn at the GWAS sample size of 5,976.

Figure S2. Regional plots of index SNPs (1).

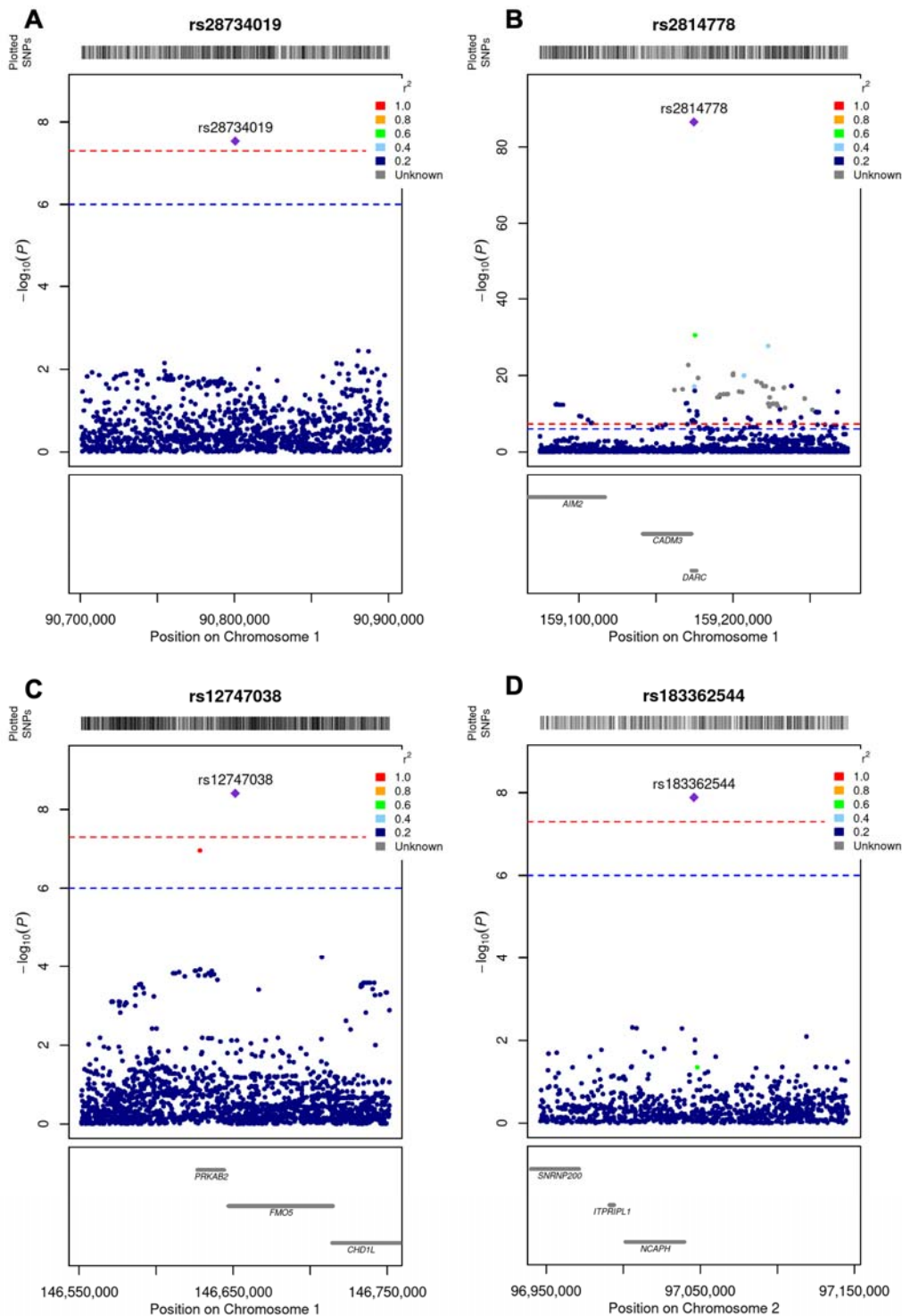


Figure S3. Regional plots of index SNPs (2).

