A genome-wide association study in individuals of African ancestry reveals the importance of the Duffy-null genotype in the assessment of clozapine-related neutropenia

Sophie E. Legge[1]* (PhD), Antonio F. Pardiñas[1]* (PhD), Marinka Helthuis[2] (MD), John A. Jansen[2] (PhD), Karel Jollie[2] (PharmD), Steven Knapper[3] (MD), James H. MacCabe[4] (MD PhD), Dan Rujescu[5] (MD PhD), David A. Collier[6] (MD PhD), Michael C O'Donovan[1] (MD PhD), Michael J Owen[1] (MD PhD), James T.R. Walters[1] (MD PhD)

*Both authors contributed equally to this work and are joint first authors

Corresponding author

Prof James T.R. Walters

Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Hadyn Ellis Building, Maindy Road, Cathays, Cardiff, CF24 4HQ

Tel: 02920 688434

Email: WaltersJT@Cardiff.ac.uk

Affiliations

[1] MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK

[2] Leyden Delta B.V., Nijmegen, Netherlands

[3] Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff, UK

[4] Department of Psychosis Studies, Institute of Psychiatry Psychology and Neuroscience, King's College London, London, UK

[5] Department of Psychiatry, Martin Luther University Halle-Wittenberg, Halle, Germany

[6] Discovery Neuroscience Research, Eli Lilly and Company Ltd, Lilly Research Laboratories, Erl Wood Manor, Surrey, UK

Abstract

Individuals of African ancestry in the US and Europe are at increased risk of developing schizophrenia and have poorer clinical outcomes. The antipsychotic clozapine, the only licensed medication for treatment-resistant schizophrenia, is under-prescribed and has high rates of discontinuation in individuals of African ancestry, due in part to increased rates of neutropenia. The genetic basis of lower neutrophil levels in those of African ancestry has not previously been investigated in the context of clozapine treatment. We sought to identify risk alleles in the first genome-wide association study of neutrophil levels during clozapine treatment, in 552 individuals with treatment-resistant schizophrenia and robustly inferred African genetic ancestry. Two genome-wide significant loci were associated with low neutrophil counts during clozapine treatment. The most significantly associated locus was driven by rs2814778 (β= -0.9, $p=4.21x10^{-21}$), a known regulatory variant in the Atypical Chemokine Receptor 1 (*ACKR1*) gene. Individuals homozygous for the C allele at rs2814778 were significantly more likely to develop neutropenia and have to stop clozapine treatment (OR=20.4, $p=3.44x10^{-7}$). This genotype, also termed 'Duffy-null', has previously been shown to be associated with lower neutrophil levels in those of African ancestry. Our results indicate the relevance of the rs2814778 genotype for those taking clozapine and its potential as a pharmacogenetic test, dependent on the outcome of additional safety studies, to assist decision-making in the initiation and on-going management of clozapine treatment.

Introduction

The incidence of psychotic disorders is increased in migrant populations[1,2]. In particular, individuals of African and African-Caribbean ethnicity in the UK and African American ethnicity in the US are at especially high risk of developing schizophrenia[3,4]. The exact causes of this are unknown, but are at least in part due to factors associated with being a member of a disadvantaged minority, such as socioeconomic stress and isolation[5,6]. Those of African and African-Caribbean ancestry have also been consistently demonstrated to have poorer clinical outcomes including higher rates of inpatient admissions involving the police or compulsion[7,8]. Furthermore, a recent study examining cases of first-episode psychosis in the UK found that in comparison with White British patients, Black African

and Black African-Caribbean patients had a worse long-term clinical course with lower rates of recovery[8].

The antipsychotic clozapine is the only licenced treatment for those with schizophrenia who fail to respond to other antipsychotics (termed treatment-resistant schizophrenia)[9,10]. While clozapine has been robustly demonstrated to reduce the risk of rehospitalisation and to be associated with better symptomatic and functional outcomes in treatment-resistant schizophrenia[11], it is widely under-prescribed. Clozapine is particularly under-utilised in those of African American ethnicity in the US and Black African or African-Caribbean ethnicity in the UK[12-14], populations that have also been consistently associated with a higher rate of clozapine discontinuation[15,16]. This is at least partly due to increased rates of neutropenia in those of African ancestry whilst taking clozapine[17,18]. Neutropenia is defined as an absolute neutrophil count less than 1500 cells/mm^3, while agranulocytosis is diagnosed if the neutrophil count is below 500 cells/mm^3 and is a severe condition that can endanger life. Clozapine treatment increases the cumulative risks of neutropenia (3.8%) and agranulocytosis (0.8%), which has led many countries to introduce regulatory systems for frequent blood monitoring whilst on the medication[19].

Recent evidence has indicated that genetics plays a role in susceptibility to neutropenia and agranulocytosis on clozapine. Genetic association studies have implicated *HLA-DQB1* and *HLA-B* variants in European and Japanese populations[20-22], and rs149104283, a SNP intronic to transcripts of hepatic transporter genes *SLCO1B3* and *SLCO1B7*[22] in individuals of European ancestry. Reflecting a general dearth of clinical and genetic research in non-European populations, all studies to date investigating clozapine-associated neutropenia have focused predominantly on individuals of European ancestry, limiting the generalizability of findings. This is particularly relevant to those of African ancestry who are known to have lower baseline neutrophil counts on average compared to those from other populations[23]. In this study, we report the first GWAS of neutrophil levels during treatment with clozapine in individuals with treatment-resistant schizophrenia and robustly inferred African genetic ancestry.

Methods

Sample

Individuals included in this study were from the CLOZUK2 sample, all of whom were prescribed clozapine in the UK with a clinician diagnosis of treatment-resistant schizophrenia. The samples were acquired anonymously in partnership with Leyden Delta (Nijmegen, Netherlands), a company that supplies and monitors clozapine in the UK, as part of the CRESTAR collaborative project (www.crestar-project.eu). The project has received UK National Research Ethics Service approval and was in accordance with the UK Human Tissue Act. All samples were anonymised and linked with blood monitoring data provided from clozapine blood monitoring databases. Full details of the CLOZUK2 sample are provided elsewhere[24].

Neutrophil data

The neutrophil count for each individual was defined as the lowest absolute neutrophil count (ANC) on record during clozapine treatment within the blood monitoring database held by Leyden Delta, who retain all historical blood results within their monitoring system. All individuals were started on clozapine after fulfilling baseline criteria including a baseline neutrophil count above 2000 cells/mm^3. For clarity, the terms "ANC" and "neutrophil count" are used interchangeably. All results of ANC <1500 cells/mm^3, indicating neutropenia and the standard threshold that triggers discontinuation of clozapine in the UK, were confirmed by either a) a consecutive ANC <1500 cells/mm^3 or b) two or more results of ANC <2000 cells/mm^3 before or after the index result. We excluded individuals who in the opinion of their treating clinician had an alternative explanation for neutropenia such as concomitant immunosuppressive medication (n = 4).

Genotype quality control and imputation

The CLOZUK2 sample was genotyped by deCODE Genetics (Reykjavik, Iceland) on the Illumina HumanOmniExpress-12 chip. PLINK v1.9[25] was used for genotype quality control following standard protocols[26]. Maximum per-individual and per-marker missingness were set at 2%, and individuals with inbreeding coefficients (F) higher than 0.2 were removed from the dataset. After this curation process, 7,287 individuals genotyped at 698,442 markers remained in the dataset.

Genotype imputation was performed using the Haplotype Reference Consortium (HRC) panel and the pipeline offered by the Michigan Imputation Server[27,28]. As this pipeline allowed for imputation of autosomes only, genotype data from the X-chromosome were imputed on the Cardiff University RAVEN cluster[29] using the SHAPEIT/IMPUTE2

algorithms[30] and a combination of the 1000 Genomes phase 3 (1KGPp3) and UK10K reference panels[31]. Both approaches to genotype imputation have been shown to produce compatible results and to perform similarly in terms of accuracy for variants with minor allele frequencies (MAF) larger than 1%[27]. After imputation, 20 million SNPs with INFO scores higher than 0.8 remained in the dataset.

Defining Genetic Ancestry

In order to select a cohort of individuals with African genetic ancestry, we stratified the CLOZUK2 individuals using Ancestry Informative Markers (AIMs), routinely employed in the field of forensic genetics[32]. The use of AIMs has been shown to be an efficient way of inferring biogeographical ancestry[33,34], which reflects the genetic association of an individual to a particular continental or sub-continental population group. As these groups are broadly defined and show relatively large genetic differentiation[35], this approach circumvents the problems associated with several other approximations to genetic ancestry[36], such as self-reported ethnicity[37] or country-of-origin[38]. Our analysis of AIMs in the CLOZUK2 identified 566 individuals of Sub-Saharan African ancestry, of whom 552 had complete neutrophil count and covariate data, which we term CLOZUK2-AFR. Details on the procedure used to select these individuals are provided in **Supplementary Methods**.

Post-imputation curation of the CLOZUK-AFR genotype data

From this sample, a total of 13.5 million SNPs were taken forward for analysis after applying a MAF filter of 1% and a Hardy-Weinberg Equilibrium (HWE) filter of $p \leq 1x10^{-6}$. HWE tests were carried out using the exact "mid-p" test implemented in the *"HardyWeinberg"* R package[39], as this test is valid for both autosomal and sex-linked markers. Relatedness was assessed using the PC-Relate approach, which identified 18 pairs of relatives ($\hat{\pi} \geq 0.2$). For analyses sensitive to confounding by including related individuals, such as contingency table tests, we excluded one random member of each of these pairs. Otherwise, all individuals were included.

Association analysis of neutrophil count

Imputed data from the CLOZUK2-AFR individuals was analysed using the linear mixed model (LMM) implemented in GCTA v1.26[40], specifically the "leave-one-chromosome-out" procedure[41]. Genotype relatedness matrices, needed to control for population stratification and family structure in LMM frameworks, were calculated directly from the

genotyped SNPs. Covariates used in the analyses included gender, age (at lowest ANC), age^2 and total days on clozapine treatment. The total number of days on clozapine treatment was not associated with neutrophil count. PLINK v1.9[25] was used to identify index SNPs in approximate linkage equilibrium (r^2=0.1) using the LD-clumping procedure, with a p-value cut-off of 10^{-4} and a distance cut-off of 3000 kb. Conditional analyses to further identify independent index SNPs were carried out using the GCTA-COJO procedure[42].

Imputation and analysis of human leukocyte antigen (HLA) alleles

Previous research into clozapine-associated neutropenia has highlighted polymorphisms of the HLA system as drivers of adverse drug reactions in schizophrenia[21]. In order to investigate these, we imputed HLA classical alleles using the software HIBAG v1.12[43] (see **Supplementary Methods** for further details). Association testing of the HLA classical alleles used linear regression of dosages weighted by imputation probabilities, following Levin et al. 2015[44]. Covariates employed matched the GWAS described before, though we added the first 5 principal components calculated by PC-AiR[45] to correct for potential population stratification.

Code availability

The code used to run the analysis is available from the authors upon request.

