Unravelling The Genetics Of Iron Status In African Populations

----

Candidate Gene Association Studies

Wanjiku N Gichohi-Wainaina

Unravelling the Genetics of Iron Status in African Populations

CANDIDATE GENE ASSOCIATION STUDIES

Wanjiku N Gichohi-Wainaina

Thesis committee

Promotors

Prof. Dr Michael B. Zimmermann

Professor of Micronutrients and International Health, Wageningen University

Professor of Human Nutrition, Swiss Federal Institute of Technology ETH Zürich, Switzerland

Prof. Dr Edith J.M. Feskens

Personal chair at the Division of Human Nutrition, Wageningen University

Co-promotors

Dr Alida Melse-Boonstra

Assistant professor, Division of Human Nutrition, Wageningen University

Dr Gordon Wayne Towers

Senior Lecturer, Nutritional Genetics, Centre of Excellence for Nutrition

North West University, Potchefstroom Campus, South Africa

Other members

Prof. Dr Sander Kersten, Wageningen University

Prof. Dr P. Eline Slagboom, Leiden University Medical Centre

Prof. Sue Fairweather-Tait, Norwich Medical School, UK

Dr Diego Moretti, ETH Zürich, Switzerland

Unravelling the Genetics of Iron Status in African Populations

CANDIDATE GENE ASSOCIATION STUDIES

Wanjiku N Gichohi-Wainaina

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 11 March 2015
at 4 p.m. in the Aula.

.

Wanjiku N Gichohi-Wainaina

Especially dedicated To my husband Martin

and My parents Charles and Lucy Gichohi

Abstract

Background: Investigating the manner in which genetic and environmental factors interact to increase susceptibility to iron deficiency, has the potential to impact on strategies to overcome iron deficiency as well as the development of biomarkers to monitor iron status in populations. Single nucleotide polymorphisms or genetic variants that may affect the composition and hence the functionality of proteins involved in iron metabolism have been the subject of recent genetic association studies. However, these investigations have not yet been carried out in African populations that differ genetically from populations of European ancestry and which bear the highest burden of iron deficiency. The overall aim of this thesis was to investigate the genetics of iron status in African populations using a candidate gene approach.

Methods: In order to evaluate the association between identified *TMPRSS6* gene variants and iron status we conducted a systematic review with meta-analyses. We primarily searched the literature using the HuGE Navigator, Pubmed and Scopus databases for primarily genome wide association studies. Fixed effects meta-analysis was used to obtain summary estimates. Associations between reported variants and iron status as well as gene-gene and variant interactions that influence iron status were investigated in a female black South African cohort (n=686; range 32–86 years) which were part of the Prospective Urban and Rural Epidemiology (PURE) study. Concentrations of haemoglobin, serum ferritin, serum transferrin receptor and body iron stores were determined. Thirty SNPs were genotyped and passed all quality criteria. To investigate whether previously identified associations in populations of European ancestry are replicated in populations of African ancestry, we conducted candidate gene association studies. Twenty iron status-associated variants in 628 Kenyans, 609 Tanzanians, 608 South Africans and 228 African Americans were genotyped and associations investigated using haemoglobin and serum ferritin as outcome measures. Finally, we assessed the effect of *TNF*-α allele variants (*TNF*$_{-1031}$, *TNF*$_{-308}$) on malaria rates, the severity of malaria as indicated by haemoglobin concentrations at the time of presentation with febrile episodes and the association between *Plasmodium* infection and haemoglobin concentration in symptomless parasite carriers. We used data from a placebo-controlled trial which consisted of 612 Tanzanian children aged 6–60 months. Cox regression models were used in the primary analysis to account for multiple episodes per child.

Results: In our systematic review we included eleven studies on Caucasian populations, four on Asian populations and one study on an African-American population. Differences in minor allele frequencies (MAF) of 8 *TMPRSS6* SNPs (rs855791, rs4820268, rs2111833, rs1421312, rs228921, rs228918, rs228919 and rs575620) across ethnic groups were observed; with the MAF of rs855791 being significantly higher in Asian populations than in Caucasians (0.55 vs 0.42). In the meta-analysis, the A allele of rs855791 was associated with lower haemoglobin and ferritin concentrations in all populations. This allele was also associated with increased serum transferrin receptor and transferrin concentrations. We observed similar associations for the G allele in rs4820268. In general, minor allele frequencies (MAF) from females in the PURE population were lower compared to those of males and females of European ancestry populations in the 1000 Genomes Project. In the *TF* gene, the SNP rs1799852 was associated with decreased serum ferritin (p=0.01) and body iron concentrations (p=0.03) and increased serum transferrin receptor (sTfR) concentrations (P=0.004), while rs3811647 was associated with transferrin receptor and body iron (both P=0.03) in a U-shaped manner. The chromosome 6 SNP allele combination (AAA) consisting of rs1799964 and rs1800629 both in *TNF* α and rs2071592 in *NFKBIL1* was associated with higher odds for low serum ferritin concentrations (serum ferritin<15μg/L; OR:1.86 (95%-CI, 1.23-2.79)). The chromosome 22 SNP allele combination (GG) consisting of rs228918 and rs228921 in the *TMPRSS6* gene was associated with lower odds for increased sTfR concentrations (sTfR>8.3mg/L; OR:0.79 (95%-CI,0.63-0.98). We successfully replicated reported significant associations with lowered haemoglobin concentrations for two loci in *TMPRSS6* namely rs2413450 and rs4820268 and with increased haemoglobin concentrations for one locus in *TF* (rs3811658) when analysing the four populations of African ancestry. When ferritin was considered as an outcome measure, we replicated associations with increased ferritin concentrations in two loci namely, rs228918 in *TMPRSS6* and rs1525892 in *TF*. No other significant associations were determined. Malaria rates were higher in Tanzanian children with the TNF_{-1031}CC genotype (rs1799964) compared to the AA genotype (crude hazard ratio (HR), 95%CI: 1.41 [1.01–1.97], adjusted HR 1.31 [0.97–1.76]y) but were lower in those with the TNF_{-308}AA genotype (rs1800629) (adjusted HR 0.13 [0.02–0.63]) compared to those harbouring the wild type homozygous genotype.

Conclusions: This thesis demonstrates previously observed associations between *TMPRSS6* gene variants and haemoglobin concentrations in European ancestry

populations are replicated in African populations. Replication of results in other loci previously associated with iron status in European ancestry populations was not achieved. Additionally, minor allele frequencies of single nucleotide polymorphisms associated with iron status are generally higher in European ancestry cohorts compared to those of African ancestry populations. The lack of association of reported variants may indicate that novel loci are responsible for the heritability of iron status in African populations. We have additionally observed that *TNF* α variants increase malaria severity. Malaria is a major cause of iron deficiency in malaria endemic areas. Our finding emphasizes that to alleviate iron deficiency in malaria endemic areas prevention and treatment of malaria is necessary. This thesis highlights the need to conduct genetic association studies in African populations where iron deficiency is of utmost public health significance. In addition, investigations into the genetics of iron status are bound to contribute towards the development of biomarkers that are useful in the determination of iron status in areas of high inflammation burden.

Contents

Chapter I

General Introduction

Background

Anaemia has been associated with increased mortality and morbidity as well as with decreased physical functioning and quality of life *[1, 2]*. The causes of anaemia are multifactorial, intertwined, and context-specific. Important known risk factors for anaemia include micronutrient deficiencies (e.g. iron, vitamin A, folate, vitamin B-12), infections (e.g. intestinal parasites, schistosomiasis, malaria, human immunodeficiency virus [HIV]), and inherited red blood cell disorders (e.g. sickle cell, α-thalassemia) [3]. On average, globally, 50% of anaemia is assumed to be attributable to iron deficiency *[4-6]*. It is known that low iron intake and low bioavailability of dietary iron are crucial in the development of iron deficiency *[7]*. In developing countries, haem iron which is highly bioavailable, is often consumed in small amounts as most of the iron is from plant sources (non-haem iron). Non-haem iron is poorly absorbed as the diets rich in this form of iron are often also high in constituents that impair iron absorption *[8]*. After digestion, the efficiency of iron absorption is determined by several factors to ensure sufficient iron is available for physiological functions while avoiding the accumulation of excessive iron. These factors include the presence of dietary components with various iron-binding strengths and iron-reducing properties, the iron status of the individual and heritable factors *[9]*. In terms of heritable factors, single nucleotide polymorphisms/mutations in the genes encoding proteins involved in iron uptake, absorption and transport may influence eventual iron status. This is because these mutations affect the functions and the quantities of proteins they encode *[10]*.

It is well established that genetics plays a significant role in iron overload but much less is reported about the genetic basis of iron deficiency. Of even greater interest is the manner in which genetic factors interact with factors such as diet to increase susceptibility to iron deficiency. Understanding the genetics of iron status is key to optimizing strategies to address iron deficiency as well as discovering appropriate biomarkers to monitor iron status in populations. Rare genetic mutations and knock-out animal models have been invaluable tools for elucidating the molecular events controlling iron metabolism *[11, 12]*. Particularly, the recent progress in genome-wide association (GWA) studies has led to the identification of many loci that are associated with iron deficiency/lowered iron status, anaemia, or both *[13-19]*. These findings highlight the pivotal role of the genes in influencing iron metabolism and thus affecting iron status. However, genetic association studies have mainly concentrated on populations of European ancestry. Little has been reported about the presence and

actions of these gene variants in the genetically diverse African populations that carry the highest burden of iron deficiency.

Prevalence of iron deficiency in African populations in context: Is genetics a potential player?

Various studies have observed a geographical disparity in iron status leading to the hypothesis that dissimilarities in genetic alterations across ethnicities may indeed contribute to differences in iron status. Globally, in preschool-age children, anaemia prevalence is 47.4%, affecting 293 million children. The highest prevalence is in Africa (67.6%) and South-East Asia (65.5%) as indicated in **Figure 1**. In addition, 468.4 million non-pregnant women are anaemic (30.2% prevalence globally) with the highest prevalence in Africa at 47.5% *[4, 5]*. The genetic determinants of iron metabolism have not been well studied in black Africans, except for haemoglobinopathies such as thalassemia and sickle-cell anaemia. Further studies are thus required in these populations to ascertain associations of previously reported single nucleotide polymorphisms to various iron status measures.

Understanding the genetics of iron deficiency is crucial due to its consequences. Specifically, it affects more people than any other condition, constituting a major public health problem. Iron deficiency anaemia (IDA) has substantial health and economic costs *[7]* and adversely affects immune status *[20, 21]*, pregnancy outcomes *[20]* and children's cognitive development *[21, 22]*. In order to appreciate the potential role of genetic mutations/variants in the iron regulatory system, it is necessary to first present a general description of iron metabolism. Regulation of iron uptake, absorption and transport is the primary determinant of iron balance as there is no physiological mechanism for iron excretion [10, 23]. Homeostatic mechanisms regulating the absorption, transport, storage and mobilization of cellular iron are therefore of critical importance.

Iron absorption

Iron absorption occurs predominantly on the apical surface of the duodenum and upper jejunum. Dietary iron changes are sensed by the liver *[24]* which sends signals to the villus enterocytes, where there is a rapid up regulation of iron transporters. The two forms of dietary iron, haem and non-haem iron, are absorbed by enterocytes non-

competitively. Haem iron is not chelated by common dietary inhibitors of non-haem iron absorption, and therefore is much more readily absorbed compared to non-haem iron. Haem iron transporter (HCP1) is responsible for haem transport across the apical membrane and it is up regulated by hypoxia and iron deficiency *[25]*. Once haem iron enters the cells, it is released from the protoporphyrin ring by haem oxygenase to release the iron II (ferrous) ions (Fe^{2+}) where after it becomes part of the common iron pool.

Non-haem iron which exists in the iron III (ferric) ionic (Fe^{3+}) state is reduced to ferrous iron before it is transported across the intestinal epithelium. This is accomplished by dietary components such as ascorbic acid, and the enzyme termed duodenal cytochrome b reductase (Dcytb) which is highly expressed in the brush border of enterocytes *[26]*. Once the insoluble Fe^{3+} is converted to Fe^{2+}, it enters the mucosal phase. The reduced Fe^{3+} is transported across the apical membrane into the cells through a divalent metal transporter (DMT1). DMT1 is expressed at the duodenal brush border where it controls uptake of dietary iron. The Fe^{2+} inside the cells can either be stored as ferritin or transported across the basolateral surface into the bloodstream. The iron inside the cell is exported out of the cell by a transport protein called ferroportin (FPN). Hephaestin, a multicopper ferroxidase, works in concert with FPN in the export of iron from the enterocytes *[27, 28]*. It oxidizes Fe^{2+} to Fe^{3+} which enables it to be taken up by transferrin (TF).

Transport, uptake and storage of iron

Transferrin (TF) receives and binds iron for delivery to the transferrin receptors of the recipient cells, after which apotransferrin is returned to the plasma to again function as an iron transporter *[29]*. Transferrin receptor 1 (TfR1) mediates the uptake of TF-bound iron and remains a major regulatory site for iron homeostasis *[30, 31]*. When iron-loaded TF attaches to TfR1, this interaction triggers the formation of intracellular TF-TfR1 containing endosomes. The proton pumps present in the endosomal membrane pump hydrogen (H+) ions into the endosome from the cytoplasm, which induces a conformational change of the TF and its receptor resulting in the release of iron. The released Fe^{3+} is converted to Fe^{2+} by six-transmembrane epithelial antigen of the prostate (STEAP) family member 3 (STEAP3), which is a metalloreductase present in the endosomes. The Fe^{2+} is then transported out of the endosomes by DMT1. Human TfR

shedding is caused by an integral membrane metalloprotease, releasing soluble TfR (sTfR) into the plasma *[32, 33]*. This process is controlled by TF, and the level of sTfR is an indicator of available functional iron, independent of the iron stores *[34]*. Thus, the higher the sTfR concentration, the lower the levels of functional iron. Protein bound forms of iron such as isoferritin and haemoglobin can also be taken up by the cells *[35]*. Haemoglobin released by intravascular haemolysis is bound by haptoglobin and then taken up by the scavenger receptor (CD163) present on monocytes and macrophages *[36]*.

Iron homeostasis

Iron is conserved and recycled within the body (20–25 mg/day), mainly for the purpose of red blood cell formation. However, there is an obligatory loss of approximately 1-2 mg of iron originating from the daily loss of blood (menstrual and other sources of bleeding), sweat, intestinal cells, skin and hair. The lost iron needs to be replenished with dietary iron. Hepcidin is involved in the systemic regulation of iron absorption. Once hepcidin is produced, it directly binds to ferroportin causing it to be internalized and degraded, thereby blocking cellular iron export *[37]*. When iron stores are adequate or high, the liver produces hepcidin which reduces iron efflux from duodenal cells and macrophages. In understanding the molecular events leading up to the production of hepcidin, variants within transmembrane protease, serine 6 (*TMPRSS6*) which encodes matriptase-2, are currently of great interest. The exact mechanism through which *TMPRSS6* action is exerted is still under investigation. It has generally been hypothesised that *TMPRSS6* polymorphisms affect hepcidin transcription, thereby altering hepcidin concentrations in response to systemic iron concentrations *[1, 38-40]*. However, two recent studies *[41, 42]* did not confirm an intermediate role for hepcidin in the SNP-iron status parameters associations. These studies instead indicate pleiotropic SNP effects on hepcidin and iron status parameters.
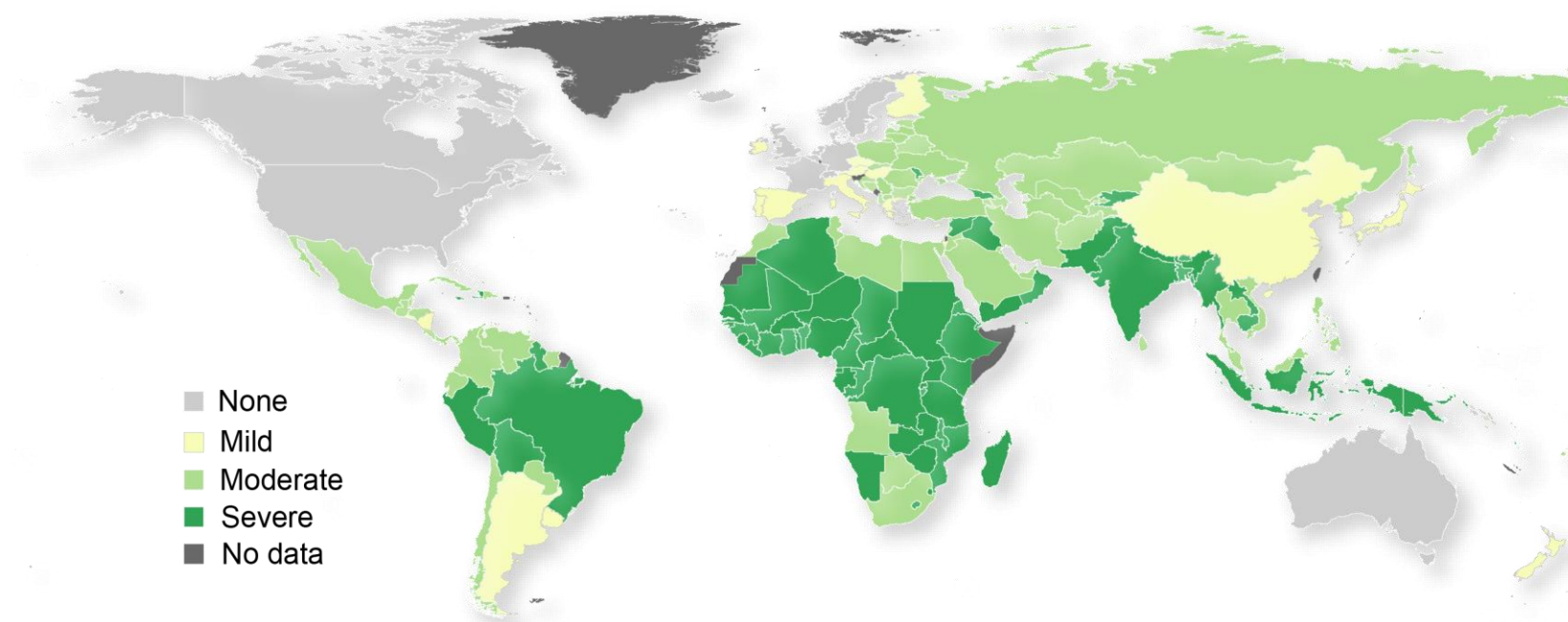
Figure 1: Prevalence of iron deficiency is based on the percentage of children 6–59 months of age with haemoglobin below 110g/L

Note: Public health significance is classified as none (<5.0%), mild (≥5.0%–<20.0%), moderate (≥20.0%–<40.0%), or severe (≥40.0%).

Source: Worldwide Prevalence of Anaemia 1993–2005: WHO Global Database of Anaemia World Health Organization, World Health Organization, 2008.http://www.who.int/vmnis/anaemia/en/)

Iron homeostasis and genetic variations

Over the past decade, heritable, overtly pathological iron deficiencies and iron overload syndromes have been accredited to mutations in several key genes that regulate iron homeostasis *[43, 44]*. Insight into the rare monogenic causes of iron deficiency and iron overload in humans is useful in identifying candidate genes that may harbour genetic variation that could influence iron status. From a public health perspective, the existence of common mutations or single nucleotide polymorphisms that lower iron status in African populations is of interest, given the high prevalence of iron deficiency. This is because these polymorphisms might heighten susceptibility to poor iron status especially in environments of low bioavailable iron diets.

Evidence for genetic influences on iron status in humans can be obtained from the classical twin design in which the resemblance of genetically identical monozygotic (MZ) twins is compared to the resemblance of dizygotic (DZ) twins for a specific trait of interest e.g. iron status. Comparing the resemblance of MZ twins for a trait or disease with the resemblance of DZ twins offers the first estimate of the extent to which genetic variation determines phenotypic variation of the trait [45].

Two twin studies have been conducted to investigate the genetics of iron status. One twin study demonstrated that the heritability of serum ferritin (SF) and soluble transferrin receptor was 44% and 54%, respectively *[10]*. Another study observed that after correction for age and body-mass index, 23% and 31% of the variance in serum iron, 66% and 49% of the variance in transferrin, 33% and 47% of the variance in transferrin saturation, and 47% and 47% of the variance in ferritin could be explained by additive genetic factors for men and women respectively [46]. Heritability estimates for iron status of 20%–30% have also been calculated from other population studies *[46-49]* and this further suggests a substantial genetic contribution to iron regulation.

Recently, GWA studies have observed associations between the presence of single nucleotide polymorphisms and iron status biomarkers. Specifically, GWA studies have consistently highlighted the significance of matriptase-2 in the control of iron homeostasis by identifying common *TMPRSS6* variants associated with abnormal haematological parameters, including haemoglobin, transferrin saturation, erythrocyte mean cell volume (MCV) and serum iron concentrations *[13, 14, 16]*. Genetic variants or SNPs present in the transferrin (*TF*), hemochromatosis (*HFE)* and transferrin receptor 2 (*TFR2)* genes have also been associated with iron status measures in population-based

studies *[15, 17, 50-53]*. These associations have however not yet been demonstrated in African populations. Additionally, with a few exceptions, GWA studies have been centred on populations of European descent which contain only a subset of human genetic variation. Since populations may vary in terms of allele frequencies and biological adaptations, it is necessary to investigate the genetics of iron deficiency in non-European populations and to establish the generality of findings obtained initially in Europeans and European-Americans. Furthermore, the high levels of genetic diversity in African populations *[54]* and their demographic history make these populations particularly informative for the investigation of the genetics of iron deficiency.

Identifying common genetic variants influencing iron status in African populations may further offer insights into variants that may also modify the severity of malaria. Anaemia due to malaria is a major public health problem *[55]* (especially in African children) that remains a challenge to treat. The key to understanding the pathophysiology of iron metabolism in malaria is the activity of the iron regulatory hormone hepcidin. Hepcidin gene expression is up regulated by (among other factors) inflammatory cytokines such as *TNF α [56]*. Single nucleotide polymorphisms in the *TNF α* locus have been identified as potential risk factors in the aetiology of malaria *[57]*. Investigating the regulation of hepcidin by inflammatory cytokines may offer opportunities to formulate better approaches to address anaemia in malaria endemic areas.

Studies on the genetics of iron status in African populations may raise several questions. In particular, will the same results observed in Europeans be detected in African populations? Will causal variants have similar allele frequencies and iron deficiency risk in African populations? How will these causal variants influence iron status and the resulting disease severity?

Expansion of studies on the genetics of iron status in African populations is thus important for the ultimate goal of characterizing risk variants beyond what can be determined from populations of European descent alone. This information will further be useful in informing strategies to address iron deficiency.

Rationale, research questions

Despite the success of recent GWA studies, there remains a lack of research on the prevalence of genetic factors and their effects on iron metabolism in African populations. As the overview makes clear, the single nucleotide polymorphisms associated with iron status in African populations are not fully understood. The unearthing of these genetic factors is especially important in understanding the heterogeneity of response to interventions designed to address iron deficiency. In general this project was aimed at harnessing the potential of genetics to address iron deficiency in African populations. However, prior to utilizing this potential, specific fundamental questions with regard to the genetics of iron status in African populations need to be answered.

The **primary research question** was therefore:

Do the single nucleotide polymorphisms previously identified as being associated with iron status in non-African populations present with similar associations in African populations?

The **secondary research questions** were:

1.   Are the minor allele frequencies of single nucleotide polymorphisms associated with iron status in African populations similar to those of non-African populations?

2.   How do the single nucleotide polymorphisms associated with iron status influence disease severity in the African population?

Outline of this thesis

Based on the above rationale, the studies undertaken in this thesis are focussed on the genetics of iron status in African populations. In **Chapter 2**, a detailed systematic review with meta analyses on the inter-ethnic differences in single nucleotide polymorphisms within *TMPRSS6* that are associated with iron status indicators is presented. In particular we focused on the magnitude of effect sizes, direction of associations and minor allele frequency differences across study populations. In the study designed in **Chapter 3**, we aimed to explore common variants and gene-gene interactions associated with iron status in a female black South African population using the candidate gene approach. In **Chapter 4**, the influence of single nucleotide polymorphisms involved in the inflammatory response on iron status and effects on eventual malaria severity were

explored using a population-based genetic approach. In **Chapter 5**, we explored the question of transferability and generalizability of association results previously obtained in European ancestry populations in several African ancestry populations. The candidate gene approach was used for this study. Finally in **Chapter 6**, we have discussed the findings of this thesis, methodological aspects that may have impacted on these findings as well as their public health relevance. Furthermore, recommendations for further research are presented.

References

1.   Adams PF, Hendershot GE, Marano MA: **Current estimates from the National Health Interview Survey, 1996.** *Vital Health Stat 10* 1999:1-203.

2.   Weiss G, Goodnough LT: **Anemia of chronic disease.** *N Engl J Med* 2005, **352**:1011-1023.

3.   WHO/CDC: **Assessing the iron status of populations. Report of a Joint WHO/CDC Technical Consultation on the Assessment of Iron Status at the Population Level. Geneva, Switzerland, 6–8 April 2004. Geneva: WHO Library Cataloguing-in-Publication Data, 1–35.** 2007.

4.   WHO, **Micronutrient deficiencies**, accessed on 30th September 2014
     http://www.who.int/nutrition/topics/ida/en/

5.   Harvest Plus, **Iron,** accessed on 30th September 2014

6.   Kassebaum NJ, Jasrasaria R, Naghavi M, Wulf SK, Johns N, Lozano R, Regan M, Weatherall D, Chou DP, Eisele TP, et al: **A systematic analysis of global anemia burden from 1990 to 2010.** *Blood* 2014, **123**:615-624.

7.   Zimmermann MB, Hurrell RF: **Nutritional iron deficiency.** *Lancet* 2007, **370**:511-520.

8.   Backstrand JR, Allen LH, Black AK, de Mata M, Pelto GH: **Diet and iron status of nonpregnant women in rural Central Mexico.** *Am J Clin Nutr* 2002, **76**:156-164.

9.   Monsen ER: **Iron nutrition and absorption: dietary factors which impact iron bioavailability.** *J Am Diet Assoc* 1988, **88**:786-790.

10. Fairweather-Tait SJ, Guile GR, Valdes AM, Wawer AA, Hurst R, Skinner J, Macgregor AJ: **The contribution of diet and genotype to iron status in women: a classical twin study.** *PLoS One* 2013, **8**:e83047.

11.   Donovan A, Andrews NC: **The molecular regulation of iron metabolism.** *Hematol J* 2004, **5:**373-380.

12.   McGregor J, McKie AT, Simpson RJ: **Of mice and men: genetic determinants of iron status.** *Proc Nutr Soc* 2004, **63:**11-20.

13.   Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, et al: **Genome-wide association study identifies variants in** *TMPRSS6* **associated with hemoglobin levels.** *Nat Genet* 2009, **41:**1170-1172.

14.   Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, et al: **Common variants in** *TMPRSS6* **are associated with iron status and erythrocyte volume.** *Nat Genet* 2009, **41:**1173-1175.

15.   Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, et al: **Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium.** *Nat Genet* 2009, **41:**1191-1198.

16.   Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, Walston JD, Fried LP, Singleton A, Guralnik J, et al: **A genome-wide association analysis of serum iron concentrations.** *Blood* 2010, **115:**94-96.

17.   McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, Adams PC, Acton RT, Murray JA, Leiendecker-Foster C, et al: **Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations.** *PLoS One* 2012, **7:**e38339.

18.   Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, Evans MK, Li Y, Taylor HA, Willis MS, 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* 2014.

19.   Ding K, de Andrade M, Manolio TA, Crawford DC, Rasmussen-Torvik LJ, Ritchie MD, Denny JC, Masys DR, Jouni H, Pachecho JA, et al: **Genetic variants that confer resistance to malaria are associated with red blood cell traits in African-Americans: an electronic medical record-based genome-wide association study.** *G3 (Bethesda)* 2013, **3:**1061-1068.

20.   Brabin BJ, Hakimi M, Pelletier D: **An analysis of anemia and pregnancy-related maternal mortality.** *J Nutr* 2001, **131:**604S-614S; discussion 614S-615S.

21. Brabin BJ, Premji Z, Verhoeff F: **An analysis of anemia and child mortality.** *J Nutr* 2001, **131:**636S-645S; discussion 646S-648S.

22. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, Carter JA: **Child development: risk factors for adverse outcomes in developing countries.** *Lancet* 2007, **369:**145-157.

23. Hurrell RF: **Bioavailability of iron.** *Eur J Clin Nutr* 1997, **51 Suppl 1:**S4-8.

24. Evstatiev R, Gasche C: **Iron sensing and signalling.** *Gut* 2012, **61:**933-952.

25. Shayeghi M, Latunde-Dada GO, Oakhill JS, Laftah AH, Takeuchi K, Halliday N, Khan Y, Warley A, McCann FE, Hider RC, et al: **Identification of an intestinal heme transporter.** *Cell* 2005, **122:**789-801.

26. McKie AT, Barrow D, Latunde-Dada GO, Rolfs A, Sager G, Mudaly E, Mudaly M, Richardson C, Barlow D, Bomford A, et al: **An iron-regulated ferric reductase associated with the absorption of dietary iron.** *Science* 2001, **291:**1755-1759.

27. Donovan A, Brownlie A, Zhou Y, Shepard J, Pratt SJ, Moynihan J, Paw BH, Drejer A, Barut B, Zapata A, et al: **Positional cloning of zebrafish ferroportin1 identifies a conserved vertebrate iron exporter.** *Nature* 2000, **403:**776-781.

28. Abboud S, Haile DJ: **A novel mammalian iron-regulated protein involved in intracellular iron metabolism.** *The Journal of biological chemistry* 2000, **275:**19906-19912.

29. Edison ES, Bajel A, Chandy M: **Iron homeostasis: new players, newer insights.** *Eur J Haematol* 2008, **81:**411-424.

30. Aisen P, Enns C, Wessling-Resnick M: **Chemistry and biology of eukaryotic iron metabolism.** *The international journal of biochemistry & cell biology* 2001, **33:**940-959.

31. Richardson DR, Ponka P: **The molecular mechanisms of the metabolism and transport of iron in normal and neoplastic cells.** *Biochimica et biophysica acta* 1997, **1331:**1-40.

32. Ohgami RS, Campagna DR, Greer EL, Antiochos B, McDonald A, Chen J, Sharp JJ, Fujiwara Y, Barker JE, Fleming MD: **Identification of a ferrireductase required for efficient transferrin-dependent iron uptake in erythroid cells.** *Nat Genet* 2005, **37:**1264-1269.