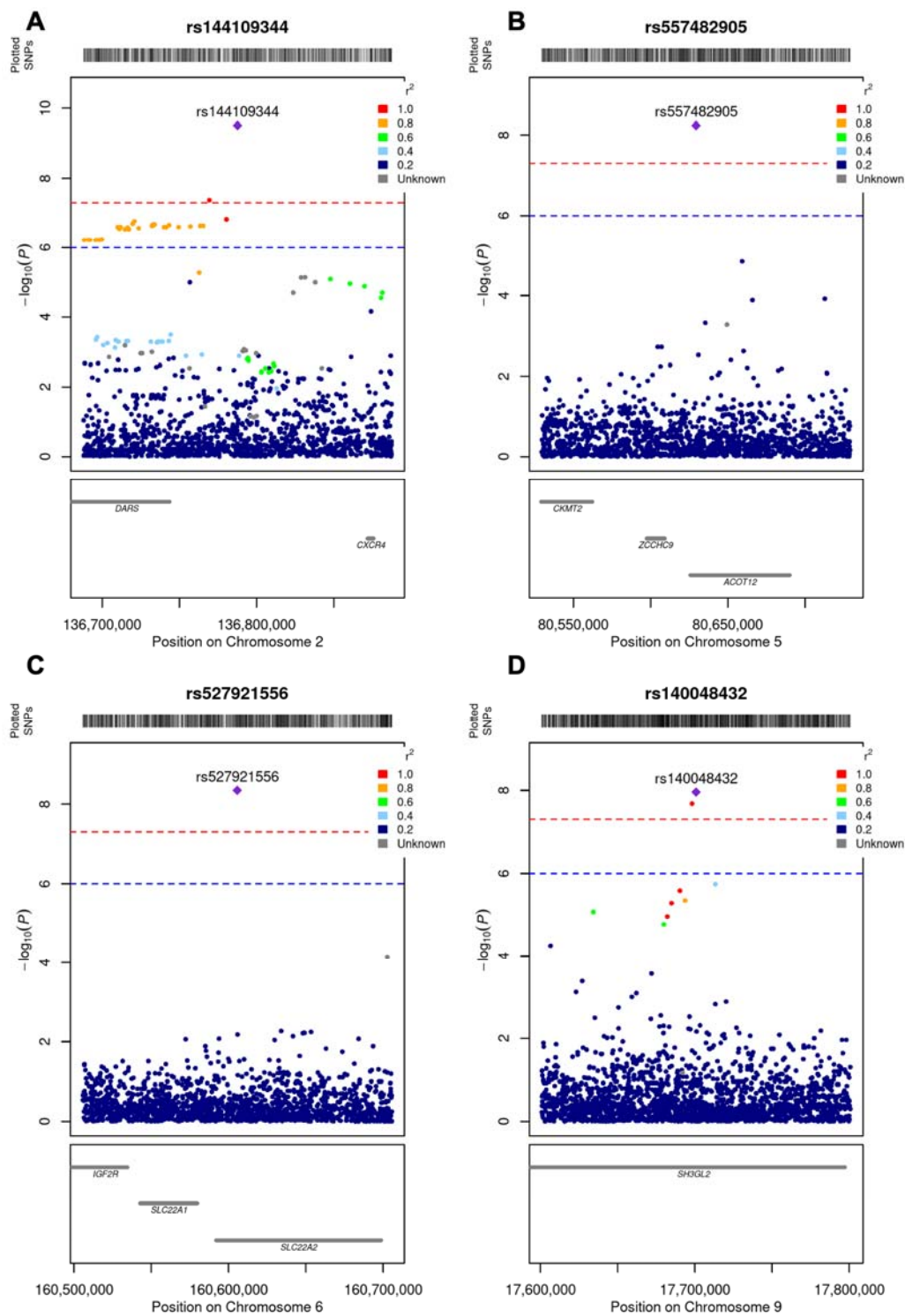


Figure S4. Regional plots of index SNPs (3).

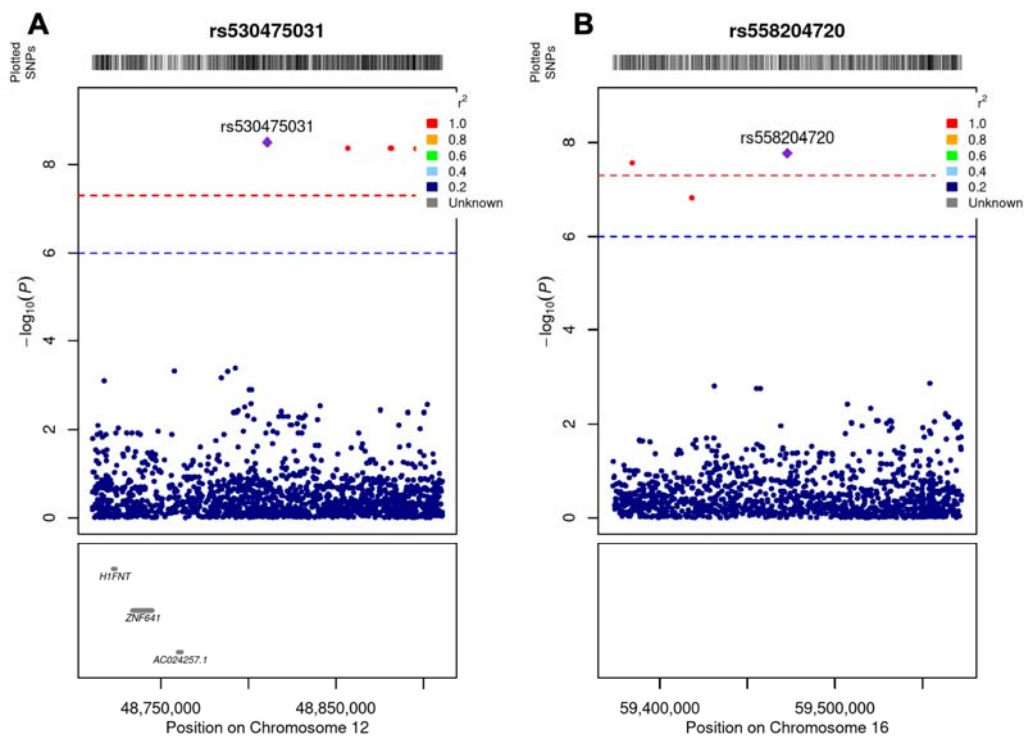


Figure S5: Forest plot of index SNPs effect sizes.