Results

Absolute neutrophil count in African individuals

A GWAS of lowest ANC during treatment with clozapine in a total of 552 individuals of African ancestry (**Figure 1**; **Supplementary Figure 1**, λ_{GC} = 0.985) identified six independent SNPs (r^2<0.1) that were associated at the genome-wide significance level of p<5x10^{-8} (**Table 1**). Five of those SNPs were in close proximity at a locus (1q23.2) tagging, among others, *ACKR1*, previously called the Duffy Antigen Receptor Complex (*DARC*). The most significantly associated SNP in that locus was rs2814778 (β=-0.86, P=4.21x10^{-21}), which is a regulatory variant in the *ACKR1* promoter region. None of the other SNPs remained genome-wide significant after a mixed model analysis conditional on rs2814778, indicating this SNP is responsible for the association signal of the entire locus (**Supplementary Figure 2**). The other genome-wide significant signal was rs77198048

(β=0.34, P=8.95x10^{-9}) an intronic variant (MAF=1.07%) of the Zinc Finger Protein 618 (*ZNF618)* gene on chromosome 9 (9q32). Our sample size had 80% power to detect a $\beta \geq 1.06$ for alleles with MAF≥0.05 at the genome-wide significance level of P<5x10^{-8}, and 98% power to detect the association of a variant with similar MAF and effect size to rs2814778.

rs2814778 genotype effect on neutrophil counts

We examined the effects of the different rs2814778 genotypes on ANC after excluding 18 related individuals who were included in the mixed model analysis, leaving 534 CLOZUK2-AFR in the sample. Of these, 419 individuals were homozygous for the C (African) allele, 106 were heterozygous and 9 homozygous for the T (European) allele. A Mann-Whitney test (P=0.099) did not show significant ANC differences between CT and TT individuals, supporting previous evidence that the C allele has a recessive effect on neutrophil counts[46]. Thus, for all further analyses we combined individuals with TC and TT genotypes. **Figure 2A** displays the distribution of neutrophil counts during treatment with clozapine by rs2814778 genotype in our African ancestry sample. Individuals with the CC genotype had a median lowest ANC of 1900/mm^3 compared to 2900/mm^3 for CT/TT individuals (P=3.55x10^{-24}).

Figure 2B is a density plot showing the ANC distribution in different CLOZUK2 ancestry subsets (**Supplementary Methods; Supplementary Table 5**), stratified by rs2814778 genotype. While the difference between CC and CT/TT individuals is clearly shown, the CC neutrophil distributions are similar among Sub-Saharan Africans and North Africans (Kolmogorov-Smirnov test P=0.941). Similarly, the neutrophil distributions of CT/TT groups show no difference between Sub-Saharan Africans and all other ancestries (Kolmogorov-Smirnov test P=0.234, see **Figure 2B**).

Given these results we then tested explicitly whether the rs2814778 Duffy-null genotype is more informative of ANC than genetic ancestry, using generalised linear modelling (**Supplementary Methods**). A model including genetic ancestry (European or African) and the GWAS covariates explained 8.29% of the variance in ANC. When the rs2814778 genotype was added to this model, the variance explained increased to 10.94%, and genetic ancestry was no longer associated with ANC. The removal of genetic ancestry resulted in a statistically equivalent model (likelihood ratio test P=0.794) indicating that the rs2814778 genotype is more informative of ANC than genetic ancestry. Adding the

second GWS SNP (rs77198048) to this model also resulted in a statistically equivalent model (likelihood ratio test P=0.182). Analogous results were obtained by fitting these models to predict neutropenia (ANC < 1500 cells/mm^3) in our sample, which yielded a maximum explained variance of 13.03%.

A total of 83 (19.81%) CC individuals had neutropenia during treatment with clozapine (ANC < 1500/mm^3) in comparison to 2 (1.74%) individuals with a T allele (**Table 2**). In both the US and UK, thresholds of ANC below which alterations in clozapine monitoring and management are indicated, have been defined based on normative values from European populations. In the UK, ANC results below 2000/mm^3 mandate closer monitoring and more regular blood testing whereas ANC < 1500/mm^3 requires clozapine treatment to be withdrawn. All study individuals who developed neutropenia had clozapine immediately discontinued, although some were later rechallenged. We used Barnard's exact test to estimate the effect size of rs2814778 on crossing these thresholds, given their important clinical implications. CC individuals are much more likely to develop an ANC < 2000/mm^3 (OR=6.84, 95% CI=4.13-13.67, P=2.90×10^{-16}), and an ANC < 1500/mm^3 (OR=20.36, 95% CI=5.37-314.28, P=3.44×10^{-7}) than T allele carriers. We could not test genotype-mediated differences at lower ANC thresholds, due to the absence of CT/TT carriers.

rs2814778 and benign ethnic neutropenia

There are regulatory mechanisms in place in the US and UK to lower the neutropenia threshold at which clozapine has to be discontinued for those deemed to have benign ethnic neutropenia (BEN), a hereditary condition characterised by mild, chronic neutropenia[47-50]. Of the 74 individuals in our sample with a formal diagnosis of BEN provided by a Consultant Haematologist, 72 (97.30%) have the CC genotype for rs2814778 **(Table 2)**. Considering the safety and clinical outcomes of the 83 individuals with a CC genotype and ANC < 1500 cells/mm^3, a total of 80 were rechallenged with clozapine. Of these, at the time of data collection 75 (93.75%) were still maintained on treatment, 4 (5.0%) had subsequently discontinued, and 1 had died (1.25%, unrelated to ANC). It is not valid to test the sensitivity and specificity of rs2814778 as a predictive test for BEN in this sample due to the lack of systematic screening for BEN.

Association of rs77198048

The second genome-wide significant polymorphism was rs77198048 on chromosome 9 (**Table 1**). This signal comes from a single intronic SNP with no known function and there

were no other SNPs in LD (r^2>0.6) from this region in our sample (or on the 1000 genomes African reference genomes). However, we were not able to identify any reasons related to population stratification or admixture (**Supplementary Figure 3**), which would suggest that this finding is an artefact. None of the 10 individuals that had the minor T allele had neutropenia (ANC < 1500 cells/mm^3), and thus we could not reliably statistically test the impact on neutropenia case/control status.

Association analysis of HLA alleles

Using the HIBAG pipeline, we were able to impute 11 *HLA-DQB1* classical alleles and 21 *HLA-B* classical alleles at MAF > 1% in the CLOZUK2-AFR sample; none were significantly associated with ANC after Bonferroni correction for multiple testing ($P<0.05/32=1.56x10^{-3}$), although the HLA-B*45:01 allele was nominally significant ($P=4.45x10^{-3}$).

Assessment of previous findings from European populations

We were unable to impute any of the following variants implicated in recent association studies of clozapine-associated neutropenia in European populations due to the risk alleles being absent or very rare in CLOZUK2-AFR; rs149104283 (intronic to transcripts of *SLCO1B3* and *SLCO1B7)*[22], the *HLA-DQB1* candidate SNP rs113332494[22], or the *HLA-DQB1* (126Q) and *HLA-B* (158T) amino acid polymorphisms[21].

Discussion

In the first genetic association study of neutrophil counts during clozapine treatment in individuals of African ancestry, we identify two genome-wide significant loci. The most significant association is attributable to rs2814778 ($P = 4.21x10^{-21}$), a regulatory variant in *ACKR1* which has previously been associated with lower neutrophil counts in individuals of African ancestry, and is thought to be causal for BEN. We demonstrate that in those taking clozapine, individuals homozygous for the C allele for rs2814778, also known as the Duffy-null genotype, are substantially more likely (OR=20.36) to be classified as having neutropenia (ANC < 1500 cells/mm^3), the threshold at which clozapine must be stopped. Our results indicate the relevance of the rs2814778 genotype for those taking clozapine and its potential as a pharmacogenetic test, dependent on the outcome of additional safety studies, to assist decision-making in the initiation and on-going management of clozapine treatment.

The rs2814778 Duffy-null genotype had a frequency of 78% in our study population of UK people with African ancestry and has an ~65% frequency in African Americans[51] (**Supplementary Table 6**). In those of African ancestry, it has been robustly implicated in white blood cell and neutrophil counts in several large meta-analyses[46,51-53]. It is also considered to be the cause of BEN[48], an hereditary condition characterized by low neutrophil counts which occurs in 25-50% of individuals with African or Middle Eastern ancestry[47-50]. In support of rs2814778 as causal for BEN, over 97% of individuals diagnosed with BEN in our study were homozygous for the C allele. For those on clozapine, we show that rs2814778 genotype is a better predictor of low ANC than ancestry. Furthermore, the distribution of neutrophil counts closely followed genotype at rs2814778 regardless of ancestry (African, European and South-West Asian) (**Figure 2B**). Given that rs2814778 is the likely cause of BEN, this study implies that BEN is not adequately diagnosed in individuals treated with clozapine. Indeed this is demonstrated by our finding that only 59% of those with the CC genotype and an ANC between 1000 and 1500 cells/mm^3 had a clinical BEN diagnosis. Indirect evidence also supports the hypothesis that BEN is under-diagnosed; BEN had a much lower frequency (14%) in our sample than expected from its prevalence of 25-50% in healthy populations of African ancestry[47-50]. Furthermore, the rates of BEN diagnosis in this sample will be overestimated given its cross-sectional nature (at the point of sample collection), which enriches for those that have been re-challenged with clozapine. Others have similarly noted under-representation of BEN in smaller samples of people of African ancestry taking clozapine[14].

The Duffy-null rs2814778 (C) allele disrupts an erythroid transcription factor GATA-1 binding site in the promoter of *ACKR1* and as a result, the erythrocytes of homozygote rs2814778 carriers do not express ACKR1 protein[54]. Erythrocytes lacking this Duffy antigen are refractory to the malaria parasite *Plasmodium vivax* infection and hence the Duffy-null variant confers an evolutionary advantage[55]. Recent experimental work in mice has shown that ACKR1 deficiencies during early haematopoiesis do not result in reduced production of neutrophils, rather neutrophils exhibit altered phenotypic characteristics that result in their preferential loss from blood by egress into tissues, particularly via migration to the spleen, thus causing neutropenia[56,57]. Importantly there is good evidence that BEN does not lead to increased rates of infection or clozapine-associated agranulocytosis[17,18,58]. In light of this, for patients with a diagnosis of BEN, clozapine monitoring thresholds in the UK are reduced to ANC > 1500 cells/mm^3 and ANC < 1000

cells/mm^3 for initiation and discontinuation, respectively. The recently implemented Clozapine Risk Evaluation and Mitigation Strategy (REMS) program in the US also allocates BEN patients separate monitoring thresholds of ANC > 1000 cells/mm^3 for initiation and ANC < 500 cells/mm^3 for discontinuation.

Our study supports the safety of separate thresholds for those with BEN; 94% of rs2814778 C homozygotes with ANC < 1500 cells/mm^3 were successfully maintained on clozapine treatment after reinstatement. However, in clinical practice, the process of diagnosing BEN is challenging, particularly for those who may be acutely psychotic at the point of clozapine initiation or as a result of clozapine withdrawal. In the UK and the US, a BEN diagnosis is made by a specialist in haematology after assessing the individual's ancestral background, drug history, and the presence of stable low neutrophil counts in the absence of infection. All this necessitates referral by psychiatrists, attendance at haematology outpatient clinics, further blood sampling and review. In light of these practical considerations, it is perhaps unsurprising that only a minority of those who have treatment-resistant schizophrenia and who are eligible for a BEN diagnosis actually receive it, and then go on to have appropriate management with clozapine[14].