33. Ohgami RS, Campagna DR, McDonald A, Fleming MD: **The Steap proteins are metalloreductases.** *Blood* 2006, **108:**1388-1394.

34. Dassler K, Zydek M, Wandzik K, Kaup M, Fuchs H: **Release of the soluble transferrin receptor is directly regulated by binding of its ligand ferritransferrin.** *J Biol Chem* 2006, **281:**3297-3304.

35. Meyron-Holtz EG, Fibach E, Gelvan D, Konijn AM: **Binding and uptake of exogenous isoferritins by cultured human erythroid precursor cells.** *Br J Haematol* 1994, **86:**635-641.

36. Kristiansen M, Graversen JH, Jacobsen C, Sonne O, Hoffman HJ, Law SK, Moestrup SK: **Identification of the haemoglobin scavenger receptor.** *Nature* 2001, **409:**198-201.

37. Lonnerdal B: **The importance and bioavailability of phytoferritin-bound iron in cereals and legume foods.** *International journal for vitamin and nutrition research Internationale Zeitschrift fur Vitamin- und Ernahrungsforschung Journal international de vitaminologie et de nutrition* 2007, **77:**152-157.

38. Finberg KE, Heeney MM, Campagna DR, Aydinok Y, Pearson HA, Hartman KR, Mayo MM, Samuel SM, Strouse JJ, Markianos K, et al: **Mutations in *TMPRSS6* cause iron-refractory iron deficiency anemia (IRIDA).** *Nat Genet* 2008, **40:**569-571.

39. Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, Khovananth K, Mudd S, Mann N, Moresco EM, et al: **The serine protease *TMPRSS6* is required to sense iron deficiency.** *Science* 2008, **320:**1088-1092.

40. De Falco L, Sanchez M, Silvestri L, Kannengiesser C, Muckenthaler MU, Iolascon A, Gouya L, Camaschella C, Beaumont C: **Iron refractory iron deficiency anemia.** *Haematologica* 2013, **98:**845-853.

41. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FC, Fleming RE, Kiemeney LA, Vermeulen SH, Swinkels DW: **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.

42. Traglia M, Girelli D, Biino G, Campostrini N, Corbella M, Sala C, Masciullo C, Vigano F, Buetti I, Pistis G, et al: **Association of HFE and *TMPRSS6* genetic variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations.** *J Med Genet* 2011, **48:**629-634.

43. Hentze MW, Muckenthaler MU, Andrews NC: **Balancing acts: molecular control of mammalian iron metabolism.** *Cell* 2004, **117:**285-297.

44. Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R, Jr., Ellis MC, Fullan A, et al: **A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis.** *Nat Genet* 1996, **13:**399-408.

45. Boomsma D, Busjahn A, Peltonen L: **Classical twin studies and beyond.** *Nat Rev Genet* 2002, **3:**872-882.

46. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, Duffy DL, Martin NG: **Effects of HFE C282Y and H63D polymorphisms and polygenic background on iron stores in a large community sample of twins.** *Am J Hum Genet* 2000, **66:**1246-1258.

47. Njajou OT, Alizadeh BZ, Aulchenko Y, Zillikens MC, Pols HA, Oostra BA, Swinkels DW, van Duijn CM: **Heritability of serum iron, ferritin and transferrin saturation in a genetically isolated population, the Erasmus Rucphen Family (ERF) Study.** *Hum Hered* 2006, **61:**222-228.

48. Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, et al: **Heritability of cardiovascular and personality traits in 6,148 Sardinians.** *PLoS Genet* 2006, **2:**e132.

49. Marroni F, Grazio D, Pattaro C, Devoto M, Pramstaller P: **Estimates of genetic and environmental contribution to 43 quantitative traits support sharing of a homogeneous environment in an isolated population from South Tyrol, Italy.** *Hum Hered* 2008, **65:**175-182.

50. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, Peltonen L, Martin NG, Montgomery GW, Whitfield JB, Visscher PM: **Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels.** *Am J Hum Genet* 2009, **84:**60-65.

51. Benyamin B, Esko T, Ried JS, Radhakrishnan A, Vermeulen SH, Traglia M, Gogele M, Anderson D, Broer L, Podmore C, et al: **Novel loci affecting iron homeostasis and their effects in individuals at risk for hemochromatosis.** *Nat Commun* 2014, **5:**4926.

52. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, Zhou D, Song P, Wang C, Meng L, 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.

53. Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, Jing L, Yu Y, Ruan S, Xie D, et al: **Association of *TMPRSS6* polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese Han population.** *Am J Clin Nutr* 2012, **95**:626-632.

54. Tishkoff SA, Williams SM: **Genetic analysis of African populations: human evolution and complex disease.** *Nat Rev Genet* 2002, **3**:611-621.

55. Spottiswoode N, Duffy PE, Drakesmith H: **Iron, anemia and hepcidin in malaria.** *Front Pharmacol* 2014, **5**:125.

56. Maliken BD, Nelson JE, Kowdley KV: **The hepcidin circuits act: balancing iron and inflammation.** *Hepatology* 2011, **53**:1764-1766.

57. Sharp P: **SNPs linking TNF with anemia.** *Blood* 2008, **112**:3923-3924.

Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (*TMPRSS6*) gene associated with iron status indicators: A systematic review with meta - analyses

Wanjiku N Gichohi-Wainaina
Dorine W Swinkels
Michael B Zimmermann
Edith J Feskens
Alida Melse-Boonstra

Abstract

Background: Transmembrane protease, serine 6 (*TMPRSS6*) is likely to be involved in iron metabolism through its pleiotropic effect on hepcidin concentrations. Recently, genome-wide association studies have identified common variants in the *TMPRSS6* gene to be linked to anaemia and low iron status. To get a more precise evaluation of identified *TMPRSS6* single nucleotide polymorphism associations with iron status in cohorts of differing continental ancestry, we conducted a systematic review with meta analyses.

Methods: We searched the literature using HuGE Navigator, Pubmed and Scopus databases for primarily genome wide association studies using *TMPRSS6* as a free term. Fixed effects meta-analysis was used to obtain summary estimates of associations.

Results: Eleven studies comprised Caucasian populations, four included an Asian population and one study included an African-American population. Differences in minor allele frequencies (MAF) of 8 *TMPRSS6* SNPs (rs855791, rs4820268, rs2111833, rs1421312, rs228921, rs228918, rs228919 and rs575620) across ethnic groups were observed; with the MAF of rs855791 significantly higher in Asian populations than in Caucasians (0.55 vs 0.42, P<0.0001). In the meta-analysis, the A allele of rs855791 was associated with lower Hb and ferritin concentrations in all populations. This allele was also associated with increased serum transferrin receptor and transferrin concentrations. We observed similar associations for the G allele in rs4820268.

Conclusions: Clear disparities in associations were found for the African-American population, although not statistically significant. Associations between *TMPRSS6* SNPs and anaemia are consistent across Caucasian and Asian populations. This study highlights the need to conduct studies in African populations where iron deficiency is of utmost public health significance.

Introduction

Iron deficiency has previously been hypothesized to originate entirely from dietary and/or other environmental factors. However, several discoveries regarding disorders of iron metabolism have indicated that there is a genetic contribution to the development of iron deficiency [1-3]. In particular, polymorphisms in the Trans Membrane Protease Serine 6 (*TMPRSS6*) gene have been implicated as influencing iron metabolism in both animal and human studies. In the mask mutant mouse, loss of the catalytic domain of matriptase-2 (protein produced due to *TMPRSS6* expression) results in an increase in hepcidin expression in the liver, lowered dietary iron absorption, and severe microcytic anaemia [4]. Similar findings have been observed in *TMPRSS6* knockout mice (*TMPRSS6* 2/2) [5] and in zebra-fish over-expressing mutant matriptase-2 [6]. Contrariwise, over-expression of wild-type matripatse-2 in human hepatoblastoma (HepG2) cells has been observed to result in suppression of hepcidin gene promoter activation [4]. Mutations in the *TMPRSS6* gene have also been implicated in iron deficiency anaemia refractory to oral iron therapy within Caucasian populations [7-9]. Further evidence of the association of *TMPRSS6* polymorphisms with iron status in persons not affected by overt genetic disorders of iron metabolism have been observed in several GWAS though not all findings are consistent across studies [10-13].

At the population level, various studies have observed a geographic disparity in iron status leading to the hypothesis that dissimilarities in genetic alterations across ethnicities may indeed contribute to differences in iron status [14-17]. Specifically, studies have shown that Asia and Africa have the highest proportion of individuals affected by anaemia according to the WHO regional prevalence estimates [18]. Furthermore, in studies conducted in the US, African-Americans appear to have lower haemoglobin concentrations and serum transferrin saturation (TS) as compared to Caucasians [19-23]. Disparities in iron status have also been observed between Mexican American and non-Hispanic white women of childbearing age [24]. It is not known, however, if these differences contain a genetic component.

In order to shed light on the contradictory findings from GWAS and to evaluate the role of ethnicity as an explanatory factor, we performed a systematic review with meta-analysis on *TMPRSS6* loci identified in cohorts of differing continental ancestry.

Methods

Search Strategy

An electronic literature search was conducted using HuGE Navigator which is a database of published population-based epidemiological studies of human genes extracted and curated from PubMed since 2001. The search term '*TMPRSS6* × [Text MesH]' was used. The latest search was conducted on 31st August 2013. In addition, the PubMed and Scopus databases were searched to include any articles that may have been published on the topic prior to 2001. Searches in both databases were conducted with *TMPRSS6* as a free term. Manual searching of reference lists of original articles was also conducted.

Eligibility Criteria

The first step in the study selection was exclusion of duplicates followed by examination of titles and abstracts obtained. Articles were included when they were: original research articles; conducted in humans; and when testing for *TMPRSS6* SNP associations with iron status measures i.e. haemoglobin (Hb), serum or plasma ferritin (SF/PF) and/or serum transferrin receptor (sTfR), was undertaken. In addition, we included studies on associations of dichotomous outcomes (anaemia, iron deficiency, or iron deficiency anaemia) with *TMPRSS6* SNPs. Animal, single patient studies and studies on disorders of iron metabolism (iron-refractory iron deficiency anaemia (IRIDA), hemojuvelin, hereditary hemochromatosis) were excluded. The full text of each remaining study was reviewed to establish eligibility, and all relevant information and data were extracted.

Data Extraction

Data extraction was conducted by one author (WGW) and repeated by two other authors (AMB and EJF) for 30% of the papers that met the inclusion criteria. The latter was done for quality control purposes. For each article, information on authors, publication year, sample size, ethnicity, health status of the population (e.g. type 2 diabetes type patients), study design, mean age, gender distribution, minor allele frequencies, genotyping platform, call rate, beta values, standard errors, confidence intervals, reported measure of variance, agreement with Hardy-Weinberg equilibrium and model

adjustments were extracted. In the case of GWA studies, information on the main and replication study was reported separately. Additionally, where several cohorts were included in a study, information on each cohort was reported separately. In the case that information provided was missing, insufficient or unclear, authors were contacted for further information. Any recalculation of values required prior to meta-analyses was performed by one author (WGW).

Statistical Methods

Meta analyses were performed on genetic variants with information from more than two studies or cohorts and results ordered by ethnicity. Results were presented at the sub-group and overall levels. The sub-groups were designated as Caucasian, Asian and Mixed, with Mixed implying that the study results presented were from more than one ethnicity.

All data for specific iron status measures i.e. Hb, SF/PF and sTfR were transformed into identical units before meta-analyses. In the case where all extracted values had different transformations, values were back transformed to comparable units to enable comparison. New standard error values were obtained by calculating the ratio between the untransformed beta values and their standard errors and applying this ratio to the recalculated beta value.

Our goal in conducting the meta analyses was to compute the common effect size for the identified populations, and not to generalize the findings to other populations. Additionally, the studies with complete information were few. For these reasons, we used the fixed effects model to assign study weights as well as combine summary statistics. We also conducted a random effect analyses to check for heterogeneity between studies.

An estimate of potential publication bias was carried out by generating funnel plots. The symmetry of the funnel plot was assessed both visually and formally by using Egger's test [25]. The chi square test was conducted to test for heterogeneity and I square values reported. We conducted an independent samples T test on summary estimates per SNP for each outcome measure to determine whether differences observed between ethnicities were statistically significant.

The R program for statistical computing version 2.15.2 *[26]* was used to perform all analyses.

Results

We found 20 articles through HuGE navigator, 84 articles through PubMed and 86 articles through Scopus and four articles through manual search of the references. In total we obtained 14 articles that contained information from various cohorts (**Figure 1 and Table 1**). Articles identified as eligible for meta analyses (n=14) contained complete information on two *TMPRSS6* variants. In addition, 11 articles on MAF comparisons and 3 articles on 4 *TMPRSS6* variants were included in the systematic review (**Figure 1**). Only one study corrected for iron intake in association analyses. Additional data on covariates considered in association analyses in the various studies can be found in **Table 1**. The majority of the articles were based on studies conducted entirely in subjects of Caucasian ethnicity (n = 11; 73%). Three studies had multiple ethnicities as part of the cohorts investigated *[11, 27, 28]*. We did not find any studies conducted among individuals from the African population (**Table 1 and 2**).No publication bias was present in the meta analyses performed except for the association between rs855791 and transferrin in the Caucasian population.

In rs855791, MAF was higher in the Asian than in Caucasian populations (0.55 vs 0.42, P<0.0001). The MAF of the rs2111833 SNP ranged from 0.20 in the Hispanic population to 0.47 in the Caucasian population, while the MAF of rs1421312 was 0.31 in the Asian population and 0.47 in the Caucasian population. MAF for rs228921, rs228918, rs228919 and rs575620 were comparable in Caucasian and Indian Asians (**Table 1**).

Table 1: Characteristics of cohorts included in the review

Cohort	Location	Ethnicity	Mean age years (SD)	*n*	Gender ratio Male/Female	Iron status marker/erythrocyte trait studied	Confounders considered in association analyses	Reference
-	China	Han (East)	62.2 (7.7)	1141	0/100	Hb, SI, SF, sTfRN, FEP, TIBC	age	[29]
-	China	Han (North)	58.2 (7.5)	354	0/100	Hb, SI, SF, sTfRN, FEP, TIBC		
-	China	Zhuang	59.4 (7.9)	630	0/100	Hb, SI, SF, sTfRN, FEP, TIBC		
Health professional follow up study (HPFS) and Nurses health study (NHS)	USA	Caucasian	43.3 (6.7)	2422	-	Ferritin, TfR, TRFN	None	[30]
KORA F3, KORA F4, Micors, InCHIANTI, Croat-Vis	Europe	Caucasian	58.9 (12.8)	~6600	-	sTfR, SI, TRFN, TRFN saturation	Age, sex	[31]

Table 1 Continued

-	China	Han	58.3(5.9)	1574	45.2/45.8	Hb, PF	age, sex, dietary iron and BMI	[32]
-	USA	Mixed	56.5(13.5)	48	0/100	Hb, SI, SF, MCV, TIBC	Age, presence or absence of pica	[28]
Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium	Europe	Caucasian	58.5 (8.3)	24731	43/57	Hb, Hct, MCH, MCHC, MCV, Erythrocyte count	None	[12]
Haematology genetics (HAEMGEN) study	Italy	Caucasian	-	9456	-	Hb, Hct, MCH, MCHC, MCV, Erythrocyte count	None	[12]
InCHIANTI	USA	Caucasian	68.3 (15.5)	1206	44/56	SI, SF, Hb, Hct, MCV, RBC count, RBC width, platelets	age& gender	[13]

Table 1 Continued

Study	Country	Ethnicity	Age	n	M/F	Measures	Adjustments	Ref
Women's Healthy Aging Study (WHAS) I	USA	Caucasian	78.3 (8.0)	375	0/100	SI, SF, Hb, Hct, MCV, RBC count, RBC width, platelets	None	[13]
WHAS II	USA	Caucasian	74 (2.7)	194	0/100	SI, SF, Hb, Hct, MCV, RBC count, RBC width, platelets	None	[13]
InCHIANTI, BLSA, SardiNIA stage 2	Italy and USA	Caucasian	59.5 (15.8)	3200	46/54	Fe, TRFN, Ferritin, sTfR	Sex, age	[33]
-	Spain	Caucasian	24.3 (4.8)	270	0/100	SI, SF, Hb, Hct, MCV, RBC count, TRFN, Transferrin saturation, TIBC	None	[34]

Table 1 Continued

Australian Twin families	Australia	Caucasian	adolescents	~2512	40/60	Fe, TRFN, Transferrin saturation, Hb, MCV	sex, age, sex*age	[10]
Nijmegen biomedical study (NBS)	NL	Caucasian	-	1832	-	Fe, TRFN, Transferrin saturation, Hb, MCV	None	[10]
London Life Sciences Population (LOLLIPOP)	UK	Caucasian	54.5 (9.8)	1599	87.3/12.7	Hb, RBC, MCV, MCH, MCHC	None	[11]
LOLLIPOP	UK	Indian Asians	53.6 (10.8)	9685	88.8/12.2	Hb, RBC, MCV, MCH, MCHC	None	[11]

Table 1 Continued

Salzburg Atherosclerosis Prevention Program in Subjects at High Individual Risk (SAPHIR)	Austria	Caucasian	Range Male-39-66 years Females-39-67 years	1726	63/37	Hb, RBC count, SI	age& gender	[35]
ValBorbera (VB) study	Italy	Caucasian	55.4 (17.8)	~1650	44/56	Hb, Hct, MCV, MCH, MCHC, SI, TRFN, ferritin, TRFN saturation	None	[36]
Hemochromatosis and Iron Overload Screening (HEIRS*) study	USA	Caucasian	60.1 (10.4)	1084	24/76	SI, UIBC, SF, sTfR, TIBC	age, sex, In(CEA), ln(CRP), ln(GGT), C282Y/H63D	[27]
HEIRS*	USA	African-American	57.5 (12.4)	221	36/64	SI, UIBC, SF, sTfR, TIBC	ln(CRP),ln(GGT),CagA	[27]

Table 1 Continued

HEIRS*	USA	Hispanic	56.3 (10.1)	239	25/75	SI, UIBC, SF, sTfR, TIBC	In(CEA)	[27]
HEIRS*	USA	Asian	55.7 (10.9)	153	36/64	SI, UIBC, SF, sTfR, TIBC	ln(GGT)	[27]

Ethnicities with similar vowels next to them were analysed and compared in the same study

*adjustments considered in association testing with serum ferritin

information not available

ln-natural log transformation

Covariate definitions: carcinoembryonic antigen (CEA), c-reactive protein (CRP), gamma-glutamyltransferase (GGT), HFE gene

C282Y/H63D genotype, CagA strain infection status (infected vs. non-infected)

haemoglobin (Hb), serum iron (SI), serum ferritin (SF), plasma ferritin (PF), transferrin (TRFN), serum transferrin (sTfRN), serum transferrin receptor (sTfR), free erythrocyte porphyrins (FEP), total iron binding capacity (TIBC), unsaturated iron-binding capacity (UIBC), hematocrit (Hct), mean corpuscular haemoglobin (MCH), mean corpuscular haemoglobin concentration (MCHC), mean corpuscular volume (MCV), red blood cell (RBC)
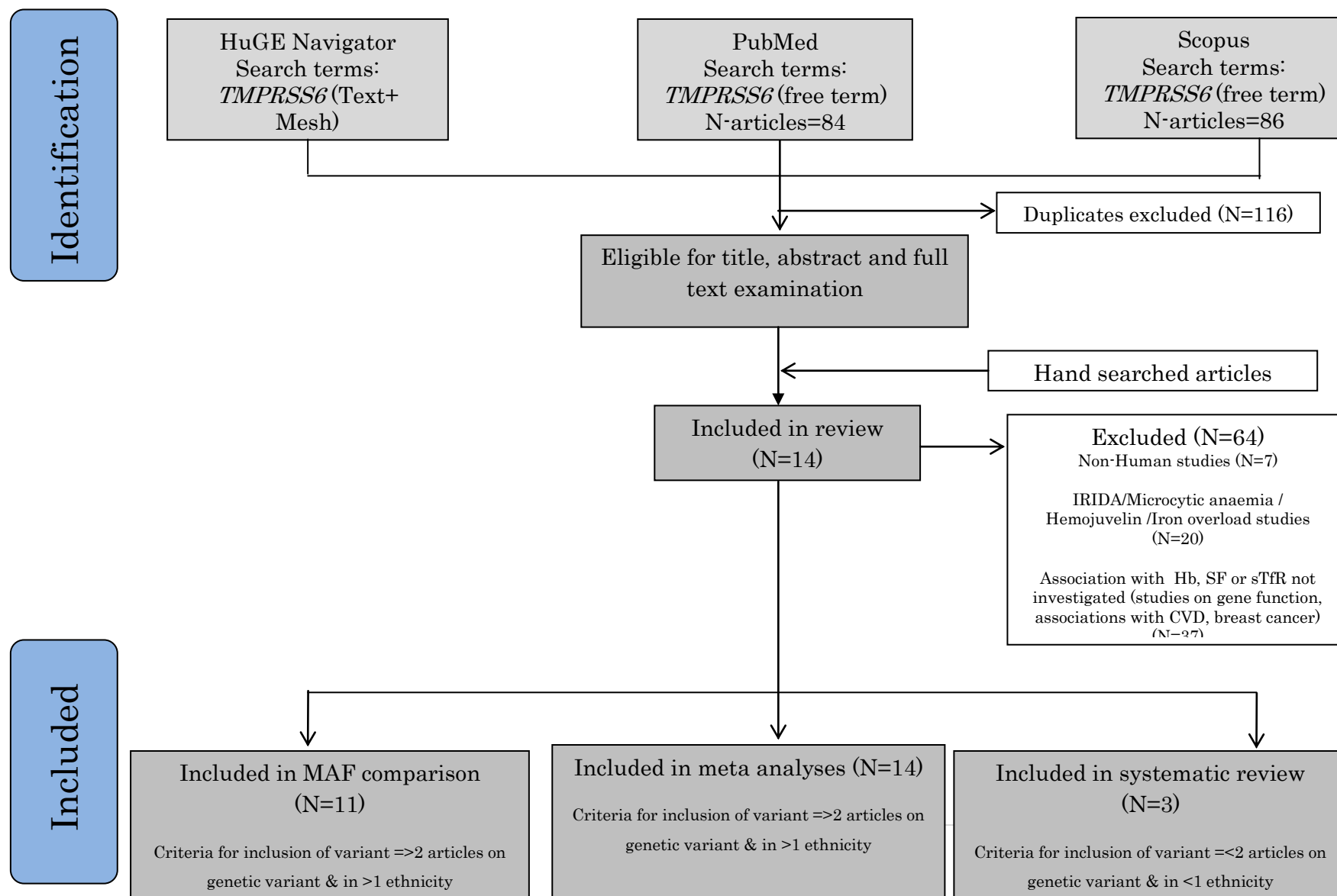
Figure 1:Study selection

Table 2: Minor allele frequencies of *TMPRSS6* SNPs in various populations

SNP	Cohort	Ethnicity	MAF	Reference
rs855791 (A)	-	Han (East)	0.53	[29]
	-	Han (North)	0.55	[29]
	-	Zhuang	0.63	[29]
	-	Han	0.53	[32]
	CHARGE+HAEMGEN	Caucasian	0.39	[12]
	InCHIANTI + BLSA +			
	WHAS I+WHAS II	Caucasian	0.41	[13]
	Australian			
	Twin families	Caucasian	0.42	[10]
	LOLLIPOP	Caucasian	0.34	[11]
	LOLLIPOP	Indian Asians	0.53	[11]
	SAPPHIR	Caucasian	0.43	[35]
	NBS	Caucasian	0.46	[35]
	-	Mixed	0.37	[28]
	ValBorbera study	Caucasian	0.45	[36]
	NBS	Caucasian	0.46	[37]

Table 2 Continued

rs4820268 (G)	-	Han (East)	0.51	[29]
		Han (North)	0.53	[29]
		Zhuang	0.61	[29]
	-	Han	0.50	[32]
	InCHIANTI+			
	BLSA+WHAS I+			
	WHAS II	Caucasian	0.46	[13]
		Caucasian	0.47	[10]
	LOLLIPOP	Caucasian	0.43	[11]
	LOLLIPOP	Indian Asians	0.55	[11]
	SAPPHIR	Caucasian	0.45	[35]
	NBS	Caucasian	0.47	[35]
rs2111833 (T)	HEIRS	Caucasian	0.34	[27]
	HEIRS	African-American	0.39	[27]
	HEIRS	Hispanic	0.20	[27]
	HEIRS	Asian	0.31	[27]

Table 2 Continued

rs1421312 (G)	HEIRS	Caucasian	0.40	[27]
	HEIRS	African-American	0.61	[27]
	HEIRS	Hispanic	0.34	[27]
	HEIRS	Asian	0.31	[27]
	InCHIANTI+			
	BLSA+ WHAS I+			
	WHAS II	Caucasian	0.39	[13]
	LOLLIPOP	Caucasian	0.46	[11]
	LOLLIPOP	Indian Asians	0.54	[11]
rs228921 (G)	LOLLIPOP	Caucasian	0.41	[11]
	LOLLIPOP	Indian Asians	0.48	[11]
rs228918 (C)	LOLLIPOP	Caucasian	0.47	[11]
	LOLLIPOP	Indian Asians	0.48	[11]
rs228919 (T)	LOLLIPOP	Caucasian	0.40	[11]
	LOLLIPOP	Indian Asians	0.48	[11]
rs5756520 (A)	LOLLIPOP	Caucasian	0.41	[11]
	LOLLIPOP	Indian Asians	0.48	[11]

Associations of rs855791 with Hb and iron status

The rs855791 (c. 2207 T>C) SNP causes the matriptase-2 valine to alanine amino acid substitution (p.Val736Ala) [38]. In our association analyses we considered the A allele which is the minor allele on the reverse strand (similar to the T allele on the forward strand). The meta analysis of the association of rs855791 with Hb indicates that each A allele (encoding for Valine [Val]) was associated with 0.11 g/dl (95% CI -0.11,-0.10) lower Hb concentrations overall, with values of -0.08 (95% CI -0.11, -0.04), -0.15 (95% CI -0.18,-0.12) and -0.11 (95% CI -0.11,-0.10) g/dl in the Caucasian, Asian and mixed populations, respectively (**Figure 2**). The difference in the effect estimates between Caucasian and Asian populations, based on 11 and 5 separate study populations, respectively, was not significant (P=0.85). Heterogeneity was high among the studies within the Caucasian population (I^2=96.2%, P=<0.0001, n=11) and in the mixed populations (I^2=96.3%, P<0.0001, n=2), but was absent for the Asian cohorts (I^2=0.00%, P=0.47, n=5).

Data from 13 study populations were available to study the associations of the A allele and serum ferritin concentration. Overall, a reduction of 3.71 ug/L (95% CI -4.12,-3.31) was observed, without significant differences between ethnicities (P>0.05) (**Figure 3**). As for Hb, heterogeneity was high in the Caucasian population (I^2=65.2%, P=0.02) but not in the Asian populations (I^2=0.00%).

Associations of rs855791 with serum transferrin receptor concentration were only conducted within the Caucasian population. Based on three study populations, the estimate shows that the A allele was associated with an increase of 0.02 mg/dl (95% CI 0.01,0.02) in serum transferrin receptor concentration (**Figure 4**). Heterogeneity between the studies was low (I^2=0.00%, P=0.93).