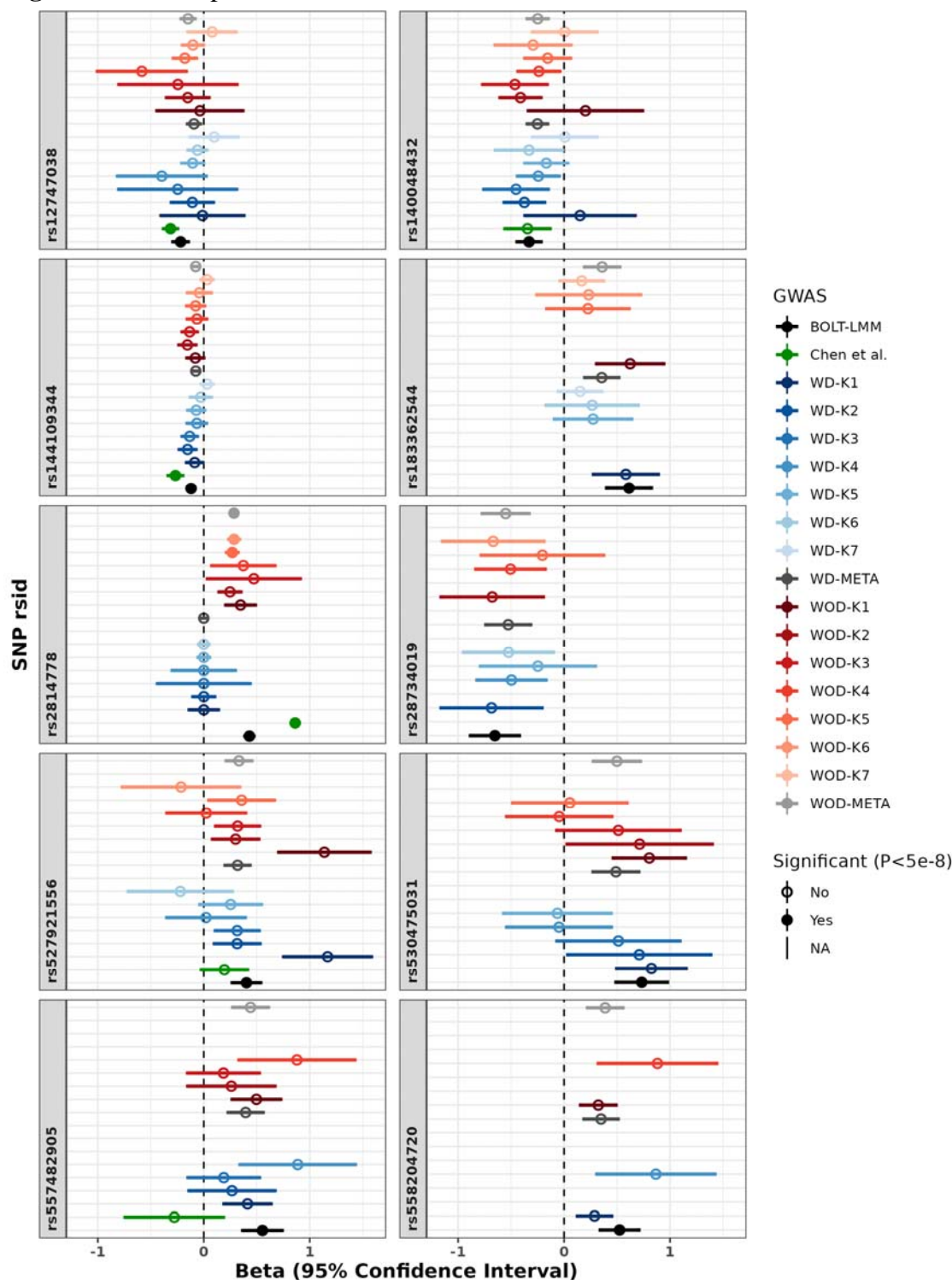


Figure S5. Forest plot of index SNPs by K-means cluster. The effect-size of each BOLT index SNP was compared to that from SNPTTEST/META, by-Kpop runs and Chen et al GWAS. Effect-sizes for each SNP across GWAS are present in the respective boxes of the figure. The x-axis indicates the effect-size (beta coefficient) of each SNP with 95% CIs,

while the y-axis is the type of GWAS (indicated by the figure legend colouring). Some effect sizes were not displayed, either due to a low minor allele count in the case of the Kpop GWAS, or due to not being present in the summary statistics, in the case of the Chen GWAS. Non-signif. WD = adjusting for rs2814778; WOD = without adjusting for rs2814778.

Figure S6. Scatter plot of GCTA-COJO effect sizes.

