

## RESEARCH ARTICLE

## Differences in the frequency of genetic variants associated with iron imbalance among global populations

Momodou W. Jallow<sup>1,2\*</sup>, Carla Cerami<sup>1</sup>, Taane G. Clark<sup>2,3</sup>, Andrew M. Prentice<sup>1</sup>, Susana Campino<sup>2\*</sup>

**1** Nutrition Theme, MRC Unit The Gambia at London School of Hygiene & Tropical Medicine, Banjul, The Gambia, **2** Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, London, United Kingdom, **3** Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, United Kingdom

\* [susana.campino@lshtm.ac.uk](mailto:susana.campino@lshtm.ac.uk) (SC); [mwjallow@mrc.gm](mailto:mwjallow@mrc.gm) (MWJ)



## OPEN ACCESS

**Citation:** Jallow MW, Cerami C, Clark TG, Prentice AM, Campino S (2020) Differences in the frequency of genetic variants associated with iron imbalance among global populations. PLoS ONE 15(7): e0235141. <https://doi.org/10.1371/journal.pone.0235141>

**Editor:** Francesc Calafell, Universitat Pompeu Fabra, SPAIN

**Received:** February 5, 2020

**Accepted:** June 9, 2020

**Published:** July 1, 2020

**Copyright:** © 2020 Jallow et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The 1000 Genomes data is publicly available ([www.internationalgenome.org](http://www.internationalgenome.org)). The Keneba Biobank genotyping data for the 13 SNPs used in this study is available in the Supporting Information files.

**Funding:** MWJ, AMP, CC were supported by the UK Medical Research Council (grant MC-A760-5QX00) and the UK Department for International Development under the MRC-DFID Concordat agreement, <https://mrc.ukri.org>. SC and TGC received funding from the MRC UK (MR/R020973/

## Abstract

Iron deficiency anaemia is a major health problem affecting approximately 1.2 billion people worldwide. Young children, women of reproductive age and pregnant women living in sub-Saharan Africa are the most vulnerable. It is estimated that iron deficiency accounts for half of anaemia cases. Apart from nutritional deficiency, infection, inflammation and genetic factors are the major drivers of anaemia. However, the role of genetic risk factors has not been thoroughly investigated. This is particularly relevant in African populations, as they carry high genetic diversity and have a high prevalence of anaemia. Multiple genetic variations in iron regulatory genes have been linked to impaired iron status. Here we conducted a literature review to identify genetic variants associated with iron imbalance among global populations. We compare their allele frequencies and risk scores and we investigated population-specific selection among populations of varying geographic origin using data from the Keneba Biobank representing individuals in rural Gambia and the 1000 Genomes Project. We identified a significant lack of data on the genetic determinants of iron status in sub-Saharan Africa. Most of the studies on genetic determinants of iron status have been conducted in Europeans. Also, we identified population differences in allele frequencies in candidate putative genetic risk factors. Given the disproportionately high genetic diversity in African populations coupled with their high prevalence of iron deficiency, there is need to investigate the genetic influences of low iron status in Sub-Saharan Africa. The resulting insights may inform the future implementation of iron intervention strategies.

## Introduction

Iron deficiency anaemia (IDA) is a major health problem affecting approximately 1.2 billion people worldwide [1]. It was estimated to account for the 7<sup>th</sup> leading cause of disability worldwide in 2017 [2]. IDA is regarded as the dominant cause of anaemia, accounting for approximately 60% of the global anaemia burden [3]. Pre-school children and women of

1) and the BBSRC UK (BB/R013063/1), <https://mrc.ukri.org> <https://bbsrc.ukri.org>. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

childbearing age in low- and middle-income countries are the most vulnerable [3,4], particularly those living in sub-Saharan Africa, where anaemia prevalence in the general population exceeds 40% [3]. This high prevalence of IDA persists despite the existence of aggressive iron supplementation programmes for vulnerable populations (women of childbearing age and children) [5–7].

Although iron supplementation can be effective in nutritional IDA, it is ineffective in non-nutritional IDA, particularly those caused by genetic factors [8]. Therefore, the identification of the major drivers of IDA in sub-Saharan Africa is required to inform new strategies. The discovery of hepcidin and other proteins involved in iron regulation have led to the identification of genetic factors associated with altered iron homeostasis [9–11]. Several genetic variants within the iron regulatory genes have been associated with imbalances in iron homeostasis, which could lead either to iron deficiency or overload [12–16]. Genetic variants leading to excess body iron occur mainly in the haemochromatosis (*HFE*) gene but are also seen in hepcidin (hepcidin antimicrobial peptide (*Hamp*)), transferrin receptor 2 (*TFR2*), solute carrier family 40 member 1 (*SLC40A1*), haemojuvelin (*HJV*) and transferrin (*TF*) genes [9–11]. These loci have important functions in the iron homeostasis pathways. For example, hepcidin regulates iron absorption and release [17]. Genetic polymorphisms in genes involved in the hepcidin suppressive pathway such as *TMPRSS6* (transmembrane protease serine 6), have been associated with low iron status [18–20] and a condition described as iron-refractory iron deficiency anaemia (IRIDA) [18–21]. Individuals with IRIDA have a hereditary form of anaemia that does not respond to oral iron supplementation [22,23]. Although IRIDA is quite rare, it may be at the extreme end of a broad continuum of disease, since *TMPRSS6* genetic variants can lead to different degrees of iron deficiency and anaemia [18–20]. In addition, SNPs in the *TF* gene, also important in iron transport to cells, have also been reported to affect iron status and lead to low iron status [24–26]. Furthermore, SNPs in the divalent metal transporter 1 (*DMT1*), the duodenal apical iron transporter encoded by the *SLC11A2* gene have been associated with an unusual syndrome characterized by microcytic anaemia and a paradoxical iron overload [27,28].

A genome-wide association study (GWAS) investigating genetic determinants of relevant haematological traits and iron status have identified variants in *TF* and *HFE*, which explain approximately 40% of variation in serum transferrin levels [26]. Also, GWASs have identified genetic variants in *TMPRSS6* associated with alterations of serum iron status, erythrocyte volume [29], and haemoglobin levels [20]. African populations have been greatly under-represented in such studies. A GWAS using an African population cohort replicated only the association of two SNPs in *TMPRSS6* with lowered haemoglobin concentration, and one SNP in *TF* with increased ferritin concentrations [30]. Differences in the frequencies of risk alleles and linkage disequilibrium patterns might explain the limited replication of association results between European, Asian and African populations. Hence, there is a need to investigate population-specific genetic variants that may affect iron status.

Here, we conducted a review of the literature to identify genetic variants that have been associated with iron imbalances, with a special focus on SNPs in *TMPRSS6*, *HAMP*, *TF*, *TFR2*, *SLC40A1* and *HFE* genes. We investigated the geographical distribution of studies and assessed the differences in allele frequency of these polymorphisms and their linkage disequilibrium patterns across global populations. We use genetic data from our Keneba Biobank in rural Gambia and from the 1000 Genomes Project. We also explored the possibility of natural selection acting on these genes and any resulting population-specific selection, as measured through large differences in allele frequencies between geographic regions. As part of this, we sought to summarize the geographical distribution of genetic determinants of iron status. The resulting insights may assist in designing future genetic association studies that are geared

towards identifying population-specific genetic risk factors affecting iron status and, ultimately, guiding population-specific iron intervention strategies.

## Methods

### Selection of SNPs

A literature search was conducted using the Human Genetic Epidemiology (HuGE) navigator, a database of published population-based human genetic epidemiology studies. This review was complemented using the PubMed site with search terms: “anaemia”, “iron”, “iron overload”, “iron deficiency anaemia”, “iron imbalance”, “hepcidin”, “genome-wide association study”, “GWAS”, “haematology traits”, and “haemochromatosis”. The search was conducted on articles published between 01 January 1999 to 31 October 2018. The assessment process included examining titles and abstracts of studies and excluding duplicates. Articles were included if they were: (1) original research papers conducted in humans; (2) tested for an association between at least one SNP in the genes commonly linked to dysregulated iron status (*TMPRSS6*, *HAMP*, *TF*, *TFR2*, *SLC40A1* and *HFE*) or iron status measures. These include iron status biomarkers (serum iron, transferrin, ferritin, soluble transferrin receptor, transferrin saturation, total iron binding capacity, unsaturated iron binding capacity and hepcidin) alone or in combination with haematology traits (haemoglobin, red blood cells, hematocrit, mean corpuscular haemoglobin and mean corpuscular hemoglobin concentration). Animal studies, case reports, commentaries and articles not written in English were excluded. Rare variants reported in a single individual or family were discarded. Information on genomic and gene location, allele ancestry, minor allele variant and the predicted consequence of each SNP were obtained from the Ensembl dataset (release 98) [31] and the dbSNP nucleotide variation database [32].

### Genotype data and statistical analysis

We obtained genotype data from the Keneba at MRCG at LSHTM [33] ( $n = 3,116$  healthy Gambian individuals) and from the 1000 Genomes project [ $n = 2,504$ ; 26 populations categorised into African (AFR,  $n = 661$ ), European (EUR,  $n = 503$ ), American (AMR,  $n = 347$ ), East Asian (EAS,  $n = 504$ ) and South Asian (SAS,  $n = 489$ )] [34]. Genotyping of the Keneba Biobank populations was performed using the Infinium 240K Human Exome Beadchip (v1.0 and v1.1). Genotype calling was performed using data-driven clustering (Genome Studio, Illumina, CA, USA).

We assessed the differences in allele frequencies for SNPs with genotype calls in both the Keneba Gambian and the pan-African populations in the 1000 Genomes Project. Linkage disequilibrium (LD) measures ( $D'$  and  $r^2$ ) were calculated using the R package Genetics [35]. The correlation between minor allele frequencies across populations was calculated using the Pearson's correlation coefficient in the R package corplot.

We calculated the allele risk score for each individual by aggregating the number of risk alleles an individual carried. To do this, from each SNP, the risk allele was assigned 1 and alternate allele assigned 0. For the genotype of each SNP, an individual was given either 0 (wild-type), 1 (heterozygote) or 2 (homozygote for the risk allele). Using this information, we determined the allele risk scores across populations for both low and high iron SNPs. For 23 SNPs it was not possible to identify the associated alleles (e.g. just a “A/T” label) or classify the direction of association (e.g. absence of regression coefficients). Also, for some SNPs (*TF* rs3811658 and rs1880669, and *TMPRSS6* rs2072860 and rs2111833) (S1 Table) we found contradictory information about their association with iron biomarkers between studies. They were all excluded from risk allele analysis. Statistic differences in the distribution of risk alleles

between populations were calculated using a Wilcoxon rank sum test in the R statistical package [36]. To allow for multiple comparisons, a Bonferroni correction was applied.