In determining the association with transferrin, overall meta-analysis of five study populations indicated that the A allele had no effect 0.00 mg/dl (95% CI -0.05,0.06) (**Figure 5**). Heterogeneity between the studies was high(I^2=99.7%, P=0.01). Publication bias was detected in the meta analyses performed (P=0.01)

Figure 2: Fixed-effects meta-analysis of observational studies evaluating association of rs855791 with haemoglobin concentration (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Subtitle Mixed refers to a cohort consisting of more than one ethnicity. Effect allele = A

Figure 3: Fixed-effects meta-analysis of observational studies evaluating association of rs855791 with ferritin (μg/L)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Subtitle Mixed refers to a cohort consisting of more than one ethnicity. Effect allele = A

Figure 4: Fixed-effects meta-analysis of observational studies evaluating association of rs855791 with serum transferrin receptor (mg/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A
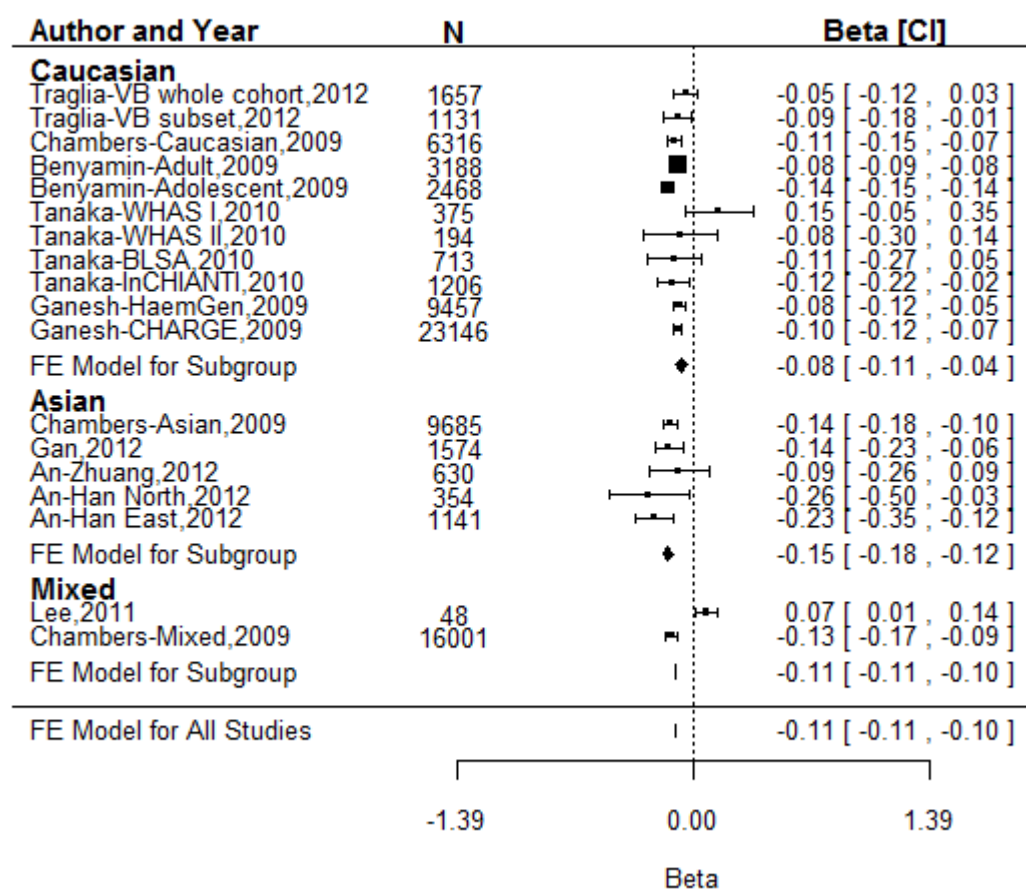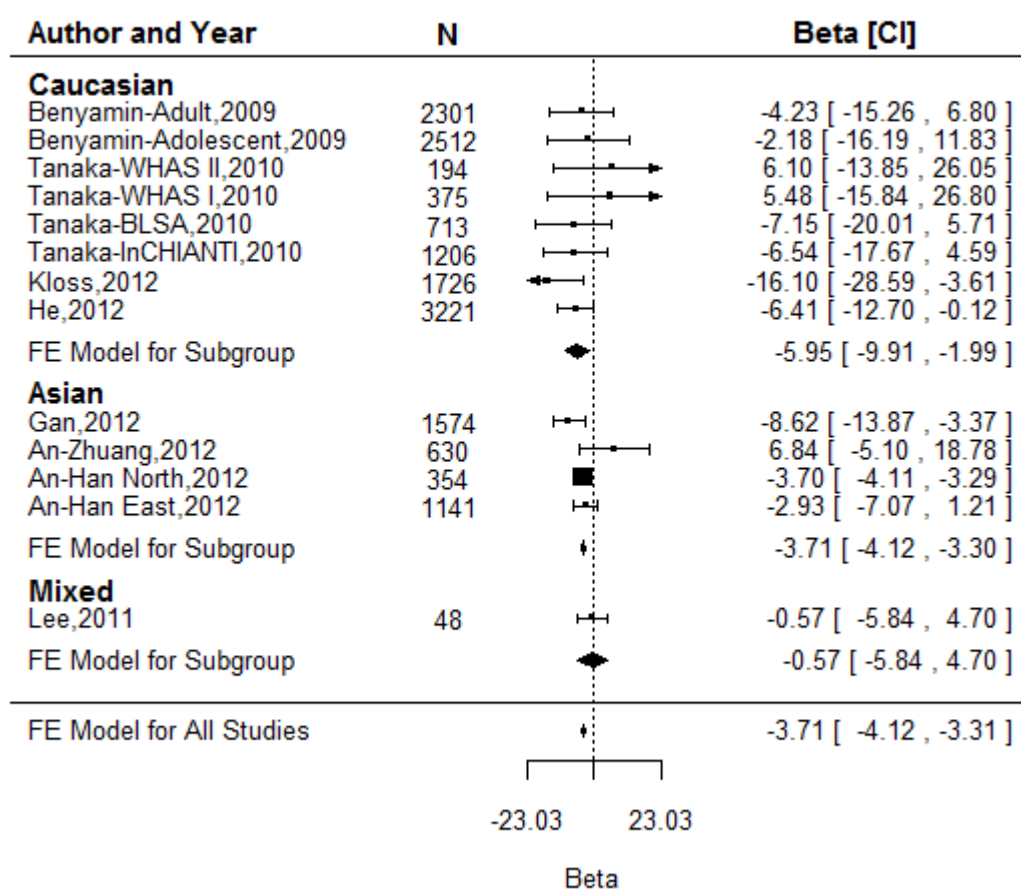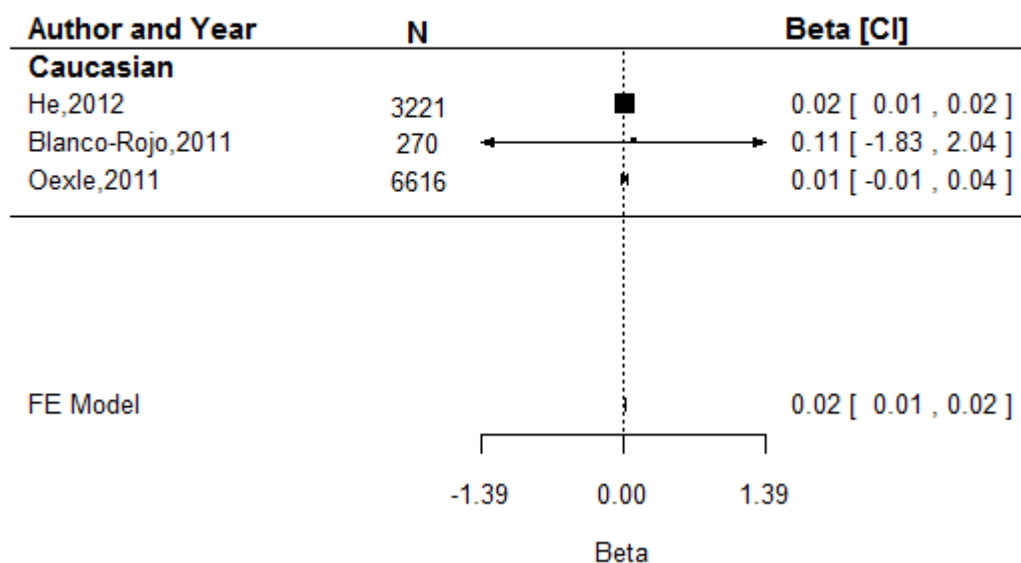
Figure 5: Fixed-effects meta-analysis of observational studies evaluating association of rs855791 with with transferrin (mg/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A

Associations of rs4820268 with Hb and iron status

The minor allele at the rs4820268 (c.1563C>T) leads to a synonymous change at nucleotide position 521 (p.Asp521Asp). In our association analyses we considered the G allele which is the minor allele on the forward strand (similar to the C allele on the reverse strand). The meta analysis of the association of rs4820268 with Hb and ferritin within the Caucasian and Asian ethnicities indicated that the G allele results in lower concentrations of Hb by 0.08 g/dl (95% CI -0.16,-0.01) and 0.16 g/dl (95% CI -0.22,-0.10) respectively (**Figure 6**). The difference in the effect estimates between four Caucasian and Asian populations was not significant (P>0.05). We observed a non-significant amount of heterogeneity within studies in both the Caucasian (I^2=6.20%, P=0.34) and Asian population (I^2=0.61%, P=0.23).

We obtained data from 13 study populations on the associations of the G allele and serum ferritin concentration. The G allele resulted in an increased ferritin concentration

of 0.12 µg/l (95% CI -6.16,6.39) in the Caucasian population and a decrease of 3.69 µg/l (95% CI -9.09,1.72) in the Asian population (**Figure 7**). There was no significant difference in the effect estimates between Caucasian and Asian populations (P>0.05). We observed a significant amount of heterogeneity in the studies within the Caucasian population (I^2=65.15%, P=0.02), but not in the Asian studies (I^2=0.00%, P=0.43).

Associations of rs4820268 with sTfR concentration were only conducted in three studies within the Caucasian population. The G allele was associated with a reduction by 0.08mg/dL (95% CI -0.10,-0.06) in the Caucasian population (**Figure 8**). There was a significant amount of heterogeneity detected (I^2=65.15%, P=0.02).

Author and Year	N		Beta [CI]
Caucasian			
Tanaka_WHAS I,2010	375		0.08 [-0.12 , 0.28]
Tanaka_WHAS II,2010	194		-0.06 [-0.28 , 0.16]
Tanaka_BLSA,2010	713		-0.11 [-0.27 , 0.05]
Tanaka_InCHIANTI,2010	1206		-0.12 [-0.22 , -0.02]
FE Model for Subgroup			-0.08 [-0.16 , -0.01]
Asian			
Gan,2012	1574		-0.13 [-0.21 , -0.05]
An_Zhuang,2012	630		-0.06 [-0.24 , 0.12]
An_Han North,2012	354		-0.34 [-0.57 , -0.10]
An_Han East,2012	1141		-0.20 [-0.32 , -0.09]
FE Model for Subgroup			-0.16 [-0.22 , -0.10]
FE Model for All Studies			-0.13 [-0.17 , -0.08]

-1.39   0.00   1.39

Beta

Figure 6: Fixed-effects meta-analysis of observational studies evaluating association of rs4820268 with haemoglobin concentration (g/dL), Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = G
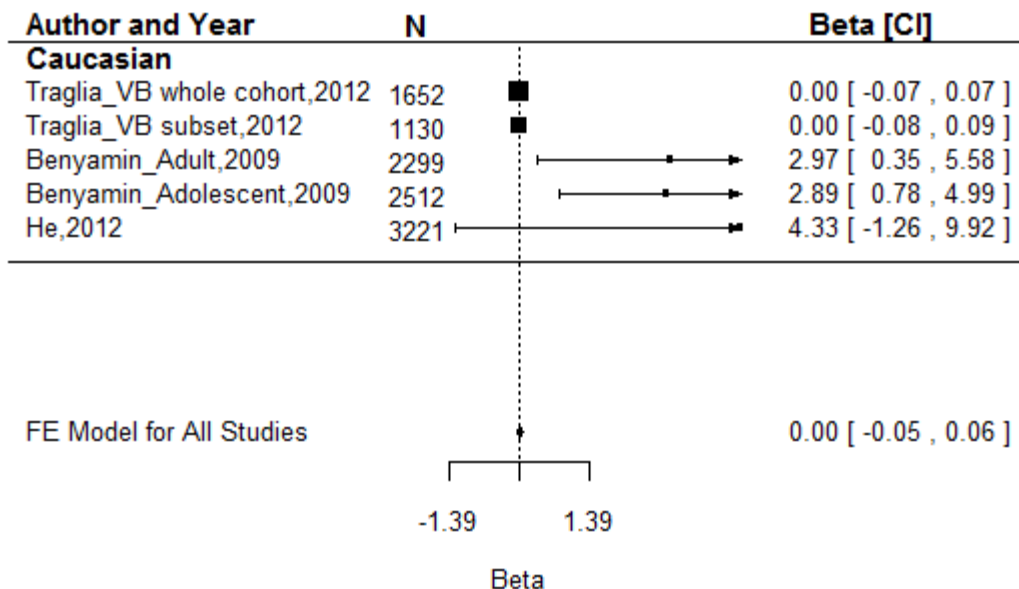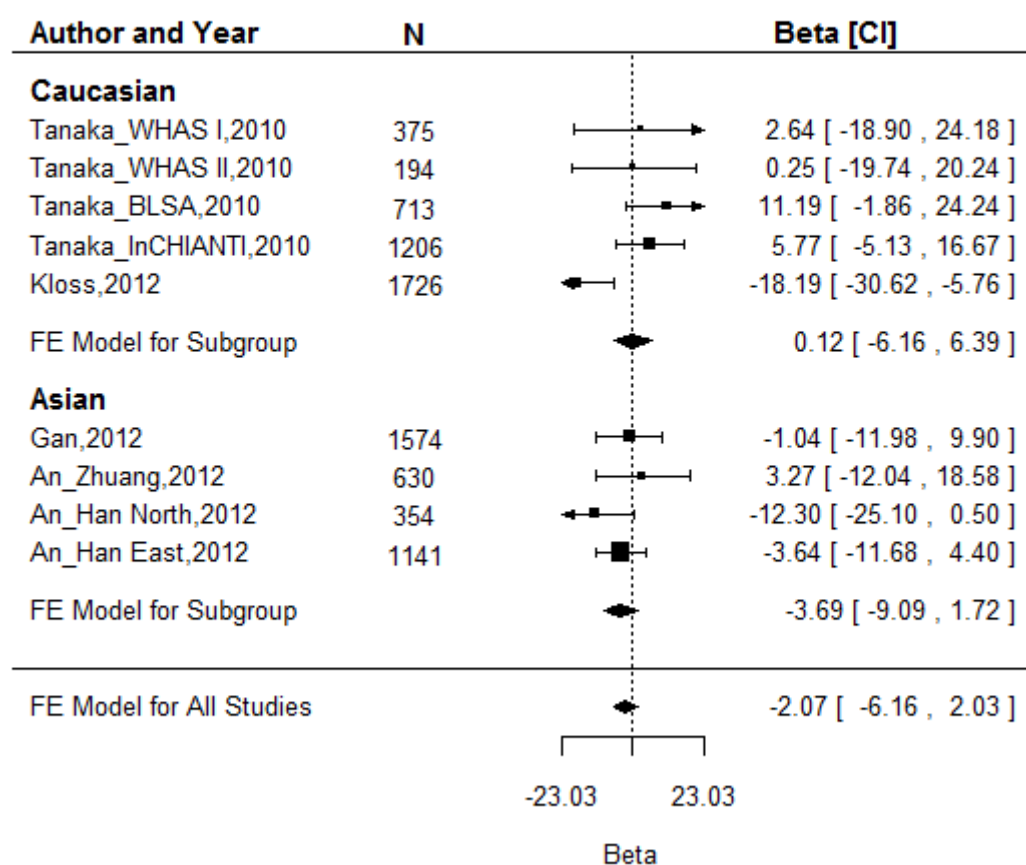
Figure 7: Fixed-effects meta-analysis of observational studies evaluating association of rs4820268 with ferritin (μg/L)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = G
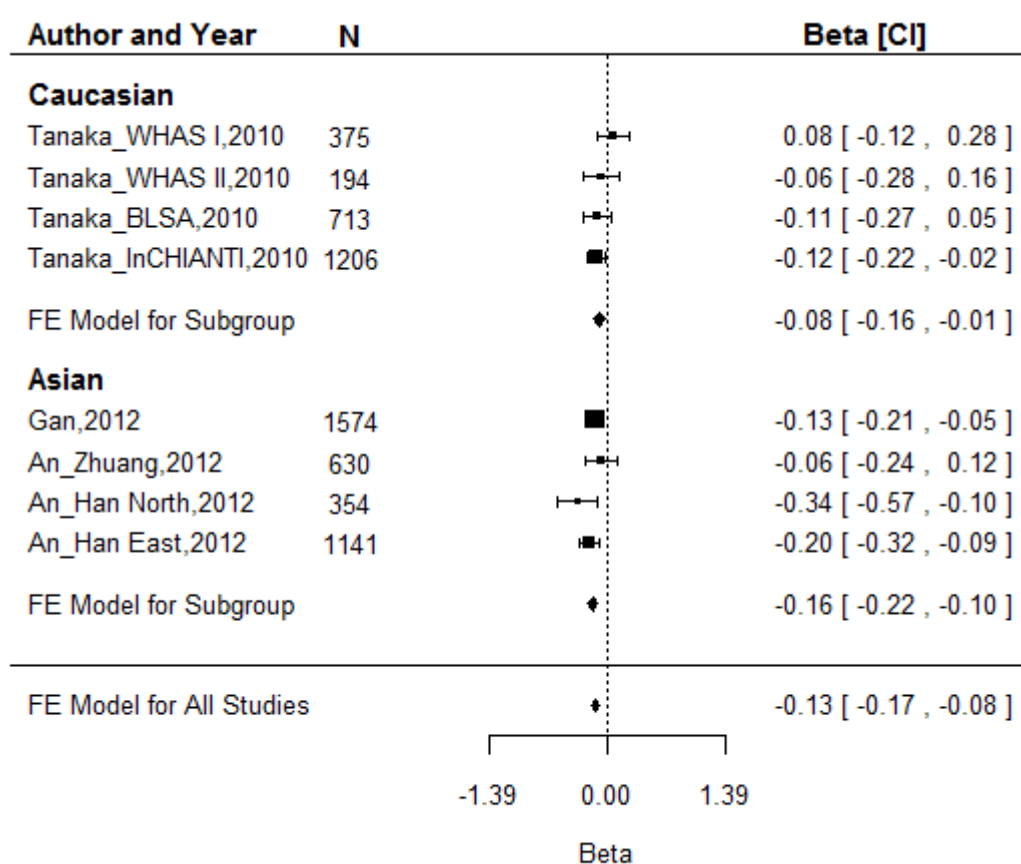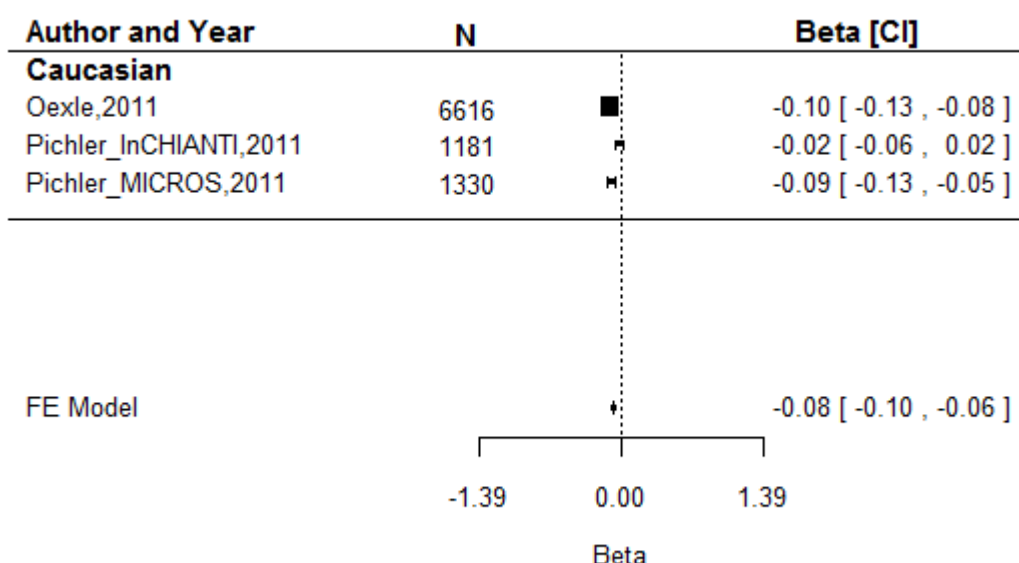
Figure 8: Fixed-effects meta-analysis of observational studies evaluating association of rs4820268 with serum transferrin receptor (mg/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = G

TMPRRS6 variants not eligible for meta-analysis

Since the associations of rs2111833 (c.1083G>A) and rs1421312 (c.659-1988T>C) with iron status among Caucasians, Asians, Hispanics and African-Americans was described in only one study by McLaren and McLachlan *[27]*, these variants were not eligible for the meta-analysis. In this study, MAF across the populations for rs211833 indicated that the T allele has the lowest frequency among the Hispanic population (20%) while the Caucasian, Asian and African-American populations had a MAF of approximately 30% (**Table 2**). When investigating the association of rs2111833 with serum ferritin, the T allele was associated with an increase in serum ferritin concentrations across all ethnicities. For rs2111833 and sTfR the estimates in each ethnicity indicated an association of the T allele with a decrease in concentrations *[27]*. However, none of the observations were significant (P>0.05).

The SNP rs1421312 was also studied in four ethnicities. The MAF of this SNP in the African-American population was 0.61, higher than the Caucasian population 0.40, Asian population 0.41, and Hispanic population, 0.34. Overall, the G allele of rs1421312 was associated with a non-significant decrease in serum ferritin concentration. However, each additional G allele resulted in a non-significant decrease of 30.75 μg/L (95% CI -159.62,98.12) in SF concentrations within the African-American population, but in increases in the other ethnic populations. For sTfR this SNP was associated with non-significant negative associations in Caucasians (-0.23, (95% CI -0.45, -0.01) and Asians (-0.25, (95% CI -0.78, 0.28) and a positive association in African-Americans (0.88, (95% CI -0.04, 1.80) and Hispanics (0.19, (95% CI -0.48, 0.86) [27].

We also identified studies in the Chinese population that have observed rs855791 and rs4820268 polymorphisms as genetic risk factors for developing anaemia, iron deficiency and iron deficiency anaemia *[29, 32]*. We additionally identified another study conducted among Danish blood donors that has observed that the T and G alleles of rs855791 and rs4820268, respectively, are not associated with low serum ferritin *[39]*.

Discussion

To our knowledge, this is the first review to focus on the association of *TMPRSS6* genetic variants with iron related parameters in different ethnic groups. Sample sizes included in the meta analyses were sufficient to investigate the research question. We observed that the A allele of rs855791 is associated with lower concentrations of Hb and ferritin across all populations. Additionally, this allele is associated with increased sTfR and transferrin concentrations. Based on our meta analyses, the overall effect estimate of the Asian studies indicates that for each A and G allele of rs855791 and rs4820268 respectively, an additional decrease in Hb values of 0.07and 0.12 g/dL is observed, as compared to the Caucasian population. Similarly, ferritin concentration is lowered by an additional 2.24 μg/l and 3.85 μg/l, respectively, in the Asian populations. The directions and magnitudes of associations were identical in both fixed effects and random effects model.

Although the observed differences in iron status may not be impressive from a clinical point of view, it may be of significance on the population level especially in populations that already have marginal iron status due to low dietary iron intake.

The exact mechanism through which *TMPRSS6* action is exerted is still under investigation. It had generally been thought that *TMPRSS6* polymorphisms affect hepcidin transcription, thereby altering hepcidin concentrations in response to systemic iron concentrations *[4, 7, 40]*. However, two recent studies *[36, 37]* did not confirm an intermediate role for hepcidin in the SNP-iron status parameters associations. These studies instead indicate pleiotropic SNP effects on hepcidin and iron status parameters. Currently, it is proposed that matriptase-2 could regulate hepcidin expression by cleaving HJV to decrease BMP-SMAD signalling *[41]*. Further studies are required to elucidate the role of hepcidin in *TMPRSS6*-iron status parameter associations.

We have observed similar directions of the effect estimates between rs855791 and rs4820268 in associations with iron status parameters. This is probably because these two SNPs are in linkage disequilibrium *[38]*. However, we also observed differences in the direction of association of the risk allele in rs1421312 with body iron, serum ferritin and serum iron in African-Americans. Similar alleles in different ethnic groups could result in variations in the expression of genes *[42]*. This further strengthens the need to conduct ethnicity specific studies when considering the role of genetics in disease outcome.

We observed that the risk allele frequencies of rs855791 and rs4820268 were as high as 63% in the Asian populations while in the Caucasian populations it was generally less than 45%. This difference in MAF across ethnicities has also been observed in the SNPs rs211833 and rs1421312. Indeed, allele frequencies for SNPs can vary greatly across ethnic groups *[43]*, and this may impact the prevalence and incidence of disease across ethnic groups in case of associations *[44, 45]*. The high frequency of the risk alleles of the rs855791 and rs4820268 SNPs in Asian populations and the stronger negative associations with Hb and ferritin predisposes this population to iron deficiency. Environmental factors such as a diets low in bioavailable iron and high inflammation burden further aggravates this predisposition. In this regard, it would have been interesting to include in this review studies conducted within the African population as well. However, at the time the search was completed, there were no studies conducted within the African population other than the one performed in the African-American

population. However, this is an admixed population *[46]* and therefore observations are not readily transferrable to the African population. Besides this, it may be that the phenotypic expression for the same alleles investigated are different within the genetically diverse African population [47-55].

Several limitations of our study need mentioning. We focussed on SNP–iron status measure associations that were investigated in at least two studies and in more than one ethnicity. Consequently, significant SNP–iron status measure associations that have only been researched in one ethnicity were not described. Additionally, since we only included SNPs described in greater than two cohorts and more than one ethnicity, we were only able to include two genetic variants from the selected SNPs in the meta-analyses. For this reason, there may be ethnic differences in other genetic variations that have not been included in this review.

Secondly, high heterogeneity was observed among the Caucasian studies. We attribute this to the variation in population characteristics and study designs. Most studies that were conducted within the Caucasian populations and have previously been designed to study other outcomes, whereas the studies within the Asian population were specifically designed to study genetic influences of iron metabolism and are fewer in number.

Thirdly, in conducting association analyses, only one study *[32]* included dietary iron as a predictor of iron status within the model. This investigation on the interactions between SNPs and the environment is crucial in order to quantify the contribution of genetic alterations to the development of iron deficiency. Further to this, only two studies considered inflammation as a potential confounder in association analyses. These studies either excluded subjects with CRP levels above a pre-defined cut off *[29, 32, 36]* or corrected for inflammation *[27]*. Correction for inflammation or exclusion of values of inflamed subjects would ensure that what is observed is a true association. Ferritin and Hb are influenced by inflammation and sTfR concentrations might also be influenced negatively by inflammatory cytokines *[56]*.

In conclusion, we observe that the risk alleles A and G in rs855791 and rs4820268 are associated with a reduction in Hb and serum ferritin concentrations and an increase in transferrin and serum transferrin receptor concentrations in most populations investigated. Our study highlights key information gaps in several areas regarding the associations of SNPs in *TMPRSS6* and iron status. First of all, more information on the

(factors involved in) regulation of *TMPRSS6* is required as well as how exactly *TMPRSS6* exerts its effects on body iron. Moreover, our study also highlights the need to conduct further studies in ethnicities or populations where iron deficiency is of public health significance. The question is whether populations with high prevalence of risk alleles may require higher recommendations for iron intake in order to maintain normal erythropoiesis. This is especially relevant in combination with unfavourable environmental conditions such as diets with low bioavailable iron and high inflammation burden. Knowledge of genetic factors influencing iron status could improve care and advice given to populations at risk for iron deficiency.

Acknowledgements

References

1. Morse AC, Beard JL, Jones BC: **A genetic developmental model of iron deficiency: biological aspects.** *Proc Soc Exp Biol Med* 1999, **220:**147-152.

2. Leboeuf RC, Tolson D, Heinecke JW: **Dissociation between tissue iron concentrations and transferrin saturation among inbred mouse strains.** *J Lab Clin Med* 1995, **126:**128-136.

3. Finberg KE: **Iron-refractory iron deficiency anemia.** *Semin Hematol* 2009, **46:**378-386.

4. Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, Khovananth K, Mudd S, Mann N, Moresco EM, et al: **The serine protease *TMPRSS6* is required to sense iron deficiency.** *Science* 2008, **320:**1088-1092.

5. Folgueras AR, de Lara FM, Pendas AM, Garabaya C, Rodriguez F, Astudillo A, Bernal T, Cabanillas R, Lopez-Otin C, Velasco G: **Membrane-bound serine protease matriptase-2 (*TMPRSS6*) is an essential regulator of iron homeostasis.** *Blood* 2008, **112:**2539-2545.

6.   Silvestri L, Pagani A, Nai A, De Domenico I, Kaplan J, Camaschella C: **The serine protease matriptase-2 (*TMPRSS6*) inhibits hepcidin activation by cleaving membrane hemojuvelin.** *Cell Metab* 2008, **8:**502-511.

7.   Finberg KE, Heeney MM, Campagna DR, Aydinok Y, Pearson HA, Hartman KR, Mayo MM, Samuel SM, Strouse JJ, Markianos K, et al: **Mutations in *TMPRSS6* cause iron-refractory iron deficiency anemia (IRIDA).** *Nat Genet* 2008, **40:**569-571.

8.   Guillem F, Lawson S, Kannengiesser C, Westerman M, Beaumont C, Grandchamp B: **Two nonsense mutations in the *TMPRSS6* gene in a patient with microcytic anemia and iron deficiency.** *Blood* 2008, **112:**2089-2091.

9.   Melis MA, Cau M, Congiu R, Sole G, Barella S, Cao A, Westerman M, Cazzola M, Galanello R: **A mutation in the *TMPRSS6* gene, encoding a transmembrane serine protease that suppresses hepcidin production, in familial iron deficiency anemia refractory to oral iron.** *Haematologica* 2008, **93:**1473-1479.

10.   Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, et al: **Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume.** *Nat Genet* 2009, **41:**1173-1175.

11.   Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, et al: **Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels.** *Nat Genet* 2009, **41:**1170-1172.

12.   Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, et al: **Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium.** *Nat Genet* 2009, **41:**1191-1198.

13.   Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, Walston JD, Fried LP, Singleton A, Guralnik J, et al: **A genome-wide association analysis of serum iron concentrations.** *Blood* 2010, **115:**94-96.

14.   Gordeuk V, Mukiibi J, Hasstedt SJ, Samowitz W, Edwards CQ, West G, Ndambire S, Emmanual J, Nkanza N, Chapanduka Z, et al.: **Iron overload in

Africa. Interaction between a gene and dietary iron content. *N Engl J Med* 1992, **326:**95-100.

15.   Moyo VM, Mandishona E, Hasstedt SJ, Gangaidzo IT, Gomo ZA, Khumalo H, Saungweme T, Kiire CF, Paterson AC, Bloom P, et al: **Evidence of genetic transmission in African iron overload.** *Blood* 1998, **91:**1076-1082.

16.   Wurapa RK, Gordeuk VR, Brittenham GM, Khiyami A, Schechter GP, Edwards CQ: **Primary iron overload in African Americans.** *Am J Med* 1996, **101:**9-18.

17.   Barton JC, Edwards CQ, Bertoli LF, Shroyer TW, Hudson SL: **Iron overload in African Americans.** *Am J Med* 1995, **99:**616-623.

18.   WHO: **Worldwide prevalence of anaemia 1993–2005.** 2008.

19.   McLaren CE, Li KT, Gordeuk VR, Hasselblad V, McLaren GD: **Relationship between transferrin saturation and iron stores in the African American and US Caucasian populations: analysis of data from the third National Health and Nutrition Examination Survey.** *Blood* 2001, **98:**2345-2351.

20.   Perry GS, Byers T, Yip R, Margen S: **Iron nutrition does not account for the hemoglobin differences between blacks and whites.** *J Nutr* 1992, **122:**1417-1424.

21.   Beutler E, Barton JC, Felitti VJ, Gelbart T, West C, Lee PL, Waalen J, Vulpe C: **Ferroportin 1 (SCL40A1) variant associated with iron overload in African-Americans.** *Blood Cells Mol Dis* 2003, **31:**305-309.

22.   Williams DM: **Racial differences of hemoglobin concentration: measurements of iron, copper, and zinc.** *Am J Clin Nutr* 1981, **34:**1694-1700.

23.   Jackson RT, Sauberlich HE, Skala JH, Kretsch MJ, Nelson RA: **Comparison of hemoglobin values in black and white male U.S. military personnel.** *J Nutr* 1983, **113:**165-171.

24.   Looker AC, Dallman PR, Carroll MD, Gunter EW, Johnson CL: **Prevalence of iron deficiency in the United States.** *JAMA* 1997, **277:**973-976.

25.   Egger M, Davey Smith G, Schneider M, Minder C: **Bias in meta-analysis detected by a simple, graphical test.** *BMJ* 1997, **315:**629-634.