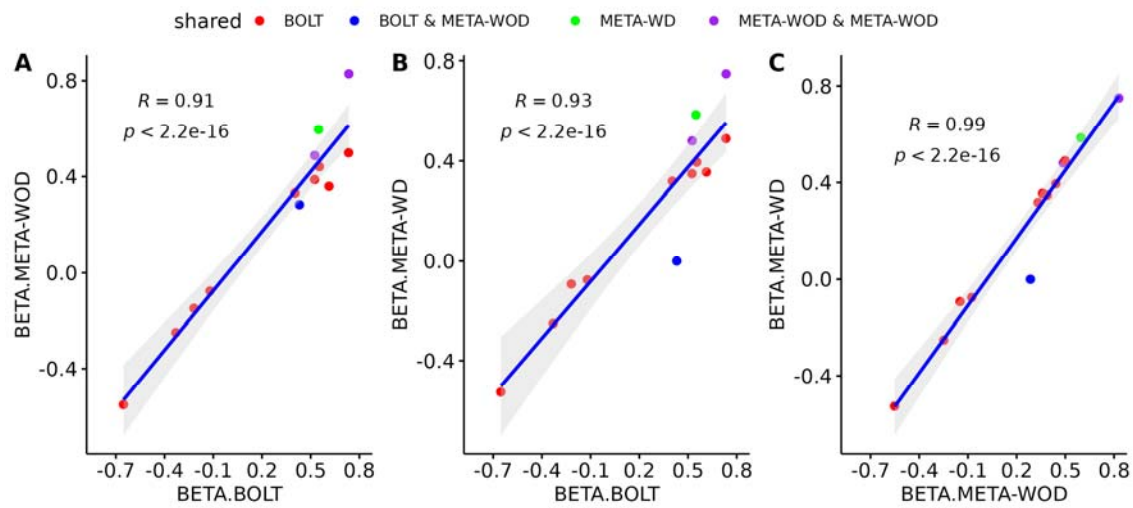


Figure S6. Scatter plot of GCTA-COJO effect sizes. Comparison of effect sizes of all GCTA-COJO independent signals of BOLT-LMM with META-WOD (A), BOLT-LMM with META-WD (B) and META-WOD with META-WD (C).

Figure S7: Description of genomic risk loci.

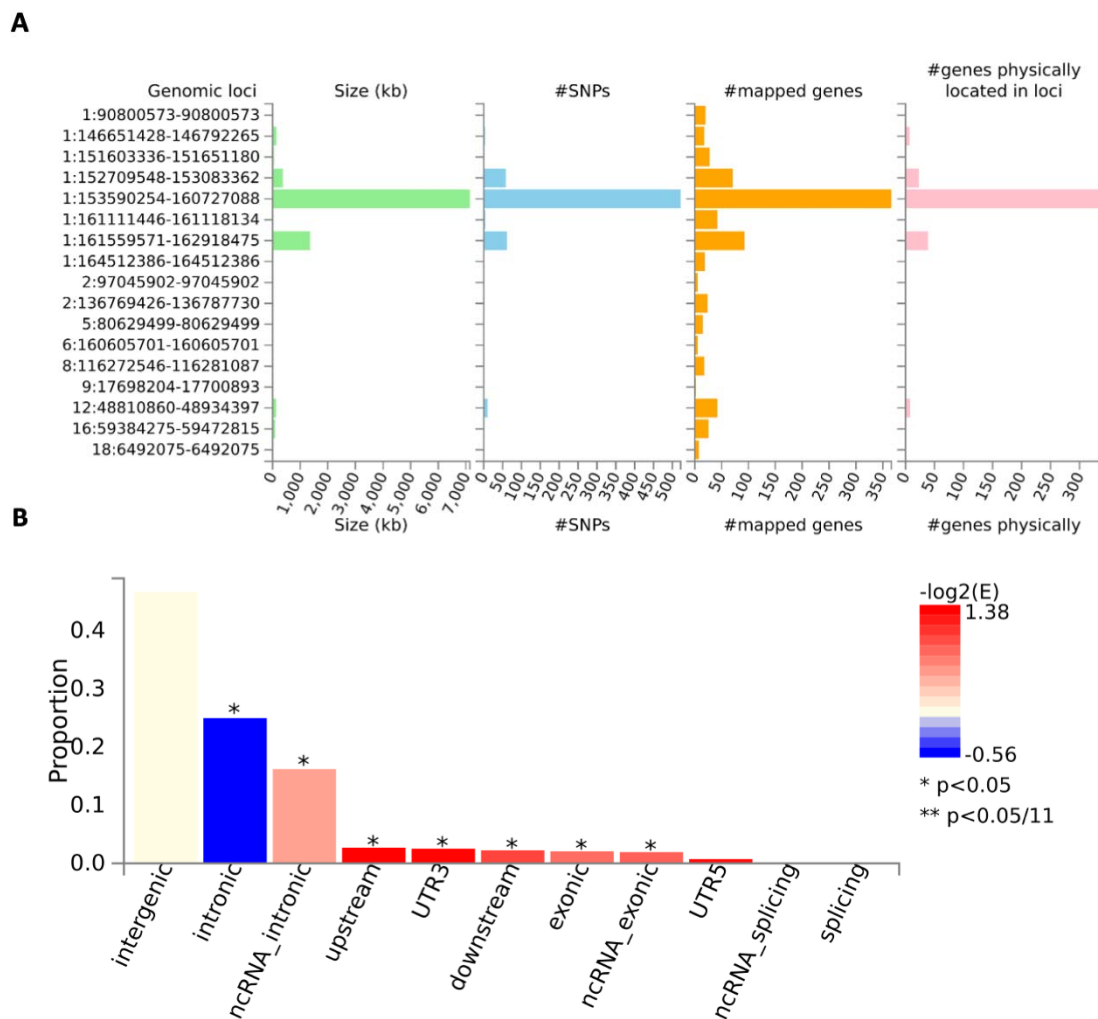


Figure S7. Description of genomic risk loci. FUMA analysis results for SNPs passing the GWAS significance threshold in the BOLT-LMM filtered GWAS.