The minor allele frequency (MAF), observed and expected heterozygosities and measures of population differentiation (global and pairwise  $F_{ST}$  to assess differences in allele frequencies) were calculated from the genotype data for all iron-associated SNPs using a combination of the R packages Adegenet [37], Hierfstat [38] and Pegas [39]. Weir & Cockerham  $F_{ST}$  values were calculated and range from 0 to 1, where a zero value implies that the two populations are interbreeding, and a value of one means that the two populations do not share any genetic diversity. Population Branch Statistic (PBS) values were calculated using the  $F_{ST}$  data from the comparison of three populations (AFR-EUR, AFR-SAS, EUR-SAS) according to methods described elsewhere [40]. To evaluate the significance of the observed  $F_{ST}$  and PBS values, the results were compared with the empirical distribution of genome-wide SNPs reported by others using individuals from several geographical locations and including data from the HapMap and HGDP [41–45].

Statistical differences between MAFs were analysed using the two-proportion Z-Test in R. The integrated Haplotype Score (iHS) [46,47] statistic was investigated using Haplotter (<http://haplotter.uchicago.edu/>) [48] and HGDP selection browsers (<http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/>) at the individual genes and surrounding regions.

## Ethics statement

The Keneba Biobank Project received ethical approval from the MRCG at LSHTM Scientific Coordinating Committee and the MRCG at LSHTM/ Gambia Government Joints Ethics Committee (SCC1185). Written informed consent was obtained from each participant.

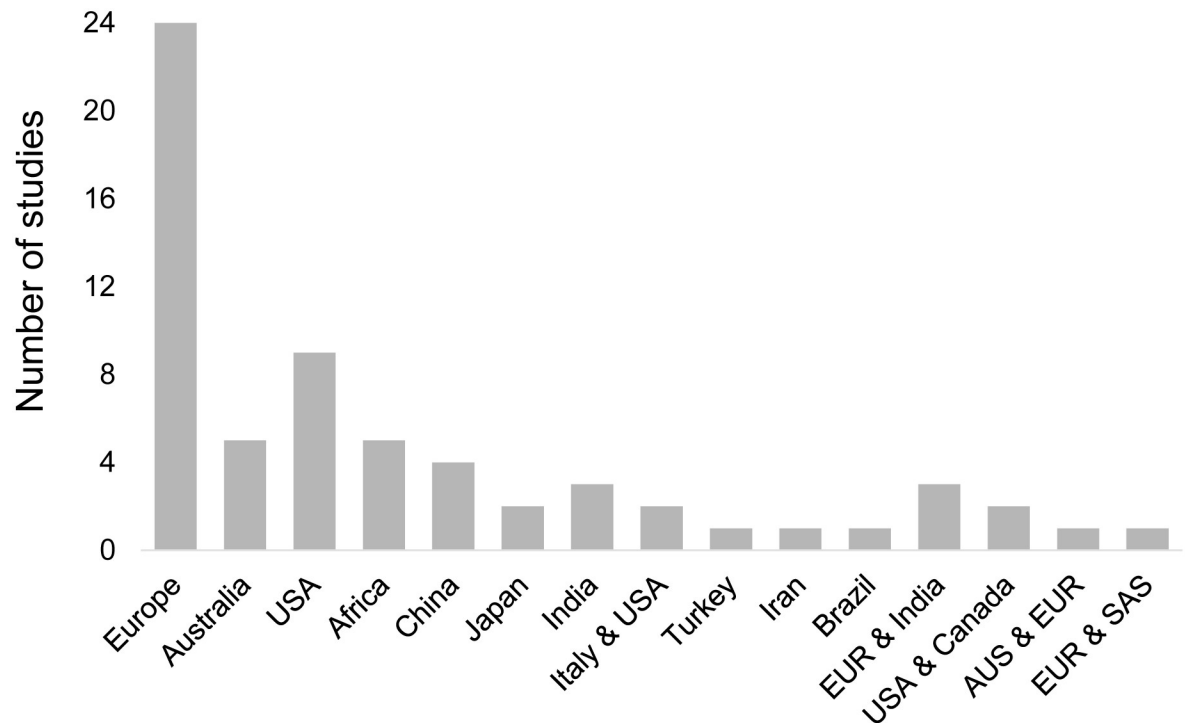
## Results

### Genetic variants associated with iron imbalances

A total of 64 studies were selected that contained data on the effects of genetic polymorphisms on the variations in iron or haematological parameters (S1 Table). The majority of the studies (59/64) were conducted in Europe, Asia and the USA (Fig 1, S2 Table). Only five studies were conducted in Africa, two in Rwanda [49,50], one in Zimbabwe [51], one in South Africa [52] and one meta-analysis across Kenya, Tanzania and South Africa [30]. Across the 64 studies, 50 SNPs were identified in six genes (*TMPRSS6*, *HAMP*, *TF*, *TFR2*, *SLC40A1* and *HFE*) (S1 Table). More than half of these SNPs were found to be associated with variation in iron or in other haematological parameters in more than one country (29 SNPs, 58%). Of these 29 SNPs, 79.3% were reported in more than one ethnic group (S2 Table). Nine SNPs lead to a missense mutation causing an amino acid change, four SNPs had synonymous variants, and the remaining SNPs are in intronic ( $n = 32$ ), regulatory or intergenic regions ( $n = 5$ ).

The highest number of SNPs were identified in the *TMPRSS6* gene region ( $n = 23$ ), where the majority were associated with IRIDA, iron deficiency or indicators of low iron status (S1 Table). The most commonly reported *TMPRSS6* SNP was rs855791, followed by rs4820268, rs2235321 and rs2235324, all associated with biomarkers of low iron status. These SNPs have been mainly reported in non-African populations. Three *TMPRSS6* SNPs (rs5756504, rs5756506 and rs1421312) were also associated with biomarkers indicating elevated iron status (S1 Table).

The *TF* gene had the second highest number of SNPs related to either low or high iron status ( $n = 18$ ). The most common of these (rs3811647) was reported by ten studies (S1 Table). This variant has been mainly associated with elevated transferrin and total iron binding capacity levels [26,53,54]. For the *SLC40A1* gene, three SNPs were selected that led to alterations in



**Fig 1. Geographical locations of the sixty-four studies that reported genetic variants associated with iron imbalance.** Nine studies involved multi-ethnic populations. AUS, Australia; EUR, Europe; SAS, South Asia.

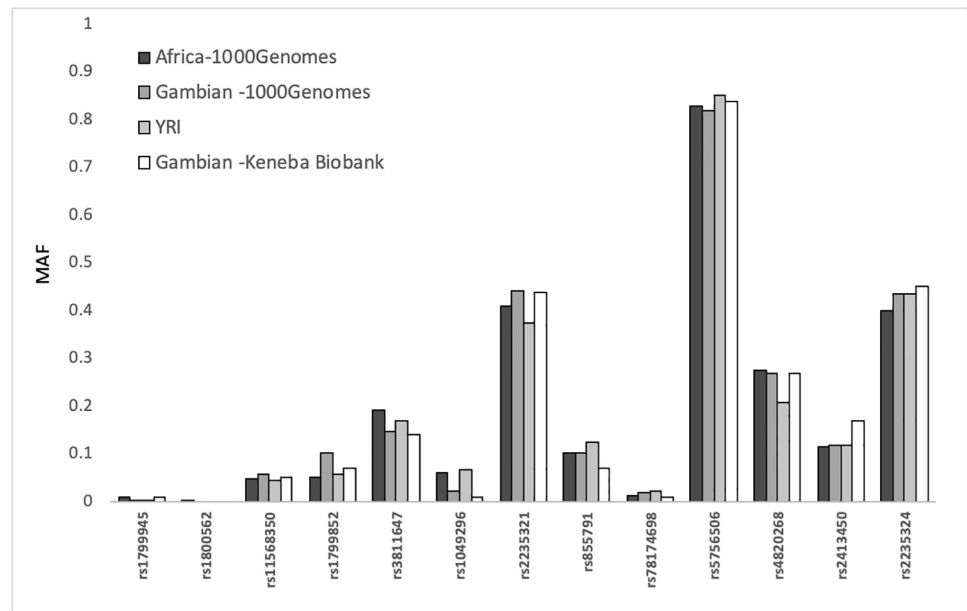
<https://doi.org/10.1371/journal.pone.0235141.g001>

iron status measures and severity of haemochromatosis [50,51,55–57]. One SNP was identified in *HAMP* (rs10421768) [30,55,58–60] and one in *TFR2* (rs7385804) (6,7,33,34,37,38), both of which were found to be associated with increases in haemoglobin and alterations serum in ferritin concentrations [30,55,56,58,61–63]. For the *HFE* gene, we found four SNPs that have been associated with alterations in haemoglobin and/or an increase in the genetic risk of hereditary haemochromatosis [13,14,19,20,24,26,29,56,62,64–74]. The most commonly reported *HFE* variant is rs1800562 (C282Y) [13,19,24,26,29,71,72], which has been widely associated with the severe form of hereditary haemochromatosis in European descents.

### Global geographic distribution of allele frequencies

We investigated the allele frequencies of the 50 SNPs across data from the Keneba Biobank at the MRCG at LSHTM in The Gambia (n = 3,116) and the 1000 Genome project (n = 2,504) [34]. The 1000 Genomes project includes data from African (AFR, n = 661; including from The Gambia), European (EUR, n = 503), American (AMR, n = 347), East Asian (EAS, n = 504) and South Asian (SAS, n = 487) populations. Only thirteen of the 50 SNPs in the *TF*, *TMPRSS6*, *HFE* and *SLC40A1* genes, were available in the Keneba Biobank, because not all the SNPs were on the Exome chip that was used for genotyping this population. When we compared the allele frequencies of the SNPs with data from The Gambians in the Keneba Biobank with the pan-African populations in the 1000 Genomes project, we observed minimal differences (Fig 2).