26. Team RC: **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.** 2013.

27. McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, Adams PC, Acton RT, Murray JA, Leiendecker-Foster C, et al: **Associations between Single Nucleotide Polymorphisms in Iron-Related Genes and Iron Status in Multiethnic Populations.** *PLoS One* 2012, **7**:e38339.

28. Lee P: **Role of matriptase-2 (*TMPRSS6*) in iron metabolism.** *Acta Haematol* 2009, **122**:87-96.

29. An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, Zhou D, Song P, Wang C, Meng L, 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.

30. He M, Workalemahu T, Manson JE, Hu FB, Qi L: **Genetic Determinants for Body Iron Store and Type 2 Diabetes Risk in US Men and Women.** *PLoS One* 2012, **7**:e40919.

31. Oexle K, Ried JS, Hicks AA, Tanaka T, Hayward C, Bruegel M, Gogele M, Lichtner P, Muller-Myhsok B, Doring A, et al: **Novel association to the proprotein convertase PCSK7 gene locus revealed by analysing soluble transferrin receptor (sTfR) levels.** *Hum Mol Genet* 2011, **20**:1042-1047.

32. Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, Jing L, Yu Y, Ruan S, Xie D, et al: **Association of *TMPRSS6* polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese Han population.** *Am J Clin Nutr* 2012, **95**:626-632.

33. Pichler I, Minelli C, Sanna S, Tanaka T, Schwienbacher C, Naitza S, Porcu E, Pattaro C, Busonero F, Zanon A, 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**:1232-1240.

34. Blanco-Rojo R, Baeza-Richer C, Lopez-Parra AM, Perez-Granados AM, Brichs A, Bertoncini S, Buil A, Arroyo-Pardo E, Soria JM, Vaquero MP: **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.

35. Kloss-Brandstatter A, Erhart G, Lamina C, Meister B, Haun M, Coassin S, Seifert M, Klein-Franke A, Paulweber B, Kedenko L, 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:**e35015.

36. Traglia M, Girelli D, Biino G, Campostrini N, Corbella M, Sala C, Masciullo C, Vigano F, Buetti I, Pistis G, et al: **Association of HFE and *TMPRSS6* genetic variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations.** *J Med Genet* 2011, **48:**629-634.

37. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FC, Fleming RE, Kiemeney LA, Vermeulen SH, Swinkels DW: **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.

38. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, Traglia M, Toniolo D, Camaschella C: ***TMPRSS6* rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals.** *Blood* 2011, **118:**4459-4462.

39. Sorensen E, Grau K, Berg T, Simonsen AC, Magnussen K, Erikstrup C, Hansen MB, Ullum H: **A genetic risk factor for low serum ferritin levels in Danish blood donors.** *Transfusion* 2012, **52:**2585-2589.

40. De Falco L, Sanchez M, Silvestri L, Kannengiesser C, Muckenthaler MU, Iolascon A, Gouya L, Camaschella C, Beaumont C: **Iron refractory iron deficiency anemia.** *Haematologica* 2013, **98:**845-853.

41. Meynard D, Babitt JL, Lin HY: **The liver: conductor of systemic iron balance.** *Blood* 2014, **123:**168-176.

42. Spielman RS, Bastone LA, Burdick JT, Morley M, Ewens WJ, Cheung VG: **Common genetic variants account for differences in gene expression among ethnic groups.** *Nat Genet* 2007, **39:**226-231.

43. Ioannidis JP, Ntzani EE, Trikalinos TA: **'Racial' differences in genetic effects for complex diseases.** *Nat Genet* 2004, **36:**1312-1318.

44.   Lan Q, Shen M, Garcia-Rossi D, Chanock S, Zheng T, Berndt SI, Puri V, Li G, He X, Welch R, et al: **Genotype frequency and F ST analysis of polymorphisms in immunoregulatory genes in Chinese and Caucasian populations.** *Immunogenetics* 2007, **59:**839-852.

45.   Myles S, Davison D, Barrett J, Stoneking M, Timpson N: **Worldwide population differentiation at disease-associated SNPs.** *BMC Med Genomics* 2008, **1:**22.

46.   Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: **Estimating African American admixture proportions by use of population-specific alleles.** *Am J Hum Genet* 1998, **63:**1839-1851.

47.   Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, et al: **The genetic structure and history of Africans and African Americans.** *Science* 2009, **324:**1035-1044.

48.   Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, et al: **A second generation human haplotype map of over 3.1 million SNPs.** *Nature* 2007, **449:**851-861.

49.   Garrigan D, Kingan SB, Pilkington MM, Wilder JA, Cox MP, Soodyall H, Strassmann B, Destro-Bisol G, de Knijff P, Novelletto A, et al: **Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data.** *Genetics* 2007, **177:**2195-2207.

50.   Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, et al: **Genotype, haplotype and copy-number variation in worldwide human populations.** *Nature* 2008, **451:**998-1003.

51.   Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, Myers RM: **Worldwide human relationships inferred from genome-wide patterns of variation.** *Science* 2008, **319:**1100-1104.

52.   Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M: **Global patterns of linkage**

disequilibrium at the CD4 locus and modern human origins. *Science* 1996, **271**:1380-1387.

53.   Tishkoff SA, Kidd KK: **Implications of biogeography of human populations for 'race' and medicine.** *Nat Genet* 2004, **36**:S21-27.

54.   Tishkoff SA, Verrelli BC: **Patterns of human genetic diversity: implications for human evolutionary history and disease.** *Annu Rev Genomics Hum Genet* 2003, **4**:293-340.

55.   Tishkoff SA, Williams SM: **Genetic analysis of African populations: human evolution and complex disease.** *Nat Rev Genet* 2002, **3**:611-621.

56.   Seiser C, Teixeira S, Kuhn LC: **Interleukin-2-dependent transcriptional and post-transcriptional regulation of transferrin receptor mRNA.** *J Biol Chem* 1993, **268**:13074-13080.

Chapter 3

Common variants in the *TF*, *TNF α* and *TMPRSS6* genes are associated with iron status in a female black South African population

Wanjiku N Gichohi-Wainaina
Alida Melse-Boonstra
Dorine W Swinkels
Michael B Zimmermann
Edith J Feskens
G Wayne Towers

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

Abstract

Background: It is unknown whether single nucleotide polymorphisms (SNP) associated with iron status in European and Asian populations have the same relationship within the African population.

Objective: This study was designed to investigate associations of reported SNPs with iron status in a female black South African cohort. We additionally aimed to investigate gene-gene interactions that influence iron status.

Methods: Blood samples were obtained from women (n=686; range 32–86 years) which were part of the Prospective Urban and Rural Epidemiology (PURE) study. Concentrations of hemoglobin, serum ferritin, serum transferrin receptor and body iron stores were determined. Thirty two single nucleotide polymorphisms in 10 genes were selected based on existing GWAS data; 30 SNPs passed quality criteria.

Results: In the *TF* gene, the heterozygote genotype (AG) of rs1799852 was associated with decreased serum ferritin (p=0.01) and body iron (p=0.03) and increased serum transferrin receptor (sTfR) concentrations (P=0.004) compared to the homozygote minor allele genotype (AA), while the heterozygote genotype (AG) of rs3811647 was significantly associated with the lowest transferrin receptor and body iron concentrations (both P=0.03) compared to the homozygote wild type (AA) and minor allele groups (GG). The chromosome 6 allele combination (AAA) consisting of rs1799964 and rs1800629 both in *TNF α* and rs2071592 in *NFKBIL1* was associated with higher odds for low serum ferritin concentrations (serum ferritin<15µg/L; OR:1.86 (95%-CI, 1.23-2.79)).

The chromosome 22 allele combination (GG) consisting of rs228918 and rs228921 in the *TMPRSS6* gene was associated with lower odds for increased sTfR concentrations (sTfR>8.3mg/L; OR:0.79 (95%-CI,0.63-0.98).

Conclusion: We conclude that various SNPs and allele combinations in the *TF, TNF α* and *TMPRSS6* genes are associated with iron status in this African population; however, these association patterns are different compared to European ancestry populations. This stresses the need for population specific genomic data.

Introduction

Anaemia, accounts for 3.7% of maternal deaths during pregnancy and childbirth in Africa *[1]*. Maternal iron deficiency anaemia escalates the risk of poor pregnancy outcomes, predicts iron deficiency in infants after four months of age and affects child development *[2, 3]*. In adults, consequences of iron deficiency (ID) include reduced physical activity *[4]* and work performance as well as lethargy *[5, 6]* which impacts greatly on the burgeoning economies of developing countries *[7]*.

Several discoveries regarding genetic disorders of iron metabolism indicate that there is a substantial genetic contribution to the development of iron deficiency *[8-10]*. In addition, iron status biomarkers (serum iron, transferrin, transferrin saturation and ferritin) present with significant heritability *[11-14]* and previous population studies have identified relevant single nucleotide polymorphisms (SNPs) and gene loci (TF, HFE, TFR2, *TMPRSS6*) to be associated with these biomarkers *[15, 16]*. HFE and *TMPRSS6* have also been determined to affect red blood cell count, haemoglobin (Hb), and erythrocyte indices *[17]* most likely by affecting iron availability *[17-19]*

Nearly all studies on the genetics of iron status, however, have been performed in populations of European ancestry. The limited data available from Asian populations have pointed to potential ethnic differences in the heritability of specific measures of iron status *[20-22]*. However, the genetic determinants of iron metabolism have hardly been studied in the black African population which bears the greatest iron deficiency burden. For example, in preschool-age children, anaemia prevalence is 47.4%, affecting 293 million children globally. The highest prevalence of iron deficiency is in Africa (67.6%) and South-East Asia (65.5%). In addition, 468.4 million non-pregnant women are anaemic (global prevalence: 30.2%) with the highest prevalence also being in Africa (47.5%) *[23]*. Furthermore, it has been determined that the highest level of genetic diversity (both nuclear and mitochondrial) in the global human population is within the African population *[24-27]*. As such, it can be expected that the genetic risk for disease in African populations would be more complex due to the greater level of possible interactions between the larger amount of genetic diversity present and the numerous differential environmental factors that have an impact on iron absorption. As a first step to unravelling this in the African population, the main aim of this study was to replicate previously reported associations between single nucleotide polymorphisms (SNPs) and

low iron status or anaemia in a female black South African cohort. We additionally aimed to investigate gene-gene interactions that influence iron status. To assess this, we genotyped 32 SNPs previously reported to be involved in the systemic and cellular mechanism iron regulation, transport and storage (**Figure 1 and 2**).

Gut Lumen

Liver **HFE** HJV TFR1/TFR2

Fe^{3+} Fe^{2+}

HEME

DMT

HEME

HO-1

Enterocyte

Nucleus

Ferritin

HEPH

FPN

Fe^{3+} Fe^{2+}

Fe^{3+} binds to Tf

Hepcidin

TFR**2**

FPN

Ferritin

Nucleus

TFR1

Ferritin FPN

TFR

Hepcidin

Increased hepcidin release:
- *TMPRSS6*
- **HFE**-TFR**2**
- Infection-*IL16*,***TNF*** and ***NF-kB***

Macrophage
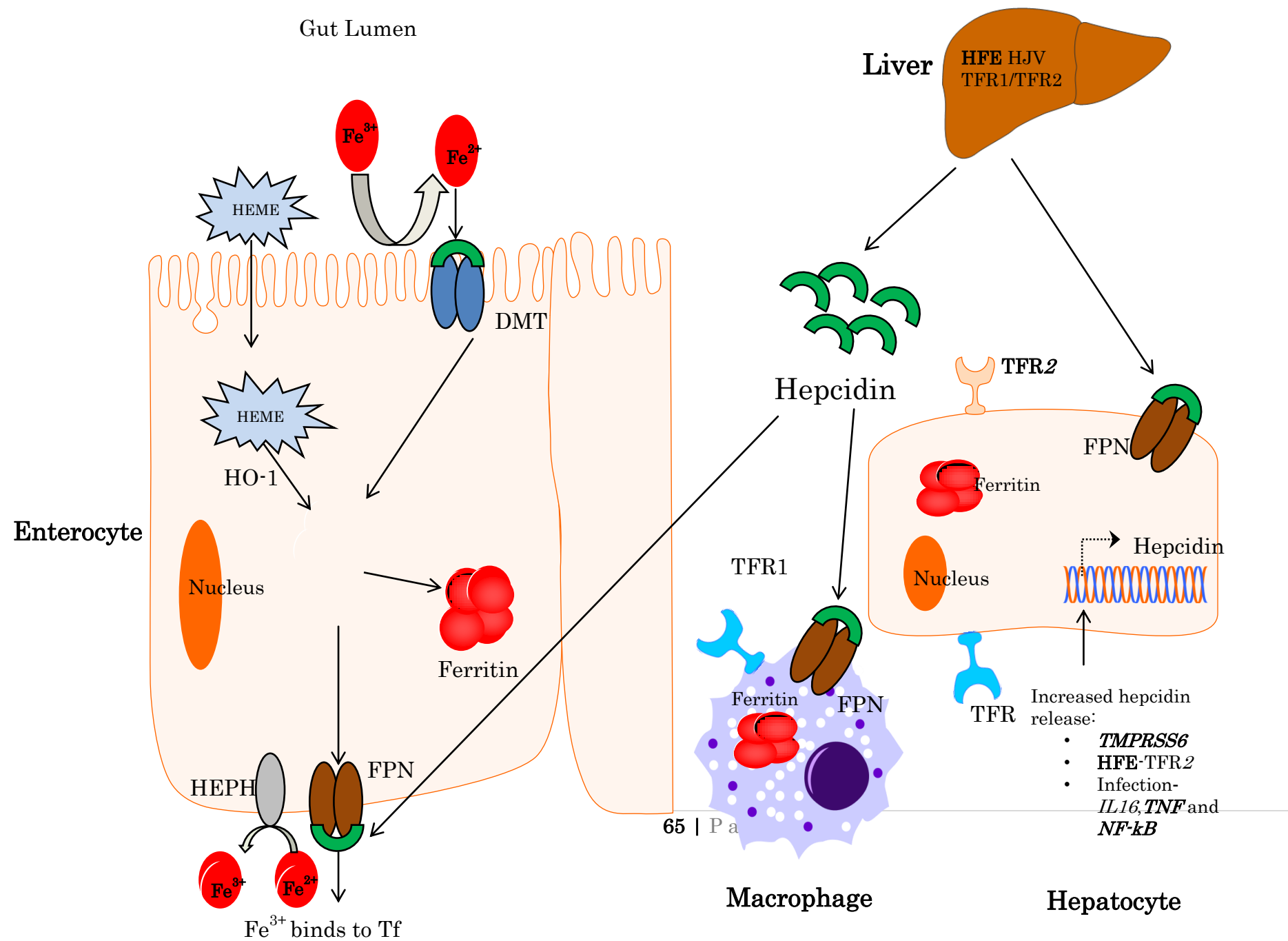
Hepatocyte

Figure 1. Iron metabolism in enterocytes, macrophages and hepatocytes

Dietary iron is present in the gut lumen in two forms - heme and non heme iron. Absorption takes place in the enterocytes. Non heme iron (Fe^{3+}) is reduced to Fe^{2+}. Fe^{2+} is then transported into the enterocytes through the divalent metal transporter 1 (DMT1). Based on the individual's iron status, iron in the enterocytes can be stored in ferritin or exported into plasma through ferroportin (FPN). Ferroportin is also involved in iron export from the macrophages and enterocytes. The exported iron is re-oxidized to Fe^{3+} by hephaestin (HEPH) before transport by transferrin (TF). In cases where transferrin within plasma is saturated, excess iron is stored within the liver. The liver plays a central role in iron metabolism due to its production and secretion of hepcidin. This hormone regulates iron absorption, recycling, and mobilization by binding to ferroportin and causing its internalization and degradation.
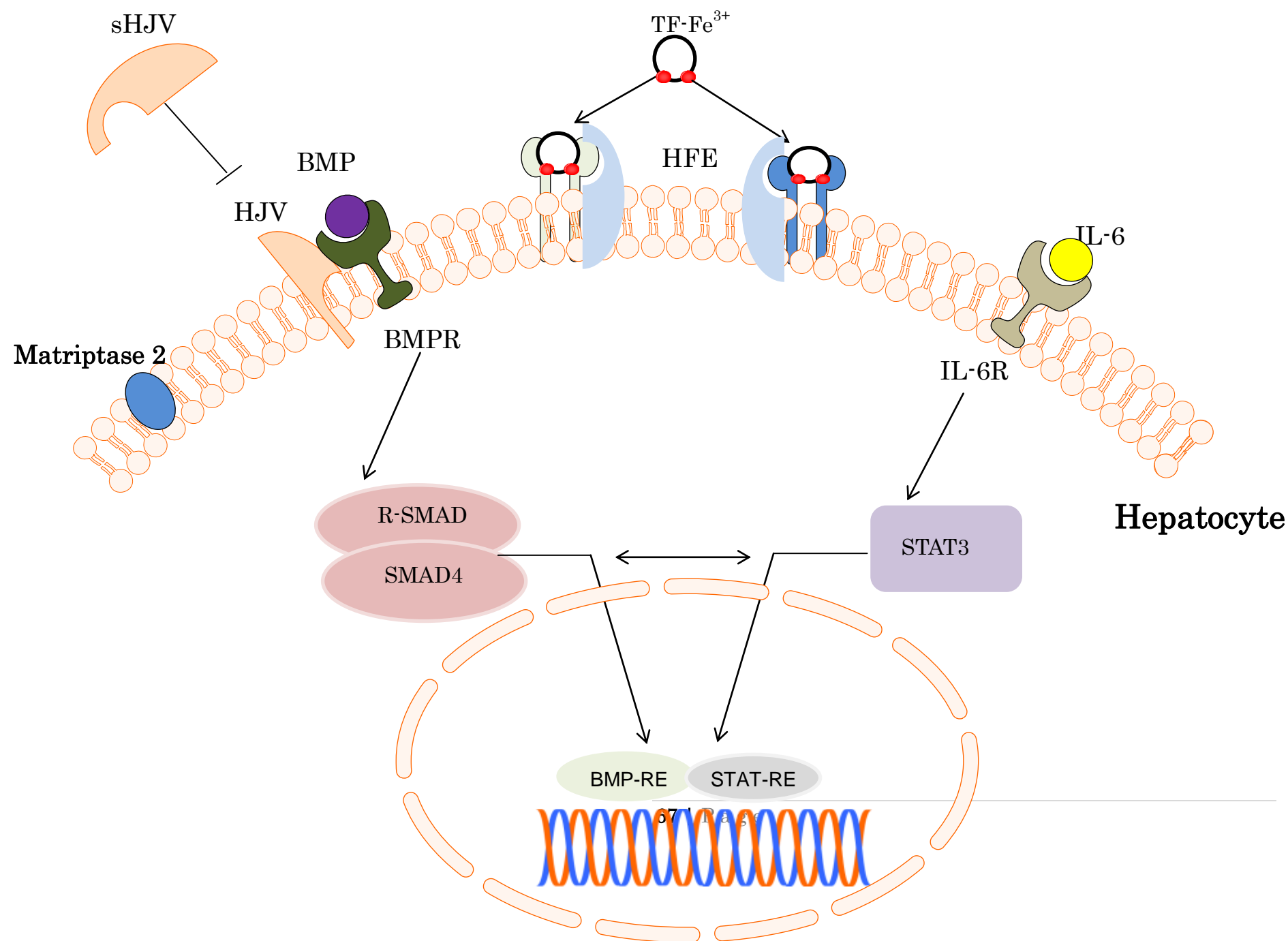
Figure 2. Regulation of hepatocyte hepcidin expression

The HAMP gene, which encodes the peptide hormone hepcidin, is activated by elevated circulating iron levels, increased body iron status as well as inflammatory cytokines such as IL-6.

Liver iron stores influence the hepatic expression of the extracellular signaling molecule bone morphogenetic protein (BMP) 6 (BMP-6). Interaction of BMP-6 with hepatocyte BMP receptors initiates intracellular signal transduction through SMAD proteins, increasing hepcidin transcription. BMP-6 signaling to hepcidin is heightened by cell-surface expression of the BMP co receptor hemojuvelin. The circulatory iron signal regulating hepcidin is via transferrin, which, on binding iron, serves as a ligand for two hepatocellular receptors: transferrin receptor 1 TfR1 and transferrin receptor 2 (TfR2). Ferri-transferrin–mediated signaling appears to be modulated by the physical interaction of these two receptors with the hemochromatosis protein HFE.

Under low iron conditions membrane bound HJV is cleaved by the protein matriptase-2 encoded by *TMPRSS6* to form soluble hemojuvelin (sHJV). Soluble hemojuvelin binds to BMP6 thus blocking the activation of BMP receptors. Inflammatory cytokines such as TNF and NFKBIL1 induce HAMP gene expression. NF-κB appears to be essential for the transcriptional activation of tumor necrosis factor (TNF) within monocytes. During hypoferremia or inflammation, inflammatory stimuli acting through TNFα suppreses HJV expression, thus inhibiting the signalling pathway from suppressing hepcidin.

Methods

Study population

The study population included in this analysis was a subset of the participants enrolled as part of the South African arm of the Prospective Urban and Rural Epidemiological (PURE) study and consisted of only women who are mainly of Tswana ancestry *[28]* as determined by self-report. Participants were only eligible for our sub study if they were apparently healthy, not pregnant, non-lactating and were not diagnosed with any chronic disease or taking drugs for any chronic disease at the time of blood collection. Additionally, we included only participants with all outcome measures determined and who were human immune deficiency virus (HIV) negative. The final sample size, after all inclusion and exclusion criteria were met, comprised of 686 individuals. Permission to conduct the larger PURE study had previously been obtained from the North West Department of Health, tribal chiefs, community leaders, employers and mayors as well as the Ethics Committee of the North–West University, Potchefstroom, South Africa (Approval number NWU-0073-10-A1). Approval for this sub study was also obtained from the Ethics Committee of the North–West University, Potchefstroom, South Africa. All participants gave written informed consent before being enrolled in the project and received pre-counselling regarding the HIV testing in groups of ten. Participants had to give special permission for the determination of HIV status as well as whether they wanted to receive the results of the test. Post-test counselling was undertaken individually with all participants that indicated that they wanted to receive their HIV status results.

Biochemical analyses

Fasting whole blood samples were collected into EDTA tubes. Fasting serum samples were similarly collected into serum tubes, after which the sample was allowed to clot for 30 minutes. Both types of tubes were in turn centrifuged at 2,000xg for 15 minutes at 10°C. Four iron status related outcomes were studied namely (Hb), serum ferritin (SF), serum transferrin receptor (sTfR) and body iron (BI) stores.

Haemoglobin was determined spectrophotometrically from whole blood using the Hemocue® HB201 haemoglobin meter (Hemocue AB, Angelholm Sweden). Whole blood was also used for the determination of HIV status making use of the First Response (PMC Medical, India) rapid HIV card test. If tested positive, the test was repeated with the Pareeshak (BHAT Bio-tech India) card test. SF and sTfR concentration were

determined quantitatively using an enzyme immunoassay procedure (Ramco Laboratories, Inc, Stafford Texas). The coefficient of variance (CV) for all assays was <10% *[29]*.

Body iron (BI, mg/kg), was assessed as follows: body iron (mg/kg)=–[log (serum transferrin receptor concentration/ferritin concentration)– 2.8229]/0.1207 as defined by Cook *et al.* *[30]*. C-reactive protein (CRP) was measured from serum via a particle-enhanced immunoturbidometric assay using Sequential Multiple Analyser Computer (SMAC) using the Konelab20iTM auto analyser (Thermo Fisher Scientific Oy, Vantaa, Finland).

Other measurements

Weight (kilograms) and height (meters) were determined by qualified level 2 accredited anthropometrists (International Society for the Advancement of Kinanthropometry (ISAK)). Standardized and calibrated equipment was used for measurements on volunteers wearing minimal clothing. Body mass index (BMI) was calculated by dividing weight in kilogram by height in meter squared.

Dietary intake was assessed by means of previously developed *[31]* and validated [32] culturally sensitive quantified food frequency questionnaires. These questionnaires were completed by the participants in collaboration with a trained field worker in the participant's preferred language. Nutrient intake was estimated by using the Foodfinder3® program (Medical Research Council, Tygerberg, South Africa) which is based on the South African Food composition tables *[33]*.

Genotyping and quality control procedures

DNA isolation was performed by using Qiagen® Flexigene™ DNA extraction kits according to the manufacturer's instructions. In order to set up a list of reference SNPs, we searched for SNPs that were previously reported to be involved in the systemic and cellular iron regulation, transport and storage. To this purpose we performed a literature search of HUGE navigator database.

All abstracts were scanned for appropriateness of use i.e. association or linkage of a SNP to the hematological parameters or key biomarkers of iron status (Hb, SF, sTfR and BI). The search strategy as well as final publications from which information was obtained are detailed in Gichohi-Wainaina et al [34]. If appropriate, the reference SNP

(rs) number of the polymorphism was recorded and included in the list to be sent to Illumina for verification of use in the array. Based on this, we selected 32 SNPs for genotyping. All 32 rs numbers were submitted to Illumina Technical Support for evaluation using the Assay Design Tool. Each SNP was scored (varying from 0 to 1) by the Assay Design Tool from Illumina based on its compatibility for use in a GoldenGate genotyping assay. SNPs with a score above 0.4 were chosen for genotyping. No SNP fell out based on these criteria. All 32 SNPs were determined using Illumina's VeraCode™ GoldenGate Genotyping Assay on a BeadXpress™ platform. Lack of a detectable signal after analysis was designated as a failed genotype call. SNP clustering was also assessed visually to determine success of genotyping. SNPs with a GenCall Score of >0.5 and samples with a call rate of ≥0.95 were included in the final analysis.

Information on MAFs from males and females of European (EUR) as well as African (AFR) ancestry was obtained from the 1000 Genomes website [35]. The minor allele frequencies of the SNPs from the females of the PURE population were then determined using the formula: MAF=minor allele count/total allele count

To investigate gene-gene interactions as well as interactions between SNPs, grouping of SNPs into the chromosomes within which they were located was undertaken. There were three chromosomes (chromosome 3, 6 and 22) with more than one genotyped SNP that were eligible for these analyses. Frequencies of allele combinations were determined by means of the software package Haploview version 4.2. Markers with a Hardy-Weinberg equilibrium p-value of ≥ 0.05 and a minor allele frequency of ≥ 1% were included in the analysis.

Statistical analysis

All outcome variables were determined to be parametric continuous quantitative traits. Natural logarithm transformations were applied to the serum ferritin and transferrin receptor concentration outcomes to improve their fit to the normal distribution. Potential disturbances in the distributions of the genotypes were tested for deviation from the assumptions of Hardy Weinberg Equilibrium (HWE) by using the chi-square test. Genotypes were coded as 0, 1 or 2 indicating the number of minor alleles. Differences in iron status across the genotype classes were assessed by using ANCOVA. In these analyses, ferritin and body iron concentrations were adjusted for inflammation by including C-reactive protein concentrations as a covariate in the model. Additionally, we adjusted for age, area of residence (rural or urban) and the proportion of dietary haem

iron intake to total iron intake in all association analyses. We did not adjust mean iron status values (per genotype) for BMI and menopausal status as there were negligible differences between the crude and adjusted models.

To investigate gene-gene interactions as well as interactions between SNPs and associations with iron status, cases of low iron stores were defined as serum ferritin concentration < 15 µg/L after correction for inflammation and controls as corrected serum ferritin concentration ≥15 µg/L. In terms of sTfR concentration, cases with iron deficiency were defined as those with sTfR> 8.3 mg/L (test kit reference value) and controls as ≤ 8.3mg/L. When body iron stores were considered as a marker of iron deficiency, cases were defined as body iron < 18 mg/kg and controls as body iron ≥ 18 mg/kg and excluding subjects with CRP> 5mg/L. Anaemia cases were defined at a Hb concentration <12 g/dl and controls at a Hb concentration of ≥12 g/dl *[35]*.

Statistical analyses were performed using version 20 of the Statistical Package for Social Sciences (SPSS Inc Chicago III). All reported P-values were two-tailed and adjustment for multiple testing was performed using the false discovery rate method. Statistical significance was defined as α < 0.05 while the q value cut off was set at 0.2 *[36]*.

Results

Following quality control analysis of the genotyping data, two SNPs i.e. rs198846 and rs5756504 failed due to low signal strength. Therefore, data was available for 30 SNPs in 12 genes for 686 individuals. **Table 1** presents the main characteristics of the study population. Included in the table are the numbers and proportion of participants that were rural. Fourteen percent (14%) of the population had a low serum ferritin concentration (<15 µg/L), whereas 61% had a sTfR concentration above the reference range (>8.3 mg/L). Anaemia was present in 8% of the study population. More than half (57%) of the study population was either overweight or obese and almost half (43%) of the study population had inflammation as indicated by a C-reactive protein concentration >5 mg/L.

The minor allele frequencies (MAF) determined in our study population were in most cases comparable with those of the males and females of the African ancestry populations investigated in the 1000 Genomes project (**Table 2**). We however observed differences in MAF at the rs482692 locus (4-fold) and rs2413450 locus (2- fold) in our

population compared to the males and females of the African ancestry in 1000 Genomes project. These differences were not as large as the differences in MAF observed between our study population and the European ancestry population in the 1000 Genomes Project. Differences in MAF between our study population and males and females of European ancestry in the 1000 Genomes project, were observed at the rs2071592 (MAF 0.63 vs. 0.30,) locus in the *NFKBIL1* gene, and the rs2413450 (MAF 0.09 vs. 0.43,), rs5756506 (MAF 0.81 vs. 0.40,) and rs855791 (MAF 0.09 vs. 0.40,) loci in the *TMPRSS6* gene (**Table 2**). Three SNPs i.e. rs2071592, rs1421312, and rs5756506, had MAFs > 0.5, indicating that the minor alleles at these loci were actually the major alleles in our study population.