Our results indicate that genotyping rs2814778 may offer a simple but sensitive alternative strategy for the diagnosis of BEN. In the context of clozapine treatment, individuals who are homozygous for the C allele and who do not show signs of compromised immune function could have revised neutrophil thresholds in line with current BEN monitoring procedures. This ability to prospectively lower acceptable neutrophil thresholds could address the underutilisation of clozapine in those of African ancestry by (i) enabling more of those suitable for clozapine to start the medication given a lower baseline threshold (ii) avoiding disruption of treatment for those who discontinue clozapine and may or may not under current arrangements be subsequently diagnosed with BEN. In addition to identifying BEN, rs2814778 could also avoid misclassification; the two individuals that had a BEN diagnosis but did not have the Duffy-null genotype in this study have likely been diagnosed incorrectly and could therefore be at increased risk of agranulocytosis. It is not known how many people are denied clozapine due to neutrophil levels that fall below the permitted threshold for initiation (ANC > 2000 cells/mm^3), however we do know that a pre-treatment analysis of neutrophil levels would not have identified the individuals in this study who would have benefitted from such a test.

Furthermore, the potential applications could extend beyond those of African ancestry given rs2814778 occurs at non trivial rates in other populations in which malaria has been historically endemic, including some parts of the Middle East, South West Asia and Oceania (**Supplementary Table 6**)[59].

Further studies are required prior to the implementation of Duffy-null genotype testing to confirm its safety and utility both in prospectively determining BEN status and also to investigate the risk of agranulocytosis when re-challenging Duffy-null patients, though it is reassuring that our cross-sectional data and other currently available evidence support the safety of such approaches[17,18,58]. We envisage that the genetic test could be conducted by clinical or commercial laboratories with DNA sampling being added to the current pre-treatment screening procedures for clozapine, which includes blood testing, once regulatory approval is in place. Recent studies suggest that current genotyping procedures for determining Duffy-null status are both robust and reproducible, and thus their implementation in the context of a clinical blood monitoring service should be straightforward[60,61].

It is important to note that the Duffy-null genotype is the main causal factor for the Duffy-null phenotype (classically termed "Fy(a-b-)")[59,62]. Recent studies have shown that direct genotyping is the best available method to identify Duffy-null individuals in the context of neutropenia[60,61], since in the isolated handful of reported cases with Duffy-null phenotype in the absence of the rs2814778 mutation, neutropenia is not observed[63]. A further challenge to using Duffy-null phenotyping arises from other *ACKR1* mutations causing weak antigen expression which can mimic the Duffy-null phenotype causing serological ambiguity, resulting in misclassification of up to 3.5% of individuals depending on population[60,62,64]. For this reason, we argue that genotyping of rs2814778 would outperform the serological typing of the Duffy-null phenotype.

The second genome-wide significant finding corresponds to an intronic variant in the *ZNF618* gene, which has been characterised as a contributor to methylation and chromatin binding of epigenetic regulators[65]. Given this association comes from a single SNP rs77198048 with no known function, we interpret this finding with caution. Nonetheless we were not able to identify any reasons to suggest that this finding is an artefact but would strongly suggest Independent replication is required to confirm this result.

The Duffy-null association we report in this study adds to existing evidence from our group and others, which together indicate three potential pathways to clozapine-associated neutropenia; (i) immune-mediated[20-22], (ii) hepatic transport and potentially drug toxicity[22], and (iii) genetic factors that cause benign neutropenia causally unrelated to clozapine. In summary, we provide novel insights into the genetic architecture of neutrophil counts during clozapine treatment in individuals with African genetic ancestry. In the first genome-wide association study of neutrophil levels during clozapine treatment, we demonstrate strong association with rs2814778, a regulatory variant in the *ACKR1* gene that has also been described as the genetic basis for BEN. We suggest that rs2814778 genotyping offers an opportunity for personalised medicine in psychiatry although crucially we recognise that further research is needed to establish the safety, acceptability, uptake, clinical utility, and the practical outcomes and cost-benefits of such a test.

References

1. Cantor-Graae, E. & Selten, J.P. Schizophrenia and migration: a meta-analysis and review. *Am J Psychiatry* 2005; **162**: 12-24.
2. McGrath, J., Saha, S., Chant, D. & Welham, J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev* 2008; **30**: 67-76.
3. Fearon, P., Kirkbride, J.B., Morgan, C., Dazzan, P., Morgan, K., Lloyd, T. *et al.* Incidence of schizophrenia and other psychoses in ethnic minority groups: results from the MRC AESOP Study. *Psychol Med* 2006; **36**: 1541-1550.
4. Bresnahan, M., Begg, M.D., Brown, A., Schaefer, C., Sohler, N., Insel, B. *et al.* Race and risk of schizophrenia in a US birth cohort: another example of health disparity? *Int J Epidemiol* 2007; **36**: 751-758.
5. Brugha, T., Jenkins, R., Bebbington, P., Meltzer, H., Lewis, G. & Farrell, M. Risk factors and the prevalence of neurosis and psychosis in ethnic groups in Great Britain. *Soc Psychiatry Psychiatr Epidemiol* 2004; **39**: 939-946.
6. Boydell, J., van Os, J., McKenzie, K., Allardyce, J., Goel, R., McCreadie, R.G. *et al.* Incidence of schizophrenia in ethnic minorities in London: ecological study into interactions with environment. *BMJ* 2001; **323**: 1336-1338.
7. Bhui, K., Stansfeld, S., Hull, S., Priebe, S., Mole, F. & Feder, G. Ethnic variations in pathways to and use of specialist mental health services in the UK. Systematic review. *Br J Psychiatry* 2003; **182**: 105-116.
8. Morgan, C., Fearon, P., Lappin, J., Heslin, M., Donoghue, K., Lomas, B. *et al.* Ethnicity and long-term course and outcome of psychotic disorders in a UK sample: the AESOP-10 study. *Br J Psychiatry* 2017; **211**: 88-94.
9. Kane, J., Honigfeld, G., Singer, J. & Meltzer, H. Clozapine for the treatment-resistant schizophrenic. A double-blind comparison with chlorpromazine. *Arch Gen Psychiatry* 1988; **45**: 789-796.
10. Leucht, S., Corves, C., Arbter, D., Engel, R.R., Li, C. & Davis, J.M. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet* 2009; **373**: 31-41.
11. Tiihonen, J., Mittendorfer-Rutz, E., Majak, M., Mehtala, J., Hoti, F., Jedenius, E. *et al.* Real-World Effectiveness of Antipsychotic Treatments in a Nationwide Cohort of 29823 Patients With Schizophrenia. *JAMA Psychiatry* 2017; **74**: 686-693.
12. Kuno, E. & Rothbard, A.B. Racial disparities in antipsychotic prescription patterns for patients with schizophrenia. *Am J Psychiatry* 2002; **159**: 567-572.
13. Kelly, D.L., Dixon, L.B., Kreyenbuhl, J.A., Medoff, D., Lehman, A.F., Love, R.C. *et al.* Clozapine utilization and outcomes by race in a public mental health system: 1994-2000. *J Clin Psychiatry* 2006; **67**: 1404-1411.
14. Whiskey, E., Olofinjana, O. & Taylor, D. The importance of the recognition of benign ethnic neutropenia in black patients during treatment with clozapine: case reports and database study. *J Psychopharmacol* 2011; **25**: 842-845.
15. Davis, M.C., Fuller, M.A., Strauss, M.E., Konicki, P.E. & Jaskiw, G.E. Discontinuation of clozapine: a 15-year naturalistic retrospective study of 320 patients. *Acta Psychiatr Scand* 2014; **130**: 30-39.
16. Moeller, F.G., Chen, Y.W., Steinberg, J.L., Petty, F., Ripper, G.W., Shah, N. *et al.* Risk factors for clozapine discontinuation among 805 patients in the VA hospital system. *Ann Clin Psychiatry* 1995; **7**: 167-173.
17. Munro, J., O'Sullivan, D., Andrews, C., Arana, A., Mortimer, A. & Kerwin, R. Active monitoring of 12,760 clozapine recipients in the UK and Ireland. Beyond pharmacovigilance. *Br J Psychiatry* 1999; **175**: 576-580.

18. Kelly, D.L., Kreyenbuhl, J., Dixon, L., Love, R.C., Medoff, D. & Conley, R.R. Clozapine underutilization and discontinuation in African Americans due to leucopenia. *Schizophr Bull* 2007; **33**: 1221-1224.
19. Myles, N., Myles, H., Xia, S., Large, M., Kisely, S., Galletly, C. *et al.* Meta-analysis examining the epidemiology of clozapine-associated neutropenia. *Acta Psychiatr Scand* 2018.
20. Saito, T., Ikeda, M., Mushiroda, T., Ozeki, T., Kondo, K., Shimasaki, A. *et al.* Pharmacogenomic Study of Clozapine-Induced Agranulocytosis/Granulocytopenia in a Japanese Population. *Biol Psychiatry* 2016; **80**: 636-642.
21. Goldstein, J.I., Fredrik Jarskog, L., Hilliard, C., Alfirevic, A., Duncan, L., Fourches, D. *et al.* Clozapine-induced agranulocytosis is associated with rare HLA-DQB1 and HLA-B alleles. *Nature Commun* 2014; **5**: 4757.
22. Legge, S.E., Hamshere, M.L., Ripke, S., Pardinas, A.F., Goldstein, J.I., Rees, E. *et al.* Genome-wide common and rare variant analysis provides novel insights into clozapine-associated neutropenia. *Mol Psychiatry* 2016.
23. Gibson, C. & Berliner, N. How we evaluate and treat neutropenia in adults. *Blood* 2014; **124**: 1251-1258; quiz 1378.
24. Pardinas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* 2018.
25. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**: 7.
26. Anderson, C.A., Pettersson, F.H., Clarke, G.M., Cardon, L.R., Morris, A.P. & Zondervan, K.T. Data quality control in genetic case-control association studies. *Nat Protoc* 2010; **5**: 1564-1573.
27. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* 2016; **48**: 1279-1283.
28. Das, S., Forer, L., Schonherr, S., Sidore, C., Locke, A.E., Kwong, A. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* 2016; **48**: 1284-1287.
29. Advanced Research Computing @ Cardiff (ARCCA). Introduction to RAVEN. (accessed: 29/03/2016).
30. van Leeuwen, E.M., Kanterakis, A., Deelen, P., Kattenberg, M.V., Genome of the Netherlands, C., Slagboom, P.E. *et al.* Population-specific genotype imputations using minimac or IMPUTE2. *Nat Protoc* 2015; **10**: 1285-1296.
31. Huang, J., Howie, B., McCarthy, S., Memari, Y., Walter, K., Min, J.L. *et al.* Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun* 2015; **6**: 8111.
32. Phillips, C., Salas, A., Sanchez, J.J., Fondevila, M., Gomez-Tato, A., Alvarez-Dios, J. *et al.* Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 2007; **1**: 273-280.
33. Bulbul, O., Filoglu, G., Zorlu, T., Altuncul, H., Freire-Aradas, A., Sochtig, J. *et al.* Inference of biogeographical ancestry across central regions of Eurasia. *Int J Legal Med* 2016; **130**: 73-79.
34. Shriver, M.D., Parra, E.J., Dios, S., Bonilla, C., Norton, H., Jovel, C. *et al.* Skin pigmentation, biogeographical ancestry and admixture mapping. *Hum Genet* 2003; **112**: 387-399.

35. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 2008; **319**: 1100-1104.
36. Tishkoff, S.A. & Kidd, K.K. Implications of biogeography of human populations for 'race' and medicine. *Nat Genet* 2004; **36**: S21-27.
37. Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *Am J Hum Genet* 2015; **96**: 37-53.
38. Avena, S., Via, M., Ziv, E., Perez-Stable, E.J., Gignoux, C.R., Dejean, C. *et al.* Heterogeneity in genetic admixture across different regions of Argentina. *PloS one* 2012; **7**: e34695.
39. Graffelman, J. Exploring Diallelic Genetic Markers: The HardyWeinberg Package. *J Stat Softw* 2015; **64**: 1-23.
40. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011; **88**: 76-82.
41. Yang, J., Zaitlen, N.A., Goddard, M.E., Visscher, P.M. & Price, A.L. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet* 2014; **46**: 100-106.
42. Yang, J., Ferreira, T., Morris, A.P., Medland, S.E., Genetic Investigation of, A.T.C., Replication, D.I.G. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012; **44**: 369-375, S361-363.
43. Zheng, X., Shen, J., Cox, C., Wakefield, J.C., Ehm, M.G., Nelson, M.R. *et al.* HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J* 2014; **14**: 192-200.
44. Levin, A.M., Adrianto, I., Datta, I., Iannuzzi, M.C., Trudeau, S., Li, J. *et al.* Association of HLA-DRB1 with Sarcoidosis Susceptibility and Progression in African Americans. *Am J Respir Cell Mol Biol* 2015; **53**: 206-216.
45. Conomos, M.P., Miller, M.B. & Thornton, T.A. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* 2015; **39**: 276-293.
46. Reich, D., Nalls, M.A., Kao, W.H., Akylbekova, E.L., Tandon, A., Patterson, N. *et al.* Reduced neutrophil count in people of African descent is due to a regulatory variant in the Duffy antigen receptor for chemokines gene. *PLoS Genet* 2009; **5**: e1000360.
47. Rajagopal, S. Clozapine, agranulocytosis, and benign ethnic neutropenia. *Postgrad Med J* 2005; **81**: 545-546.
48. Manu, P., Sarvaiya, N., Rogozea, L.M., Kane, J.M. & Correll, C.U. Benign Ethnic Neutropenia and Clozapine Use: A Systematic Review of the Evidence and Treatment Recommendations. *J Clin Psychiatry* 2016; **77**: e909-916.
49. Haddy, T.B., Rana, S.R. & Castro, O. Benign ethnic neutropenia: what is a normal absolute neutrophil count? *J Lab Clin Med* 1999; **133**: 15-22.
50. Thobakgale, C.F. & Ndung'u, T. Neutrophil counts in persons of African origin. *Curr Opin Hematol* 2014; **21**: 50-57.
51. Reiner, A.P., Lettre, G., Nalls, M.A., Ganesh, S.K., Mathias, R., Austin, M.A. *et al.* Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). *PLoS Genet* 2011; **7**: e1002108.
52. Nalls, M.A., Wilson, J.G., Patterson, N.J., Tandon, A., Zmuda, J.M., Huntsman, S. *et al.* Admixture mapping of white cell count: genetic locus responsible for lower white blood cell count in the Health ABC and Jackson Heart studies. *Am J Hum Genet* 2008; **82**: 81-87.

53. The Charge Consortium Hematology Working Group. Meta-analysis of rare and common exome chip variants identifies S1PR4 and other loci influencing blood cell traits. *Nat Genet* 2016; **48**: 867-876.
54. Davis, M.B., Walens, A., Hire, R., Mumin, K., Brown, A.M., Ford, D. *et al.* Distinct Transcript Isoforms of the Atypical Chemokine Receptor 1 (ACKR1)/Duffy Antigen Receptor for Chemokines (DARC) Gene Are Expressed in Lymphoblasts and Altered Isoform Levels Are Associated with Genetic Ancestry and the Duffy-Null Allele. *PloS one* 2015; **10**: e0140098.
55. Pierron, D., Heiske, M., Razafindrazaka, H., Pereda-Loth, V., Sanchez, J., Alva, O. *et al.* Strong selection during the last millennium for African ancestry in the admixed population of Madagascar. *Nat Commun* 2018; **9**: 932.
56. Duchene, J., Novitzky-Basso, I., Thiriot, A., Casanova-Acebes, M., Bianchini, M., Etheridge, S.L. *et al.* Atypical chemokine receptor 1 on nucleated erythroid cells regulates hematopoiesis. *Nat Immunol* 2017; **18**: 753-761.
57. Permanyer, M., Bosnjak, B. & Forster, R. Dual role for atypical chemokine receptor 1 in myeloid cell hematopoiesis and distribution. *Cell Mol Immunol* 2018; **15**: 399-401.
58. Richardson, C.M., Davis, E.A., Vyas, G.R., DiPaula, B.A., McMahon, R.P. & Kelly, D.L. Evaluation of the Safety of Clozapine Use in Patients With Benign Neutropenia. *J Clin Psychiatry* 2016; **77**: e1454-e1459.
59. Howes, R.E., Patil, A.P., Piel, F.B., Nyangiri, O.A., Kabaria, C.W., Gething, P.W. *et al.* The global distribution of the Duffy blood group. *Nat Commun* 2011; **2**: 266.
60. Meyer, S., Vollmert, C., Trost, N., Bronnimann, C., Gottschalk, J., Buser, A. *et al.* High-throughput Kell, Kidd, and Duffy matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry-based blood group genotyping of 4000 donors shows close to full concordance with serotyping and detects new alleles. *Transfusion* 2014; **54**: 3198-3207.
61. Lopez, G.H., Morrison, J., Condon, J.A., Wilson, B., Martin, J.R., Liew, Y.W. *et al.* Duffy blood group phenotype-genotype correlations using high-resolution melting analysis PCR and microarray reveal complex cases including a new null FY*A allele: the role for sequencing in genotyping algorithms. *Vox Sang* 2015; **109**: 296-303.
62. Hoher, G., Fiegenbaum, M. & Almeida, S. Molecular basis of the Duffy blood group system. *Blood Transfus* 2018; **16**: 93-100.
63. Rios, M., Chaudhuri, A., Mallinson, G., Sausais, L., Gomensoro-Garcia, A.E., Hannon, J. *et al.* New genotypes in Fy(a-b-) individuals: nonsense mutations (Trp to stop) in the coding sequence of either FY A or FY B. *Brit J Haematol* 2000; **108**: 448-454.
64. Langhi, D.M., Jr. & Bordin, J.O. Duffy blood group and malaria. *Hematology* 2006; **11**: 389-398.
65. Liu, Y.D., Zhang, B., Kuang, H., Korakavi, G., Lu, L.Y. & Yu, X.C. Zinc Finger Protein 618 Regulates the Function of UHRF2 (Ubiquitin-like with PHD and Ring Finger Domains 2) as a Specific 5-Hydroxymethylcytosine Reader. *J Biol Chem* 2016; **291**: 13679-13688.

Figure 1

Manhattan plot of the lowest ANC GWAS in the CLOZUK2-AFR sample. The genome-wide significant peak at chromosome 1 corresponds to the *ACKR1* locus mentioned in the text.

Figure 2

A: Histogram of lowest neutrophil count in the CLOZUK2-AFR sample, stratified by rs2814778 genotype. B: Scaled density plots of ANC in the different CLOZUK2 biogeographical ancestry subsets, stratified by rs2814778 genotype.

Table 1

SNP	CHR	BP	A1	A1 Frequency	Beta	SE	P-value
rs2814778	1	159174683	C	88.38%	-0.856	0.091	4.21×10^{-21}
rs4971072	1	155273869	G	92.25%	-0.700	0.106	3.61×10^{-11}
rs260913	1	163968414	C	4.45%	0.788	0.131	1.89×10^{-9}
rs12128479	1	162268123	G	1.63%	1.205	0.203	2.98×10^{-9}
rs77198048	9	116789254	T	1.07%	1.625	0.282	8.95×10^{-9}
rs12143237	1	162480145	A	2.90%	0.877	0.153	1.11×10^{-8}

Genome-wide significant SNPs from the ANC GWAS in the CLOZUK2-AFR sample. Columns represent; variant name (SNP), chromosome (CHR), base position (BP), risk allele (A1), frequency of the risk allele (A1 Frequency), beta (Beta), standard error (SE) and association P-value.

Table 2

	CC		TC/TT	
	N (% total CC)	**N BEN (%)**	**N (% total TC/TT)**	**N BEN (%)**
ANC < 500	1 (0.24%)	1 (100.00%)	0 (0.00%)	-
500 ≥ ANC < 1000	19 (4.53%)	14 (73.68%)	0 (0.00%)	-
1000 ≥ ANC < 1500	63 (15.04%)	37 (58.73%)	2 (1.74%)	1 (50.00%)
1500 ≥ ANC < 2000	149 (35.56%)	17 (11.40%)	16 (13.91%)	0 (0.00%)
ANC ≥ 2000	187 (44.63%)	3 (1.60%)	97 (84.35%)	1 (1.03%)
Total	419 (100.00%)	72 (17.18%)	115 (100.00%)	2 (1.73%)

ANC and BEN diagnosis in CLOZUK2-AFR, stratified by rs2814778 genotype.

The individual with an ANC < 500 cells/mm^3 had an ANC of 400 cells/mm^3 15 weeks after onset of clozapine treatment. They were rechallenged three days after discontinuing clozapine and although did not have any further agranulocytosis results, they did develop additional amber (ANC < 1500 cells/mm^3) results. The patient was still being successfully maintained on clozapine treatment when the data was extracted.