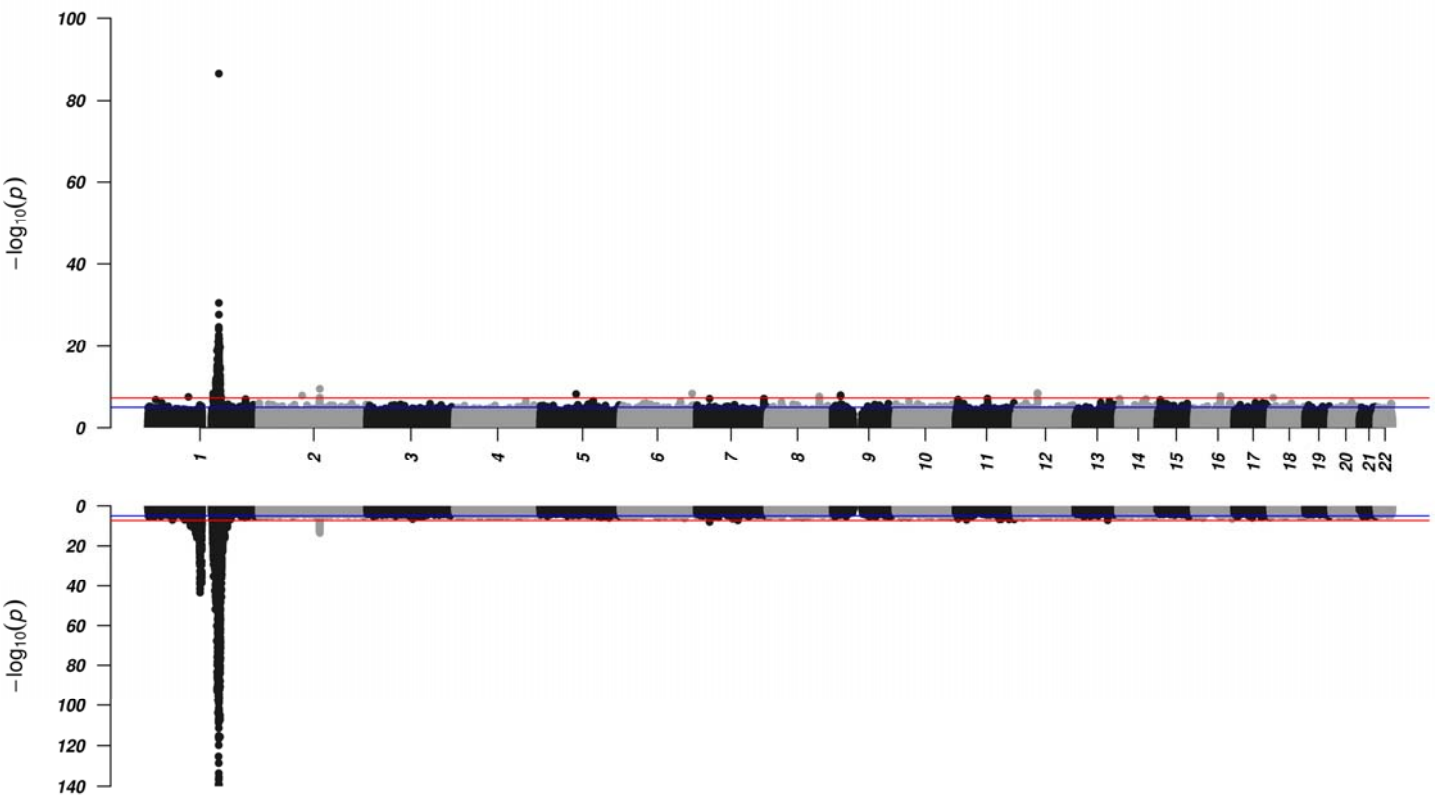


Figure S8. Comparison of GWAS results for neutrophil count in Africans. Manhattan plot of BOLT-LMM neutrophil count GWAS from our study (top) mirrored with another Manhattan plot generated using summary statistics from a GWAS of neutrophil count done in people of African ancestry (Chen et al, reference).