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

Table 1: General characteristics and iron status indicators of the South African PURE Study female population [1]

	Rural *n* (%)	Total *n*	Value for Total N
Age, y	400 (59)	678	50 (49,50)
Dietary Intake			
Energy, kJ	384 (58)	658	7097 (6977,7229)
Iron, mg	384 (58)	658	12 (11,12)
Heme iron (% of total iron)	384 (58)	658	13 (13,14)
BMI, kg/m^2	400 (62)	645	27 (27,28)
Menopausal status, *n* (% menopausal)	392 (59)	664	369 (56)
BMI categories			
< 18 kg/m^2 (underweight), *n* (%)	399 (62)	644	66 (10)
18-25 kg/m^2 (normal weight), *n* (%)	399 (62)	644	213 (33)
25-30 kg/m^2 (overweight), *n* (%)	399 (62)	644	156 (24)
>30 kg/m^2 (obese), *n* (%)	399 (62)	644	209 (32)
Presence of inflammation			
Serum C-reactive protein (CRP), mg/L $^\$$	388 (59)	663	8.4 (8.0,8.8)
> 5 mg/L, *n* (%)	388 (59)	663	288 (43)
Iron status markers			
Serum transferrin receptor, mg/L $^\$$	400 (59)	678	9.2 (9.1,9.3)
> 8.3 mg/L, n (%)	400 (59)	678	414 (61)
Serum ferritin, μg/L $^\$$		686	84.5 ± 3.7
<15 μg/L, *n*, (%)	393 (59)	668	72 (11)
<15 μg//L), corrected for inflammation, *n* (%)$^\infty$.	388 (64)	603	60 (10)
<15 μg/L, excluding CRP > 5 mg/L, *n* (%)	222 (60)	369	53 (14)
Body iron store, mg/kg body weight	400 (59)	678	31.4 (31.2,31.6)
<18 mg/kg, n (%)	400 (59)	678	15 (2)
Hemoglobin, g/dl	301 (60)	474	13.8 (13.1,14.5)
< 12 g/dl (anaemia), n (%)	301 (60)	474	40 (8)

[1]Values are n (%) or the geometric mean (-1 SD, +1 SD) for the total sample.
$^\infty$Values adjusted for inflammation according to Cook et al [41]. BMI=Body mass index

Table 2: Minor allele frequency of SNPs of the South African PURE Study female population in comparison to African and European populations in the 1000 Genomes project

Locus/SNP	Chromosome	Base Pair	Minor/Major allele	MAF PURE	MAF 1000AFR	MAF 1000EUR
OPRD1						
rs482692[†]	1	29154255	A/G	0.18	0.04	0.13
SRPRB						
rs1830084	3	133508464	T/A	0.15	0.14	0.35
TF						
rs1358024[†]	3	133484188	A/G	0.01	0.01	0.19
rs1525892	3	133484712	A/G	0.29	0.25	0.35
rs1799852	3	133475722	A/G	0.03	0.05	0.14
rs1799899	3	133475812	A/G	0.00	0.01	0.08
rs1867504	3	133410661	A/G	0.28	0.23	0.49
rs3811647	3	133484029	A/G	0.28	0.20	0.35
rs3811658	3	133476852	A/G	0.15	0.13	0.35
rs4525863	3	133436136	A/C	0.17	0.27	0.33
rs7638018	3	133495461	G/A	0.15	0.15	0.35
rs8177248	3	133479626	A/G	0.05	0.08	0.35
TNF α						
rs1799964	6	31542308	G/A	0.19	0.17	0.22
rs1800629	6	31543031	A/G	0.17	0.10	0.14
HFE						
rs1800562	6	26093141	A/G	0.00	0.00	0.05
NFKBIL1						
rs2071592	6	31515340	A/T	0.63	0.51	0.30
TFR2						
rs7385804	7	100235970	C/A	0.28	0.33	0.39

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

Locus/SNP	Chromosome	Base Pair	Minor/Major allele	MAF PURE	MAF 1000AFR	MAF 1000EUR
CUBN						
rs10904850	10	16997707	A/G	0.29	0.11	0.33
HAMP						
rs10421768	19	35772899	G/A	0.15	0.19	0.23
KIAA1468						
rs9948708	18	59865121	A/G	0.45	0.44	0.44
KCTD17						
rs2235320	22	37453256	A/C	0.00	0.01	0.13
TMPRSS6						
rs1421312	22	37487810	G/A	0.68	0.58	0.41
rs2111833	22	37480797	A/G	0.27	0.37	0.38
rs228918	22	37506680	G/A	0.43	0.43	0.42
rs228921	22	37506876	G/A	0.43	0.43	0.43
rs2413450	22	37470224	A/G	0.09	0.18	0.43
rs4820268	22	37469591	G/A	0.18	0.31	0.44
rs5756506	22	37467392	C/G	0.81	0.80	0.40
rs5756520[†]	22	37508507	A/G	0.37	0.46	0.42
rs855791	22	37462936	A/G	0.09	0.13	0.40

†= SNP not in Hardy Weinberg Equilibrium. 1000AFR, 1000 Genomes African population; 1000EUR, 1000 genomes European population, *CUBN*, cubilin (intrinsic factor-cobalamin receptor; *HAMP*, hepcidin antimicrobial peptide; *HFE*, hemochromatosis; *KCTD17*, potassium channel tetramerization domain containing 17; *NFKBIL1*, MAF= minor allele frequency; nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1; *OPRD1*, opioid receptor, delta 1; PURE, PURE Study population; *SRPRB*, signal recognition particle receptor, B subunit; *TF*, transferrin; *TFR2*, transferrin receptor 2; *TMPRSS6*, transmembrane protease, serine 6; *TNF α*, tumor necrosis factor.

Table 3 displays the significant associations observed between SNPs investigated and iron status parameters. These significant associations were only determined between SNPs in the TF gene and SF, sTfR and BI. None of the SNPs in TF were significantly associated with Hb. None of the SNPs in other genes were significantly associated with Hb and iron status markers (**Supplemental Tables 1-3**). The heterozygous genotype class of rs1799852 in the *TF* gene, with a MAF of 0.03 in our study population, presented with lower serum ferritin concentration (AG vs. GG; 54.0 µg/L vs 87.9 µg/L, P=0.02) and body iron stores (29.4 mg/kg vs. 31.6 mg/kg, P=0.005), and higher serum transferrin concentrations (10.2 mg/L vs 9.1 mg/L, P=0.03) as compared to the homozygous wild type genotype class. Only the association with body iron stores remained significant after adjustment using the false discovery rate method (FDR). For rs3811647 in the same gene, the MAF was 0.28 and mean serum concentration of transferrin receptor in both homozygous groups as compared to the heterozygote group were higher (9.4 in each case vs. 8.8 mg/L P=0.01) and mean body iron store was lower (31.0/31.1 vs. 32.1 mg/kg, P=0.03). In a multiple linear regression model, rs1799852 and rs3811647 explained 13% of the variation in sTfR concentrations after adjustment for age, area of residence, and proportion of haem iron intake.

Table 3: Genotype distribution of SNPs in the *TF* gene and significant associations with iron status measures in the South African PURE Study female population [1,2,3]

Outcome Measure	SNP	*P*[4]	Genotype	N	Mean (95% CI)
Serum ferritin, µg/L	rs1799852	0.01	GG	603	87.9 (79.8, 96.8)
			AG	46	54.0 (38.0, 76.7)
			AA	0	-
Serum transferrin receptor, mg/L	rs1799852	0.03	GG	619	9.1 (8.8, 9.3)
			AG	46	10.2 (9.2, 11.3)
			AA	0	-
	rs3811647	0.03	GG	355	9.4 (9.1, 9.8)
			AG	249	8.8 (8.4, 9.1)
			AA	61	9.4 (8.6, 10.2)
Body iron stores, mg/kg	rs1799852	0.004*	GG	619	31.6 (31.2, 32.0)
			AG	46	29.4 (27.9, 30.8)
			AA	0	-
	rs3811647	0.03	GG	355	31.1 (30.5, 31.5)
			AG	249	32.1 (31.5, 32.7)
			AA	61	31.1 (29.8, 32.4)

[1]For each SNP, top to bottom rows represent values for the homozygote wild type, heterozygotes and homozygote minor allele genotypes respectively. [2]All associations are adjusted for age, area of residence (rural or urban), ratio of heme iron intake. [3]Associations involving serum ferritin and body iron are additionally adjusted for serum CRP. [4]*P* refers to the p value of the SNP association with the corresponding outcome measure. *Polymorphism is significantly associated with outcome measure after adjustment using the false discovery rate (FDR method) (q-values cut off =0.20). *TF*, transferrin.

Gene-gene interaction analyses revealed SNP combinations in chromosome 6 consisting of rs1799964 and rs1800629 in the tumour necrosis factor (*TNF) gene*, and rs2071592 in the *NFKBIL1 gene*, of which the AAA variant was associated with higher odds for low serum ferritin concentrations (serum ferritin<15µg/L; OR:1.86 (95%-CI, 1.23-2.79)) as indicated in **Table 4**). The chromosome 22 SNP allele combination (GG) consisting of rs228918 and rs228921 in the *TMPRSS6* gene was associated with lower odds for increased sTfR concentrations (sTfR >8.3mg/L; OR:0.79 (95%-CI,0.63-0.98).

Table 4: Allele combinations associated with iron deficiency within PURE Study female population [1,2,3]

Chr	SNPs	Allele combinations	Low SF SF<15 µg/L		High sTfR sTfR>8.3 mg/L		Body iron stores		Anemia Hb<12g/dL	
			Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)	Frequency	OR (95% CI)
6	rs1799964; rs1800629; rs2071592	AGA	0.46	1 (ref)	0.46	1 (ref)	0.46	1 (ref)	0.46	1 (ref)
		GGT	0.19	0.80 (0.54,1.16)	0.19	0.92 (0.70,1.21)	0.19	0.69 (0.28,1.69)	0.19	0.67 (0.37,1.22)
		AGT	0.17	0.83 (0.56,1.24)	0.17	0.94 (0.71,1.25)	0.17	0.69 (0.28,1.69)	0.17	1.30 (0.71, 2.38)
		AAA	0.17	1.86 (1.23, 2.79)*	*0.17*	1.22 (0.91,1.63)	0.17	2.37 (0.96, 5.88)	0.17	1.19 (0.62, 2.26)
22	rs2111833; rs228918	GG	0.73	1 (ref)	-	-	-	-	-	-
		AA	0.27	1.08 (0.74, 1.56)	-	-	-	-	-	-

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

Chr	SNPs	Genotype	Freq	OR (CI)	Freq	OR (CI)	Freq	OR (CI)	Freq	OR (CI)
	rs228918; rs2413450	GG	0.91	1 (ref)	-	-	-	-	-	-
		AA	0.09	0.82 (0.46,1.47)	-	-	-	-	-	-
	rs228918; rs228921	AA	-	-	0.57	1 (ref)	0.56	1 (ref)	0.58	1 (ref)
		GG	-	-	0.43	0.79 (0.63, 0.98)$	0.44	1.12 (0.55, 2.27)	0.42	1.11 (0.70, 1.77)
3	rs3811647; rs3811658	GG	0.72	1 (ref)	0.72	1 (ref)	0.72	1 (ref)	0.72	1 (ref)
		AA	0.15	0.86 (0.56,1.33)	0.15	0.94 (0.69,1.27)	0.15	0.61 (0.21,1.76)	0.15	0.81 (0.43, 1.53)
		AG	0.13	0.77 (0.49,1.19)	0.13	0.82 (0.60,1.13)	0.13	0.75 (0.26, 2.15)	0.13	0.92 (0.49, 1.74)
	rs7638018; rs81772248	AG	0.86	1 (ref)	0.86	1 (ref)	0.86	1 (ref)	-	-
		GG	0.09	1.13 (0.68,1.87)	0.09	0.86 (0.56,1.33)	0.09	0.96 (0.33, 2.76)	-	-
		GA	0.05	0.70 (0.36,1.39)	0.05	0.77 (0.49,1.19)	0.05	1.05 (0.25, 4.48)	-	-
	rs1525892; rs1830084	GT	-	-	-	-	-	-	0.68	1 (ref)
		AT	-	-	-	-	-	-	0.16	0.81 (0.43, 1.53)
		AA	-	-	-	-	-	-	0.15	0.92 (0.49, 1.74)

Cases and controls are defined as: [1]corrected serum ferritin concentration <15 µg/L (cases) and ≥15 µg/L and serum CRP <5 mg/L (controls); [2]serum transferrin receptor >8.3 mg/L (cases) and ≤ 8.3mg/L (controls); or [3]body iron store <18 mg/kg (cases) and ≥18 mg/kg (controls). *Allele combination is associated with higher odds for low serum ferritin concentrations. $ Allele combination is associated with lower odds for increased sTfR concentrations. BI, body iron stores; Chr, chromosome; SF, serum ferritin; SNPs, single nucleotide polymorphisms; sTfR, serum transferrin receptor.

Discussion

We report here the first study focusing on the genetic aetiology of iron status in a population of adult black South African women. The most striking finding from our study is the general lack of associations between the investigated SNPs and iron status indicators in our study population, whereas others have reported such associations in European ancestry and Asian populations. For example, rs855791 in the *TMPRSS6* gene has repeatedly been reported to be associated with lower serum ferritin concentrations and anaemia *[15, 17, 18, 21, 22, 37, 38]*. We did not find such an association, and the SNP was also hardly present in the females in our study population with a MAF of 0.09 compared to a MAF of ~0.40 in males and females of European ancestry. Our cohort of women had lower iron status compared to the European ancestry populations previously investigated. It would thus imply that all factors being equal, we should be able to detect associations in our cohort with lower numbers than previous GWA studies. However the low MAFs of the risk alleles in our cohort would require large numbers to detect associations.

We observed several other significant differences in MAF of investigated SNPs between our female study population and the male and female Europeans included in the 1000 Genomes project. The observed differences may be due to certain environmental conditions acting through selective pressure to alter the frequency of the genetic variants in the European population *[39]*. This has resulted in specific allele frequencies in the European ancestry population that are different from those in the African population. These differences in MAF between populations could theoretically result in dissimilarities in the prevalence of a disease outcome *[28]*, in this case iron deficiency, and could in case of strong gene-disease associations call for population specific strategies to address this. MAFs of the females in our study population on the other hand were similar to those of the males and females in the African population in the 1000 Genomes Project. The similarity is probably because MAFs reflected in the 1000 genomes project represent a wide variation of African populations and would therefore capture the great diversity within the population *[34]*.

We did, however, observe some significant associations between SNPs and iron status, particularly in the *TF* gene. The *TF* gene encodes for transferrin, a glycoprotein that binds free iron in blood plasma. On the cell surface, ferric transferrin binds to TfR1 to allow iron entry into the cell. Additionally, TfR2 helps sense and regulate circulating

iron levels in the body by controlling the levels of hepcidin. This feedback mechanism between transferrin and transferrin receptor 1 may explain why SNPs in the *TF* gene could be associated with sTfR. Genetic variants within *TF* have previously been studied for their association with iron status *[16, 40-44]*. The A allele of rs1799852, a non-synonymous coding SNP in exon 17, has been associated with a reduction in serum transferrin concentrations (and thus indirectly higher sTfR concentrations) *[16, 45]*. In our study we observed an association between this SNP and increased sTfR and lower body iron stores, which is in line with the expectations. However, the SNP was rare in our study population (MAF, 0.03) and therefore the importance of this finding from a public health perspective is limited.

Heterozygous individuals at the rs3811647 locus (AG), which is located in intron 11 of the TF gene had the lowest mean sTfR concentrations compared to those harbouring the homozygous wildtype (GG) or mutant genotypes (AA). Higher serum transferrin concentrations in the AA genotype group have previously been observed in a group of menstruating women *[45]*. In another study, women that harboured the A allele had lower transferrin saturation, which may indicate a reduction in iron transport to tissues. In order to explain this discrepancy, larger studies are required to confirm this finding. Additionally, it would be necessary to conduct further mechanistic and clinical studies.

In a study by Benyamin et al *[16]*, three SNPs that included rs1800562 (in the HFE gene), and both rs1799852 and rs3811647 in the *TF* gene together explained 40% of the genetic variation in sTfR based on analyses performed on an Australian population of European decent. In our African population, 13% of the variation in sTfR could be explained by the two SNPs in the *TF* gene. The lower explanation of variance in our study may be caused by the low MAF of rs1799852 in our study population. Additionally this could indicate that these may not be the most important SNPs explaining variation in sTfR in this African population.

Gene-gene interaction analyses further identified a SNP allele combination i.e. AAA (frequency 0.17) consisting of rs1799964, rs1800629, and rs2071592, which was associated with higher odds for low serum ferritin concentrations (SF<15μg/L) as well as for low body iron stores (BI <18 mg/kg). This allele combination has previously been associated with an increased risk of iron deficiency in a cohort of West African children *[46]*. The SNPs in the identified combination are present in the tumour necrosis factor (rs1799964 and rs1800629) and *NFKBIL1* (rs2071592) genes, which are both involved in inflammatory processes. The rs1800629 SNP increases *TNF α* gene transcription *[47]*,

which would block recycling of macrophage iron [48], inhibit erythropoiesis *[49]* and has been suggested to reduce absorption of dietary iron *[50]*. The *NFKBIL1* gene in turn is hypothesized to exert an influence on the expression of the *TNF α* gene, leading to higher circulating concentrations of tumour necrosis factor *[51]*. The SNP rs1800629 has previously been associated with iron deficiency as part of a haplotype together with other *TNF α* SNPs *[46]*, as well as with an increased risk of severe anaemia in low-birth-weight children in Kenya *[52]* and increased risk of severe malarial anaemia in Gambian children *[53]*. It has been suggested that this haplotype would only influence iron status under pressure of infectious diseases *[46]*. However, our study was not conducted in a malaria endemic area and comprised of apparently healthy adult population. Nevertheless, although we excluded subjects with diagnosed HIV infection, there may still have been undetected infected individuals in the study population. In addition, over half of the subjects were either overweight or obese, which induces low grade inflammation. Indeed, 43% of the subjects had raised C-reactive protein concentrations, and therefore disease induced inflammatory processes may underlie our findings.

We observed a combination of single nucleotide polymorphisms consisting of rs228918 and rs228921 in the *TMPRSS6* gene that was associated with lower odds for increased sTfR concentrations (sTfR>8.3mg/L), reflecting low iron status. Both SNPs have previously been associated with increased haemoglobin concentrations *[18]*. Genome wide association studies have convincingly proven that single nucleotide polymorphisms (SNPs) in *TMPRSS6* are associated with iron parameters *[37, 53]*, especially iron and transferrin saturation but the combinations in the present study have not been described before. The exact mechanism by which *TMPRSS6* and its encoded protein matriptase-2 action is exerted is still under investigation. One study proposed that matriptase-2 regulates hepcidin expression by cleaving HJV to decrease BMP-SMAD signalling *[54, 55]*. However, two other studies suggest that the influence of these SNPs is not mediated by hepcidin *[56, 57]*. Further studies are required to ascertain the role of hepcidin in *TMPRSS6*-iron status parameter associations.

The strength of our study is the inclusion of potential effect-modifiers as co-variates in the association analyses, such as haem iron intake, age, inflammation and area of residence. This is not commonly done in association studies, but strengthens the confidence in our findings. Moreover, haemoglobinopathies, which are commonly prevalent in East and Central Africa, are uncommon in the black South African population *[58, 59]* and therefore do not distort the findings.

We observed a considerable discrepancy in the prevalence of low iron status when estimated using SF (14%) or sTfR (61%). This may be due to the high prevalence of inflammation, which increases SF concentrations as an acute phase protein. Despite correction for high CRP concentrations, undetected late convalescence of inflammation may still have impacted SF concentrations in our study population. Overweight-related low-grade inflammation may also have played a role [60]. In the presence of inflammation, transferrin receptor concentration is usually regarded as a better indicator of iron deficiency [61], but can also be raised due to certain pathologic conditions [62]. An alternative to assess iron deficiency in this case would be a bone marrow examination (BME) that establishes the absence of stainable iron in the bone marrow. BME is, however, invasive, expensive, and requires technical expertise, so that it cannot be performed routinely in a field setting. This dilemma stresses once more the need to develop an appropriate biomarker to estimate iron deficiency in the presence of inflammation.

In conclusion, we determined several genetic variants and single nucleotide polymorphism combinations in the *TF, TNF α* and *TMPRSS6* genes to be associated with low iron status in a study population of black South African women. However, many of the associations between SNPs and iron status previously described for other populations were not replicated in this cohort. This may be explained both by heterogeneity in genetic makeup as well as environmental differences between populations, especially the extent of inflammation. This calls for population specific genome studies rather than inferring genetic data across populations. Discovery and validation of additional genetic variants associated with iron status are likely to uncover new mechanisms and pathways that affect iron status, offering better insights into iron deficiency especially in relation to infectious diseases and inflammatory processes in African populations.

Acknowledgements

Finally we thank the participants of the PURE study as well as the team members of the PURE-SA study without whom this work would not have been possible.

References

1. Khan KS, Wojdyla D, Say L, Gulmezoglu AM, Van Look PF: **WHO analysis of causes of maternal death: a systematic review.** *Lancet* 2006, **367**:1066-1074.

2. Brabin BJ, Hakimi M, Pelletier D: **An analysis of anemia and pregnancy-related maternal mortality.** *J Nutr* 2001, **131**:604S-614S; discussion 614S-615S.

3. Perez EM, Hendricks MK, Beard JL, Murray-Kolb LE, Berg A, Tomlinson M, Irlam J, Isaacs W, Njengele T, Sive A, Vernon-Feagans L: **Mother-infant interactions and infant development are altered by maternal iron deficiency anemia.** *J Nutr* 2005, **135**:850-855.

4. Haas JD, Brownlie Tt: **Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship.** *J Nutr* 2001, **131**:676S-688S; discussion 688S-690S.

5. Brownlie Tt, Utermohlen V, Hinton PS, Giordano C, Haas JD: **Marginal iron deficiency without anemia impairs aerobic adaptation among previously untrained women.** *Am J Clin Nutr* 2002, **75**:734-742.

6. Brutsaert TD, Hernandez-Cordero S, Rivera J, Viola T, Hughes G, Haas JD: **Iron supplementation improves progressive fatigue resistance during dynamic knee extensor exercise in iron-depleted, nonanemic women.** *Am J Clin Nutr* 2003, **77**:441-448.

7. S Horton JR: **The Economics of Iron Deficiency** *Food Policy* 2003:pp. 51–75.

8. Leboeuf RC, Tolson D, Heinecke JW: **Dissociation between tissue iron concentrations and transferrin saturation among inbred mouse strains.** *J Lab Clin Med* 1995, **126**:128-136.

9. McLaren CE, Li KT, Gordeuk VR, Hasselblad V, McLaren GD: **Relationship between transferrin saturation and iron stores in the African American and US Caucasian populations: analysis of data from the third National Health and Nutrition Examination Survey.** *Blood* 2001, **98**:2345-2351.

10. Morse AC, Beard JL, Jones BC: **A genetic developmental model of iron deficiency: biological aspects.** *Proc Soc Exp Biol Med* 1999, **220**:147-152.

11. Pilia G, Chen WM, Scuteri A, Orru M, Albai G, Dei M, Lai S, Usala G, Lai M, Loi P, et al: **Heritability of cardiovascular and personality traits in 6,148 Sardinians.** *PLoS Genet* 2006, **2**:e132.

12. Marroni F, Grazio D, Pattaro C, Devoto M, Pramstaller P: **Estimates of genetic and environmental contribution to 43 quantitative traits support sharing of a homogeneous environment in an isolated population from South Tyrol, Italy.** *Hum Hered* 2008, **65**:175-182.

13. Njajou OT, Alizadeh BZ, Aulchenko Y, Zillikens MC, Pols HA, Oostra BA, Swinkels DW, van Duijn CM: **Heritability of serum iron, ferritin and transferrin saturation in a genetically isolated population, the Erasmus Rucphen Family (ERF) Study.** *Hum Hered* 2006, **61**:222-228.

14. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, Duffy DL, Martin NG: **Effects of HFE C282Y and H63D polymorphisms and polygenic background on iron stores in a large community sample of twins.** *Am J Hum Genet* 2000, **66**:1246-1258.

15. Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, et al: **Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume.** *Nat Genet* 2009, **41**:1173-1175.

16. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, Peltonen L, Martin NG, Montgomery GW, Whitfield JB, Visscher PM: **Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels.** *Am J Hum Genet* 2009, **84**:60-65.

17. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, et al: **Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium.** *Nat Genet* 2009, **41**:1191-1198.

18. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, et al: **Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels.** *Nat Genet* 2009, **41**:1170-1172.

19. Soranzo N, Sanna S, Wheeler E, Gieger C, Radke D, Dupuis J, Bouatia-Naji N, Langenberg C, Prokopenko I, Stolerman E, et al: **Common variants at 10 genomic loci influence hemoglobin A(1)(C) levels via glycemic and nonglycemic pathways.** *Diabetes* 2010, **59**:3229-3239.

20.   Lok CY, Merryweather-Clarke AT, Viprakasit V, Chinthammitr Y, Srichairatanakool S, Limwongse C, Oleesky D, Robins AJ, Hudson J, Wai P, et al: **Iron overload in the Asian community.** *Blood* 2009, **114**:20-25.

21.   An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, Zhou D, Song P, Wang C, Meng L, 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.

22.   Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, Jing L, Yu Y, Ruan S, Xie D, et al: **Association of** *TMPRSS6* **polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese Han population.** *Am J Clin Nutr* 2012, **95**:626-632.

23.   WHO: **Worldwide prevalence of anaemia 1993–2005.** 2008.

24.   Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC: **Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups.** *Am J Hum Genet* 1995, **57**:133-149.

25.   Jorde LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, Seielstad MT, Batzer MA: **The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data.** *Am J Hum Genet* 2000, **66**:979-988.

26.   Garrigan D, Kingan SB, Pilkington MM, Wilder JA, Cox MP, Soodyall H, Strassmann B, Destro-Bisol G, de Knijff P, Novelletto A, et al: **Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data.** *Genetics* 2007, **177**:2195-2207.

27.   Long JC, Li J, Healy ME: **Human DNA sequences: more variation and less race.** *Am J Phys Anthropol* 2009, **139**:23-34.

28.   Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S, Group PI-W: **The Prospective Urban Rural Epidemiology (PURE) study: examining the impact of societal influences on chronic noncommunicable diseases in low-, middle-, and high-income countries.** *Am Heart J* 2009, **158**:1-7 e1.

29.   Aderibigbe OR, Pisa PT, Mamabolo RL, Kruger HS, Vorster HH: **The relationship between indices of iron status and selected anthropometric cardiovascular disease risk markers in an African population: the THUSA study.** *Cardiovasc J Afr* 2011, **22**:249-256.

30.   Cook JD, Flowers CH, Skikne BS: **The quantitative assessment of body iron.** *Blood* 2003, **101**:3359-3364.

31. MacIntyre UE, Venter CS, Vorster HH: **A culture-sensitive quantitative food frequency questionnaire used in an African population: 1. Development and reproducibility.** *Public Health Nutr* 2001, **4:**53-62.

32. MacIntyre UE, Venter CS, Vorster HH: **A culture-sensitive quantitative food frequency questionnaire used in an African population: 2. Relative validation by 7-day weighted records and biomarkers.** *Public Health Nutr* 2001, **4:**63-71.

33. Langenhoven M, Kruger, M., Gouws, E. & Faber, M., Medical Research Council **Food Composition Tables** 1998.

34. 1000 genomes project website (Accessed on January 10, 2014) http://browser.1000 Genomes Projectenomes.org/index.html

35. WHO: **Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity** 2011.

36. Smith NL, Hindorff LA, Heckbert SR, Lemaitre RN, Marciante KD, Rice K, Lumley T, Bis JC, Wiggins KL, Rosendaal FR, Psaty BM: **Association of genetic variations with nonfatal venous thrombosis in postmenopausal women.** *JAMA* 2007, **297:**489-498.

37. Delbini P, Vaja V, Graziadei G, Duca L, Nava I, Refaldi C, Cappellini MD: **Genetic variability of *TMPRSS6* and its association with iron deficiency anaemia.** *Br J Haematol* 2010, **151:**281–284.

38. 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.**

39. Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admetlla A, Pattini L, Nielsen R: **Signatures of environmental genetic adaptation pinpoint pathogens as the main selective pressure through human evolution.** *PLoS Genet* 2011, **7:**e1002355.

40. Lee PL, Halloran C, Beutler E: **Polymorphisms in the transferrin 5' flanking region associated with differences in total iron binding capacity: possible implications in iron homeostasis.** *Blood Cells Mol Dis* 2001, **27:**539-548.

41. Lee PL, Halloran C, Trevino R, Felitti V, Beutler E: **Human transferrin G277S mutation: a risk factor for iron deficiency anaemia.** *Br J Haematol* 2001, **115:**329-333.

42. Lee PL, Ho NJ, Olson R, Beutler E: **The effect of transferrin polymorphisms on iron metabolism.** *Blood Cells Mol Dis* 1999, **25:**374-379.

43.   Sarria B, Navas-Carretero S, Lopez-Parra AM, Perez-Granados AM, Arroyo-Pardo E, Roe MA, Teucher B, Vaquero MP, Fairweather-Tait SJ: **The G277S transferrin mutation does not affect iron absorption in iron deficient women.** *Eur J Nutr* 2007, **46:**57-60.

44.   Delanghe J, Verstraelen H, Pynaert I, Debels L, Taes Y, Verhasselt B, De Henauw S, Temmerman M: **Human transferrin G277S mutation and iron deficiency in pregnancy.** *Br J Haematol* 2006, **132:**249-250.

45.   Blanco-Rojo R, Baeza-Richer C, Lopez-Parra AM, Perez-Granados AM, Brichs A, Bertoncini S, Buil A, Arroyo-Pardo E, Soria JM, Vaquero MP: **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.

46.   Atkinson SH, Rockett KA, Morgan G, Bejon PA, Sirugo G, O'Connell MA, Hanchard N, Kwiatkowski DP, Prentice AM: **Tumor necrosis factor SNP haplotypes are associated with iron deficiency anemia in West African children.** *Blood* 2008, **112:**4276-4283.

47.   Gonzalez S, Rodrigo L, Martinez-Borra J, Lopez-Vazquez A, Fuentes D, Nino P, Cadahia V, Saro C, Dieguez MA, Lopez-Larrea C: **TNF-alpha -308A promoter polymorphism is associated with enhanced TNF-alpha production and inflammatory activity in Crohn's patients with fistulizing disease.** *Am J Gastroenterol* 2003, **98:**1101-1106.

48.   Alvarez-Hernandez X, Liceaga J, McKay IC, Brock JH: **Induction of hypoferremia and modulation of macrophage iron metabolism by tumor necrosis factor.** *Lab Invest* 1989, **61:**319-322.

49.   Moldawer LL, Marano MA, Wei H, Fong Y, Silen ML, Kuo G, Manogue KR, Vlassara H, Cohen H, Cerami A, et al.: **Cachectin/tumor necrosis factor-alpha alters red blood cell kinetics and induces anemia in vivo.** *FASEB J* 1989, **3:**1637-1643.