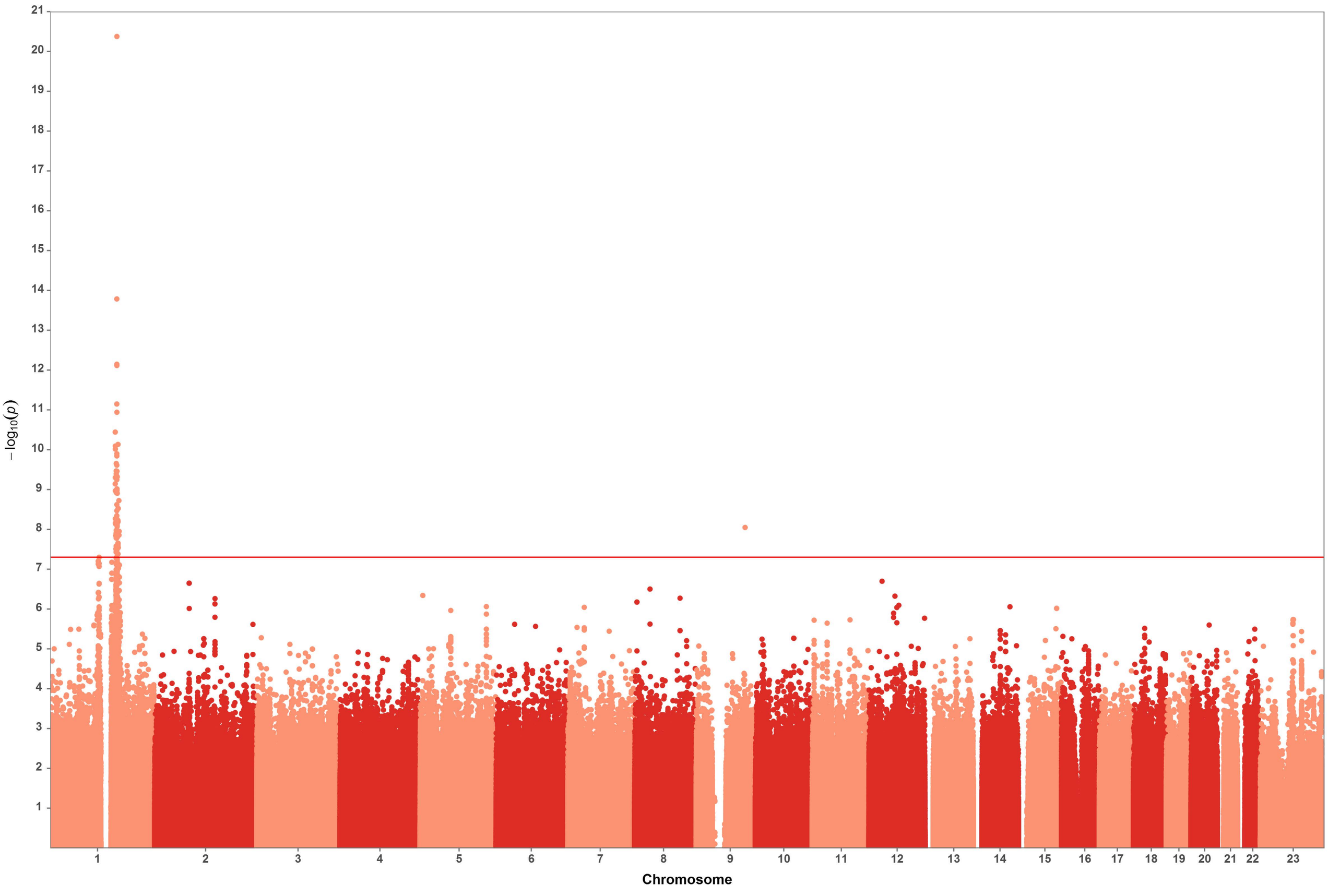
Acknowledgements

This project was supported by Medical Research Council (MRC) Centre (MR/L010305/1), Program (G0800509), and Project (MR/L011794/1) grants to Cardiff University. The project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 279227 (CRESTAR Consortium; http://www.crestar-project.eu/). This publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained therein.

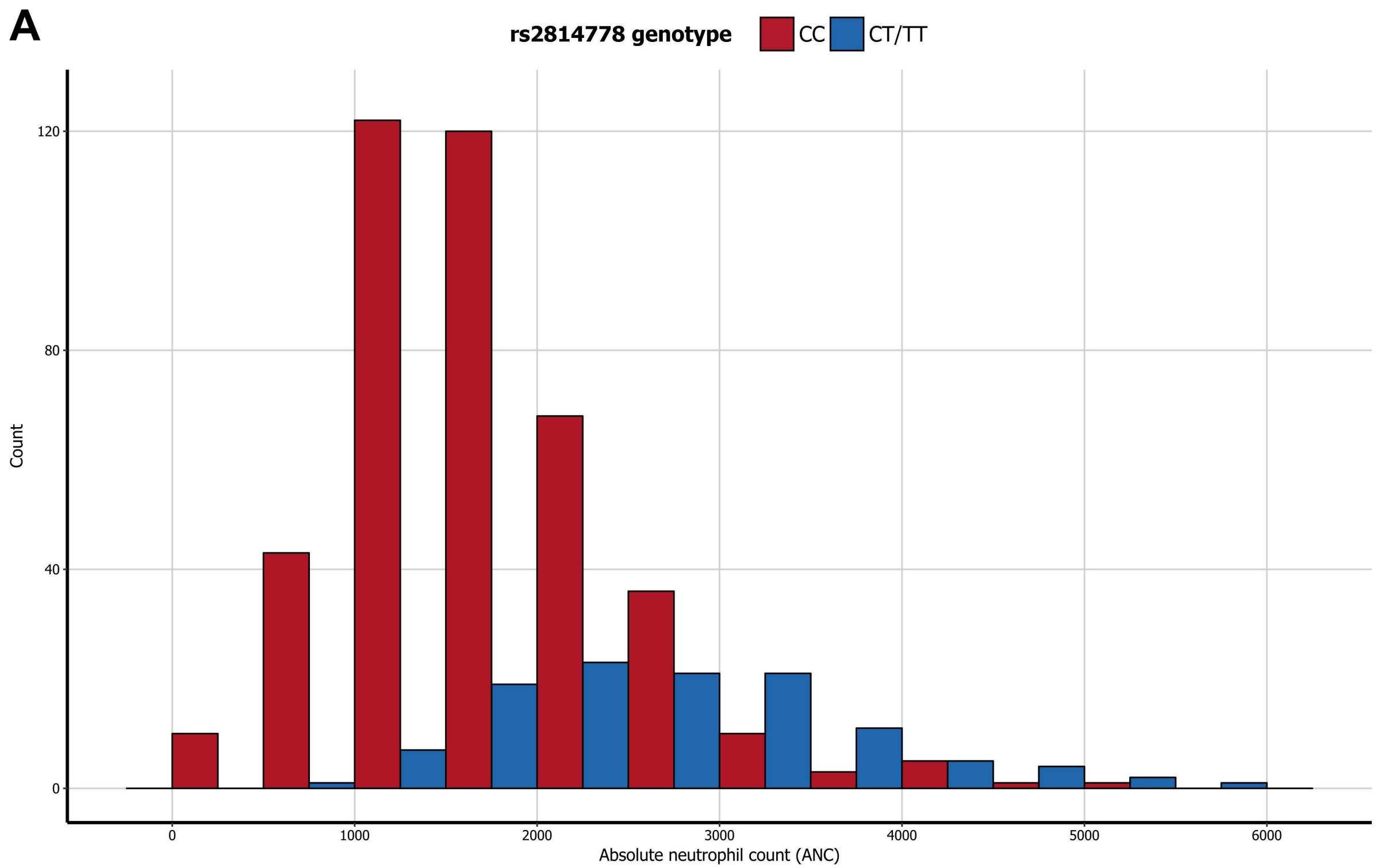
We acknowledge Leyden Delta and Magna Laboratories, UK for supporting the CLOZUK2 sample collection, anonymisation and data preparation (Andy Walker and Anouschka Colson). We acknowledge deCODE genetics (Hreinn Stefansson and colleagues) for genotyping of the CLOZUK2 sample. We acknowledge the MRC Centre laboratory staff (particularly Lucinda Hopkins, Lesley Bates, and Catherine Bresner) at Cardiff University, for laboratory sample management and Wayne Lawrence and Mark Einon, at Cardiff University, for support with the use and setup of computational infrastructures.

Conflict of interest

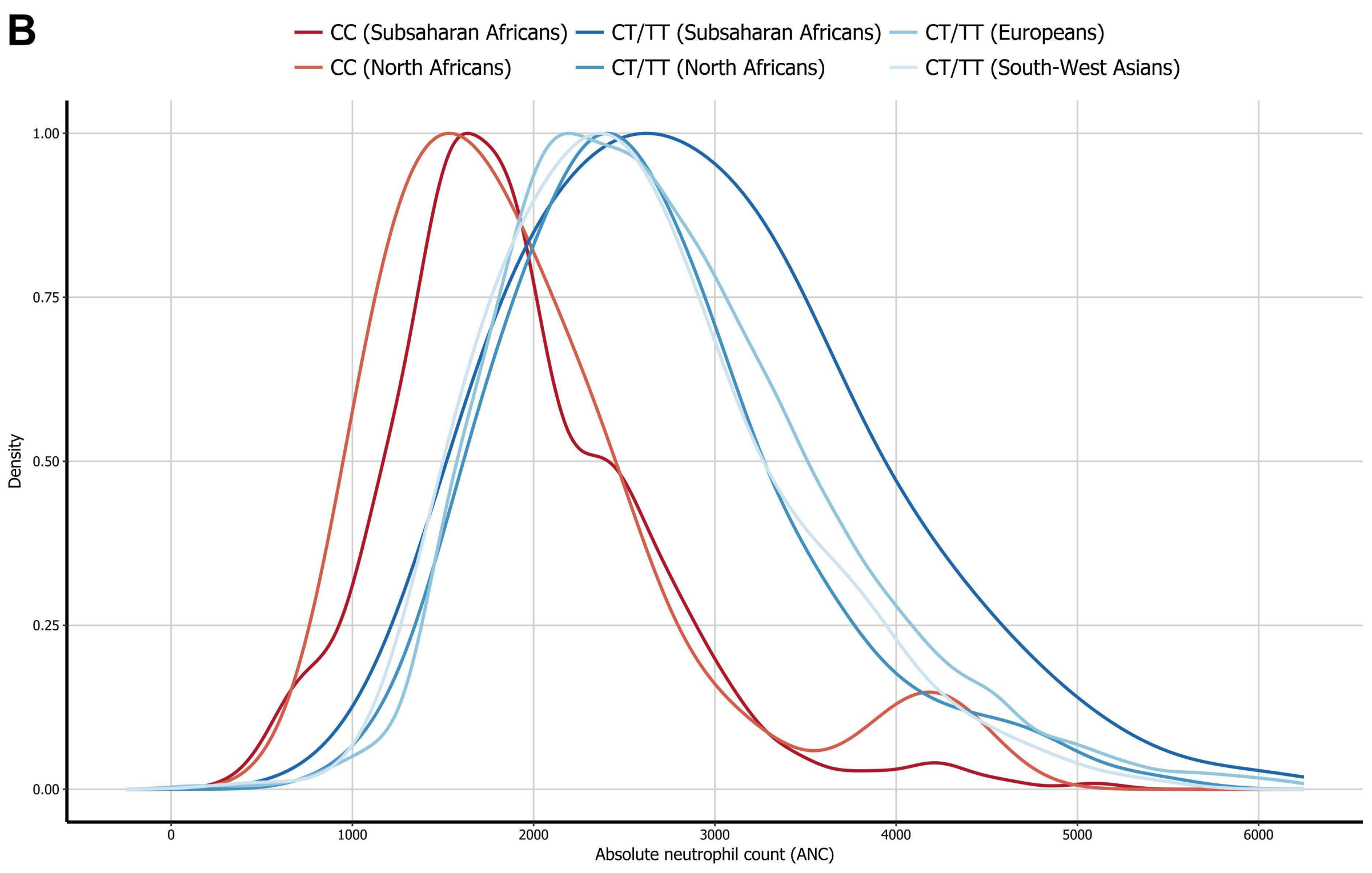
D. A. C. is a full-time employee and stockholder of Eli Lilly and Company. M. H., J. J. & K. J. are full-time employees of Leyden Delta B.V. The remaining authors declare no conflicts of interest.