For the majority of SNPs, the MAFs in the African populations were very different to other worldwide populations (Figs 3 and 4). The greatest allele frequency differences were observed in rs1439816 in *SLC40A1*, and in several SNPs in *TMPRSS6* (including rs855791 and



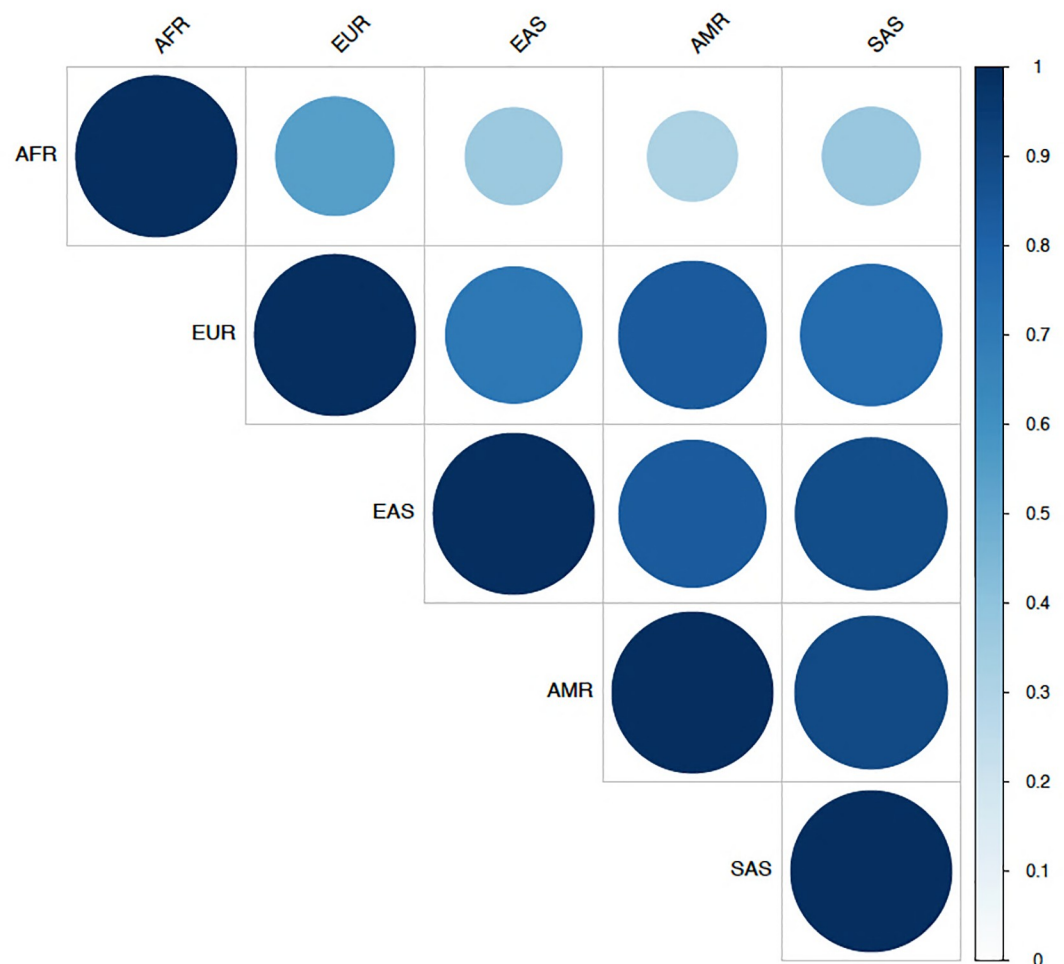
**Fig 2. Minor allele frequencies (MAF) of 13 SNPs across African populations.** Comparing MAF between the two Gambian datasets, Yoruba (YRI) from Nigeria and overall African populations included in the 1000 Genomes Project. The minor alleles were defined by the 1000 Genomes Project.

<https://doi.org/10.1371/journal.pone.0235141.g002>

rs855788). The intronic variant rs1439816 in the *SLC40A1* gene has a MAF of ~20% in the non-African populations but reaches >73% frequency in Africa (S1 Table). The missense variant A736V (*TMPRSS6* rs855791) is the most reported SNP associated with iron deficiency and has a MAF of ~50% across all non-African populations, but in Africa it only reaches 10% (7% in the MRCG Keneba Biobank population) (Fig 4). The intronic variant rs855788 in *TMPRSS6* has a MAF of ~30% across non-African populations, contrasting with a frequency in excess of 86% in the African populations (Fig 4).

From the selected SNPs, several in African (n = 10 SNPs) and East Asian (n = 11 SNPs) populations have fixed ancestral alleles or low MAF (<5%) (Fig 4). These SNPs include four missense variants, with the lowest overall MAF or with fixed ancestral alleles in several populations (associated with low iron: *TMPRSS6* rs78174698 and *TF* rs1799899; associated with increased serum ferritin: *SLC40A1* rs11568350; associated with haemochromatosis: *HFE* rs1800562). The *TMPRSS6* rs78174698 (P555S) MAF is low overall (<2%) across most populations, except in South Asia where the minor allele is >10%. The minor allele for rs1799899 (G277S) is rare in Africa and East Asia (<0.2%), and only reaches >4% MAF in European, American and South Asian populations. For *SLC40A1* rs11568350 (Q248H), the minor allele reaches 5% in Africans, including in both The Gambian populations in the two datasets analysed. In the other global populations, the ancestral allele is almost fixed. The variant A allele of rs1800562 (C282Y) has the highest frequency in European populations (4.3% and 5.3% in Caucasians from Europe in the 1000 Genome Project and in the HapMap CEU population, which have ancestry from Northern and Western Europe, respectively). The frequency of this variant is extremely low in Africans (0.2% in the 1000 Genomes project) and it was not detected (MAF = 0) in the Keneba Biobank population.

We also investigated the population-specific linkage disequilibrium (LD) patterns between SNPs in the candidate genes. There were blocks of high LD in the non-African population, and the overall levels of LD were lower in the African populations (Fig 5, S1 Fig), including in



**Fig 3. Correlation of minor allele frequencies between different geographic regions.** Correlation coefficients were obtained by pairwise comparisons of each of the 50 SNPs identified across two populations. They are coloured according to the value using a gradient from white (representing 0 for no correlation) to dark blue (1 for perfect correlation). The minor allele variant was defined by the 1000 Genomes Project. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

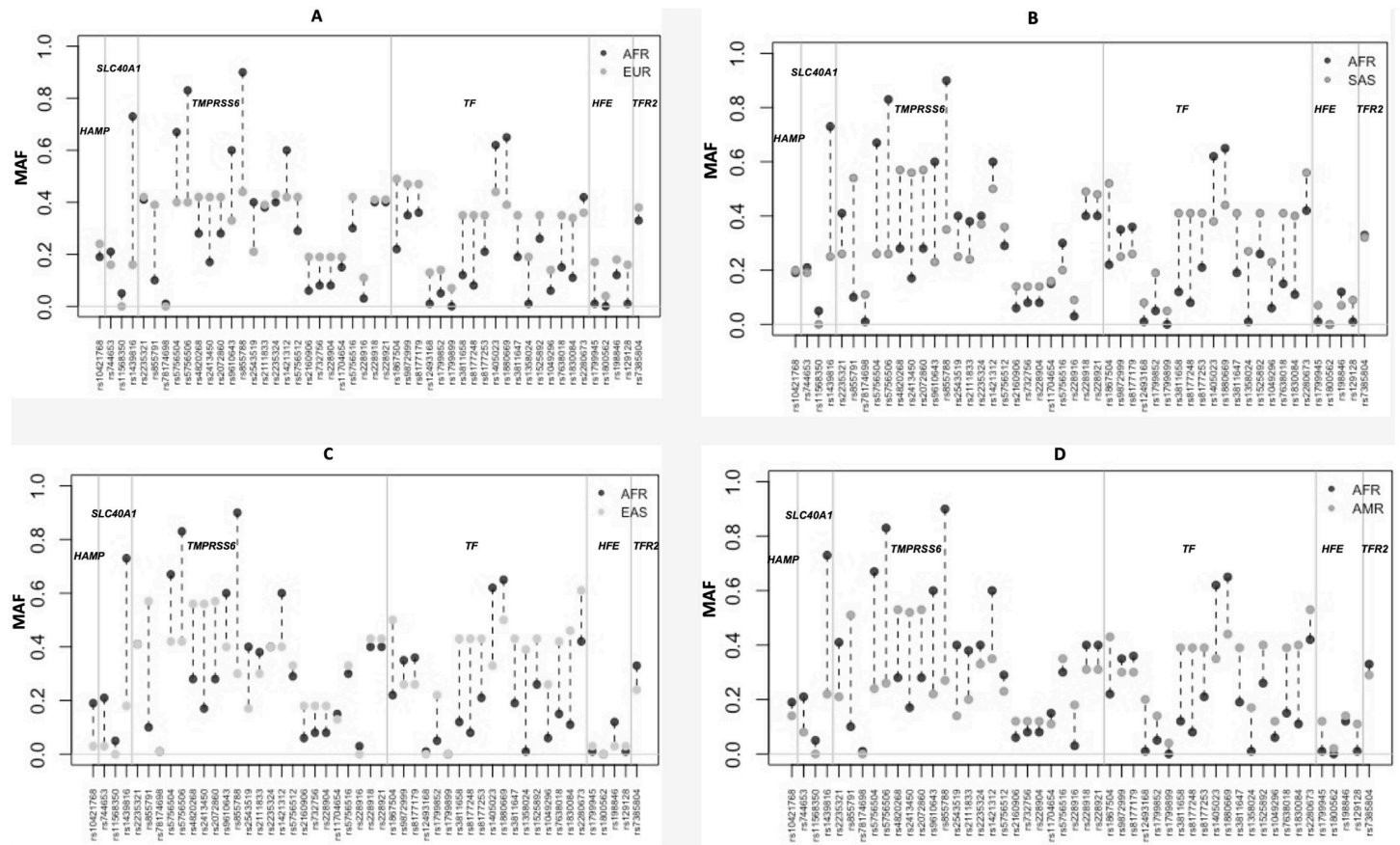
<https://doi.org/10.1371/journal.pone.0235141.g003>

The Gambia. In contrast, the SNPs in the *TF* gene still showed a pattern of high LD in the African populations.

### Distribution and frequency of iron imbalance risk alleles

To investigate if any population had an over- or under-representation of risk alleles leading to iron imbalances, we first classified the alleles as protective or susceptible based on previous associations with low or high iron status or related biomarkers (S1 Table). A total of 23 SNPs were included in the risk allele analysis (see Methods for exclusion criteria). Eleven SNPs had alleles that were clearly associated with low iron, iron deficiency anaemia and/or IRIDA (SNPs in *TMPRSS6* (rs855791, rs2235321, rs2235324, rs4820268, rs2413450, rs228916, rs228918 and rs228921) and *TF* (rs3811647, rs1799899 and rs8177253) (S1 Table).