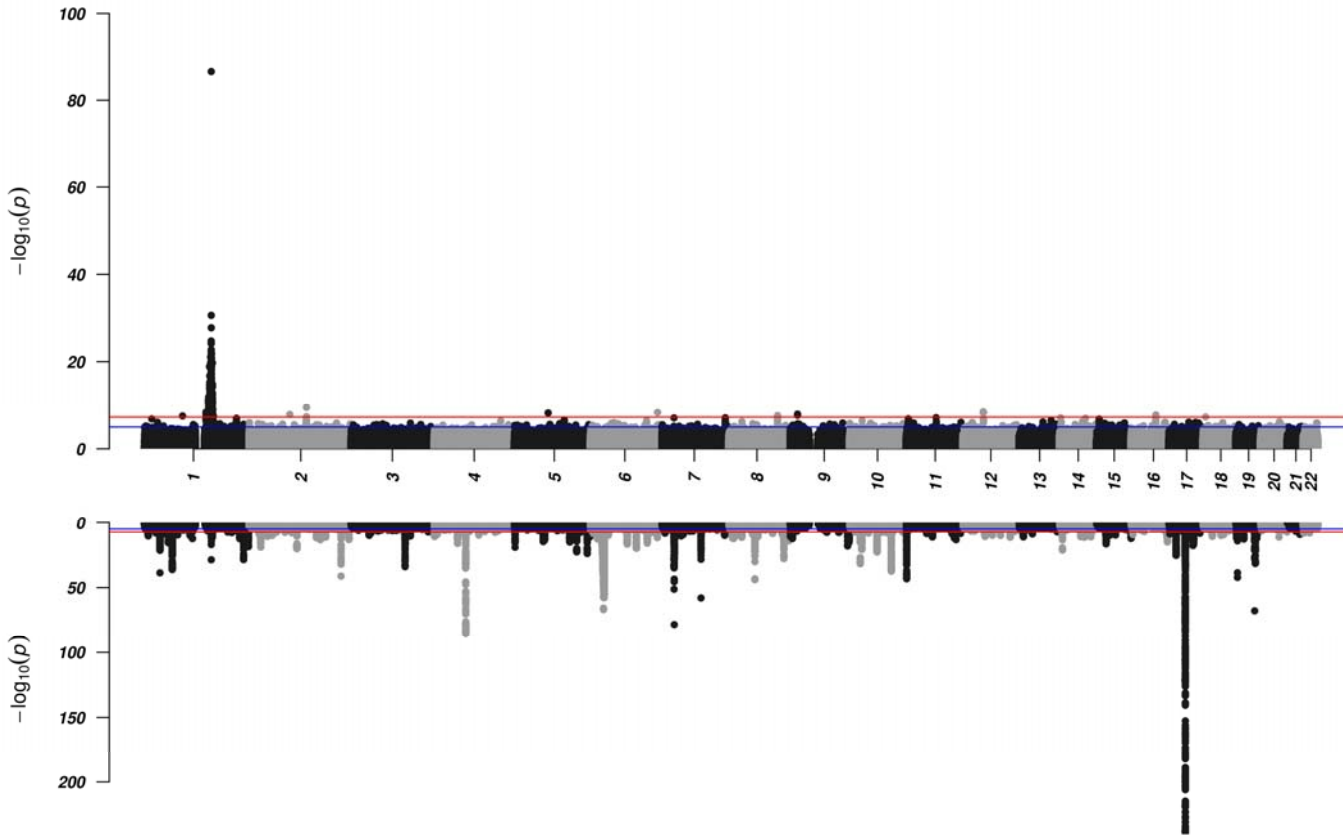


Figure S9. Comparison of GWAS results for neutrophil count in Europeans. Manhattan plot of BOLT-LMM neutrophil count GWAS from our study (top) mirrored with another Manhattan plot generated a GWAS of neutrophil count done in people of European ancestry in UK Biobank by Astle et al 10.1016/j.cell.2016.10.042.