50.   Johnson D, Bayele H, Johnston K, Tennant J, Srai SK, Sharp P: **Tumour necrosis factor alpha regulates iron transport and transporter expression in human intestinal epithelial cells.** *FEBS Lett* 2004, **573:**195-201.

51.   Schutze S, Wiegmann K, Machleidt T, Kronke M: **TNF-induced activation of NF-kappa B.** *Immunobiology* 1995, **193:**193-203.

52.   Aidoo M, McElroy PD, Kolczak MS, Terlouw DJ, ter Kuile FO, Nahlen B, Lal AA, Udhayakumar V: **Tumor necrosis factor-alpha promoter variant 2 (TNF2) is**

associated with pre-term delivery, infant mortality, and malaria morbidity in western Kenya: Asembo Bay Cohort Project IX. *Genet Epidemiol* 2001, **21**:201-211.

53. McGuire W, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D: **Variation in the TNF-alpha promoter region associated with susceptibility to cerebral malaria.** *Nature* 1994, **371**:508-510.

54. Meynard D, Babitt JL, Lin HY: **The liver: conductor of systemic iron balance.** *Blood* 2014, **123**:168-176.

55. Fleming RE, Ponka P: **Iron overload in human disease.** *N Engl J Med* 2012, **366**:348-359.

56. Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FC, Fleming RE, Kiemeney LA, Vermeulen SH, Swinkels DW: **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**:593-598.

57. Traglia M, Girelli D, Biino G, Campostrini N, Corbella M, Sala C, Masciullo C, Vigano F, Buetti I, Pistis G, et al: **Association of HFE and *TMPRSS6* genetic variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations.** *J Med Genet* 2011, **48**:629-634.

58. Wonkam A, Ponde C, Nicholson N, Fieggen K, Ramessar R, Davidson A: **The burden of sickle cell disease in Cape Town.** *S Afr Med J* 2012, **102**:752-754.

59. Williams TN, Weatherall DJ: **World distribution, population genetics, and health burden of the hemoglobinopathies.** *Cold Spring Harb Perspect Med* 2012, **2**:a011692.

60. Cepeda-Lopez AC, Osendarp SJ, Melse-Boonstra A, Aeberli I, Gonzalez-Salazar F, Feskens E, Villalpando S, Zimmermann MB: **Sharply higher rates of iron deficiency in obese Mexican women and children are predicted by obesity-related inflammation rather than by differences in dietary iron intake.** *Am J Clin Nutr* 2011, **93**:975-983.

61. Dao MC, Meydani SN: **Iron biology, immunology, aging, and obesity: four fields connected by the small peptide hormone hepcidin.** *Adv Nutr* 2013, **4**:602-617.

62. Karlsson T: **Pathologic conditions associated with elevated plasma soluble transferrin receptor levels in elderly iron replete anemic patients.** *Int J Hematol* 2010, **91**:723-724.

Supplemental Appendices

Supplemental Table 1. Single nucleotide polymorphisms and genotype distribution in chromosome 22 and associations with iron status indicators within the PURE Study female population [1,2,3,4]

Locus	SNP	Genotype	n	Hb Mean (95% CI)	P	n	SF Mean (95% CI)	P	n	sTfR Mean (95% CI)	P	n	BI Mean (95% CI)	P
TMPRSS6	rs1421312	AA	59	13.8 (13.4, 14.3)	0.88	80	78.3 (60.1, 103)	0.68	81	8.8 (8.1, 9.4)	0.42	81	31.3 (30.2, 32.4)	0.76
		AG	191	13.9 (13.7, 14.1)		266	88.9 (76.2, 103)		274	9.2 (8.8, 9.6)		274	31.6 (31.0, 32.2)	
		GG	220	13.8 (13.6, 14.0)		302	83.3 (72.6, 95.6)		309	9.3 (8.9, 9.6)		309	31.3 (30.7, 31.9)	
TMPRSS6	rs2111833	GG	256	13.8 (13.6, 14.0)	0.72	344	86.3 (75.9, 98.2)	0.50	352	9.1 (8.8, 9.5)	0.96	352	31.5 (30.9, 32.0)	0.56
		AG	182	13.9 (13.6, 14.1)		259	80.6 (69.5, 93.5)		265	9.2 (8.8, 9.6)		265	31.2 (30.6, 31.8)	
		AA	33	14.0 (13.5, 14.6)		46	100 (70.5, 143)		48	9.1 (8.2, 10.0)		48	32.1 (30.6, 33.5)	
KCDT17	rs2235320	CC	468	13.8 (13.7, 14.0)	0.96	644	85.3 (77.7, 93.7)	0.11	660	9.2 (8.9, 9.4)	0.38	660	31.4 (31.0, 31.8)	0.08
		AC	2	13.8 (11.5, 16.0)		3	27.4 (6.9, 109)		3	10.9 (7.4, 16.3)		3	26.6 (20.9, 32.3)	
		AA	0	-		0	-		0	-		0	-	
TMPRSS6	rs228918	AA	169	13.8 (13.6, 14.1)	0.82	214	83.9 (71.3, 98.9)	0.73	222	9.3 (8.9, 9.7)	0.69	222	31.3 (30.6, 32.0)	0.68
		AG	210	13.8 (13.6, 14.0)		302	88.3 (77.0, 101)		308	9.1 (8.7, 9.4)		308	31.6 (31.0, 32.2)	
		GG	92	13.9 (13.6, 14.3)		132	80.2 (65.0, 98.8)		134	9.2 (8.6, 9.7)		134	31.2 (30.3, 32.1)	
TMPRSS6	rs228921	AA	169	13.8 (13.6, 14.1)	0.82	214	84.0 (71.3, 99.0)	0.77	222	9.3 (8.9, 9.7)	0.70	222	31.3 (30.6, 32.0)	0.72
		AG	210	13.8 (13.6, 14.0)		303	87.7 (76.5, 101)		309	9.1 (8.7, 9.4)		309	31.6 (31.0, 32.1)	
		GG	92	13.9 (13.6, 14.3)		132	80.6 (65.0, 98.8)		134	9.2 (8.6, 9.7)		134	31.2 (30.3, 32.1)	
TMPRSS6	rs2413450	GG	384	13.9 (13.7, 14.1)	0.40	532	85.4 (77.0, 94.7)	0.28	545	9.1 (8.9, 9.4)	0.56	545	31.5 (31.0, 31.9)	0.28
		AG	82	13.6 (13.3, 14.0)		111	79.4 (63.3, 99.6)		113	9.4 (8.8, 10.0)		113	31.1 (30.2, 32.0)	
		AA	5	13.5 (12.1, 14.9)		6	178 (67.2, 472)		7	8.3 (6.4, 10.8)		7	34.1 (30.4, 37.9)	

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

	SNP	Genotype	N	Hb	P	N	SF	P	N	sTfR	P	N	BI	P
TMPRSS6	rs4820268	AA	312	13.9 (13.7, 14.1)	0.25	435	86.8 (77.4, 97.3)	0.58	446	9.1 (8.9, 9.4)	0.95	7	31.5 (31.0, 32.0)	0.71
		AG	142	13.7 (13.4, 14.0)		193	79.4 (66.9, 94.3)		197	9.2 (8.7, 9.6)		197	31.2 (30.5, 31.9)	
		GG	16	14.3 (13.5, 15.1)		20	101 (58.8, 172)		21	9.4 (8.1, 10.9)		21	32.0 (29.8, 34.1)	
TMPRSS6	rs5756506	GG	25	13.6 (13.0, 14.2)	0.45	29	96.6 (62.0, 151)	0.66	31	8.0 (7.0, 9.0)	0.07	31	32.1 (30.3, 33.8)	0.68
		CG	127	13.7 (13.5, 14.0)		180	80.0 (66.9, 95.6)		186	9.2 (8.8, 9.7)		186	31.2 (30.5, 32.0)	
		CC	319	13.9 (13.7, 14.1)		440	86.2 (77.0, 96.7)		448	9.2 (8.9, 9.5)		448	31.5 (31.0, 31.9)	
TMPRSS6	rs5756520	GG	136	13.7 (13.4, 14.0)	0.32	172	85.1 (71.1, 102)	0.70	179	9.4 (9.0, 9.9)	0.42	179	31.3 (30.6, 32.0)	0.63
		AG	219	13.9 (13.7, 14.2)		309	90.4 (79.0, 104)		315	9.0 (8.7, 9.4)		315	31.7 (31.1, 32.2)	
		AA	26	13.6 (12.9, 14.2)		36	76.9 (51.6, 115)		36	9.1 (8.1, 10.2)		36	31.1 (29.5, 32.8)	
TMPRSS6	rs855791	GG	412	13.9 (13.7, 14.0)	0.53	566	86.2 (78.0, 95.3)	0.68	578	9.1 (9.0, 9.4)	0.95	578	31.5 (31.1, 31.9)	0.83
		AG	35	13.7 (13.1, 14.2)		47	74.1 (52.3, 105)		51	9.3 (8.7, 10.2)		51	31.1 (29.7, 32.4)	
		AA	24	13.6 (12.9, 14.2)		36	77.0 (53.5, 119)		36	9.2 (8.1, 10.3)		36	31.2 (29.6, 32.9)	

[1]For each SNP, top to bottom rows represent values for the Homozygote wild type, Heterozygotes and Homozoygote minor allele genotypes respectively. [2]Geometric means and confidence intervals presented are adjusted for age, area of residence (rural or urban) and proportion of dietary heme iron intake. [3]comparisons involving serum ferritin and body iron were additionally adjusted for inflammation by including serum C-reactive protein concentrations as a covariate in the model. [4]*P* refers to the p value of the SNP association with the corresponding outcome measure. *Polymorphism is significantly associated with outcome measure before adjustment for multiple testing (P-values < 0.05). BI, body iron (mg/kg); Hb, Hemoglobin (g/dL); SF, serum ferritin (µg/L); sTfR, soluble transferrin receptor (mg/L); *TMPRSS6*, transmembrane protease serine 6.

Supplemental Table 2. Single nucleotide polymorphisms and genotype distribution in chromosome 3 and associations with iron status indicators within the PURE Study female population [1,2,3,4]

Locus	SNP	Genotype	n	Hb Mean (95% CI)	P	n	SF Mean (95% CI)	P	n	sTfR Mean (95% CI)	P	n	BI Mean (95% CI)	P
TF	rs1358024	GG	465	13.8 (13.7, 14.0)	0.2	636	85.1 (77.4, 93.6)	0.88	652	9.2 (8.9, 9.4)	0.94	652	31.4 (31.0, 31.8)	0.89
		AG	5	14.7 (13.3, 16.1)		11	71.3 (34.7, 147)		11	9.4 (7.7, 11.6)		11	30.7 (27.7, 33.7)	
		AA	1	16.0 (12.8, 19.2)		2	95.4 (17.6, 517)		2	8.7 (5.4, 14.2)		2	31.9 (24.9, 38.9)	
TF	rs1525892	GG	220	13.8 (13.6, 14.0)	0.86	307	78.6 (68.6, 90.1)	0.27	316	9.3 (9.0, 9.7)	0.41	316	31.1 (30.6, 31.7)	0.27
		AG	187	13.9 (13.7, 14.1)		262	92.8 (80.1, 108)		266	9.0 (8.6, 9.4)		266	31.8 (31.2, 32.4)	
		AA	42	13.8 (13.3, 14.3)		51	82.1 (58.7, 115)		54	9.3 (8.5, 10.2)		54	31.3 (30.0, 32.7)	
TF	rs1799852	GG	437	13.8 (13.7, 14.0)	0.16	603	87.9 (79.8, 96.8)	0.01*	619	9.1 (8.8, 9.3)	0.03*	619	31.6 (31.2, 32.0)	0.004*
		AG	34	14.2 (13.7, 14.8)		46	54.0 (38.0, 76.7)		46	10.2 (9.2, 11.3)		46	29.4 (27.9, 30.8)	
		AA	0	-			-		0	-		0	-	
TF	rs1799899	GG	467	13.8 (13.7, 14.0)	-	643	84.4 (76.7, 92.7)	-	659	9.2 (8.9, 9.4)	-	659	31.4 (31.0, 31.8)	-
		AG	0	-		0	-		0	-		0	-	
		AA	0	-		0	-		0	-		0	-	
SRPRB	rs1830084	AA	325	13.8 (13.6, 14.0)	0.24	457	84.2 (75.3, 94.2)	0.52	469	9.1 (8.8, 9.4)	0.62	469	31.4 (30.9, 31.9)	0.66
		AT	124	14.1 (13.8, 14.4)		163	88.2 (73.1, 107)		167	9.4 (8.9, 9.9)		167	31.5 (30.8, 32.3)	
		TT	12	13.8 (12.9, 14.7)		15	60.6 (32.6, 113)		15	8.7 (7.3, 10.4)		15	30.3 (27.7, 32.9)	
TF	rs1867504	GG	242	13.9 (13.7, 14.1)	0.84	332	83.8 (73.4, 95.5)	0.96	341	9.2 (8.9, 9.6)	0.5	341	31.3 (30.7, 31.8)	0.79
		AG	193	13.8 (13.6, 14.0)		268	86.1 (74.6, 99.6)		278	9.1 (8.8, 9.5)		278	31.5 (30.9, 32.1)	
		AA	36	13.7 (13.2, 14.3)		51	86.0 (60.0, 123)		46	8.7 (7.9, 9.6)		46	31.8 (30.3, 33.2)	

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

TF	rs3811647	GG	251	13.8 (13.6, 14.0)	0.88	346	77.7 (68.4, 88.3)	0.07	355	9.4 (9.1, 9.8)	0.03*	355	31.0 (30.5, 31.5)	0.03*
		AG	173	13.9 (13.7, 14.1)		245	97.9 (84.1, 114)		249	8.8 (8.4, 9.1)		249	32.1 (31.5, 32.7)	
		AA	47	13.8 (13.3, 14.3)		58	79.0 (57.7, 108)		61	9.4 (8.6, 10.2)		61	31.1 (29.8, 32.4)	
TF	rs3811658	GG	344	13.8 (13.6, 14.0)	0.6	477	86.0 (77.1, 96.0)	0.87	490	9.2 (8.9, 9.5)	0.93	490	31.4 (31.0, 31.9)	0.96
		AG	111	14.0 (13.7, 14.3)		151	81.1 (66.8, 98.6)		154	9.2 (8.7, 9.7)		154	31.3 (30.5, 32.1)	
		AA	16	13.9 (13.1, 14.7)		21	87.7 (52.1, 148)		21	8.9 (7.7, 10.3)		21	31.5 (29.4, 33.7)	
TF	rs4525863	CC	333	13.9 (13.8, 14.1)	0.24	446	84.7 (75.6, 94.8)	0.94	456	9.3 (9.0, 9.6)	0.14	456	31.3 (30.9, 31.8)	0.70
		AC	125	13.6 (13.4, 13.9)		184	84.6 (70.9, 101)		190	8.9 (8.5, 9.4)		190	31.5 (30.8, 32.3)	
		AA	13	13.8 (12.9, 14.7)		19	93.9 (54.1, 163)		19	8.2 (7.0, 9.6)		19	32.2 (29.9, 34.5)	
TF	rs7638018	AA	337	13.8 (13.6, 14.0)	0.76	470	85.0 (76.2, 94.9)	0.99	483	9.1 (8.9, 9.4)	0.88	483	31.4 (31.0, 31.9)	0.99
		AG	118	13.9 (13.6, 14.2)		158	84.2 (69.5, 102)		161	9.2 (8.8, 9.8)		161	31.4 (30.6, 32.2)	
		GG	16	13.9 (13.1, 14.7)		21	87.7 (52.0, 148)		21	8.9 (7.7, 10.3)		21	31.5 (29.4, 33.7)	
TF	rs8177248	GG	427	13.8 (13.7, 14.0)	0.39	581	86.1 (78.1, 95.1)	0.34	596	9.2 (8.9, 9.5)	0.65	590	31.4 (31.0, 31.8)	0.64
		AG	41	13.9 (13.1, 14.1)		63	71.2 (52.7, 96.3)		64	9.0 (8.3, 9.8)		64	30.9 (29.6, 32.1)	
		AA	3	13.9 (13.0, 16.6)		5	137 (47.2, 401)		5	8.0 (5.9, 10.9)		4	32.3 (27.3, 37.2)	

[1]For each SNP, top to bottom rows represent values for the Homozygote wild type, Heterozygotes and Homozoygote minor allele genotypes respectively. [2]Geometric means and confidence intervals presented are adjusted for age, area of residence (rural or urban) and proportion of dietary heme iron intake. [3]comparisons involving serum ferritin and body iron were additionally adjusted for inflammation by including serum C-reactive protein concentrations as a covariate in the model. [4]*P* refers to the p value of the SNP association with the corresponding outcome measure. *Polymorphism is significantly associated with outcome measure before adjustment for multiple testing (P-values < 0.05). BI, body iron (mg/kg); Hb, Hemoglobin (g/dL); SF, serum ferritin (µg/L); sTfR, soluble transferrin receptor (mg/L); *SRPRB*, signal recognition particle receptor, B subunit *TF*, transferrin.

Supplemental Table 3. Single nucleotide polymorphisms, genotype distribution in other chromosomes and associations with iron status indicators within the PURE Study female population [1,2,3,4]

Chr #	Locus	SNP	Genotype	n	Hb Mean (95% CI)	P	n	SF Mean (95% CI)	P	n	sTfR Mean (95% CI)	P	n	BI Mean (95% CI)	P
1	OPRD1	rs482692	GG	332	13.8 (13.6, 14.0)	0.83	449	81.4 (72.7, 91.1)	0.34	463	9.2 (8.9, 9.5)	0.67	463	31.2 (30.8, 31.7)	0.28
			AG	106	13.8 (13.5, 14.1)		152	96.1 (79.1, 117)		154	9.0 (8.5, 9.5)		154	31.9 (31.1, 32.7)	
			AA	31	14.0 (13.4, 14.6)		43	80.3 (55.8, 116)		43	9.5 (8.5, 10.5)		43	31.1 (29.6, 32.6)	
6	TNF α	rs1799964	AA	314	13.9 (13.7, 14.1)	0.43	439	89.6 (79.8, 101)	0.26	439	9.2 (8.9, 9.5)	0.83	439	31.6 (31.1, 32.1)	0.49
			AG	138	13.8 (13.6, 14.1)		190	77.9 (65.5, 92.6)		196	9.1 (8.7, 9.6)		196	31.1 (30.4, 31.8)	
			GG	19	13.4 (12.6, 14.1)		30	68.7 (44.4, 106)		30	8.9 (7.8, 10.0)		30	30.8 (29.0, 32.6)	
6	TNF α	rs1800629	GG	332	13.8 (13.7, 14.0)	0.76	440	84.4 (75.4, 94.6)	0.19	452	9.2 (8.9, 9.5)	1.00	452	31.4 (30.9, 31.9)	0.24
			AG	128	13.9 (13.6, 14.1)		193	82.2 (69.2, 97.5)		197	9.2 (8.7, 9.6)		197	31.3 (30.6, 32.0)	
			AA	11	13.5 (12.5, 14.5)		16	147 (80.6, 266)		16	9.1 (7.7, 10.8)		16	33.4 (30.9, 35.9)	
6	HFE	rs1800562	GG	471	13.8 (13.7, 14.0)	-	649	84.9 (77.3, 93.2)	-	665	9.2 (8.9, 9.4)	-	665	31.4 (31.0, 31.8)	-
			AG	0	-		0	-		0	-		0	-	
			AA	0	-		0	-		0	-		0	-	
6	NFKBIL1	rs2071592	TT	54	13.7 (13.3, 14.2)	0.79	82	74.0 (56.9, 96.3)	0.18	83	9.0 (8.4, 9.7)	0.68	83	31.0 (29.9, 32.0)	0.26
			AT	225	13.9 (13.7, 14.1)		310	80.8 (70.6, 92.6)		320	9.3 (8.9, 9.6)		320	31.2 (30.6, 31.8)	
			AA	192	13.8 (13.6, 14.1)		257	94.2 (81.1, 109)		262	9.1 (8.7, 9.5)		262	31.8 (31.2, 32.4)	
7	TFR2	rs7385804	AA	243	13.8 (13.6, 14.0)	0.77	331	77.6 (68.0, 88.5)	0.14	339	9.2 (8.9, 9.5)	0.97	339	31.1 (30.5, 31.6)	0.17
			AC	188	13.9 (13.7, 14.1)		262	93.3 (80.5, 108)		269	9.1 (8.8, 9.5)		269	31.8 (31.1, 32.4)	
			CC	37	13.8 (13.3, 14.4)		53	97.0 (70.0, 135)		54	9.2 (8.4, 10.1)		54	31.9 (30.5, 33.2)	

Common variants in the TF, TNF α and TMPRSS6 genes are associated with iron status in a female black South African population

				n	Hb	P	n	SF	P	n	sTfR	P	n	BI	P
10	*CUBN*	rs10904850	GG	236	13.8 (13.6, 14.0)	0.62	330	82.6 (72.4, 94.2)	0.84	338	9.2 (8.9, 9.6)	0.62	338	31.3 (30.7, 31.8)	0.81
			AG	195	13.9 (13.7, 14.2)		268	87.4 (75.5, 101)		273	9.0 (8.7, 9.4)		273	31.6 (31.0, 32.2)	
			AA	40	13.8 (13.3, 14.3)		51	87.3 (62.4, 122)		54	9.5 (8.6, 10.4)		54	31.4 (30.0, 32.8)	
19	*HAMP*	rs10421768	AA	348	13.8 (13.7, 14.0)	0.43	470	82.4 (73.8, 91.1)	0.84	481	9.3 (9.0, 9.6)	0.84	481	31.2 (30.8, 31.7)	0.73
			AG	115	13.9 (13.6, 14.2)		166	94.5 (78.5, 117)		171	8.8 (8.4, 9.3)		171	31.9 (31.2, 32.7)	
			GG	8	13.7 (12.6, 14.9)		13	65.3 (33.6, 116)		13	8.7 (7.2, 10.5)		13	30.6 (27.9, 33.4)	
18	near *TMPRSS6*	rs9948708	GG	150	13.7 (13.5, 14.0)	0.97	198	82.6 (68.7, 96.5)	0.33	206	9.2 (8.8, 9.7)	0.19	206	31.2 (30.5, 31.9)	0.32
			AG	228	13.9 (13.7, 14.1)		316	87.4 (76.0, 99.5)		321	9.1 (8.7, 9.4)		321	31.5 (32.0, 32.1)	
			AA	93	14.0 (13.7, 14.3)		135	87.3 (69.5, 105)		138	9.2 (8.7, 9.8)		138	31.4 (30.5, 32.2)	

[1]For each SNP, top to bottom rows represent values for the Homozygote wild type, Heterozygotes and Homozygote minor allele genotypes respectively. [2]Geometric means and confidence intervals presented are adjusted for age, area of residence (rural or urban) and proportion of dietary heme iron intake. [3]comparisons involving serum ferritin and body iron were additionally adjusted for inflammation by including serum C-reactive protein concentrations as a covariate in the model. [4]P refers to the p value of the SNP association with the corresponding outcome measure. *Polymorphism is significantly associated with outcome measure before adjustment for multiple testing (P-values < 0.05). BI, body iron (mg/kg); *CUBN*, cubilin (intrinsic factor-cobalamin receptor); *HAMP*, hepcidin antimicrobial peptide; Hb, Hemoglobin (g/dL); *HFE*, hemochromatosis; *NFKBIL1*, nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1; *OPRD1*, opioid receptor, delta 1; SF, serum ferritin (µg/L); sTfR, soluble transferrin receptor (mg/L); *TF*, transferrin; *TFR2*, transferrin receptor 2; *TMPRSS6*, transmembrane protease, serine 6; *TNF α*, tumor necrosis factor.

Chapter 4

Associations between common variants in
iron-related genes with haematological
traits in populations of African ancestry

Wanjiku N Gichohi-Wainaina
Toshiko Tanaka
G Wayne Towers
Hans Verhoef
Jacobien Veenemans
Elise Talsma
Jan Harryvan
Mark Boekschoten
Edith J Feskens
Alida Melse-Boonstra

Associations between common variants in iron-related genes with haematological traits in populations of African ancestry

Abstract

Background: Large genome-wide association (GWA) studies of European ancestry individuals have identified multiple genetic variants influencing iron status. Studies on the generalizability of these associations to African ancestry populations have not been conducted. These studies are important given interethnic differences in iron status and the disproportionate burden of iron deficiency among African ancestry populations.

Methods: We tested the associations of 20 previously identified iron status-associated single nucleotide polymorphisms (SNPs) in 628 Kenyans, 609 Tanzanians, 608 South Africans and 228 African Americans. In each study, we examined the associations present between 20 SNPs with ferritin and haemoglobin, adjusting for age, sex and CRP levels.

Results: We replicated previously reported associations with lowered haemoglobin concentrations for two SNPs (rs2413450, rs4820268) in *TMPRSS6* and increased haemoglobin concentrations with a locus in TF (rs3811658). Associations with increased ferritin concentrations were also confirmed for two loci, in *TMPRSS6* (rs228918) and in TF (rs1525892).

Conclusions: Only 4 of the 20 single nucleotide polymorphisms reported in large GWA studies of European ancestry individuals were associated with iron status in the African ancestry cohorts included in our study. While there is now evidence for the associations of a number of genetic variants with iron status in both European and African ancestry populations, the considerable lack of concordance highlights the importance of continued ancestry-specific studies to elucidate the genetic underpinnings of iron status in ethnically diverse populations.

Introduction

Body iron status is influenced by a combination of genes as well as environmental factors like diet, blood loss, pregnancy, alcohol intake and infections [1, 2]. In order to maintain iron levels within normal ranges and thus prevent iron excess or iron deficiency, iron homeostasis has evolved as a complex trait and a tightly coordinated process of interaction between genes and environmental factors. Whether genetic influences on iron metabolism may affect iron status at the public health level, is a subject of recent interest. This is because segments of populations that are genetically prone to iron deficiency can be targeted for interventions.

Understanding the genetic underpinnings of low iron status is particularly relevant for African ancestry populations, who experience a disproportionately high burden of low iron status. In particular, it has been observed that in the US, African-Americans have lower haemoglobin concentrations and serum transferrin saturation (TS), and higher serum ferritin (SF) concentrations compared to Caucasians [3]. Furthermore, according to WHO regional prevalence estimates, Africa has the highest proportion of individuals affected by anaemia worldwide [3].

Previous studies on iron status report heritability estimates ranging from of 20% to 30%, supporting a genetic component to iron status [4, 5]. Genome-wide association (GWA) studies have identified common variants, some in genes associated with Mendelian forms of iron deficiency, to be in association with variability in iron status in largely European ancestry populations [6-9]. Given the greater levels of linkage disequilibrium among European-ancestry individuals [10], it is expected that index single nucleotide polymorphisms (SNPs) tag larger regions of the genome than they would among African ancestry individuals. Since African populations are exposed to different environmental factors that further influence iron status, this may hamper the potential to generalize the results from European ancestry populations to African ancestry populations. Recently, genetic studies on iron status have also been conducted in the African American population [11-15]. This population has varying degrees of admixture that may also complicate transferability of these results to African populations [16]. Conclusive inferences about the potential generalization of previous findings to African ancestry populations have been precluded by the lack of data from African cohorts. In the present replication study, we investigated 20 SNPs previously reported to be

associated with iron status in GWA studies *[6, 7, 9, 13, 17]* with haematological indicators in several African ancestry populations.

Methods

Study area and population

Four cohorts of African ancestry were selected; from Kenya (n=628), Tanzania (n=609); South Africa (n=686) and the United States of America (n=228). The four cohorts differed largely in age, gender and burden of inflammation as described in more detail below.

Kenyan cohort

Samples were obtained from both a cross sectional survey and from the baseline data of an efficacy trial designed to assess the efficacy of bio-fortified cassava to improve vitamin A deficiency [18]. Haematological parameters were measured prior to the start of the trial. Ancestry was determined by self-report in both studies and consisted of children of Kamba (Bantu) ancestry. Cross sectional samples were collected from children aged 6-12 years (n=375) who were located in the Kibwezi and Makindu districts in Eastern Kenya. Trial samples were collected from children aged 5-13 years (n=437) in three primary schools in the Kibwezi district, Eastern Kenya. Kibwezi and Makindu that are low malarial transmission areas [19]. For the trial, children with low retinol binding protein concentrations (0.70 µmol/L) were selected. For both studies. subjects were apparently healthy and without fever (ear drum temperature < 37.5°C) upon examination by the research physician and children with inflammation (C-reactive protein concentration > 8mg/L), low haemoglobin concentrations (<70 g/L), or malaria were excluded. The cross sectional survey as well as the trial were approved by ethical committees in Kenya and the Netherlands. Informed consent was obtained from parents and assent from children.

Tanzanian cohort

Children in this cohort were part of a randomized placebo controlled trial [20]. Haematological parameters as well as all other mesures were measured prior to the start of the trial and ancestry was self-reported. People in this area originally belonged to the Wazigua and Wabondei Bantu tribes, but settlement of migrant plantation workers in the past has resulted in the admixture of many tribes with different origins. Samples were obtained from children aged 6-60 months (n=612) with height-for-age z-score ≤ -1.5 SD in a rural area of the Handeni district in North-Eastern Tanzania, which has intense year-round *Plasmodium falciparum* transmission. Children with severe wasting (weight-for-age z-score < -3 SD), haemoglobin concentration < 70 g/L, malaria, or signs of chronic illness, were excluded. The study was approved by the Ethical Review Committee of Wageningen University, The Netherlands (Approval number: NL: 04/07) and the National Health Research Ethics Review sub-Committee, Dar es Salaam, Tanzania. Written individual informed consent was obtained from parents or primary caretakers.

South African cohort

This study population was a subset of Tswana speaking pre- and post-menopausal women (n=686) enrolled as part of the South African arm of the Prospective Urban and Rural Epidemiological (PURE) study. Haematological measures utilized in this study were collected in a follow up study conducted in 2010. The participants were from the North-West Province in South Africa, aged 35 years or older and were mainly of Tswana (Bantu) ancestry as determined by self-report. Women were only eligible for this sub-study if they were apparently healthy, not pregnant, non-lactating and were at the time of blood collection not diagnosed with any chronic disease or taking drugs for any chronic disease. Additionally, we included only participants with all outcome measures determined and who were HIV negative. HIV testing was undertaken with the appropriate pre- and post-test counselling as required by South African law. This study was approved by the Ethics Committee of the North-West University and signed informed consent forms were received from all participants.