The South and East Asian populations had the highest number of low iron risk alleles, whereas, Africans had the lowest and were significantly different from the other populations



**Fig 4. The differences in minor allele frequencies of SNPs in the six genes investigated, across different geographic regions.** The comparisons were made between Africans and other global populations (A) Africa vs. Europe; (B) Africa vs. South Asia; (C) Africa vs. East Asia; (D) Africa vs. America. The thick grey lines indicated borders between SNPs in different genes: *HAMP*, *SLC40A1*, *TMPRSS6*, *TF*, *HFE* and *TFR2*. The minor alleles were defined according to the 1000 Genomes Project database [34]. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

<https://doi.org/10.1371/journal.pone.0235141.g004>

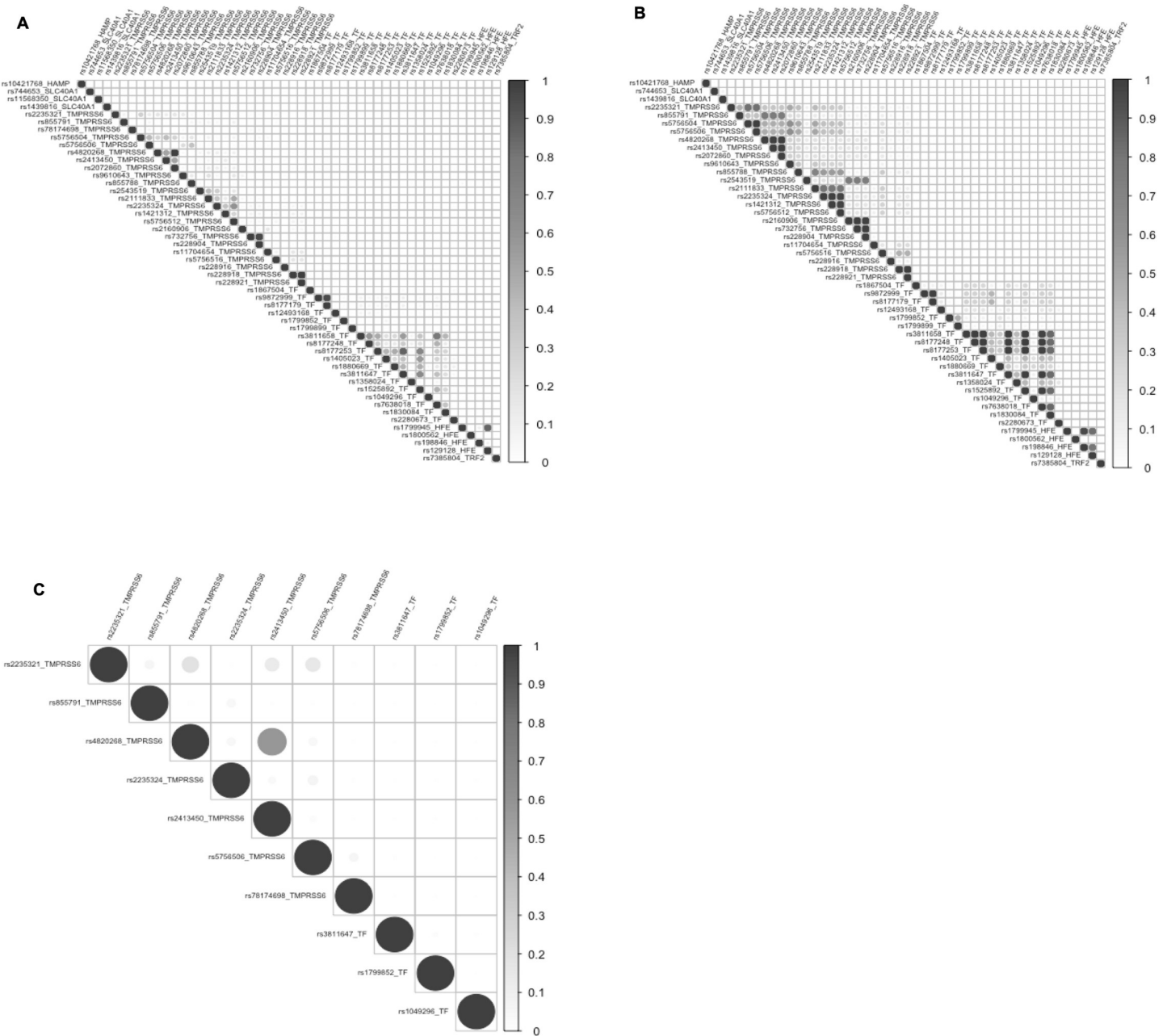
(Fig 6A,  $P < 0.0001$ ). The American and European populations had similar number of low iron risk alleles, but lower than the Asian populations ( $P < 0.0001$ ) (Fig 6A).

Out of the eleven SNPs we found to be associated with low iron, it was only possible to compare six using the Keneba Biobank data, as data on the remaining SNPs were not available. The number of low iron risk alleles of the Gambians in the Keneba Biobank and the overall Africans in the 1000 Genomes were similar (Fig 6B). However, the low iron risk alleles in the Gambian and overall African populations were significantly lower compared to the other populations ( $P < 2 \times 10^{-16}$ ) (Fig 6B).

Twelve SNPs were clearly associated with high iron or related biomarker (SNPs in *HAMP*, *TMPRSS6*, *TF*, *SLC40A1*, *TRF2* and close to *HFE*) with their risk alleles indicated (S1 Table). Three out of these twelve high iron associated SNPs were in or close to the *HFE* gene (rs1799945, rs1800562 and rs198846). These three SNPs were associated with haemochromatosis. Since haemochromatosis is predominantly common in those of European descent and rare in other populations, we analysed these SNPs separately. The European populations have the highest number of high iron risk alleles, significantly different from the other populations ( $P < 0.00850$ ) (Fig 7).

Data for two of the SNPs (rs1799945 and rs1800562) were available for the Keneba Biobank population, but the frequency of risk alleles was low (1% and 0%, respectively). Therefore, we

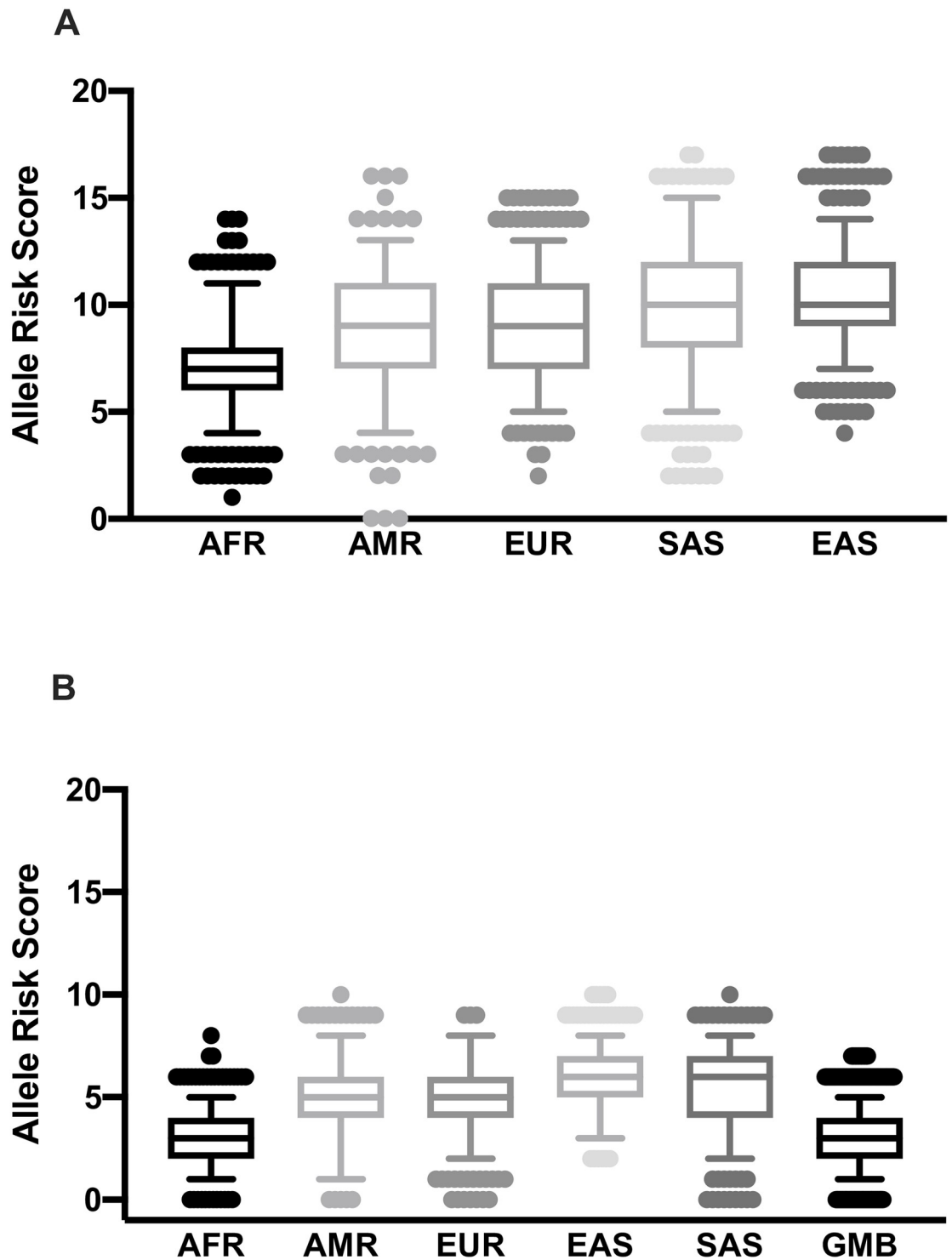




**Fig 5. Linkage disequilibrium (LD) plots in SNPs in *HAMP*, *SLC40A1*, *TMPRSS6*, *TF*, *HFE* and *TFR2* genes.** LD plot showing  $r^2$  values in SNPs associated with iron imbalances in: (A) African populations, (B) European populations and (C) Gambian population in the Keneba Biobank.

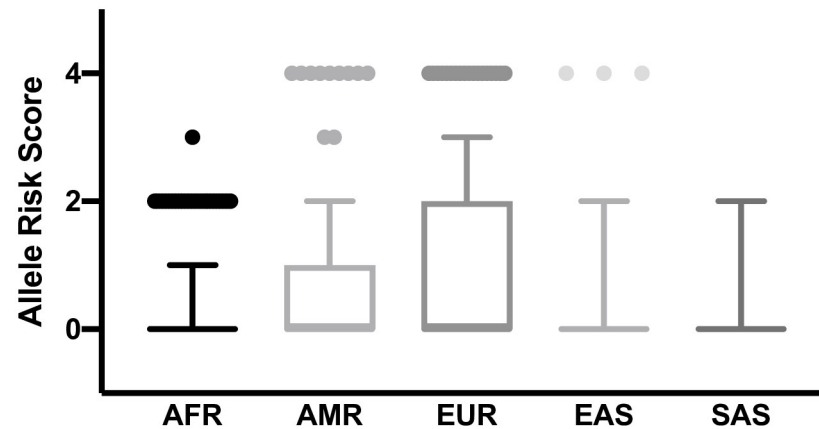
<https://doi.org/10.1371/journal.pone.0235141.g005>

could not compare the frequencies of risk alleles of these SNPs between the Keneba Biobank population and the 1000 Genomes project populations. Furthermore, we compared the frequencies of the high iron risk alleles of the remaining nine SNPs associated with elevated iron status in other genes. The African population in the 1000 Genomes Project had a significantly lower number of high iron risk alleles than the other populations ( $P < 0.0001$ ) (Fig 8A). The distributions between the other populations were similar. From these nine SNPs, genotype data for three SNPs (*TMPRSS6* rs5756506, *TF* rs1799852 and *SLC40A1* rs11568350 (Q248H)



**Fig 6. Distribution of the number of low iron risk alleles across global populations.** (A) Distribution of the number of low iron risk alleles in eleven SNPs associated with low iron status across five populations. (B) Distribution of the number of low iron risk alleles in six SNPs with genotype data in the MRC Keneba Biobank population. Designation of the allele (risk or not) was determined by their previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

<https://doi.org/10.1371/journal.pone.0235141.g006>



**Fig 7. Distribution of the number of risk alleles for haemochromatosis among global populations.** Designation of the risk allele was determined by previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

<https://doi.org/10.1371/journal.pone.0235141.g007>

were available for the Gambians in the Keneba Biobank. When we compare the frequencies of the high iron risk alleles at these three SNPs across populations (Fig 8A), Gambians in the Keneba Biobank and pan-African populations have the lowest number of combined risk alleles for high iron (Fig 8B).