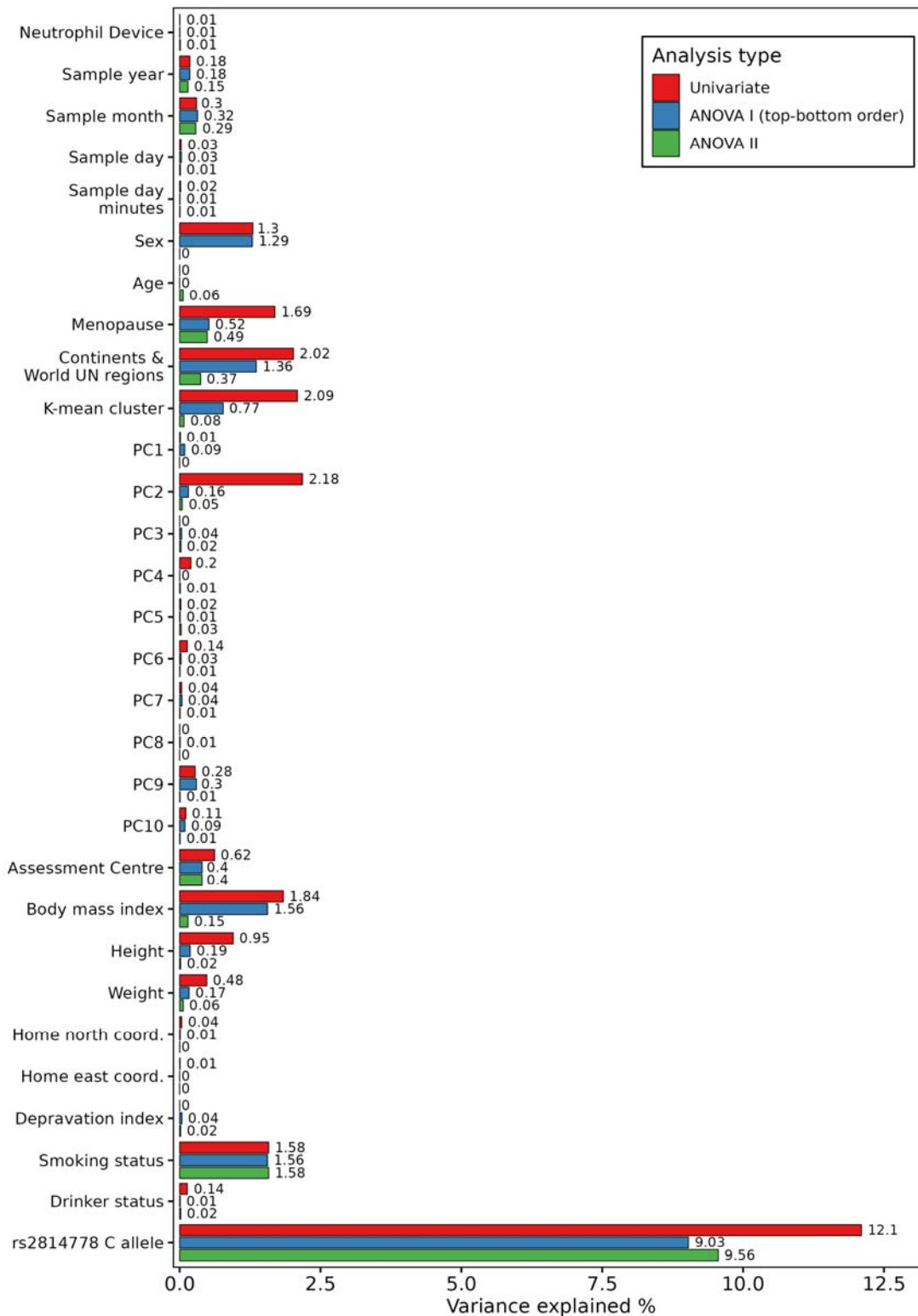


Figure S10: Proportion of variance explained by traits on neutrophil count. The x-axis indicates the trait studied, while the y-axis indicates the proportion of variance explained (PVE) on neutrophil count by that trait. The PVE of each trait was studied in a univariable manner (red), adjusted for only the other traits above in an ANOVA I hierarchical manner (blue), and adjusted between all other traits in a ANOVA II manner.