21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Chromosome

A


rs2814778 genotype
CC
CT/TT
120
80
40
0
Count
0
1000
2000
3000
4000
5000
6000
Absolute neutrophil count (ANC)

B


CC (Subsaharan Africans)
CT/TT (Subsaharan Africans)
CT/TT (Europeans)
CC (North Africans)
CT/TT (North Africans)
CT/TT (South-West Asians)
1.00
0.75
0.50
0.25
0.00
Density
0
1000
2000
3000
4000
5000
6000
Absolute neutrophil count (ANC)

SUPPLEMENTARY METHODS

BIOGEOGRAPHICAL ANCESTRY INFERENCE **2**

Identification of a subset of ancestry-informative markers (AIMs) 2

Building a classification model for biogeographical ancestry 2

Validation on an independent sample of known ancestry 3

Inferring biogeographical ancestry in CLOZUK2 3

INDEPENDENCE WITH THE CLOZUK1 SAMPLE USED IN PREVIOUS STUDIES **4**

IMPUTATION OF HUMAN LEUKOCYTE ANTIGEN (HLA) CLASSICAL ALLELES **4**

STATISTICAL MODELLING OF NEUTROPHIL COUNTS **4**

SUPPLEMENTARY TABLE 1 ..... **5**

SUPPLEMENTARY TABLE 2 ..... **6**

SUPPLEMENTARY TABLE 3 ..... **7**

SUPPLEMENTARY TABLE 4 ..... **8**

SUPPLEMENTARY TABLE 5 ..... **9**

SUPPLEMENTARY TABLE 6 ..... **10**

SUPPLEMENTARY FIGURE 1 ..... **12**

SUPPLEMENTARY FIGURE 2 ..... **13**

SUPPLEMENTARY FIGURE 3 ..... **14**

REFERENCES **15**

Biogeographical ancestry inference

Identification of a subset of ancestry-informative markers (AIMs)

Given the disparity of existing forensic AIM panels[1] and their modest coverage of SNPs present in Illumina genotyping chips[2], we developed our own set of AIMs using publicly available genotype data. To maximise the range of ancestries that a sample could be assigned to, we first obtained the data from the Human Genome Diversity Project (HGDP) sample genotyped by Li et al. 2008[3], which contains 52 worldwide populations, each represented by 5-46 individuals (see Supplementary Table 1 of Leutenegger et al. 2011[4] for full details). We excluded the 10 individuals with Uyghur ethnicity from our analyses, as this population is at the geographic border between South East and South West Asia, and shows heavy admixture from both regions[5]. Genotype data of the remaining 930 samples was merged and restricted to SNPs in common with the CLOZUK2 sample, which left 349,242 SNPs available for analysis in the HGDP individuals.

To define the AIMs we used the F_{ST}-based procedure of Kersbergen et al. 2009[6], which requires defining the desired ancestry groups a priori. Thus, we divided the HGDP sample in the biogeographical categories used by Li et al. 2008[3] (**Supplementary Table 1**), and calculated pairwise between-group F_{ST} metrics for each SNP using PLINK v1.9. Given that the CLOZUK2 sample was recruited in the UK, and most of the individuals within it are expected to have European ancestry[7], we retained only the 6 pairwise comparisons that involved European populations. Then, for each of these, we selected all the SNPs in the top 2.5 percentile of the F_{ST} values[3], assuming that this metric follows a beta distribution[8]. The resulting SNPs were then pruned (r^2=0.4) using the linkage disequilibrium (LD) structure of their corresponding non-European population (i.e. the SNPs selected from the Europe-Africa pairwise comparison were LD-pruned using the African samples as reference). This resulted in 16,114 LD-independent SNPs, which were included in the AIM panel. We note that this number is much larger than the SNP sets used in routine forensic practice, and could potentially be reduced further by applying more stringent marker selection criteria[2]. However, we consider this panel to be suitable for our purposes, as these involve inferring the ancestry of samples that have already been genotyped, rather than efficiently genotyping samples *de novo*.

Building a classification model for biogeographical ancestry

In order to automate the ancestry inference, we trained a simple machine learning model using linear discriminant analysis (LDA), which has been used successfully in similar scenarios[9,10]. In order to reduce the AIMs to a small set of independent variables we used the principal component analysis (PCA) implemented in EIGENSTRAT v6.12[11] on the HGDP AIM SNPs. Then, to retain a number of principal

components proportional to the level of population structure in the sample, we used a Tracy-Widom test[12]. The first 32 PCs were nominally significant in this analysis, and were used to train the LDA model. By using 10-fold cross-validation we determined that LDA could retrieve the original ancestries of each HGDP sample with an accuracy of up to 99.74%.

Validation on an independent sample of known ancestry

For validation purposes we retrieved the data from the Affymetrix Human Origins (AHO) project, described in Lazaridis et al. 2014[13]. This is a collection of samples from 200 worldwide populations, with an emphasis on small and indigenous ethnic groups. From the publicly available genotype data we excluded 841 HGDP individuals that had been also genotyped in this project. Also, given that the Indian subcontinent was only sparsely covered, we added the genotypes of 168 Indians from the South Asian Genome Project (SAGP) sample genotyped by Chambers et al. 2014[14]. This left a total of 1,276 individuals in AHO+SAGP, which were merged with the 16,114 HGDP AIMs. In order to take full advantage of the AIM panel, we did not restrict the merged dataset to overlapping SNPs, leaving those AIMs not present in AHO+SAGP as missing values. Afterwards we ran the EIGENSTRAT PCA algorithm on the "projection" mode, with the HGDP samples as references. This mode runs PCA on the reference data only, using afterwards a least-squares method to infer the coordinates in each PC of all the other samples, considering the non-missing SNP data[15]. The resulting PCA coordinates are effectively unbiased by the AHO+SAGP genotypes, and thus are valid to run the classification model trained before. Based on the first 32 PCs, LDA correctly classified 75.69% to 100% of the AHO+SAGP samples, depending on the ancestry (**Supplementary Table 2**). In fact, as the AHO samples included some biogeographical ancestries not properly represented in the HGDP categories (such as "Central Asian" or "North African") we noted that our algorithm classified them in those ancestries that were genetically more similar (see Extended Data Figure 3 of Lazaridis et al. 2014[13]). Though we regard this performance as good for our purposes, this is conservative as a validation approach, as only 4,200 of the AHO+SAGP SNPs overlapped with our AIM panel. Such modest concordance is due to AHO and HGDP having been genotyped on different array platforms.

Inferring biogeographical ancestry in CLOZUK2

In order to classify the CLOZUK2 individuals we merged the CLOZUK2 genotype data with the 16,114 HGDP AIMs, and repeated the generation of PCs using the EIGENSTRAT "projection" mode, as described above. As in a sample of this characteristics we expect admixed people to be present, we extracted individual classification probabilities from the LDA model, and retained only individuals that

achieved a classification probability for a single biogeographical ancestry of 90% or more (**Supplementary Table 3**).

Independence with the CLOZUK1 sample used in previous studies

To ensure independence with our previous report of genetic variants involved in clozapine-associated neutropenia[16], genetic relatedness was estimated with the individuals of the CLOZUK1 dataset, included in said study. For this, the CLOZUK2-AFR and CLOZUK1 genotypes were merged, and relatedness estimated using PC-Relate[17], controlling for the first 5 ancestry-specific principal components calculated by PC-AiR[18]. This approach been found to result in reliable calculations of kinship coefficients ($\hat{\pi}$), even in the presence of strong population structure[17]. As expected, given that the Legge et al. 2016 study[16] was focused on a European population sample, no duplicates or close relatives ($\hat{\pi} \geq 0.2$) were found.

Imputation of Human Leukocyte Antigen (HLA) classical alleles

Due to the complex LD structure of this area of the genome, specialised protocols are necessary to impute HLA alleles. While the SNP2HLA method has been used in previous research on European populations[19], it lacks the necessary reference data to be used in a Sub-Saharan African sample. Thus, we used the machine learning approach implemented in HIBAG v1.12[20], which can impute 4-digit classical HLA alleles from genotype data. For this we used the provided HLARES African reference sample[21], and retained only the alleles that passed a MAF threshold of 1%. Allele frequencies were also cross-checked with those recently published from the Kenyan Luo population[22], and no discrepancies larger than 5% were found (**Supplementary Table 4**).

Statistical modelling of neutrophil counts

In order to estimate the relative importance of variables defined as predictors of ANC, we fitted generalised linear models using the "gamlss" R package[23]. A log-normal distribution was assumed for the ANC outcome[24], and the covariates used in the GWAS were used to define a baseline model. This model was extended by adding two binary covariates representative of genetic ancestry: "European" and "African" (which included "Sub-Saharan African" and "Middle Eastern / North African"). Finally, one covariate representative of the genome-wide significant finding from the GWAS (rs2814778) was also added to define a full model. Variances explained by these models were calculated with the Nagelkerke's pseudo-R^2 formula. Given that all the models we defined were nested, goodness-of-fit was assessed using likelihood ratio tests.

Supplementary Table 1

Individuals in the HGDP data used to train our ancestry classification model. We note the "South West Asian" category was called "Central/South Asia" in Li et al. 2008. We have renamed it to more properly reflect its geographical provenance, as all of its samples were recruited in Pakistan.

Biogeographical ancestry category	***N***
Sub-Saharan African	101
Middle Eastern	163
European	157
South West Asian	190
South East Asian	228
Oceanian	27
American	64

Supplementary Table 2

Classification accuracy of our LDA model in the AHO+SAGP dataset. "Accuracy" indicates the proportion of AHO+SAGP samples that were correctly classified in a given category. Note that, as expected, misclassifications between ancestries are proportional to their genetic similarity[13].

	Inferred ancestry						
Original ancestry	*Sub-Saharan African*	*Middle Eastern*	*European*	*South West Asian*	*South East Asian*	*Oceanian*	*American*
Sub-Saharan African	*124*	*19*	*0*	*0*	*0*	*0*	*0*
North African	*0*	*42*	*1*	*0*	*0*	*0*	*0*
Middle Eastern	*0*	*109*	*27*	*8*	*0*	*0*	*0*
European	*0*	*40*	*332*	*11*	*0*	*0*	*0*
Central Asian	*0*	*0*	*9*	*57*	*174*	*0*	*0*
South West Asian	*0*	*1*	*0*	*210*	*9*	*0*	*0*
South East Asian	*0*	*0*	*0*	*0*	*46*	*0*	*0*
Oceanian	*0*	*0*	*0*	*0*	*0*	*3*	*0*
American	*11**	*0*	*1**	*0*	*0*	*0*	*42*
ACCURACY	*86.71%*	*75.69%*	*86.68%*	*95.46%*	*100.00%*	*100.00%*	*77.78%*

* These samples were recruited in the United States of America, and thus it is possible that they bear substantial admixture from outside that continent.