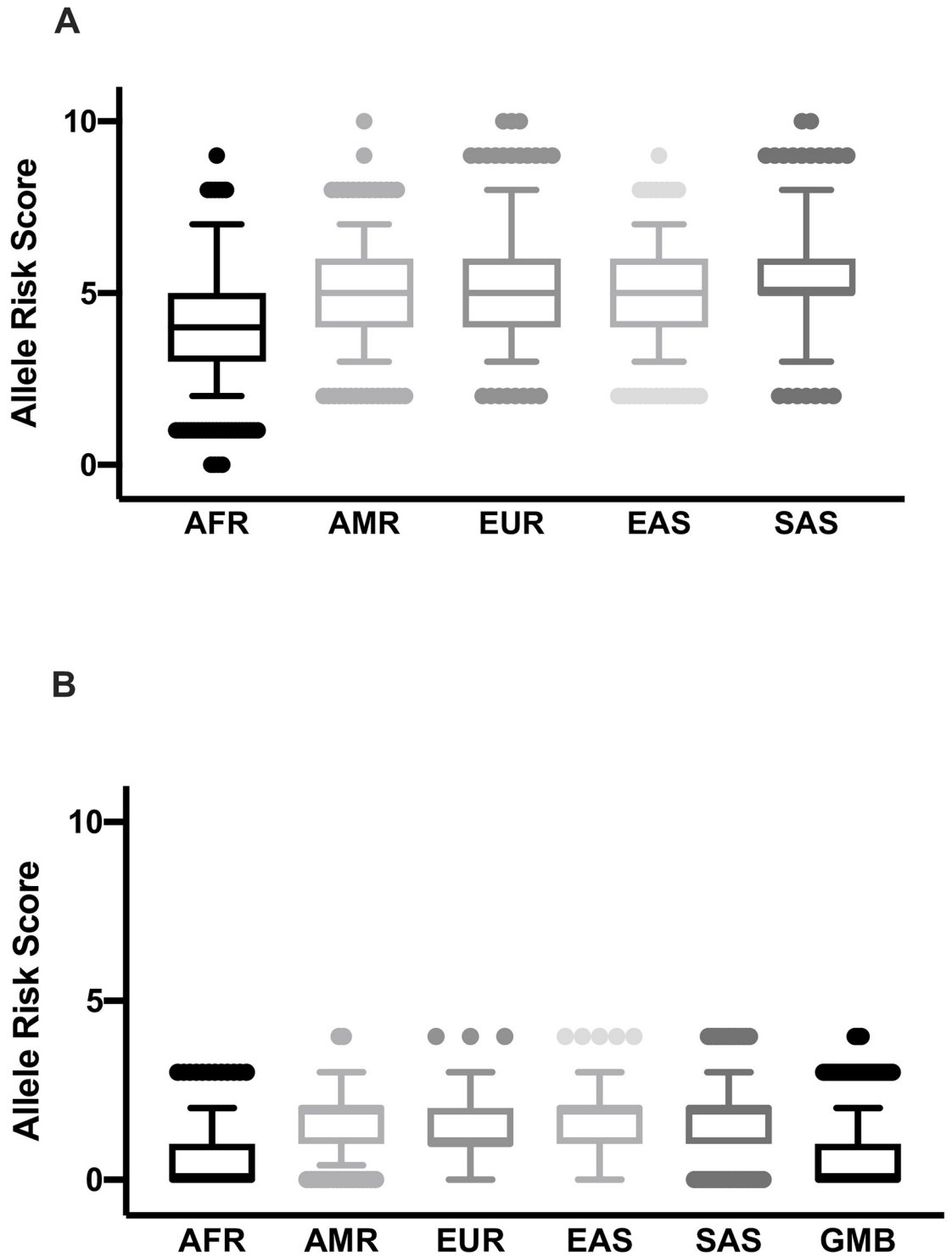
### Global population differentiation

We calculated the global and pairwise fixation index ( $F_{ST}$ ) across the 5 populations to assess population divergence for all iron-associated SNPs. The overall  $F_{ST}$  across the populations was 0.076. The pairwise  $F_{ST}$  between the continental groups shows that African versus non-African populations had the greatest allele frequency differentiation ( $F_{ST} > 0.09$ ; Table 1).

We then investigated the individual SNPs driving the differentiation between African and other populations (Fig 9). The variants with the highest  $F_{ST}$  ( $> 0.3$ ) and highest allele frequency differences were rs1439816 in *SLC40A1* and rs855791, rs855788 and rs5756506 in *TMPRSS6* (Fig 9). The average  $F_{ST}$  values for the set of SNPs in each population was less than 0.065. The highest  $F_{ST}$  values we observed lay within the top 5% of the distribution of empirical global  $F_{ST}$  values described by others (95% percentile  $F_{ST} > 0.28$ ) [43–45].

We also calculated the Population Branch Statistic (PBS) values, an  $F_{ST}$ -based test involving the comparison of three populations, to investigate if the differentiation between populations could be driven by positive selection [40]. We used African, European and South Asian populations and observed that the PBS analysis reaffirms the  $F_{ST}$  results. In particular, the highest PBS values were present in: *SLC40A1* (rs1439816: AFR = 0.27, EUR = 0.05, SAS = 0.0) and *TMPRSS6* (rs855791: AFR = 0.16, EUR = 0.0, SAS = 0.07; rs855788: AFR = 0.29, EUR = 0, SAS = 0.04; rs5756506: AFR = 0.25, EUR = 0.0, SAS = 0.07) (S3 Table). These values are above the top five-percentile threshold of genome-wide PBS values (PBS > 0.156) described by others [41,42].

Finally, we investigated if any signals of recent positive selection could be detected in these genes by using the Integrated Haplotype Score (iHS) values from the Haplotter and HGDP selection browsers. The iHS statistic is based on the LD surrounding a positively selected allele compared with the LD around the alternative variant in the same position [46]. A positive iHS score (iHS > 2) means that the haplotypes on the ancestral allele background are longer than those with the derived allele [46]. A negative iHS score (iHS < -2) means that the haplotypes



**Fig 8. Distribution of the number of high iron risk alleles across global populations.** (A) Distribution of the number of high iron risk alleles across five populations. (B) Distribution of the number of high iron risk alleles in three SNPs in the MRC Keneba Biobank (Gambian) and other populations. Designation of the allele (risk or not) was determined by their previously published information as presented in S1 and S2 Tables. AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian.

<https://doi.org/10.1371/journal.pone.0235141.g008>

**Table 1. Pairwise  $F_{ST}$  values between populations.**

	EUR	EAS	AMR	SAS
EAS	0.0317			
AMR	0.0248	0.0232		
SAS	0.0263	0.0154	0.0130	
AFR	0.0992	0.1465	0.1507	0.1425

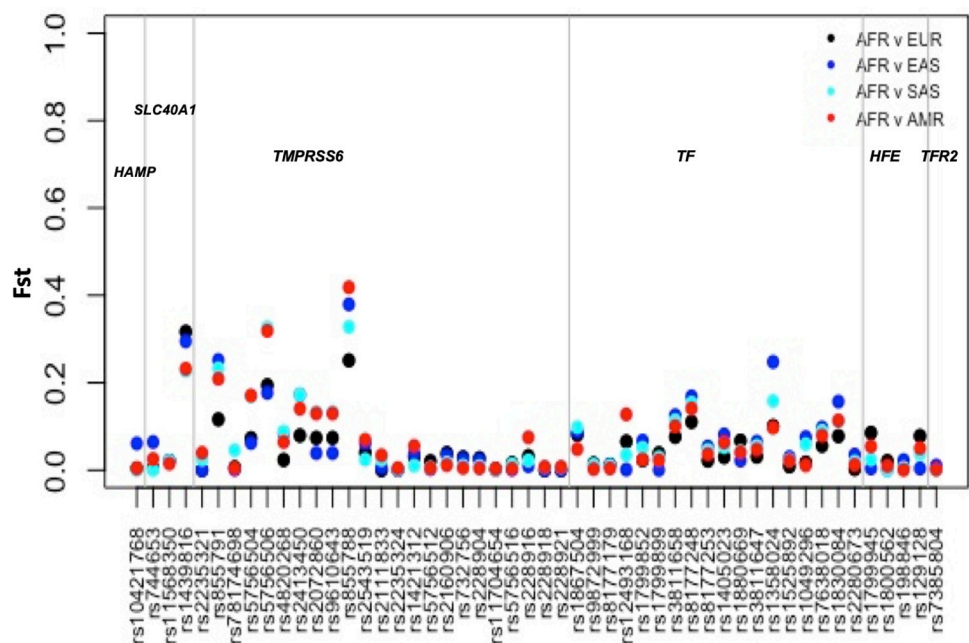
AFR, African; EUR, European; AMR, American; EAS, East Asian; SAS, South Asian;  $F_{ST}$ , fixation index.

<https://doi.org/10.1371/journal.pone.0235141.t001>

on the derived allele background are longer and are under selection. No clear evidence of selection was shown in the genomic regions containing *HAMP*, *TMPRSS6*, *TF* and *TRF2* ( $iHS < 1$ ). However, values of  $iHS$  scores close to 2 were found for the regions containing *SLC40A1* (e.g. rs1439816:  $iHS = 1.8$  (East Asian-Hapmap ASN),  $iHS = 2$  (European HapMap CEU) and *HFE* (e.g. rs198846:  $iHS = 1.8$  ASN), suggesting a high frequency of longer haplotypes with the ancestral allele. Other studies have suggested that the *HEF* locus could be under positive selection in both European and Asian populations [75].

### Discussion

In this study we identified a significant lack of data on the genetic influences of iron status in African populations. This finding highlights a critical gap since African populations have high genetic diversity, and information from other populations may not be transferable to Africans [76,77]. African-specific studies on the genetic influences of iron status will help increase our understanding of the role played by genetic risk factors in the prevalence of anaemia in sub-Saharan Africa.