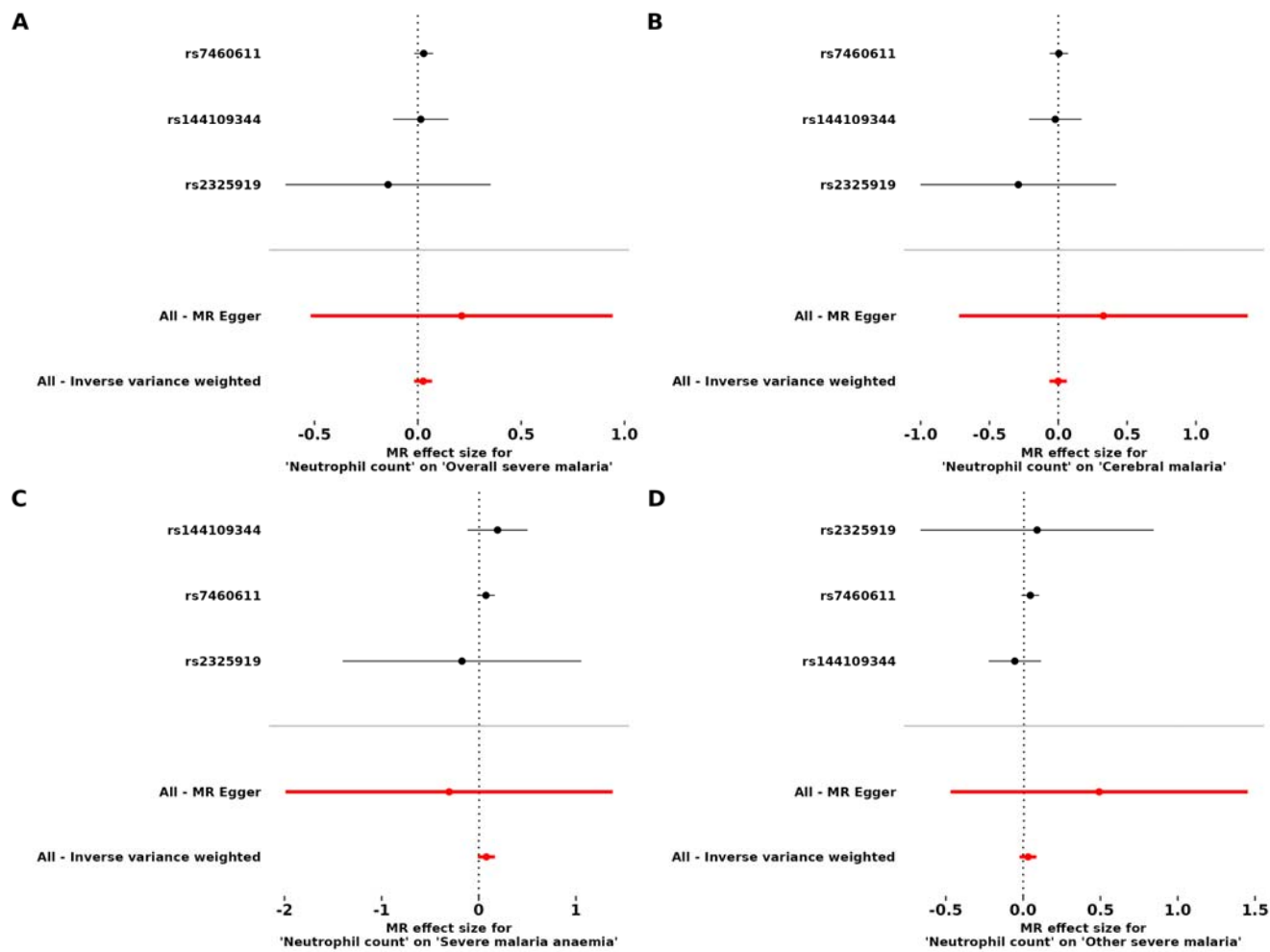


Figure S11: Single-SNP MR analysis of neutrophil count on severe malaria and its subtypes. SNPs proxying for neutrophil count are shown on the x-axis. The effect of each SNP proxying for neutrophil count is displayed on the y-axis, along with 95% CIs. The MR-Egger and IVW MR methods are shown below the single-SNP analysis.

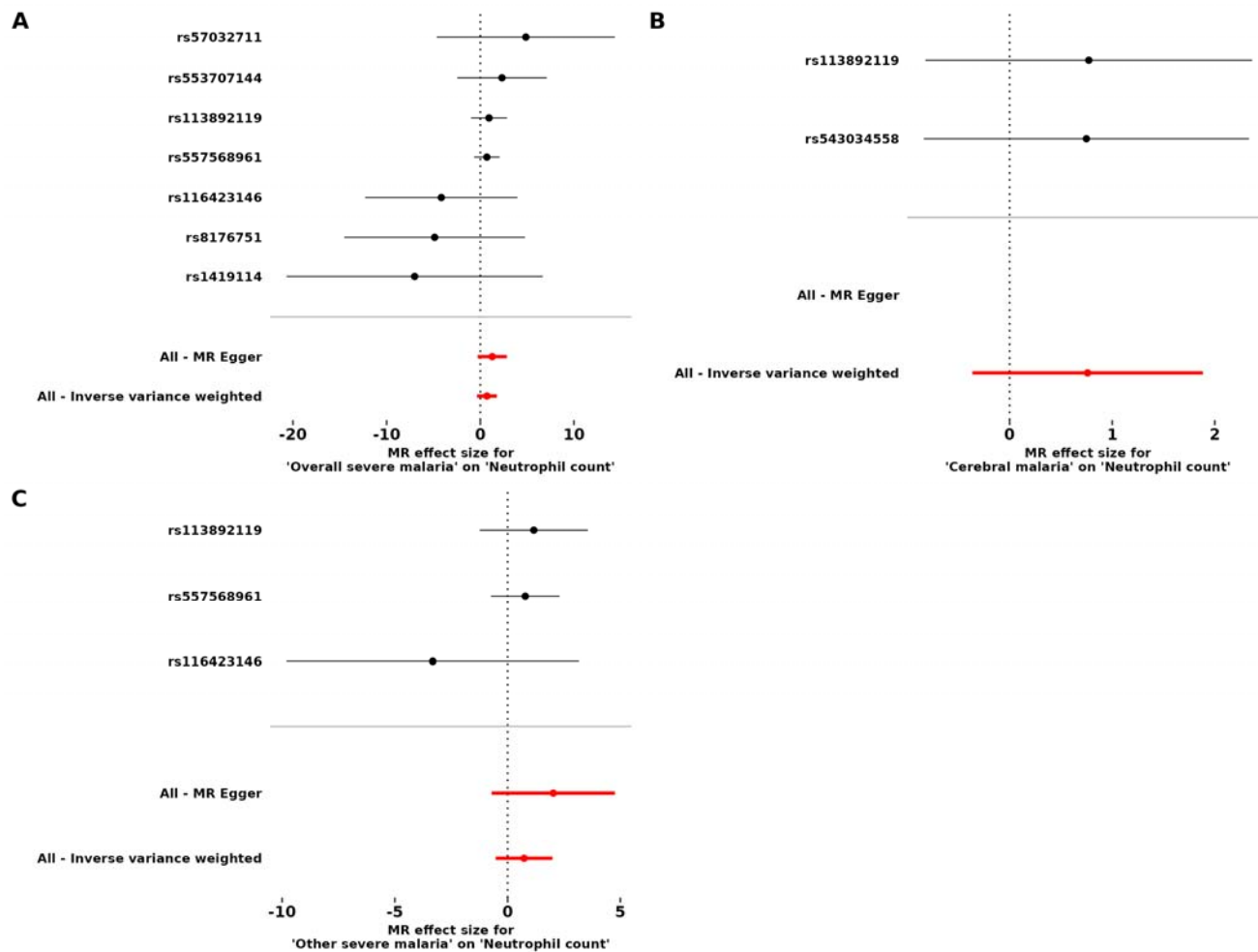


Figure S12: Single-SNP MR analysis of severe malaria on neutrophil count. SNPs proxying for liability to severe malaria are shown on the x-axis. The effect of each SNP proxying for liability to severe malaria is displayed on the y-axis, along with 95% CIs. The MR-Egger and IVW MR methods are shown below the single-SNP analysis.