Supplementary Table 3

Ancestry classification of the Individuals in the CLOZUK2 sample. Given the results of the LDA model validation (see text), the "Middle Eastern" category was recognised to encompass individuals from the North African biogeographical region. Individuals in the "admixed" category did not achieve 90% classification probability for any of the other categories.

Biogeographical ancestry category	***N***
Sub-Saharan African	566
Middle Eastern / North African	204
European	5900
South West Asian	351
South East Asian	58
Oceanian	0
American	0
Admixed	208

Supplementary Table 4

HLA 4-digit allele frequencies in the CLOZUK2-AFR sample and in the Luo population from Kenya, assessed by Arlehamn et al. 2017[22]. Only alleles with a frequency greater than 1% in CLOZUK2-AFR are shown.

HLA 4-digit allele	Allele frequency in CLOZUK2-AFR	Allele frequency in Luo
*B*07:02*	8.11%	3.50%
*B*08:01*	1.22%	3.50%
*B*14:01*	1.00%	N/A
*B*14:02*	1.56%	3.00%
*B*15:03*	7.22%	6.00%
*B*15:10*	6.67%	4.50%
*B*15:16*	1.89%	0.50%
*B*18:01*	2.78%	3.50%
*B*35:01*	7.22%	3.00%
*B*39:10*	1.67%	N/A
*B*42:01*	7.22%	11.50%
*B*44:03*	6.22%	2.50%
*B*45:01*	4.00%	8.00%
*B*49:01*	2.56%	1.00%
*B*51:01*	1.78%	3.00%
*B*52:01*	4.67%	N/A
*B*53:01*	14.56%	10.00%
*B*57:02*	1.22%	1.50%
*B*57:03*	3.67%	2.00%
*B*58:01*	4.11%	7.50%
*B*58:02*	4.56%	8.50%
*DQB1*02:01*	10.67%	10.50%
*DQB1*02:02*	9.83%	7.50%
*DQB1*03:01*	15.91%	20.00%
*DQB1*03:02*	3.09%	1.00%
*DQB1*03:19*	3.55%	N/A
*DQB1*04:02*	6.44%	11.00%
*DQB1*05:01*	16.10%	17.00%
*DQB1*05:02*	2.80%	0.50%
*DQB1*06:02*	25.37%	22.50%
*DQB1*06:03*	2.99%	2.00%
*DQB1*06:09*	1.03%	5.50%

Supplementary Table 5

ANC in the different CLOZUK2 biogeographical ancestry subsets (AFR = Sub-Saharan African, MES = Middle Eastern / North African, EUR= European, SAS = South West Asian), stratified by genotype. Note that a random member of each pair of genetically related individuals ($\hat{\pi} \geq 0.2$) inside each population has been excluded from these counts.

	CC		TC/TT			
	AFR	**MES**	**AFR**	**MES**	**EUR**	**SAS**
ANC < 0.5	1 (0.24%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	7 (0.12%)	0 (0.00%)
0.5 ≥ ANC < 1	19 (4.53%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	18 (0.31%)	1 (0.30%)
1 ≥ ANC < 1.5	63 (15.04%)	10 (25.00%)	2 (1.74%)	3 (1.90%)	96 (1.68%)	7 (2.08%)
1.5 ≥ ANC < 2	149 (35.56%)	13 (32.50%)	16 (13.91%)	22 (13.92%)	744 (13.02%)	54 (16.07%)
ANC ≥ 2	187 (44.63%)	17 (42.50%)	97 (84.35%)	133 (84.18%)	4851 (84.87%)	274 (81.55%)
Total	419 (100.00%)	40 (100.00%)	115 (100.00%)	158 (100.00%)	5716 (100.00)	336 (100.00%)

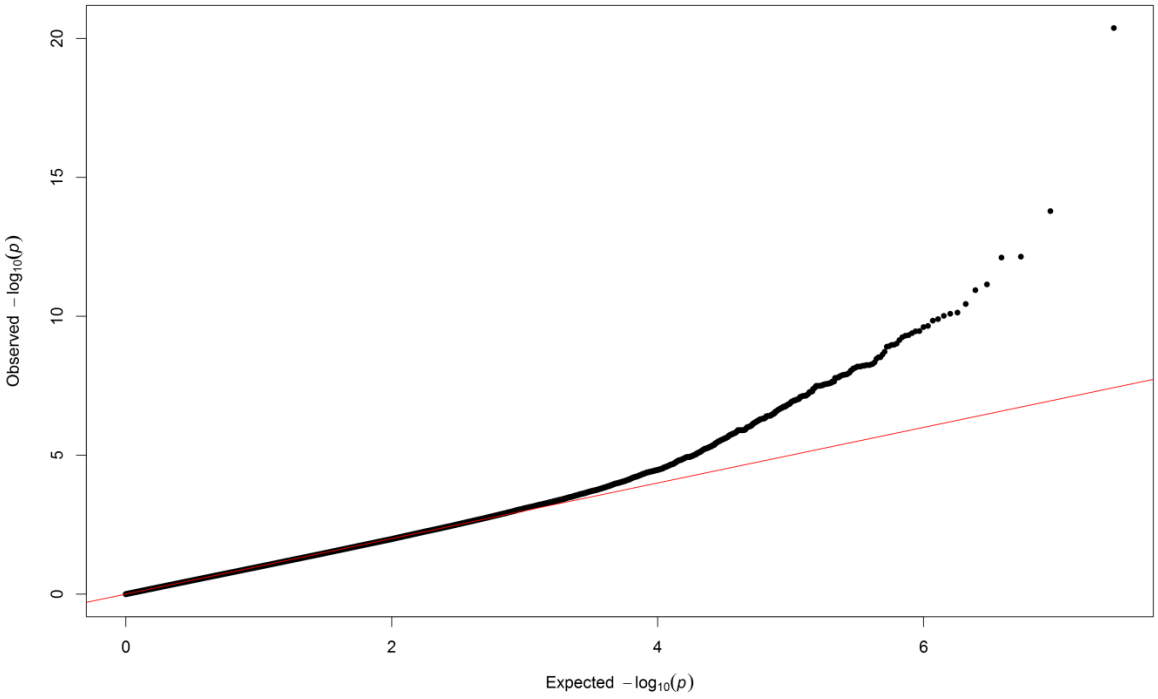
Supplementary Table 6

Frequency of rs2814778 in human populations with a C allele frequency greater than 0.05 indicating polygenicity, taken from The Allele Frequency Database (ALFRED, http://alfred.med.yale.edu, update 23/01/2018)[25]. ALFRED is a free, web accessible, expert-curated compilation of allele frequency data on DNA sequence polymorphisms in anthropologically defined human populations.

Geographic region	Population	Sample Size	C allele freq	T allele freq
Africa	Algerian	70	0.214	0.786
Africa	Bantu speakers	40	0.952	0.048
Africa	Berber	60	0.217	0.783
Africa	Biaka	208	0.995	0.005
Africa	Chagga	88	0.989	0.011
Africa	Esan	198	1	0
Africa	Ethiopian Jews	116	0.821	0.179
Africa	Ghanaian	70	0.986	0.014
Africa	Hausa	126	1.000	0.000
Africa	Ibo	94	1	0
Africa	Ivoirian	66	0.985	0.015
Africa	Libya	258	0.182	0.818
Africa	Lisongo	14	1	0
Africa	Luhya	706	0.994	0.006
Africa	Malinke	226	1	0
Africa	Masai	324	0.935	0.065
Africa	Mbuti	104	0.981	0.019
Africa	Mende	170	1	0
Africa	Moroccans	178	0.175	0.825
Africa	Mozabite	112	0.214	0.786
Africa	San	14	1	0
Africa	Sandawe	78	0.974	0.026
Africa	Sierra Leone	90	1	0
Africa	Somali	108	0.861	0.139
Africa	Tunisian	766	0.206	0.794
Africa	Yoruba	750	0.997	0.003
Africa	Zaramo	78	1	0
Asia	Arabs (U.A.E.)	138	0.424	0.576
Asia	Balochi	100	0.180	0.820
Asia	Bedouin	98	0.35	0.65
Asia	Brahui	50	0.06	0.94
Asia	Druze	202	0.064	0.936
Asia	Kachari	30	0.067	0.933
Asia	Kuwaiti	26	0.308	0.692
Asia	Makrani	32	0.312	0.688
Asia	Palestinian	228	0.327	0.673
Asia	Saudi	196	0.648	0.352
Asia	Timorese	234	0.089	0.911
Asia	Yemenite Jews	130	0.507	0.493
Europe	Cypriot	118	0.068	0.932
Europe	Sephardic Jews	52	0.231	0.769

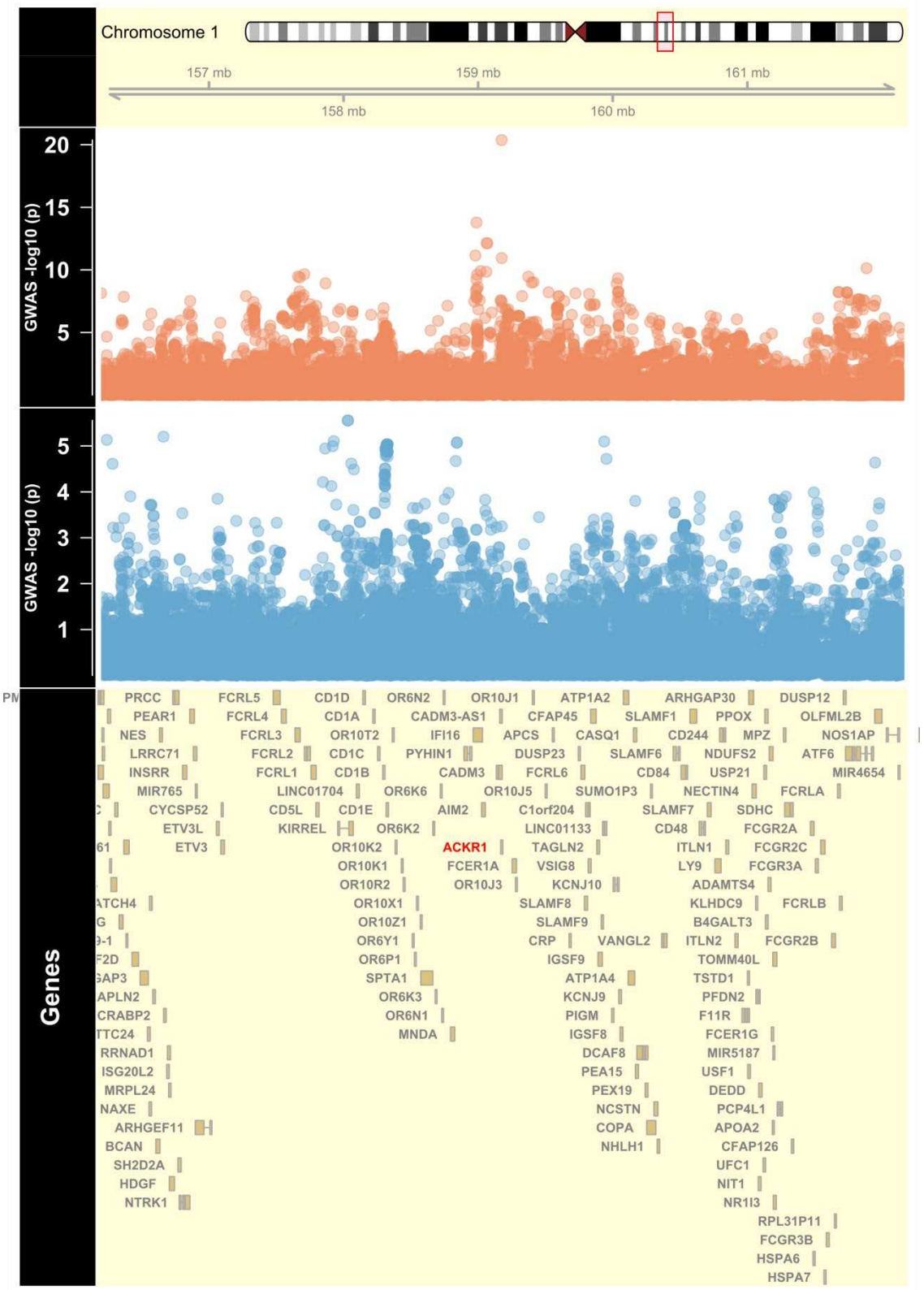
Europe	Spanish (Andalusian)	126	0.278	0.722
North America	African Americans	692	0.826	0.174
North America	Afro-Caribbeans	192	0.885	0.115
North America	Afro-Ecuadorian	58	0.672	0.328
North America	Colombian	318	0.074	0.926
North America	Guihiba	20	0.05	0.95
North America	Jamaican	90	0.856	0.144
North America	Puerto Rican	318	0.129	0.871
Oceania	Papuan New Guinean	42	0.071	0.929
Oceania	Samoans	16	0.125	0.875