African American cohort

This cohort consisted of participants from the Baltimore Longitudinal Study on Aging (BLSA). The BLSA is a population-based study aimed at evaluating contributors of healthy aging in the older population residing predominantly in the Baltimore-Washington DC area *[21]*. Started in 1958, participants aged 17-94 years were examined every one to four years depending on their age. Currently, there are approximately 1,100 active participants enrolled in the study. For the present study, 228 subjects of African ancestry with complete data on both genetic and haematological traits were available and were used for the analysis. The BLSA has continuing approval from the Institutional Review Board (IRB) of Medstar Research Institute as well as from study participants.

Iron related measurements

Kenyan and Tanzanian cohorts

For the cross sectional study study (Kenyan cohort), serum concentrations of CRP were determined by immunoturbidometric assays on a Cobas Integra 800 system (Roche Diagnostics, Mannheim, Germany). Serum ferritin concentrations were determined by enzymatic immunoassay (Ramco Laboratories, Inc, Stafford Texas). For the Kenyan efficacy trial as well as Tanzanian cohort, concentrations of ferritin and concentrations of CRP were measured on a Beckman Coulter UniCel DxC 880i analyzer as per manufacturer's instructions. Haemoglobin concentration (Hb) was assessed using a hematology analyzer (Celltac-α, MEK-6410K, Nihon Kohden Corporation, Tokyo, Japan).

South African cohort

Serum concentrations of CRP were measured from serum via a particle-enhanced immunoturbidometric assay using Sequential Multiple Analyser Computer (SMAC) using the Konelab20iTM auto analyser (Thermo Fisher Scientific Oy, Vantaa, Finland). Whole blood was used for the determination of HIV status making use of the First Response (PMC Medical, India) rapid HIV Card Test. If tested positive, the test was repeated with the Pareeshak (BHAT Bio-tech India) card test. Serum ferritin was determined using an enzyme immunoassay procedure (Ramco Laboratories, Inc, Stafford

Texas). Hb concentrations were determined spectrophotometrically from whole blood using the Hemocue® HB201 haemoglobin meter (Hemocue AB, Angelholm Sweden).

African American cohort

Serum concentrations of CRP were measured using the BNII nephelometer from Dade Behring utilizing a particle enhanced immunonepholometric assay.
Serum ferritin was measured using an immunoassay type two-stage sandwich method using two antiferritin antibodies (Advia Centaur, Bayer). Haemoglobin was assessed using SYSMEX XE-series (Sysmex Corporation).

SNP selection, genotyping and quality control procedures

SNP selection

To set up the assay it was necessary to develop a list of reference SNPs (rs) numbers from Genbank of the SNPs that are to be included in the analysis. This list was setup by performing a literature search of Pubmed, Sciencedirect, Ebscohost, HUGE navigator and Google Scholar using the following keywords:

a.   Iron, ferritin, transferrin
b.   Anaemia
c.   Haemochromatosis
d.   Human
e.   SNP, Polymorphism, Mutation
f.   Genetic

All abstracts were scanned for appropriateness of use, that is, association or linkage of a SNP to the haematological parameters or key biomarkers of iron status (Hb, SF, serum transferrin receptor (sTfR) and body iron (BI)). If appropriate, the reference SNP (rs) number of the polymorphism was recorded and included in the list to be sent to Illumina for verification of use in the array. Based on this, we selected 32 SNPs for genotyping. Once SNPs were selected, genotyping was performed for the Kenyan, Tanzanian and South African cohorts. These data were complemented with the data we obtained from the African American cohort in the BLSA.

Genotyping and quality control procedures

For the Kenyan, Tanzanian and South African cohorts, all 32 rs numbers were submitted to Illumina Technical Support for evaluation using the Assay Design Tool. Each SNP was scored (varying from 0 to 1) by the same, based on its compatibility for use in a GoldenGate genotyping assay. SNPs with a score above 0.4 were chosen for genotyping. No SNP fell out based on these criteria. Genotyping for Kenyan (efficacy trial only), South African and Tanzanian cohorts was conducted using the BeadXpress™ platform. Once genotyping was performed, several quality control procedures were applied. Lack of a detectable signal after analysis was designated as a failed SNP. SNP clustering was also assessed visually to determine success of genotyping. SNPs with a GenCall Score of >0.5 and samples with a call rate of ≥0.95 were included in the final analysis. Genotyping was performed for only three SNPs in *TMPRSS6* (rs2413450, rs4820268, rs228918), for the Kenyan pilot study samples using Applied Biosystems TaqMan Assays. Data obtained with this method were merged with the data from the Kenyan efficacy trial. Eventually, we considered 20 SNPs that had passed all quality control procedures in all four cohorts for association analyses (**Supplementary table 1**).

For the African American cohort, genome-wide genotyping was conducted using the Illumina Infinium HumanHap 550K. Genotyping was completed for 284 subjects, and imputation of ~2.5 million HapMap SNPs was conducted using 501,764 SNPs that passed quality control (minor allele frequency > 1%, genotyping completeness > 99%, and Hardy Weinberg-equilibrium > 0.0001). Imputation was conducted with MACH using the HapMap YRI sample (HapMap data release rel#23a NCBI B36) as a reference. The candidate iron-homeostasis SNPs that were either directly genotyped or imputed (with an imputation quality of greater than $r^2>0.3$) were used for the present analysis.

Statistical analyses

Outcome variables (haemoglobin and serum ferritin concentrations) were determined to be continuous quantitative traits. Anaemia was defined using different cut offs for each cohort as cut offs are age dependent [22]. Cut offs were as follows: Hb concentration <11.5 g/dL (Kenyan cohort), <11 g/dL (Tanzanian cohort), <12 g/dL for the South African and African American cohorts, while iron deficiency was defined as serum ferritin cut-off <15µg/l for Kenyan, South African and African American samples, and plasma ferritin <12µg/l for Tanzanian samples. Log_{10} transformations were applied to the ferritin

outcomes to improve their fit to the normal distribution. Results from associations involving ferritin were then presented as back transformed data. Potential disturbances in the distributions of the genotypes were tested for deviation from the assumptions of Hardy Weinberg equilibrium (HWE) by using the Chi-square test. To compare minor allele frequencies (MAFs) across cohorts, a spider web graph was generated using Microsoft Excel (Version 2010), including European ancestry populations from the 1000 Genomes Project *[30]*. Linear regression was used to assess associations between SNPs and outcome measures. For association testing, genotypes were coded as 0, 1, and 2 reflecting the number of copies of the minor allele (additive genetic model), while imputed genotypes were coded using fractional counts between 0 and 2 according to the estimated number of copies of each minor allele. All associations were adjusted for age and gender. Additionally, associations involving ferritin concentrations were adjusted for inflammation by including CRP concentrations as a covariate in the model. Statistical analyses were performed using version 20 of the Statistical Package for Social Sciences (SPSS Inc Chicago III) or SAS version 9.2 (Cary, NC). All reported P-values were two-tailed, and adjustment for multiple testing was performed using the false discovery rate (FDR) method. The R program for statistical computing version 2.15.2 *[23]* was used to perform all meta-analyses and chi square tests conducted to test for between studies. Comparisons of effect sizes and directions were made within and across populations, and meta-analyses were performed to generate summary effect estimates with and without including the African American cohort. Statistical significance for all analyses was defined as $\alpha < 0.05$, while the q values cut-off was set at 0.05.

Results

There were differences in age between the cohorts, as the studies from Kenya and Tanzania were pre- and school going children, while the studies from South Africa and the United States were of older adults. With the exception of the South African cohort which was all women, the sex distribution was approximately 50% (Table 1). Prevalence of inflammation was below 5% in the Kenyan and African American cohort and 33% and 43% in the Tanzanian and South African cohort, respectively. Iron deficiency was highest in the Kenyan cohort (37%) and lowest in the African American cohort (3%-**Table 1**). Anaemia was highest in the Tanzanian cohort at 32%, and lowest in the Kenyan cohort at 7%.

Table 1: General characteristics of the study participants in the four cohorts

	n	Kenyan	n	Tanzanian	n	South African	N	African American
Age, y	628	9.0 (2.1)	609	2.7 (1.3)	686	50.0 (10.0)	228	67.4 (11.0)
Gender (% female)	621	52	609	51	686	100	228	59
Haemoglobin (g/dL)	626	13.9 (1.6)	607	10.3 (12.8)	481	13.8 (1.6)	228	13.0 (1.4)
Ferritin (µg/L)$	623	17.8 (17.0,18.6)	603	31.6 (31.6, 39.8)	686	85.1 (77.6, 93.3)	221	103.4 (96.6)
Inflammation (%)	620	3	609	33	290	43	228	5
Anaemia (%)	626	7	609	32	481	8	227	19
Iron deficiency (%)	623	37	606	18	686	11	227	3

$ Values are geometric means (confidence intervals) or otherwise means (±SD), Inflammation cut-off CRP>5mg/L, ID-SF cut-off <15µg/l for Kenyan and South African samples and PF <12µg/l for Tanzanian samples, Anaemia cut-off Hb concentration <11.5 g/dL for Kenyan cohort, <11 g/dL for Tanzanian cohort, <12 g/dL for the South African and African American cohorts Hb = Haemoglobin, CRP = C-reactive protein

In comparing minor allele frequencies across cohorts, SNPs genotyped were generally more common in European ancestry cohorts based on data from the 1000 Genomes Project (Figure 2). We examined the differences in MAFs of TMPRSS6 SNPs between Europeans and Africans using data from the 1000 Genomes Project. In particular, for the commonly investigated rs4820268 SNP, the frequency of the allele associated with lower iron status was 27%, 22%, 18%, and 30% in the Kenyan, Tanzanian, South African and African American cohorts respectively.

The MAF of this SNP in the African American cohort was closest to that of the Africans (31%) in the 1000 Genomes Project. In comparison to the Europeans (41%, 1000 Genomes Project), all the African ancestry cohort have lower MAFs for this SNP. For the SNP rs2413450, MAFs were ~12% in the Kenyan, Tanzanian and South African cohorts and 24% in the African American cohort. These frequencies were comparable to those of Africans in the 1000 Genomes Project (17%), however, they were lower than what has been observed in Europeans within the 1000 Genomes Project (42%).

Minor allele frequencies of rs228918 were (~40% across our African ancestry cohorts and comparable to those of the Europeans (41%) and Africans (40%) in the 1000 Genomes Project. On the other hand rs1421312 was more common in the African cohorts (Kenya=58%, Tanzanian=66% and South Africa=68%) compared to the Europeans in the 1000 Genomes Project (41%), but had a comparable

MAF to that of the Africans in the 1000 Genomes Project (58%). For the same SNP (rs1421312), the African American cohort had a MAF of 43%. For rs4820258 and rs2413450, the minor allele is the risk allele while for rs1421312 and rs228918 there is no clear evidence whether the minor allele is the risk allele.
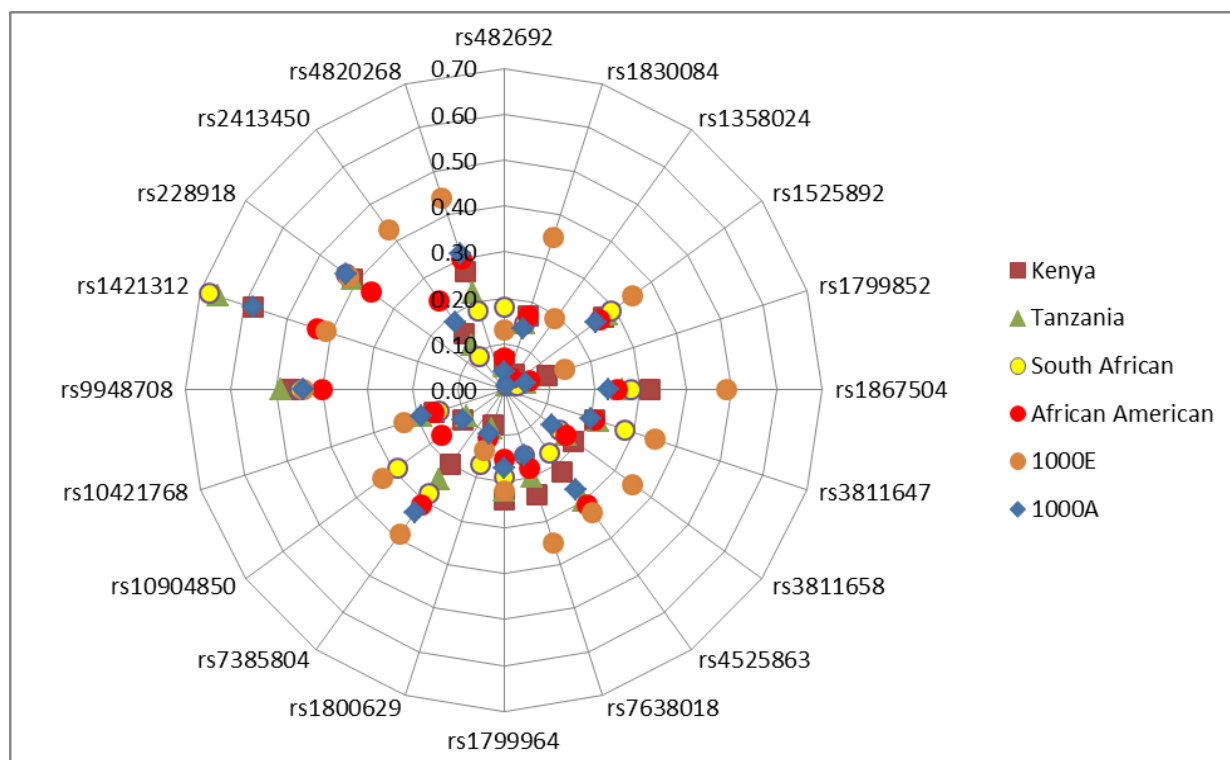
Figure 1: Minor allele frequency of SNPs in comparison to African and European population in 1000 Genomes Project
MAF 1000 A, Minor allele frequency Africans- 1000 genomes; MAF 1000 E, Minor allele frequency Europeans- 1000 Genomes Project

In linear regression analyses with haemoglobin concentrations as the dependent outcome, inverse associations were observed for the A allele of rs10904850 in *CUBN* (-β=0.38, 95 CI 0.07, 0.69g/dL; P=0.02) as well as the G allele of rs4820268 in *TMPRSS6* (β=-0.28 , 95 CI 0.22, 0.34 µg/L; P=0.01) in the Kenyan children (**Table 2**). Conversely, the minor allele A of rs10421768 in *HAMP* was positively associated with Hb concentrations (β=0.37 , 95 CI 0.10, 0.64 µg/L; P=0.01) in this population.

After meta-analyses of data from all three African cohorts, the minor allele (T) of rs3811658 in *TF* was associated with an higher Hb concentrations of 0.36 g/dL (P=0.02) while minor alleles of both rs2413450 (A) and rs4820268 (G) in TMPRSS6 were associated with a lower Hb concentrations of 0.23 g/dL and 0.18 g/dL respectively (P=0.01 for both associations). These associations remained similar after including the African American cohort in the meta-analysis. There was no heterogeneity between studies for these SNPs.

Table 2: Associations of single nucleotide polymorphism with iron status parameters in four cohorts of African ancestry with meta-analyses

Trait	Gene & SNP	Chr	Minor Allele	Ke β (se)	P	Tz β (se)	P	SA β (se)	P	Met_African β (se)	P	P (Het)	AA β (se)	P	Met_All β (se)	P	P (Het)
Hb g/dL	*TF*																
	rs3811658	3	T	0.01(0.13)	0.95	1.57(0.99)	0.11	0.10(0.04)	0.49	**0.36(0.04)**	**0.02**	0.26	0.08(0.16)	0.61	**0.09(0.03)**	**0.01**	0.45
	CUBN																
	rs10904850	10	A	**-0.38(0.16)**	**0.02**	0.91(1.19)	0.44	0.07(0.11)	0.56	-0.11(0.10)	0.14	0.05	0.03(0.16)	0.86	-0.07(0.14)	0.63	0.10
	HAMP																
	rs10421768	19	A	**0.37(0.14)**	**0.01**	0.46(0.94)	0.62	-0.01(0.15)	0.93	0.19(0.18)	0.29	0.17	0.14(0.12)	0.12	0.17(0.10)	0.09	0.30
	TMPRSS6																
	rs2413450	22	A	-0.20(0.12)	0.10	-0.65(1.13)	0.57	-0.23(0.18)	0.20	**-0.23(0.10)**	**0.01**	0.93	-0.09(0.14)	0.55	**-0.19(0.08)**	**0.02**	0.84
	rs4820268	22	G	**-0.28(0.11)**	**0.01**	-0.26(0.87)	0.76	-0.05(0.14)	0.74	**-0.18(0.07)**	**0.01**	0.43	-0.09(0.13)	0.48	**-0.16(0.07)**	**0.04**	0.55
Ferritin µg/L	*TF*																
	rs1525892	3	A	1.01(1.06)	0.80	1.05(1.07)	0.46	1.09(1.08)	0.26	1.05(0.62)	0.05	1.00	1.10(1.10)	0.37	**1.07(0.54)**	**0.04**	1.00
	rs1799852	3	A	-0.98(1.09)	0.80	1.20(1.15)	0.18	**-0.62(1.2)**	**0.01**	-0.15(0.68)	0.41	0.34	1.23(1.17)	0.21	0.18(0.60)	0.76	0.37
	TFR2																
	rs7385804	7	C	**-0.86(1.07)**	**0.02**	-0.95(1.08)	0.48	1.15(1.08)	0.06	1.03(0.64)	0.05	0.27	1.07(1.10)	0.52	0.09(0.58)	0.88	0.32
	TNF																
	rs1800629	6	A	1.02(1.10)	0.85	1.05(1.12)	0.62	1.03(1.09)	0.74	1.03(0.64)	0.10	1.00	**-0.76(1.15)**	**0.03**	0.61(0.56)	0.27	0.60
	TMPRSS6																
	rs228918	22	G	-0.99(1.04)	0.90	1.05(1.07)	0.55	-1.02(1.07)	0.75	-0.33(0.68)	0.62	0.32	1.10(1.10)	0.25	**0.01(0.60)**	**0.02**	0.29

SNP=Single nucleotide polymorphism; Ke=Kenyan cohort; Tz=Tanzanian cohort; SA=South African cohort; AA=African American cohort; Met-African=meta analyses of beta and SE values from African cohorts; Met_All= meta analyses of beta and SE values from African cohorts; Direction=direction of association based on beta value Hb=Haemoglobin, Values in bold are significant. Results are presented as mean (SE) change per copy of the minor allele from regression analysis using additive genetic model and with adjustment for age and sex for associations involving Hb, and age and sex and CRP concentrations for associations involving ferritin. Ferritin values are Log_{10} back transformed data. [†]Values not adjusted for gender as all participants were women, Values in bold are significant before FDR correction for multiple testing. Variance explained ranged from < 1% to 3 %, with the highest being 3% for rs10421768 in the Kenyan population

The minor allele of rs1799852 in *TF* was associated with lower SF concentrations in the South African cohort (0.62 µg/L; P=0.01). Additionally, the minor alleles of rs7385804 in *TFR2* and of rs1800629 were associated with decreased ferritin concentrations in the Kenyan and African American cohorts (0.86 µg/L, P=0.02; and -0.76 µg/L, P=0.03, respectively). After meta-analyses of data from all three African cohorts, no significant associations were observed, although, associations with the minor alleles of rs1525892 and rs1867504 in *TF* and of rs7385804 in *TFR2* were marginally significant (P=0.05). After including the African American cohort in the meta-analysis, the minor alleles of rs1525892 in *TF* and rs228918 in *TMPRSS6* were both significantly associated with increased serum ferritin concentrations (1.7 µg/L, P=0.02; and 0.01 µg/L, P= 0.04, respectively). Complete information on associations of all 20 SNPs with concentrations of haemoglobin and serum ferritin can be found in **Supplementary table 1.** None of the associations for either Hb or SF as outcome measures were significant after FDR correction for multiple testing.

Discussion

We report here the first candidate gene association study focussing on the genetics of iron status in selected African ancestry populations. We have observed significant associations with lower Hb concentrations involving the minor alleles A of rs2413450 and G of rs4820268, both in the *TMPRSS6* gene, and the minor allele T in rs3811658 in the *TF* gene. In associations with SF concentrations, the minor alleles A in rs1525892 of *TF* and G in rs228918 of *TMPRSS6* were associated with increased ferritin concentrations. Additionally, there were marginally significant associations for the minor alleles of rs1525892 and rs1867504 in *TF* and rs7385804 in *TFR2* with increased serum ferritin concentrations. We further observed that many of the candidate SNPs genotyped in our cohorts, were more common in the European ancestry cohort in the 1000 Genomes Project as compared to our African ancestry cohorts.

The most consistent association in terms of direction as well as statistical significance was observed for rs2413450 and rs4820268 in the *TMPRSS6* gene. Similar observations have previously been made within populations of European and Asian ancestry [6, 9, 24, 25]. The exact mechanism through which *TMPRSS6* action is exerted is still under investigation. It has generally been hypothesised that *TMPRSS6* polymorphisms affect hepcidin transcription, thereby influencing hepcidin concentrations in response to

systemic iron concentrations *[26-28]*. However, two recent studies *[29, 30]* did not confirm an intermediate role for hepcidin in the SNP-iron status parameter associations. Currently, it is proposed that matriptase-2 could regulate hepcidin expression by cleaving HJV to decrease BMP-SMAD signalling *[31]*.

We also observed associations between rs3811658 in the TF gene and increased Hb concentrations. This SNP is in linkage disequilibrium with rs3811647 that has previously been negatively associated with serum iron, serum transferrin, serum transferrin receptor, body iron and serum ferritin concentrations and positively associated with serum transferrin concentrations, as well as transferrin saturation *[8, 32]*. Also, the SNP has been associated with increased risk of iron deficiency in a population of men and women of European descent. As we would have expected to see negative associations with haemoglobin concentrations based on the previous observations, this finding requires further investigation. Previously, three variants in TF (rs3811647, rs1799852 and rs2280673) plus the HFE C282Y mutation have been observed to explain approximately 40% of genetic variation in serum transferrin. This once again indicates the importance of the *TF* locus in iron metabolism. We however only observed significant associations of rs1799852 with serum ferritin in the South African cohort. Furthermore, we did not have information on associations of rs2280673 and HFE C282Y. As stated above, transferrin (TF) is a crucial biological carrier of iron in blood plasma *[33]*, thus, severe mutations in the transferrin gene lead to atransferrinaemia *[34]*, and have been determined to cause microcytic anaemia accompanied by hepatic accumulation of iron in humans [35] and mice *[36]*. Further investigations are required to determine whether less severe mutations in transferrin, which cause some loss in iron binding capacity, may lead to iron deficiency.

We observed that none of the associations for either Hb or SF as outcome measures were significant after FDR correction for multiple testing. This indicates that the study was probably underpowered to detect associations.

A large portion of iron deficiency undoubtedly results from non-genetic factors. Environmental factors such as diets low in bioavailable iron *[37]* as well as a high inflammation burden *[38]*, are probably the foremost factors predisposing to iron deficiency. In this regard, it would have been more informative if iron status were corrected for iron intakes. We attempted to correct for this in the African American as well as South African cohorts, and there was negligible difference between the crude and adjusted regression models. This was, however, not undertaken for the Kenyan and

Tanzanian cohorts where this consideration may have been even more vital, as diets in African populations are typically low in bioavailable iron *[39]*. To account for other environmental and individual influences on iron status, we have adjusted for major biological confounders that affect Hb and ferritin concentrations including age, gender *[40]* and acute phase proteins CRP [38] to increase precision. Age and gender are a physiological characteristic that influences iron status *[41, 42]*, and since all association analyses were corrected for age and gender, variation in Hb and ferritin due to these factors was accounted for.

It is noteworthy that we did observe population specific significant associations that were not replicated in all cohorts. This was particularly observed in the Kenyan cohort that consisted of school going children (mean age 9.0 years). In this cohort we specifically highlight the association of the A allele of rs10421768 in *HAMP* with Hb. The HAMP gene encodes hepcidin, which is a 25-amino acid peptide produced and secreted mainly by hepatocytes and is a major regulator of systemic iron homeostasis *[31]*. No functional studies to investigate the role of the assessed *HAMP* polymorphism on hepcidin expression of healthy subjects have previously been performed. There is therefore a need for further investigation to examine how this polymorphism affects hepcidin concentrations. Knowledge of this would be useful to assess responsiveness to iron interventions designed to alleviate iron deficiency.

A limitation inherent to our study is that our cohorts were smaller than those used in GWA studies and therefore may have lacked power to confirm previously observed associations. The sample size for detecting associations between disease and SNP markers is affected by prevalence of the condition under study, allele frequency, linkage disequilibrium (LD), inheritance models (e.g., additive, dominant, and multiplicative models), and effect size of the genetic variants (e.g., odds ratio, relative risk, etc.) *[43-45]*. For example, assuming an additive model, 780 individuals are required to conduct association analyses for a SNP explaining 1% of the variation in the outcome measure, with $\alpha = 0.05$ and power of 80 %. Most of the SNPs investigated in this study explained less than 1% of the variation in iron status thereby requiring a larger sample size. All of our cohorts were not specifically designed to study the genetics of iron metabolism and therefore all these factors were not taken into account. However our study does offer an initial indication of what genetic loci would be important for iron status in African ancestry cohorts.

Another important consideration not taken into account in our analyses is the linkage disequilibrium structure in African ancestry populations. The linkage structure between Europeans and individuals of African descent is different as there is greater variation within the genome of African ancestry populations *[46]*. It is thus possible that actual disease markers are not in LD with the SNPs we investigated in our African cohorts therefore explaining why only a few significant associations were observed.

Previously, studies on the genetics of iron metabolism have been carried out mainly in populations of European ancestry. The assumption in this case would then be that European ancestry cohorts might productively be used for discovery of risk variants in non-Europeans. We have attempted to replicate previous findings from European ancestry populations to African ancestry populations with limited success. This may partly be due to the greater genetic variation within the African ancestry populations *[47]* meaning European ancestry populations are not sufficient to fully uncover the variants underlying iron deficiency within African ancestry populations. Additionally, populations vary by ethnicity in terms of allele frequencies [48] and biological adaptations due to various environmental factors. For these reasons, population specific studies on the genetics of iron metabolism are warranted. Specifically, GWA studies in African ancestry individuals should reveal more novel genetic variants with modest effects and/or allele frequencies. It should also be taken into account that to accurately quantify the variation in iron status measures explained by genetic factors, other known factors influencing iron status should be assessed and corrected for in analyses. Finally, functional studies on the effect of discovered variants on iron status are also required as a next step to confirm the findings as well as elucidate further the action of alleles in relation to iron status.

References

1. Andrews NC: **Iron homeostasis: insights from genetics and animal models.** *Nat Rev Genet* 2000, **1**:208-217.
2. Whitfield JB, Zhu G, Heath AC, Powell, Lw, Martin NG: **Effects of alcohol consumption on indices of iron stores and of iron stores on alcohol intake markers.** *Alcohol Clin Exp Res* 2001, **25**:1037-1045.
3. Organization WH: **Worldwide prevalence of anaemia 1993–2005.** 2008.
4. Njajou OT, Alizadeh BZ, Aulchenko Y, Zillikens MC, Pols HA, Oostra BA, Swinkels DW, van Duijn CM: **Heritability of serum iron, ferritin and transferrin saturation in a genetically isolated population, the Erasmus Rucphen Family (ERF) Study.** *Hum Hered* 2006, **61**:222-228.
5. Whitfield JB, Cullen LM, Jazwinska EC, Powell LW, Heath AC, Zhu G, Duffy DL, Martin NG: **Effects of HFE C282Y and H63D polymorphisms and polygenic background on iron stores in a large community sample of twins.** *Am J Hum Genet* 2000, **66**:1246-1258.
6. Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, Walston JD, Fried LP, Singleton A, Guralnik J, et al: **A genome-wide association analysis of serum iron concentrations.** *Blood* 2010, **115**:94-96.
7. Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, et al: **Common variants in TMPRSS6 are associated with iron status and erythrocyte volume.** *Nat Genet* 2009, **41**:1173-1175.
8. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, Peltonen L, Martin NG, Montgomery GW, Whitfield JB, Visscher PM: **Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels.** *Am J Hum Genet* 2009, **84**:60-65.
9. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, et al: **Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels.** *Nat Genet* 2009, **41**:1170-1172.
10. Need AC, Goldstein DB: **Next generation disparities in human genomics: concerns and remedies.** *Trends Genet* 2009, **25**:489-494.
11. McLaren CE, McLachlan S, Garner CP, Vulpe CD, Gordeuk VR, Eckfeldt JH, Adams PC, Acton RT, Murray JA, Leiendecker-Foster C, et al: **Associations between single nucleotide polymorphisms in iron-related genes and iron status in multiethnic populations.** *PLoS One* 2012, **7**:e38339.
12. Lo KS, Wilson JG, Lange LA, Folsom AR, Galarneau G, Ganesh SK, Grant SF, Keating BJ, McCarroll SA, Mohler ER, 3rd, et al: **Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans.** *Hum Genet* 2011, **129**:307-317.
13. Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, Li J, Lu Y, Yanek LR, Keating B, et al: **Genome-wide association analysis of red blood cell traits in African Americans: the COGENT Network.** *Hum Mol Genet* 2013, **22**:2529-2538.
14. Ding K, de Andrade M, Manolio TA, Crawford DC, Rasmussen-Torvik LJ, Ritchie MD, Denny JC, Masys DR, Jouni H, Pachecho JA, et al: **Genetic variants that confer resistance to malaria are associated with red blood cell traits in African-Americans: an electronic medical record-based genome-wide association study.** *G3 (Bethesda)* 2013, **3**:1061-1068.

15.   Li J, Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, Mentch FD, Guo Y, Kim C, Xia Q, et al: **GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children.** *Hum Mol Genet* 2013, **22**:1457-1464.

16.   Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, Forrester T, Allison DB, Deka R, Ferrell RE, Shriver MD: **Estimating African American admixture proportions by use of population-specific alleles.** *Am J Hum Genet* 1998, **63**:1839-1851.

17.   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**.

18.   Talsma EF VH, Brouwer ID, Mburu-de Wagt AS, Hulshof PJM , Melse-Boonstra A: **Proxy markers of serum retinol concentration, used alone and in combination, to assess population vitamin A status in Kenyan children: a cross-sectional study.** *BMC Med* 2015.