**Fig 9. Pairwise  $F_{ST}$  values for iron related SNPs between African and non-African populations.** This figure illustrates the comparison of  $F_{ST}$  scores between African and other global populations. AFR, African; AMR, American; EUR, European; EAS, East Asian; SAS, South Asian;  $F_{ST}$ , fixation index.

<https://doi.org/10.1371/journal.pone.0235141.g009>

We used genotype data of populations from the Keneba Biobank at MRCG at LSHTM, The Gambia [33] and the 1000 Genome project [34] to describe the minor allele frequencies and differences in risk alleles in SNPs associated with iron imbalances or iron biomarkers. The allele frequencies of the available SNPs from the Gambian participants in the Keneba Biobank population were very similar to the Gambian population in the 1000 Genomes project. Both the Keneba Biobank population and 1000 Genomes Project included Gambians from the same ethnic group the Mandinka [33,34], which is the largest ethnic group in The Gambia. However, several other ethnic groups live in The Gambia, including Fula and Wolof ethnic groups [78]. Variability in disease risk and nutrition status between the Fula and the Mandinka ethnic groups has been reported [79]. This finding is consistent with the inter-population genetic variability within African populations, which may also influence differences in disease susceptibility. Thus, future work could investigate the genetic diversity in the genes related to iron imbalances in non-Mandinka ethnic groups in The Gambia to determine their possible effect on impaired iron status.

Substantial differences in minor allele frequencies were observed when comparing the African versus non-African populations. The major differences occur in SNPs in *SLC40A1* and *TMPRSS6* genes. *SLC40A1* encodes ferroportin, a transmembrane transport protein which is the only known mammalian iron exporter [80]. The *SLC40A1* Q248H variant (rs11568350) is rare globally except in populations of African ancestry populations, where it reaches frequencies of ~5% [34]. The Q248H variant is associated with increased serum ferritin, decreased hepcidin concentrations and the risk of iron-loading in African populations [57,81]. Also, *SLC40A1* Q248H is associated with modest protection against anaemia and iron deficiency in African children [51,82].

We found significant differences in allelic frequencies for variants in the *TMPRSS6* gene which encodes for Matriptase-2, a type II transmembrane serine protease that negatively regulates hepcidin synthesis [23,83]. Impaired matriptase-2 activity leads to inappropriately raised hepcidin levels [84,85], which results in restricted iron absorption and release from storage sites [17]. Several SNPs in *TMPRSS6* had allele frequencies that are significantly different between African and non-African populations. These variants include rs855791, which has a low MAF (<10%) in African populations and reaches more than 35% in other populations. *TMPRSS6* rs855791 is associated with iron deficiency anaemia and IRIDA, with elevated hepcidin, reduced iron and reduced haemoglobin indices [20,21,84,86,87]. Differences in allele frequencies between continents have been described in many other genetic markers across the genome using data from the 1000 Genomes project [88,89] Therefore, the observed large allele frequency differences in SNPs associated with iron differences could be the result of demographic differences.

To understand if the differences in the observed allele frequencies could lead to differences in over- or under-representation of risk alleles leading to iron imbalances, we explored the frequencies of the combined risk alleles across the genes. We found that African populations, including the Gambian population from the Keneba Biobank, had a significantly lower number of alleles associated with the risk of anaemia or low iron. Similarly, we observed a lower number of risk alleles associated with high iron, or iron overload in Africans. This observation is likely because most of the studies were conducted in non-African populations. However, it is also possible that these differences are due to natural selection processes to balance the environmental risk factors to which African populations are exposed. For example, malnutrition and infections (e.g. helminths and malaria parasites) can lead to anaemia or limit iron overload which can increase susceptibility to certain infections (e.g. bacterial). It is possible that the allele frequency differences between populations we described have occurred through founder effects as humans migrated out of Africa rather than through selective pressure. Possible

signals of selection have only been observed for one SNP in *SLC40A1* and three SNPs in *TMPRSS6*, which have the highest  $F_{ST}$  and PBS values in Africa.

Our study has limitations. These include the potential for bias in the SNPs selection from the literature as there is an overrepresentation of studies related to genetics of iron imbalances in European and Asian populations. Also, it was difficult to ascertain the risk allele for several variants either because they were not described by the original study and/or the different studies used different genotyping platforms. In addition, although some risk alleles have been confirmed in more than one ethnic group (46% of the SNPs), for other SNPs it is possible that the alleles have different effects across populations and this could affect the risk allele analysis. Overall, our study highlights a major gap in genetic studies in Africa and the need to perform genetic studies in African populations.

We also observed a lower linkage disequilibrium between SNPs in African populations. For example, the *TMPRSS6* rs4820268 is in strong LD with *TMPRSS6* rs855791 in Europeans [90], but we found that these two SNPs are in weak LD in the Keneba Biobank population. This should be taken into account when performing association studies and selecting tag SNPs. In this setting, it may be easier to fine-scale map “causal” variants, but more difficult to identify the novel putative loci in a GWAS. Also, as iron imbalances can be due to multiple factors, it is critical to complement genetic studies with detailed meta-data collection, including detailed nutritional status, iron biomarkers, and clinical histories. Alternatively the effects of the variants can be studied prospectively using recall-by-genotype methods [91] that can also interrogate the dynamic responses to, for instance, the administration of iron supplements. Follow-up GWAS and candidate gene studies will be important to understand the genetic underpinning the geographic variation in the prevalence of iron imbalances disorders.

In conclusion, this study identified a substantial disparity in allele frequencies of genetic variants associated with iron, between Africans and other populations. We also, identified the scarcity of data on the genetic influences of iron status in Africa. Given the high burden of iron deficiency in sub-Saharan Africa, particularly in child-bearing women and children, comprehensive mapping of the genetic influences on iron status may help lay the foundation for future studies and assist in developing future iron intervention strategies.

## Supporting information

**S1 Fig. Linkage disequilibrium (LD) plots in SNPs in *HAMP*, *SLC40A1*, *TMPRSS6*, *TF*, *HFE* and *TFR2* genes.** LD plot showing D prime values in SNPs associated with iron imbalances in (A) African populations, (B) European populations and (C) Gambian population in the Keneba Biobank.

(DOCX)

**S1 Table. Details of the fifty SNPs identified in the six genes that are associated with iron imbalance.**

(DOCX)

**S2 Table. Details of populations where each SNP was reported and the associated phenotypes.**

(DOCX)

**S3 Table. Population Branch Statistic (PBS) values involving the comparison of three populations.**

(DOCX)

### S4 Table. Genotyping data for 13 SNPs for the Gambian population in the Keneba Biobank.

(XLSX)

## Acknowledgments

The Authors wish to thank the staff of MRCG at LSHTM Keneba Field Station for contribution to the Keneba Biobank sample collection and Dr. Branwen J Hennig for leading the Biobank sample collection. The Population of West Kiang District, The Gambia for participating in the Biobank project. We also thank Ms. K Pearce at the Core Genomics of the Institute of Child Health, UCL, and our colleagues of the Mal-ED consortium at the Centre for Public Health Genomic, University of Virginia, USA, in particular Drs J Mychaleckyj and U Nayak, for their involvement in the generation of the genetic data. We also thank Dr Neneh Sallah, LSHTM for helping to review the SNPs selection Table.

## Author Contributions

**Conceptualization:** Momodou W. Jallow, Carla Cerami, Susana Campino.

**Data curation:** Momodou W. Jallow, Susana Campino.

**Formal analysis:** Momodou W. Jallow, Taane G. Clark, Susana Campino.

**Funding acquisition:** Andrew M. Prentice, Susana Campino.

**Methodology:** Susana Campino.

**Resources:** Carla Cerami, Andrew M. Prentice.

**Supervision:** Carla Cerami, Susana Campino.

**Writing – original draft:** Momodou W. Jallow, Susana Campino.

**Writing – review & editing:** Momodou W. Jallow, Carla Cerami, Taane G. Clark, Andrew M. Prentice, Susana Campino.

## References

1. Camaschella C. Iron deficiency. *Blood*. 2019; 133: 30–39. <https://doi.org/10.1182/blood-2018-05-815944> PMID: 30401704
2. WHO. Global Burden of Disease Study 2017. *Lancet*. 2017; 1–7.
3. Kassebaum NJ. The Global Burden of Anemia. *Hematol Oncol Clin North Am*. Elsevier Inc; 2016; 30: 247–308. <https://doi.org/10.1016/j.hoc.2015.11.002> PMID: 27040955
4. Khaskheli M-N, Baloch S, Sheeba A, Baloch S, Khaskheli FK. Iron deficiency anaemia is still a major killer of pregnant women. *Pakistan J Med Sci*. 2016; 32: 630–4. <https://doi.org/10.12669/pjms.323.9557> PMID: 27375704
5. Teshome EM, Andang PEA, Osoti V, Terwel SR, Otieno W, Demir Y, et al. Daily home fortification with iron as ferrous fumarate versus NaFeEDTA: a randomised, Kenyan children. 2017; 1–16. <https://doi.org/10.1186/s12916-017-0839-z>
6. Pena-Rosas JP, De-Regil LM, Dowswell T, Viteri FE. Intermittent oral iron supplementation during pregnancy. *Cochrane database Syst Rev*. 2012; 7: CD009997. [www.cochranelibrary.com](http://www.cochranelibrary.com)
7. Petry N, Jallow B, Sawo Y, Darboe MK, Barrow S, Sarr A, et al. Micronutrient Deficiencies, Nutritional Status and the Age and Non-Pregnant Women of Reproductive Age.
8. Heeney MM, Finberg KE. Iron-Refractory Iron Deficiency Anemia (IRIDA). *Hematol Oncol Clin North Am*. 2014; 28: 637–652. <https://doi.org/10.1016/j.hoc.2014.04.009> PMID: 25064705
9. Mleczko-Sanecka K, Roche F, Da Silva AR, Call D, D'Alessio F, Ragab A, et al. Unbiased RNAi screen for hepcidin regulators links hepcidin suppression to proliferative Ras/RAF and nutrient-dependent