Supplementary Figure 1



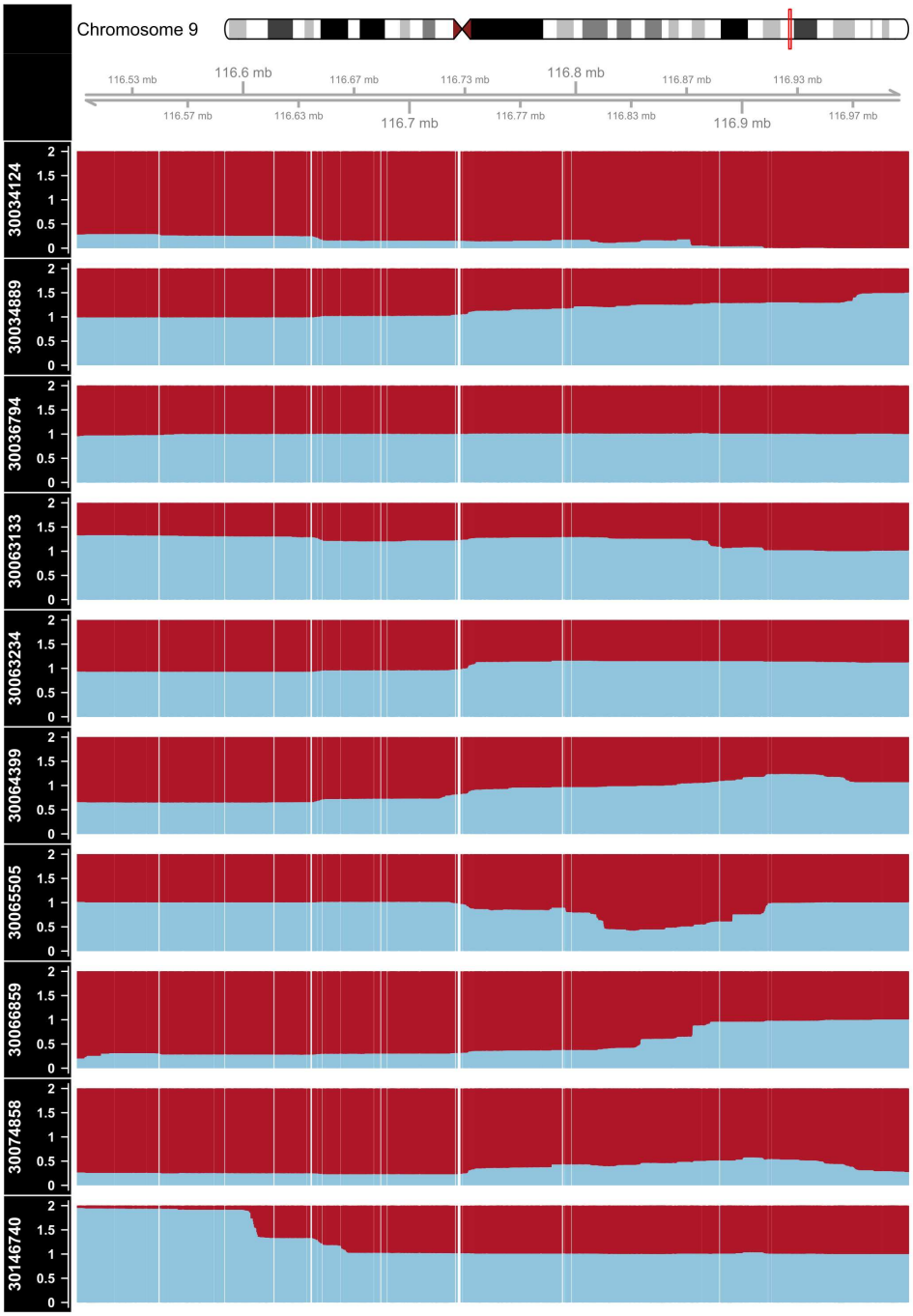
QQ plot (λ_{GC} = 0.985) of the lowest ANC GWAS in CLOZUK2-AFR.

Supplementary Figure 2



Detail of the 1q23.2 region, showing the raw ANC GWAS p-values before (upper plot, red) after (lower plot, blue) conditioning on rs2814778 genotype counts. The *ACKR1* gene is highlighted in red.

Supplementary Figure 3



High-resolution local admixture plot of a segment of chromosome 9 produced by the ELAI software[26], showing ten individuals that are T/A heterozygotes for the rs77198048 SNP. Proportions of European (blue) and African (red) ancestry were inferred using 1000 Genomes EUR/AFR superpopulation genotype data, and 10 ELAI runs over a 5 Mb region (chr9:114000000-119000000). Prior admixture date between EUR/AFR was set to 80 generations[27]. No introgressed European haplotype can be consistently detected around the 116.78 mb mark, which harbours our variant of interest. Note that while some of the individuals might bear a certain degree of European admixture, this extends to segments much larger than the average African LD block[28] (153kb), and thus should be properly accounted by the mixed-model association procedure[29].

References

1. Soundararajan U, Yun L, Shi M, Kidd KK. Minimal SNP overlap among multiple panels of ancestry informative markers argues for more international collaboration. *Forensic Sci Int Genet* 2016; **23**: 25-32.

2. Phillips C, Parson W, Lundsberg B, et al. Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. *Forensic Sci Int Genet* 2014; **11**: 13-25.

3. Li JZ, Absher DM, Tang H, et al. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* 2008; **319**(5866): 1100-4.

4. Leutenegger A-L, Sahbatou M, Gazal S, Cann H, Genin E. Consanguinity around the world: what do the genomic data of the HGDP-CEPH diversity panel tell us? *Eur J Hum Genet* 2011; **19**(5): 583-7.

5. Xu S, Jin L. A Genome-wide Analysis of Admixture in Uyghurs and a High-Density Admixture Map for Disease-Gene Discovery. *Am J Hum Genet* 2008; **83**(3): 322-36.

6. Kersbergen P, van Duijn K, Kloosterman AD, den Dunnen JT, Kayser M, de Knijff P. Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans. *BMC Genetics* 2009; **10**(1): 69.

7. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *bioRxiv* 2016.

8. Weir BS, Hill WG. Estimating F-statistics. *Annu Rev Genet* 2002; **36**(1): 721-50.

9. Bulbul O, Filoglu G, Zorlu T, et al. Inference of biogeographical ancestry across central regions of Eurasia. *Int J Legal Med* 2016; **130**(1): 73-9.

10. Yang N, Li H, Criswell LA, et al. Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine. *Hum Genet* 2005; **118**(3): 382-92.

11. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**(8): 904-9.

12. Patterson NJ, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006; **2**(12): e190.

13. Lazaridis I, Patterson N, Mittnik A, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* 2014; **513**(7518): 409-13.

14. Chambers JC, Abbott J, Zhang W, et al. The South Asian Genome. *PLOS ONE* 2014; **9**(8): e102645.

15. Skoglund P, Malmström H, Omrak A, et al. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. *Science* 2014; **344**(6185): 747-50.

16. Legge SE, Hamshere ML, Ripke S, et al. Genome-wide common and rare variant analysis provides novel insights into clozapine-associated neutropenia. *Mol Psychiatry* 2017; **22**: 1502–8.

17. Conomos Matthew P, Reiner Alexander P, Weir Bruce S, Thornton Timothy A. Model-free Estimation of Recent Genetic Relatedness. *Am J Hum Genet* 2016; **98**(1): 127-48.

18. Conomos MP, Miller MB, Thornton TA. Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. *Genet Epidemiol* 2015; **39**(4): 276-93.

19. Jia X, Han B, Onengut-Gumuscu S, et al. Imputing Amino Acid Polymorphisms in Human Leukocyte Antigens. *PLOS ONE* 2013; **8**(6): e64683.

20. Zheng X, Shen J, Cox C, et al. HIBAG-HLA genotype imputation with attribute bagging. *Pharmacogenomics J* 2014; **14**(2): 192-200.

21. Dilthey AT, Moutsianas L, Leslie S, McVean G. HLA* IMP—an integrated framework for imputing classical HLA alleles from SNP genotypes. *Bioinformatics* 2011; **27**(7): 968-72.

22. Arlehamn CSL, Copin R, Leary S, et al. Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 100 Luo infants from the Boro area of Nyanza Province, Kenya. *Hum Immunol* 2017; **78**(4): 325-6.

23. Stasinopoulos DM, Rigby RA. Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *2007* 2007; **23**(7): 46.

24. Dale DC, Alling DW, Wolff SM. Application of time series analysis to serial blood neutrophil counts in normal individuals and patients receiving cyclophosphamide. *Br J Haematol* 1973; **24**(1): 57-64.

25. Rajeevan H, Soundararajan U, Kidd JR, Pakstis AJ, Kidd KK. ALFRED: an allele frequency resource for research and teaching. *Nucleic acids research* 2012; **40**(Database issue): D1010-5.

26. Guan Y. Detecting structure of haplotypes and local ancestry. *Genetics* 2014; **196**(3): 625-42.

27. Pickrell JK, Patterson N, Loh P-R, et al. Ancient west Eurasian ancestry in southern and eastern Africa. *Proceedings of the National Academy of Sciences* 2014; **111**(7): 2632-7.

28. Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. *Human molecular genetics* 2003; **12**(7): 771-6.

29. Thornton TA, Bermejo JL. Local and global ancestry inference and applications to genetic association analysis for admixed populations. *Genet Epidemiol* 2014; **38**(S1): S5-S12.