- mTOR signaling. *Blood*. 2014; 123: 1574–1585. <https://doi.org/10.1182/blood-2013-07-515957> PMID: 24385536
10. Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Ward DM, et al. Hepcidin regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science*. 2004; 306: 2090–3. <https://doi.org/10.1126/science.1104742> PMID: 15514116
  11. Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, et al. The serine protease TMPRSS6 is required to sense iron deficiency. *Science*. 2008; 320: 1088–92. <https://doi.org/10.1126/science.1157121> PMID: 18451267
  12. Gozzelino R, Arosio P. Iron homeostasis in health and disease. *Int J Mol Sci*. 2016; 17: 2–14. <https://doi.org/10.3390/ijms17010130> PMID: 26805813
  13. McLaren CE, Garner CP, Constantine CC, McLachlan S, Vulpe CD, Snively BM, et al. Genome-wide association study identifies genetic loci associated with iron deficiency. *PLoS One*. 2011; 6. <https://doi.org/10.1371/journal.pone.0017390> PMID: 21483845
  14. Athiyarath R, Shaktivel K, Abraham V, Singh D, Bondu JD, Chapla A, et al. Association of genetic variants with response to iron supplements in pregnancy. *Genes Nutr*. Springer Berlin Heidelberg; 2015; 10: 25. <https://doi.org/10.1007/s12263-015-0474-2> PMID: 26024779
  15. Valenti L, Rametta R, Dongiovanni P, Motta BM, Canavesi E, Pelusi S, et al. The A736V TMPRSS6 Polymorphism Influences Hepatic Iron Overload in Nonalcoholic Fatty Liver Disease. *PLoS One*. 2012; 7. <https://doi.org/10.1371/journal.pone.0048804> PMID: 23144979
  16. Lo KS, Wilson JG, Lange LA, Folsom AR, Galarneau G, Ganesh SK, et al. Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans. *Hum Genet*. 2011; 129: 307–17. <https://doi.org/10.1007/s00439-010-0925-1> PMID: 21153663
  17. Ganz T. Systemic iron homeostasis. 2013; 1721–1741. <https://doi.org/10.1152/physrev.00008.2013> PMID: 24137020
  18. Soranzo N, Spector TD, Mangino M, Kühnel B, Rendon A, Teumer A, et al. A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium. *Nat Genet*. Nature Publishing Group; 2009; 41: 1182–1190. <https://doi.org/10.1038/ng.467> PMID: 19820697
  19. Ganesh SK, Zakai N A, van Rooij FJ A, Soranzo N, Smith A V, Nalls M a, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. *Nat Genet*. Nature Publishing Group; 2009; 41: 1191–1198. <https://doi.org/10.1038/ng.466> PMID: 19862010
  20. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, et al. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. *Nat Genet*. Nature Publishing Group; 2009; 41: 1170–2. <https://doi.org/10.1038/ng.462> PMID: 19820698
  21. Bhatia P, Singh A, Hegde A, Jain R, Bansal D. Systematic evaluation of paediatric cohort with iron refractory iron deficiency anaemia (IRIDA) phenotype reveals multiple TMPRSS6 gene variations. *Br J Haematol*. 2017; 177: 311–318. <https://doi.org/10.1111/bjh.14554> PMID: 28169443
  22. Finberg KE, Whittlesey RL, Fleming MD, Andrews NC, Dc W. Down-regulation of Bmp / Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis Down-regulation of Bmp / Smad signaling by Tmprss6 is required for maintenance of systemic iron homeostasis. 2012; 115: 3817–3826.
  23. De Falco L, Sanchez M, Silvestri L, Kannengiesser C, Muckenthaler MU, Iolascon A, et al. Iron refractory iron deficiency anemia. *Haematologica*. 2013; 98: 845–853. <https://doi.org/10.3324/haematol.2012.075515> PMID: 23729726
  24. Blanco-Rojo R, Baeza-Richer C, López-Parra AM, Pérez-Granados AM, Brichs A, Bertoncini S, et al. Four variants in transferrin and HFE genes as potential markers of iron deficiency anaemia risk: an association study in menstruating women. *Nutr Metab (Lond)*. 2011; 8: 69. <https://doi.org/10.1186/1743-7075-8-69> PMID: 21978626
  25. Lee PL, Halloran C, Trevino R, Felitti V, Beutler E, Scripps T. Human transferrin G277S mutation: a risk factor for iron deficiency anaemia. 2001;
  26. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, et al. Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. *Am J Hum Genet*. The American Society of Human Genetics; 2009; 84: 60–5. <https://doi.org/10.1016/j.ajhg.2008.11.011> PMID: 19084217
  27. Barrios M, Moreno-Carralero M-I, Cuadrado-Grande N, Baro M, Vivanco J-L, Morán-Jiménez M-J. The homozygous mutation G75R in the human SLC11A2 gene leads to microcytic anaemia and iron overload. *Br J Haematol*. 2012; 157: 514–6. <https://doi.org/10.1111/j.1365-2141.2012.09043.x> PMID: 22313374
  28. Kloss-Brandstätter A, Erhart G, Lamina C, Meister B, Haun M, Coassin S, et al. Candidate gene sequencing of SLC11A2 and TMPRSS6 in a family with severe anaemia: Common SNPs, rare

- haplotypes, no causative mutation. *PLoS One*. 2012; 7: 1–8. <https://doi.org/10.1371/journal.pone.0035015> PMID: 22509377
29. Benyamin B, Ferreira MAR, Willemsen G, Gordon S, Middelberg RPS, McEvoy BP, et al. Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume. *Nat Genet*. 2009; 41: 1173–5. <https://doi.org/10.1038/ng.456> PMID: 19820699
  30. Gichohi-Wainaina WN, Tanaka T, Towers GW, Verhoef H, Veenemans J, Talsma EF, et al. Associations between Common Variants in Iron-Related Genes with Haematological Traits in Populations of African Ancestry. Samuels DC, editor. *PLoS One*. 2016; 11: e0157996. <https://doi.org/10.1371/journal.pone.0157996> PMID: 27332551
  31. Hunt SE, McLaren W, Gil L, Thormann A, Schuilenburg H, Sheppard D, et al. Ensembl variation resources. 2018; 1–12. <https://doi.org/10.1093/database/bay119> PMID: 30576484
  32. Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001; 29: 308–311. <https://doi.org/10.1093/nar/29.1.308> PMID: 11125122
  33. Hennig BJ, Unger SA, Dondeh BL, Hassan J, Hawkesworth S, Jarjou L, et al. Cohort profile: The Kiang West Longitudinal Population Study (KWLPs)-a platform for integrated research and health care provision in rural Gambia. *Int J Epidemiol*. 2017; 46: 1–12.
  34. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015; 526: 68–74. <https://doi.org/10.1038/nature15393> PMID: 26432245
  35. Warnes G, Gorjanc G, Leisch F, Man M. Package ‘genetics’. 2019;
  36. R Core Team. A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
  37. Jombart T, Kamvar ZN, Collins C, Lustrik R, Beugin M-P, Knaus BJ, et al. Package “adegenet” Encoding UTF-8 Title Exploratory Analysis of Genetic and Genomic Data. 2018;
  38. Goudet J, Jombart T. Estimating and Tests of Hierarchical F-Statistics. *Evolution (N Y)*. 1998; 52: 950. <https://doi.org/10.2307/2411227>
  39. Paradis E, Jombart T, Kamvar ZN, Knaus B, Schliep K, Potts A, et al. Population and Evolutionary Genetics Analysis System. 2019;
  40. Yi X, Liang Y, Huerta-Sanchez E, Jin X, Cuo ZXP, Pool JE, et al. Sequencing of 50 human exomes reveals adaptation to high altitude. *Science (80-)*. American Association for the Advancement of Science; 2010; 329: 75–78. <https://doi.org/10.1126/science.1190371> PMID: 20595611
  41. Gouveia MH, Bergen AW, Borda V, Nunes K, Leal TP, Ogwang MD, et al. Genetic signatures of gene flow and malaria-driven natural selection in Sub-Saharan populations of the “endemic burkitt lymphoma belt.” *PLoS Genet*. Public Library of Science; 2019; 15: e1008027. <https://doi.org/10.1371/journal.pgen.1008027>
  42. Crawford JE, Amaru R, Song J, Julian CG, Racimo F, Cheng JY, et al. Natural Selection on Genes Related to Cardiovascular Health in High-Altitude Adapted Andeans. *Am J Hum Genet*. Cell Press; 2017; 101: 752–767. <https://doi.org/10.1016/j.ajhg.2017.09.023> PMID: 29100088
  43. Ding K, Kullo IJ. Geographic differences in allele frequencies of susceptibility SNPs for cardiovascular disease. *BMC Med Genet*. BioMed Central; 2011; 12: 55. <https://doi.org/10.1186/1471-2350-12-55> PMID: 21507254
  44. Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, et al. Worldwide human relationships inferred from genome-wide patterns of variation. *Science (80-)*. Science; 2008; 319: 1100–1104. <https://doi.org/10.1126/science.1153717> PMID: 18292342
  45. Myles S, Davison D, Barrett J, Stoneking M, Timpson N. Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics*. Springer Nature; 2008; 1: 22. <https://doi.org/10.1186/1755-8794-1-22> PMID: 18533027
  46. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A Map of Recent Positive Selection in the Human Genome. Hurst L, editor. *PLoS Biol*. Public Library of Science; 2006; 4: e72. <https://doi.org/10.1371/journal.pbio.0040072> PMID: 16494531
  47. Pybus M, Olio GMD, Luisi P, Uzkudun M, Pavlidis P, Laayouni H, et al. 1000 Genomes Selection Browser 1. 0: a genome browser dedicated to signatures of natural selection in modern humans. 2014; 42: 903–909. <https://doi.org/10.1093/nar/gkt1188> PMID: 24275494
  48. Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage [Internet]. *Science*. 2006. pp. 1614–1620. <https://doi.org/10.1126/science.1124309> PMID: 16778047
  49. Danquah I, Gahutu J-B, Zeile I, Musemakweri A, Mockenhaupt FP. Anaemia, iron deficiency and a common polymorphism of iron-regulation, *TMPRSS6* rs855791, in Rwandan children. *Trop Med Int Health*. 2014; 19: 117–22. <https://doi.org/10.1111/tmi.12216> PMID: 24175968

50. Masaisa F, Breman C, Gahutu JB, Mukiibi J, Delanghe J, Philippé J. Ferroportin (SLC40A1) Q248H mutation is associated with lower circulating serum hepcidin levels in Rwandese HIV-positive women. *Ann Hematol.* 2012; 91: 911–916. <https://doi.org/10.1007/s00277-011-1400-3> PMID: 22249207
51. Kasvosve I, Gomo ZAR, Nathoo KJ, Matibe P, Mudenge B, Loyevsky M, et al. Effect of ferroportin Q248H polymorphism on iron status in. 2018; 1102–1106.
52. Gichohi-Wainaina W. N. Melse-Boonstra A. Swinkels D. W. Zimmermann M. B. Feskens E. J. Towers GW. Common variants and haplotypes in the TF, TNF- alpha, and TMPRSS6 genes are associated with iron status in a female black South. *J Nutr* 2015. 2015; 145: 945–953.
53. McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, et al. Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations. *PLoS One.* 2012; 7. <https://doi.org/10.1371/journal.pone.0038339> PMID: 22761678
54. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. *Hum Mol Genet.* 2011; 20: 1232–40. <https://doi.org/10.1093/hmg/ddq552> PMID: 21208937
55. Radio FC, Majore S, Aurizi C, Sorge F, Biolcati G, Bernabini S, et al. Hereditary hemochromatosis type 1 phenotype modifiers in Italian patients. The controversial role of variants in HAMP, BMP2, FTL and SLC40A1 genes. *Blood Cells Mol Dis.* Elsevier B.V.; 2015; 55: 71–5. <https://doi.org/10.1016/j.bcmd.2015.04.001> PMID: 25976471
56. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, et al. Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis. *Nat Commun.* 2014; 5: 4926. <https://doi.org/10.1038/ncomms5926> PMID: 25352340
57. Rivers CA, Barton JC, Gordeuk VR, Acton RT, Speechley MR, Snively BM, et al. Association of ferroportin Q248H polymorphism with elevated levels of serum ferritin in African Americans in the Hemochromatosis and Iron Overload Screening (HEIRS) Study. *Blood Cells, Mol Dis.* 2007; 38: 247–252. <https://doi.org/10.1016/j.bcmd.2006.12.002> PMID: 17276706
58. Javaheri-Kermani M, Farazmandfar T, Ajami A, Yazdani Y. Impact of hepcidin antimicrobial peptide on iron overload in tuberculosis patients. *Scand J Infect Dis.* 2014; 46: 693–696. <https://doi.org/10.3109/00365548.2014.929736> PMID: 25134646
59. Andreani M, Radio FC, Testi M, De Bernardo C, Troiano M, Majore S, et al. Association of hepcidin promoter c.-582 A>G variant and iron overload in thalassemia major. *Haematologica.* 2009; 94: 1293–1296. <https://doi.org/10.3324/haematol.2009.006270> PMID: 19734422
60. Jackson HA, Carter K, Darke C, Guttridge MG, Ravine D, Hutton RD, et al. HFE mutations, iron deficiency and overload in 10 500 blood donors. *Br J Haematol.* 2001; 114: 474–484. <https://doi.org/10.1046/j.1365-2141.2001.02949.x>
61. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, et al. TMPRSS6, but not TF, TFR2 or BMP2 variants are associated with increased risk of iron-deficiency anemia. *Hum Mol Genet.* 2012; 21: 2124–2131. <https://doi.org/10.1093/hmg/dds028> PMID: 22323359
62. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, et al. Identification of a common variant in the TFR2 gene implicated in the physiological regulation of serum iron levels. *Hum Mol Genet.* Oxford University Press; 2011; 20: 1232–40. <https://doi.org/10.1093/hmg/ddq552> PMID: 21208937
63. Piao W, Wang L, Zhang T, Wang Z, Shangguan S, Sun J, et al. A single-nucleotide polymorphism in transferrin is associated with soluble transferrin receptor in Chinese adolescents. *Asia Pac J Clin Nutr.* 2017; 26: 1170–1178. <https://doi.org/10.6133/apjcn.112016.04> PMID: 28917245
64. Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, et al. Genome-wide admixture and association study of serum iron, ferritin, transferrin saturation and total iron binding capacity in African Americans. *Hum Mol Genet.* 2015; 24: 572–581. <https://doi.org/10.1093/hmg/ddu454> PMID: 25224454
65. Beutler E, Felitti V, Gelbart T, Waalen J. Haematological effects of the C282Y HFE mutation in homozygous and heterozygous states among subjects of northern and southern European ancestry. *Br J Haematol.* 2003; 120: 887–893. <https://doi.org/10.1046/j.1365-2141.2003.04215.x> PMID: 12614226
66. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FCGJ, Fleming RE, Kiemeny L A LM, et al. Associations of common variants in HFE and TMPRSS6 with iron parameters are independent of serum hepcidin in a general population: a replication study. *J Med Genet.* 2013; 50: 593–8. <https://doi.org/10.1136/jmedgenet-2013-101673> PMID: 23794717
67. De Falco L, Tortora R, Imperatore N, Bruno M, Capasso M, Girelli D, et al. The role of TMPRSS6 and HFE variants in iron deficiency anemia in celiac disease. *Am J Hematol.* 2018; 93: 383–393. <https://doi.org/10.1002/ajh.24991> PMID: 29194702
68. Pichler I, Del Greco M F, Gögele M, Lill CM, Bertram L, Do CB, et al. Serum iron levels and the risk of Parkinson disease: a Mendelian randomization study. *PLoS Med.* 2013; 10: e1001462. <https://doi.org/10.1371/journal.pmed.1001462> PMID: 23750121

69. Garewal G, Das R, Ahluwalia J, Marwaha RK. Prevalence of the H63D mutation of the HFE in north India: Its presence does not cause iron overload in beta thalassemia trait. *Eur J Haematol*. 2005; 74: 333–336. <https://doi.org/10.1111/j.1600-0609.2004.00390.x> PMID: 15777346
70. Sørensen E, Rigas AS, Thørmøer LW, Burgdorf KS, Pedersen OB, Petersen MS, et al. Genetic factors influencing ferritin levels in 14,126 blood donors: Results from the Danish Blood Donor Study. *Transfusion*. 2016; 56: 622–627. <https://doi.org/10.1111/trf.13397> PMID: 26597663
71. Kullo IJ, Ding K, Jouni H, Smith CY, Chute CG. A genome-wide association study of red blood cell traits using the Electronic Medical Record. *PLoS One*. 2010; 5: 1–9. <https://doi.org/10.1371/journal.pone.0013011> PMID: 20927387
72. Gordeuk VR, Lovato L, Barton JC, Mph MV, McLaren G, Acton RT, et al. Dietary iron intake and serum ferritin concentration in 213 patients homozygous for the HFE C282Y hemochromatosis mutation. 2012; 26: 345–349.
73. Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, et al. Genome-wide association analysis of red blood cell traits in African Americans: The cogent network. *Hum Mol Genet*. 2013; 22: 2529–2538. <https://doi.org/10.1093/hmg/ddt087> PMID: 23446634
74. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, et al. Effects of HFE C282Y and H63D Polymorphisms and Polygenic Background on Iron Stores in a Large Community Sample of Twins. 2000; 1246–1258.
75. Ye K, Cao C, Lin X, Brien KOO, Gu Z. Natural selection on HFE in Asian populations contributes to enhanced non-heme iron absorption. *BMC Genet*. *BMC Genetics*; 2015; 1–11.
76. Gomez F, Hirbo J, Tishkoff SA. Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harb Perspect Biol*. 2014; 6: a008524. <https://doi.org/10.1101/cshperspect.a008524> PMID: 24984772
77. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet*. 2008; 9: 403–33. <https://doi.org/10.1146/annurev.genom.9.081307.164258> PMID: 18593304
78. World Population View. Gambia Population 2020 [Internet]. 2020.
79. Wilson AL, Bradley J, Kandeh B, Salami K, D'Alessandro U, Pinder M, et al. Is chronic malnutrition associated with an increase in malaria incidence? A cohort study in children aged under 5 years in rural Gambia. *Parasites and Vectors*. *Parasites & Vectors*; 2018; 11: 1–11. <https://doi.org/10.1186/s13071-018-3026-y> PMID: 30081945
80. Lane DJR, Merlot AM, Huang ML-H, Bae D-H, Jansson PJ, Sahni S, et al. Cellular iron uptake, trafficking and metabolism: Key molecules and mechanisms and their roles in disease. *Biochim Biophys Acta—Mol Cell Res*. 2015; 1853: 1130–1144. <https://doi.org/10.1016/j.bbamcr.2015.01.021>
81. Gordeuk VR, Caleffi A, Corradini E, Ferrara F, Jones RA, Castro O, et al. Iron overload in Africans and African-Americans and a common mutation in the SCL40A1 (ferroportin 1) gene. *Blood Cells Mol Dis*. 31: 299–304. [https://doi.org/10.1016/s1079-9796\(03\)00164-5](https://doi.org/10.1016/s1079-9796(03)00164-5) PMID: 14636642
82. Muriuki JM, Mentzer AJ, Kimita W, Ndungu FM, MacHaria AW, Webb EL, et al. Iron Status and Associated Malaria Risk among African Children. *Clin Infect Dis*. 2019; 68: 1807–1814. <https://doi.org/10.1093/cid/ciy791> PMID: 30219845
83. Wahedi M, Wortham AM, Kleven MD, Zhao N, Jue S, Enns CA, et al. Matriptase-2 suppresses hepcidin expression by cleaving multiple components of the hepcidin induction pathway. *J Biol Chem*. 2017; 292: 18354–18371. <https://doi.org/10.1074/jbc.M117.801795> PMID: 28924039
84. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. *Blood*. 2011; 118: 4459–4462. <https://doi.org/10.1182/blood-2011-06-364034> PMID: 21873547
85. De Falco L, Silvestri L, Kannengiesser C, Morán E, Oudin C, Rausa M, et al. Functional and clinical impact of novel TMPRSS6 variants in iron-refractory iron-deficiency anemia patients and genotype-phenotype studies. *Hum Mutat*. 2014; 35: 1321–9. <https://doi.org/10.1002/humu.22632> PMID: 25156943
86. Delbini P, Vaja V, Graziadei G, Duca L, Nava I, Refaldi C, et al. Genetic variability of TMPRSS6 and its association with iron deficiency anaemia. *Br J Haematol*. 2010; 151: 281–284. <https://doi.org/10.1111/j.1365-2141.2010.08349.x> PMID: 20738301
87. Pei SN, Ma MC, You HL, Fu HC, Kuo CY, Rau KM, et al. TMPRSS6 rs855791 polymorphism influences the susceptibility to iron deficiency anemia in women at reproductive age. *Int J Med Sci*. 2014; 11: 614–619. <https://doi.org/10.7150/ijms.8582> PMID: 24782651
88. Elhaik E. Empirical Distributions of FST from Large-Scale Human Polymorphism Data. Mailund T, editor. *PLoS One*. *Public Library of Science*; 2012; 7: e49837. <https://doi.org/10.1371/journal.pone.0049837>

89. Hofer T, Ray N, Wegmann D, Excoffier L. Large allele frequency differences between human continental groups are more likely to have occurred by drift during range expansions than by selection. *Ann Hum Genet.* Blackwell Publishing Ltd; 2009; 73: 95–108. <https://doi.org/10.1111/j.1469-1809.2008.00489.x> PMID: [19040659](https://pubmed.ncbi.nlm.nih.gov/19040659/)
90. Yu F, Keinan A, Chen H, Ferland RJ, Hill RS, Mignault AA, et al. Detecting natural selection by empirical comparison to random regions of the genome. *Hum Mol Genet.* 2009; 18: 4853–67. <https://doi.org/10.1093/hmg/ddp457> PMID: [19783549](https://pubmed.ncbi.nlm.nih.gov/19783549/)
91. Corbin LJ, Tan VY, Hughes DA, Wade KH, Paul DS, Tansey KE, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nat Commun.* 2018; 9: 711. <https://doi.org/10.1038/s41467-018-03109-y> PMID: [29459775](https://pubmed.ncbi.nlm.nih.gov/29459775/)