19.   World Malaria report, Accessed on 25th February 2015 [http://www.who.int/malaria/en/]

20.   Veenemans J, Milligan P, Prentice AM, Schouten LR, Inja N, van der Heijden AC, de Boer LC, Jansen EJ, Koopmans AE, Enthoven WT, et al: **Effect of supplementation with zinc and other micronutrients on malaria in Tanzanian children: a randomised trial.** *PLoS Med* 2011, **8**:e1001125.

21.   Shock NW, Greulich RC, Andres R, Arenberg D, Paul T. Costa J, Lakata EG, Tobin JD: **Normal Human Aging: The Baltimore Longitudinal Study of Aging.** 1984.

22.   Vitamin and Mineral Nutrition Information System. Geneva WHO, 2011 (WHO/NMH/NHD/MNM/11.1): **Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity**23.   Team RC: **R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.**

24.   An P, Wu Q, Wang H, Guan Y, Mu M, Liao Y, Zhou D, Song P, Wang C, Meng L, 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.

25.   Gan W, Guan Y, Wu Q, An P, Zhu J, Lu L, Jing L, Yu Y, Ruan S, Xie D, et al: **Association of TMPRSS6 polymorphisms with ferritin, hemoglobin, and type 2 diabetes risk in a Chinese Han population.** *Am J Clin Nutr* 2012, **95**:626-632.

26.   Finberg KE, Heeney MM, Campagna DR, Aydinok Y, Pearson HA, Hartman KR, Mayo MM, Samuel SM, Strouse JJ, Markianos K, et al: **Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA).** *Nat Genet* 2008, **40**:569-571.

27.   Du X, She E, Gelbart T, Truksa J, Lee P, Xia Y, Khovananth K, Mudd S, Mann N, Moresco EM, et al: **The serine protease TMPRSS6 is required to sense iron deficiency.** *Science* 2008, **320**:1088-1092.

28.   De Falco L, Sanchez M, Silvestri L, Kannengiesser C, Muckenthaler MU, Iolascon A, Gouya L, Camaschella C, Beaumont C: **Iron refractory iron deficiency anemia.** *Haematologica* 2013, **98**:845-853.

29.   Galesloot TE, Geurts-Moespot AJ, den Heijer M, Sweep FC, Fleming RE, Kiemeney LA, Vermeulen SH, Swinkels DW: **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.

30.   Traglia M, Girelli D, Biino G, Campostrini N, Corbella M, Sala C, Masciullo C, Vigano F, Buetti I, Pistis G, et al: **Association of HFE and TMPRSS6 genetic**

variants with iron and erythrocyte parameters is only in part dependent on serum hepcidin concentrations. *J Med Genet* 2011, **48**:629-634.

31.   Meynard D, Babitt JL, Lin HY: **The liver: conductor of systemic iron balance.** *Blood* 2014, **123**:168-176.

32.   McLaren CE, Garner CP, Constantine CC, McLachlan S, Vulpe CD, Snively BM, Gordeuk VR, Nickerson DA, Cook JD, Leiendecker-Foster C, et al: **Genome-wide association study identifies genetic loci associated with iron deficiency.** *PLoS One* 2011, **6**:e17390.

33.   Lee PL, Halloran C, Trevino R, Felitti V, Beutler E: **Human transferrin G277S mutation: a risk factor for iron deficiency anaemia.** *Br J Haematol* 2001, **115**:329-333.

34.   Aslan D, Crain K, Beutler E: **A new case of human atransferrinemia with a previously undescribed mutation in the transferrin gene.** *Acta Haematol* 2007, **118**:244-247.

35.   Beutler E, Gelbart T, Lee P, Trevino R, Fernandez MA, Fairbanks VF: **Molecular characterization of a case of atransferrinemia.** *Blood* 2000, **96**:4071-4074.

36.   Trenor CC, 3rd, Campagna DR, Sellers VM, Andrews NC, Fleming MD: **The molecular defect in hypotransferrinemic mice.** *Blood* 2000, **96**:1113-1118.

37.   Jonker FA, Boele van Hensbroek M: **Anaemia, iron deficiency and susceptibility to infections.** *J Infect* 2014.

38.   Trey JE, Kushner I: **The acute phase response and the hematopoietic system: the role of cytokines.** *Crit Rev Oncol Hematol* 1995, **21**:1-18.

39.   Zimmermann MB, Chaouki N, Hurrell RF: **Iron deficiency due to consumption of a habitual diet low in bioavailable iron: a longitudinal cohort study in Moroccan children.** *Am J Clin Nutr* 2005, **81**:115-121.

40.   Coad J, Conlon C: **Iron deficiency in women: assessment, causes and consequences.** *Curr Opin Clin Nutr Metab Care* 2011, **14**:625-634.

41.   Garn SM, Ryan AS, Owen GM, Abraham S: **Income matched black-white hemoglobin differences after correction for low transferrin saturations.** *Am J Clin Nutr* 1981, **34**:1645-1647.

42.   Yip R, Johnson C, Dallman PR: **Age-related changes in laboratory values used in the diagnosis of anemia and iron deficiency.** *Am J Clin Nutr* 1984, **39**:427-436.

43.   Scherag A, Muller HH, Dempfle A, Hebebrand J, Schafer H: **Data adaptive interim modification of sample sizes for candidate-gene association studies.** *Hum Hered* 2003, **56**:56-62.

44.   Gordon D, Levenstien MA, Finch SJ, Ott J: **Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case-control association studies.** *Pac Symp Biocomput* 2003:490-501.

45.   Pfeiffer RM, Gail MH: **Sample size calculations for population- and family-based case-control association studies on marker genotypes.** *Genet Epidemiol* 2003, **25**:136-148.

46.   Tishkoff SA, Kidd KK: **Implications of biogeography of human populations for 'race' and medicine.** *Nat Genet* 2004, **36**:S21-27.

47.   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.

48.   1000 genomes project website (Accessed on 1st October 2014)

        http://browser.1000genomes.org/index.html

Supplemental Appendices

Supplementary Table 1: Associations of single nucleotide polymorphism with iron status parameters

Trait	Gene & SNP	Chr	Minor Allele	Ke β (se)	P	Tz β (se)	P	SA† β (se)	P	Met_African β (se)	P	P (Het)	African American β (se)	P	Met_All β (se)	P	P (Het)
Hb g/dL	*OPRD1*																
	rs482692	1	A	-0.20(0.20)	0.33	-2.56(1.33)	0.06	0.09(0.12)	0.45	-0.06(0.16)	0.70	0.31	-0.56(0.28)	0.05	0.22(0.20)	0.27	0.03
	SRPRB																
	rs1830084	3	T	-0.14(0.14)	0.32	0.16(1.04)	0.88	-0.18(0.14)	0.20	-0.16(0.10)	0.11	0.93	0.11(0.16)	0.47	-0.08(0.08)	0.32	0.54
	TF																
	rs1358024	3	A	0.19(0.27)	0.49	3.52(4.68)	0.45	0.87(0.54)	0.11	0.39(0.31)	0.10	0.42	-0.21(0.36)	0.57	0.17(0.20)	0.41	0.35
	rs1525892	3	A	0.12(0.11)	0.29	0.11(0.81)	0.90	0.04(0.12)	0.74	0.08(0.08)	0.16	0.89	0.12(0.14)	0.37	0.09(0.07)	0.18	0.96
	rs1799852	3	A	-0.32(0.18)	0.07	1.64(1.62)	0.31	0.41(0.28)	0.14	0.09(0.36)	0.80	0.05	0.05(0.25)	0.84	0.03(0.22)	0.87	0.10
	rs1867504	3	A	-0.02(0.12)	0.90	-0.12(0.82)	0.88	-0.06(0.12)	0.59	-0.04(0.08)	0.31	0.97	0.04(0.14)	0.75	-0.02(0.07)	0.79	0.96
	rs3811647	3	A	0.08(0.13)	0.53	0.53(0.88)	0.55	0.02(0.11)	0.83	0.05(0.08)	0.26	0.81	0.10(0.15)	0.51	0.06(0.07)	0.40	0.92
	rs3811658	3	T	0.01(0.13)	0.95	1.57(0.99)	0.11	0.10(0.04)	0.49	**0.36(0.04)**	**0.02**	0.26	0.08(0.16)	0.61	**0.09(0.03)**	**0.01**	**0.45**
	rs4525863	3	A	0.17(0.13)	0.17	-0.68(0.8)	0.39	-0.24(0.14)	0.09	-0.07(0.19)	0.36	0.07	-0.20(0.13)	0.14	-0.10(0.13)	0.43	0.09
	rs7638018	3	G	0.06(0.12)	0.63	0.34(0.94)	0.72	0.07(0.14)	0.62	0.07(0.09)	0.22	0.96	-0.18(0.16)	0.26	0.09(0.08)	0.23	0.93
	TNF																
	rs1799964	6	G	0.04(0.13)	0.74	0.65(0.89)	0.47	-0.16(0.13)	0.21	0.10(0.07)	0.15	0.40	-0.05(0.17)	0.75	-0.05(0.08)	0.52	0.62
	rs1800629	6	A	-0.03(0.18)	0.86	2.02(1.27)	0.11	-0.02(0.14)	0.87	-0.08(0.07)	0.25	0.28	0.18(0.20)	0.39	0.04(0.10)	0.71	0.35
	TFR2																
	rs7385804	7	C	-0.12(0.13)	0.36	1.16(0.85)	0.17	0.09(0.12)	0.43	0.01(0.08)	0.53	0.26	0.45(0.38)	0.25	0.05(0.08)	0.55	0.26
	CUBN																
	rs10904850	10	A	**-0.38(0.16)**	**0.02**	0.91(1.19)	0.44	0.07(0.11)	0.56	-0.11(0.10)	0.14	0.05	0.03(0.16)	0.86	-0.07(0.14)	0.63	0.10
	HAMP																
	rs10421768	19	A	**0.37(0.14)**	**0.01**	0.46(0.94)	0.62	-0.01(0.15)	0.93	0.19(0.18)	0.29	0.17	0.14(0.12)	0.12	0.17(0.10)	0.09	0.30
	KIAA1468																

Trait	SNP	Chr	Allele	β(SE)	P	β(SE)	P	β(SE)	P	β(SE)	P		β(SE)	P	β(SE)	P	
	rs9948708	18	A	0.03(0.11)	0.82	0.03(0.73)	0.97	0.14(0.1)	0.18	0.09(0.07)	0.20	0.76	0.01(0.12)	0.91	0.07(0.06)	0.28	0.83
	TMPRSS6																
	rs1421312	22	G	-0.11(0.10)	0.27	0.09(0.76)	0.91	-0.05(0.11)	0.64	-0.18(0.11)	0.10	0.90	-0.14(0.12)	0.24	-0.10(0.06)	0.12	0.94
	rs228918	22	G	0.14(0.10)	0.16	-0.86(0.74)	0.25	0.08(0.1)	0.44	0.10(0.07)	0.15	0.39	-0.08(0.12)	0.49	0.05(0.06)	0.37	0.31
	rs2413450	22	A	-0.20(0.12)	0.10	-0.65(1.13)	0.57	-0.23(0.18)	0.20	**-0.23(0.10)**	**0.01**	0.93	-0.09(0.14)	0.55	**-0.19(0.08)**	**0.02**	**0.84**
	rs4820268	22	G	**-0.28(0.11)**	**0.01**	-0.26(0.87)	0.76	-0.05(0.14)	0.74	**-0.18(0.07)**	**0.01**	0.43	-0.09(0.13)	0.48	**-0.16(0.07)**	**0.04**	**0.55**
Ferritin µg/L	*OPRD1*																
	rs482692	1	A	-0.95(1.10)	0.63	-0.99(1.12)	0.94	1.08(1.08)	0.45	-0.27(0.39)	0.70	0.31	-0.78(1.20)	0.16	-0.37(0.56)	0.51	0.47
	SRPRB																
	rs1830084	3	T	-0.95(1.07)	0.45	-0.58(1.29)	0.47	-1.00(1.10)	0.99	-0.87(0.66)	0.09	1.00	1.02(1.12)	0.79	-0.38(0.57)	0.47	0.53
	TF																
	rs1358024	3	A	-1.00(1.14)	0.97	-0.86(1.50)	0.72	-0.92(1.32)	0.76	-0.94(0.75)	0.11	1.00	-0.98(1.29)	0.94	-0.95(0.65)	0.14	1.00
	rs1525892	3	A	1.01(1.06)	0.80	1.05(1.07)	0.46	1.09(1.08)	0.26	1.05(0.62)	0.05	1.00	1.10(1.10)	0.37	**1.07(0.54)**	**0.04**	1.00
	rs1799852	3	A	-0.98(1.09)	0.80	1.20(1.15)	0.18	**-0.62(1.2)**	**0.01**	-0.15(0.68)	0.41	0.34	1.23(1.17)	0.21	0.18(0.60)	0.76	0.37
	rs1867504	3	A	1.04(1.06)	0.54	1.05(1.07)	0.52	1.03(1.08)	0.69	1.04(0.62)	0.05	1.00	1.05(1.10)	0.62	1.04(0.54)	0.05	1.00
	rs3811647	3	A	1.03(1.07)	0.66	1.02(1.07)	0.85	1.10(1.07)	0.19	1.05(0.62)	0.09	1.00	1.07(1.10)	0.45	1.05(0.54)	0.05	1.00
	rs3811658	3	T	1.05(1.07)	0.48	1.05(1.10)	0.63	-1.02(1.09)	0.82	0.36(0.69)	0.60	0.30	-0.98(1.12)	0.78	0.04(0.59)	0.95	0.32
	rs4525863	3	A	1.09(1.06)	0.18	1.07(1.07)	0.26	-0.98(1.09)	0.83	0.41(0.68)	0.54	0.30	-0.98(1.10)	0.78	0.07(0.59)	0.90	0.30
	rs7638018	3	G	1.00(1.06)	0.97	-0.96(1.10)	0.62	1.00(1.09)	0.99	0.37(0.64)	0.57	0.34	1.02(1.12)	0.76	0.52(0.55)	0.34	0.49
	TNF																
	rs1799964	6	G	-0.92(1.06)	0.17	-0.99(1.07)	0.92	-0.88(1.08)	0.12	-0.93(0.62)	0.13	1.00	1.05(1.12)	0.62	-0.47(0.54)	0.39	0.49
	rs1800629	6	A	1.02(1.10)	0.85	1.05(1.12)	0.62	1.03(1.09)	0.74	1.03(0.64)	0.10	1.00	**-0.76(1.15)**	**0.03**	0.61(0.56)	0.27	0.60
	TFR2																
	rs7385804	7	C	**-0.86(1.07)**	**0.02**	-0.95(1.08)	0.48	1.15(1.08)	0.06	1.03(0.64)	0.05	0.27	1.07(1.10)	0.52	0.09(0.58)	0.88	0.32
	CUBN																
	rs10904850	10	A	-0.92(1.08)	0.31	1.02(1.12)	0.81	1.04(1.08)	0.61	0.37(0.66)	0.29	0.34	1.07(1.12)	0.47	0.55(0.96)	0.17	0.49
	HAMP																
	rs10421768	19	A	1.02(1.07)	0.73	1.02(1.10)	0.72	1.06(1.10)	0.53	1.03(0.63)	0.10	1.00	1.00(1.12)	0.99	1.02(0.55)	0.06	1.00

SNP	Chr	Allele	Ke beta(SE)	Ke P	Tz beta(SE)	Tz P	SA beta(SE)	SA P	Met-African beta(SE)	Met-African P	P(Het)	AA beta(SE)	AA P	Met_All beta(SE)	Met_All P	P(Het)
KIAA1468																
rs9948708	18	A	-0.98(1.05)	0.73	1.02(1.07)	0.82	1.03(1.07)	0.71	0.34(0.68)	0.62	0.30	-0.91(1.07)	0.24	0.03(0.57)	0.96	0.33
TMPRSS6																
rs1421312	22	G	-0.94(1.02)	0.22	1.02(1.07)	0.68	1.01(1.07)	0.90	0.33(0.67)	0.60	0.90	-0.89(1.07)	0.15	0.03(0.55)	0.56	0.34
rs228918	22	G	-0.99(1.04)	0.90	1.05(1.07)	0.55	-1.02(1.07)	0.75	-0.33(0.68)	0.62	0.32	1.10(1.10)	0.25	**0.01(0.60)**	**0.02**	0.29
rs2413450	22	A	1.08(1.05)	0.11	1.05(1.10)	0.73	1.02(1.12)	0.90	1.05(0.63)	0.09	1.00	-0.93(1.10)	0.49	0.56(0.55)	0.30	0.49
rs4820268	22	G	-0.99(1.04)	0.83	-0.96(1.07)	0.61	-0.95(1.09)	0.60	-0.97(0.62)	0.12	1.00	-0.98(1.10)	0.82	-0.97(0.54)	0.07	1.00

SNP=Single nucleotide polymorphism; Ke=Kenyan cohort; Tz=Tanzanian cohort; SA=South African cohort; AA=African American cohort; P(Het)=P value of heterogeneity testing; Met-African=meta analyses of beta and SE values from African cohorts; Met_All= meta analyses of beta and SE values from African cohorts; Direction=direction of association based on beta value Hb=Haemoglobin, Values in bold are significant

Results are presented as mean (SE) change per copy of the minor allele from regression analysis using additive genetic model and with adjustment for age and sex for associations involving Hb and age and sex and CRP concentrations for associations involving ferritin

† Values not adjusted for gender as all participants were women

Values in bold are significant before FDR correction for multiple testing

Variance explained ranged from < 1% to 3 %, with the highest being 3% for rs10421768 in the Kenyan population
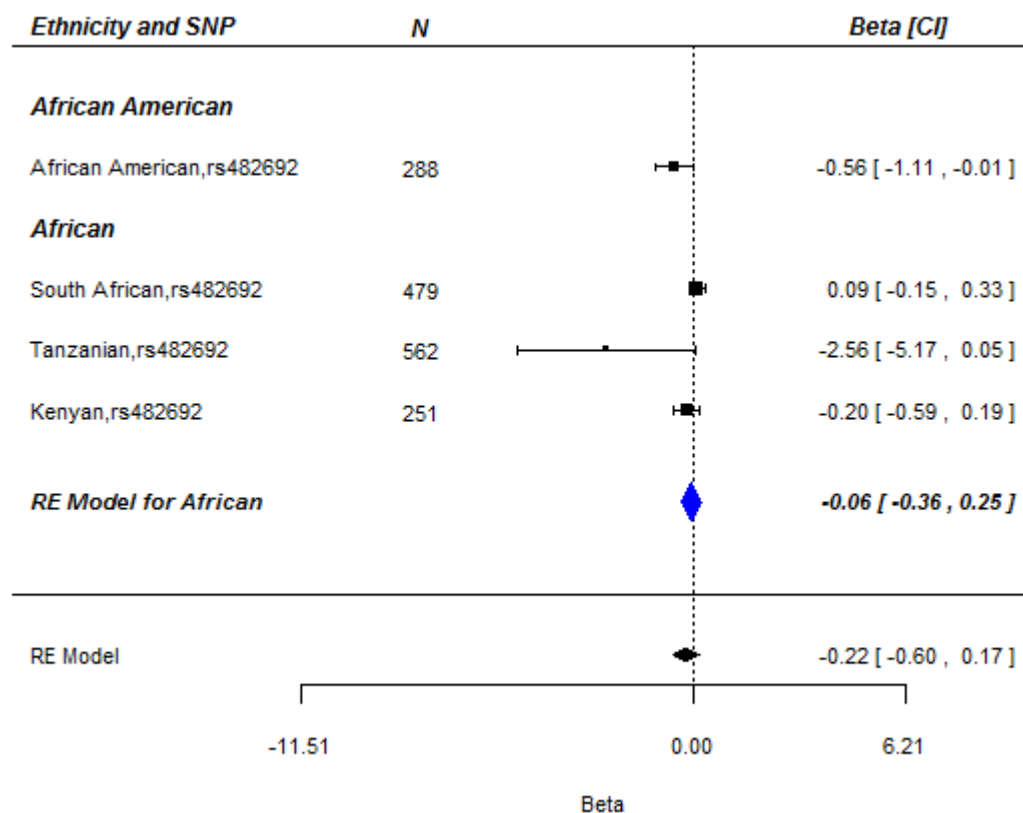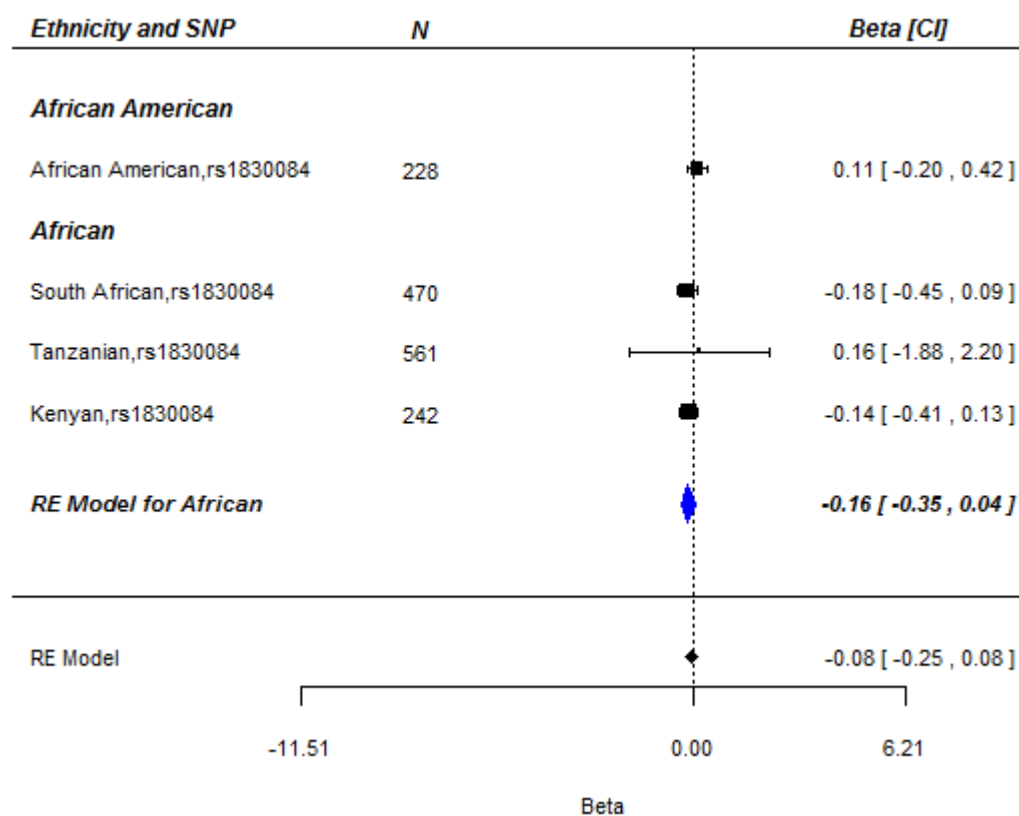
Supplementary Figure 1a: *OPRD1* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A
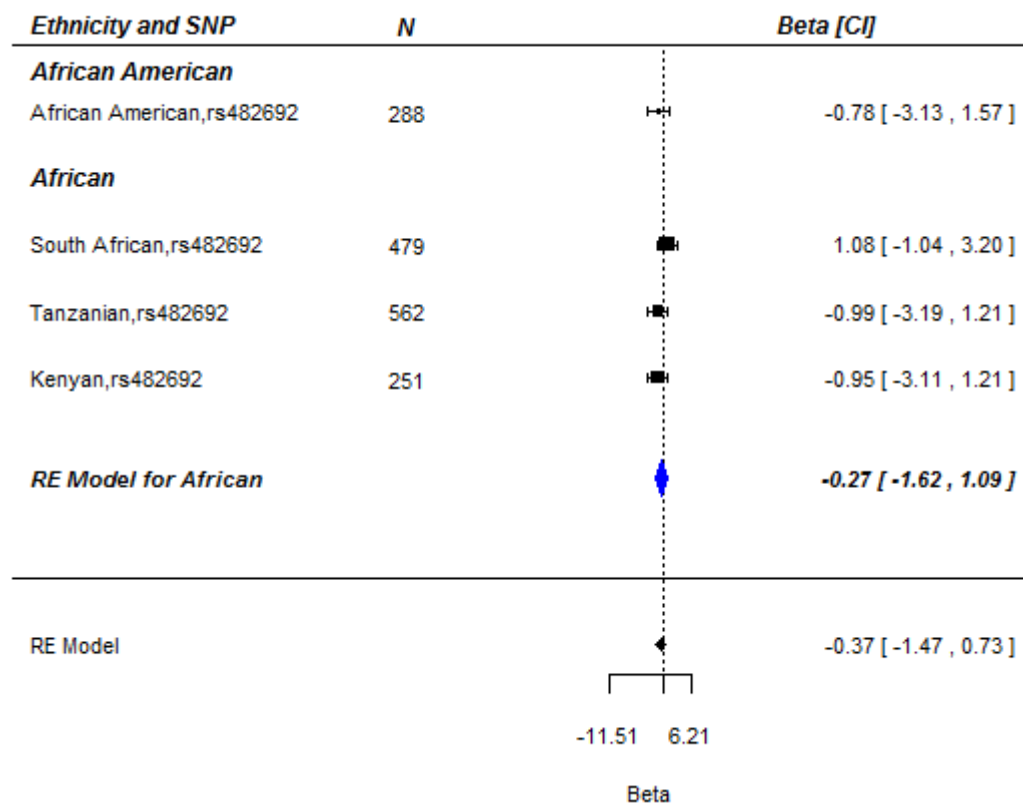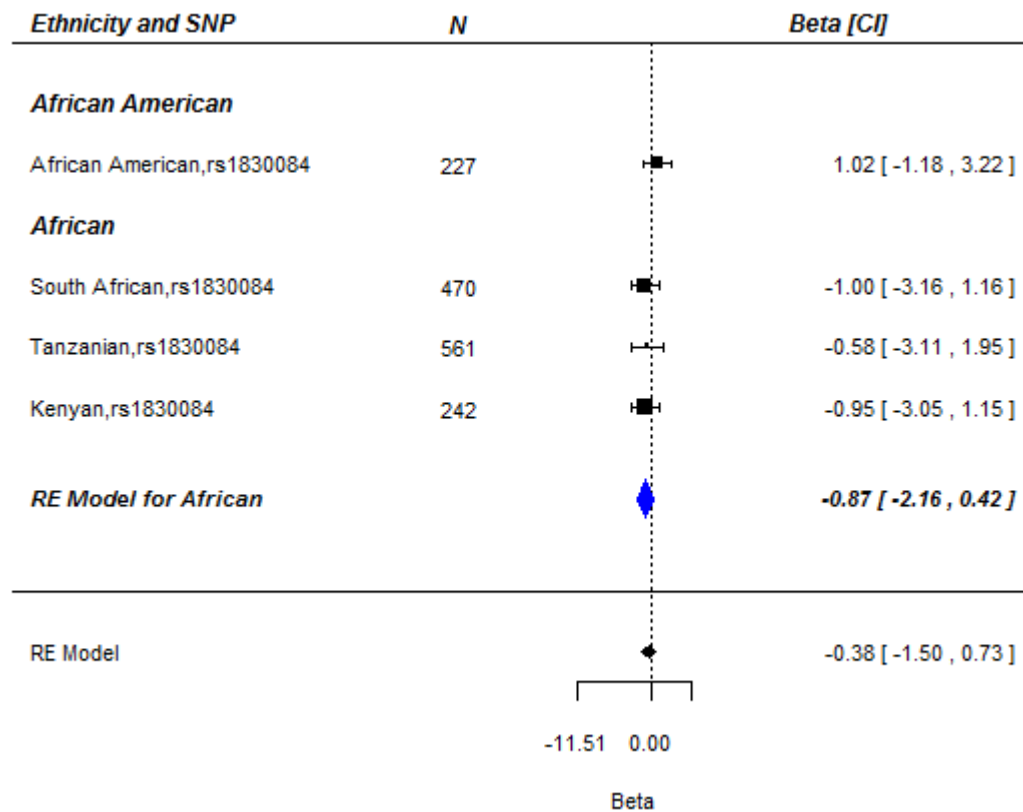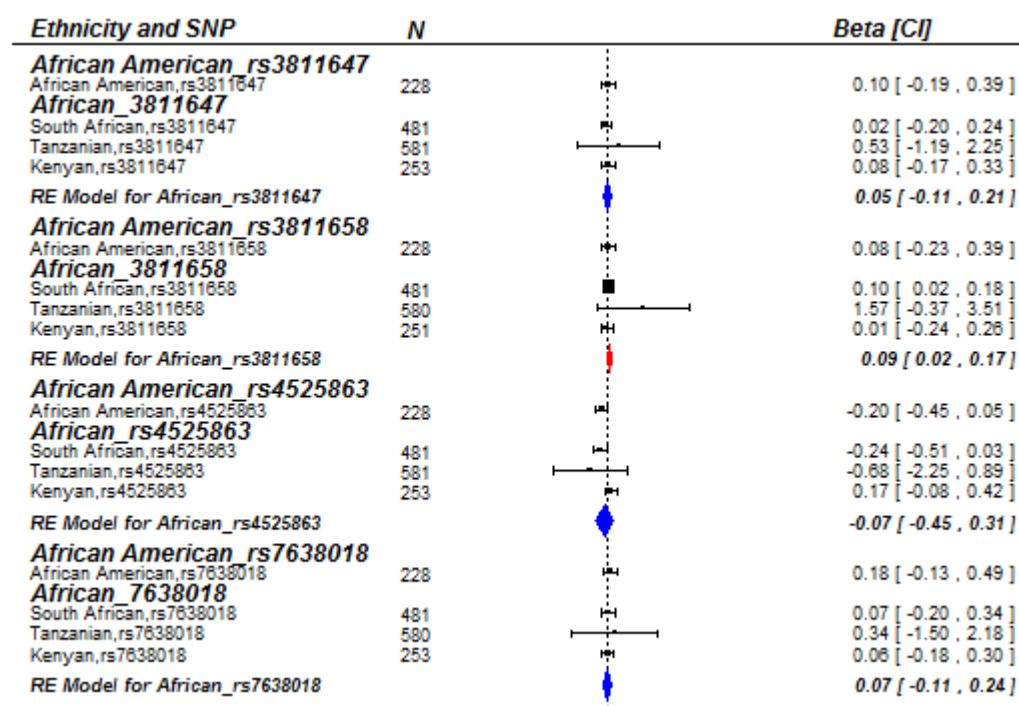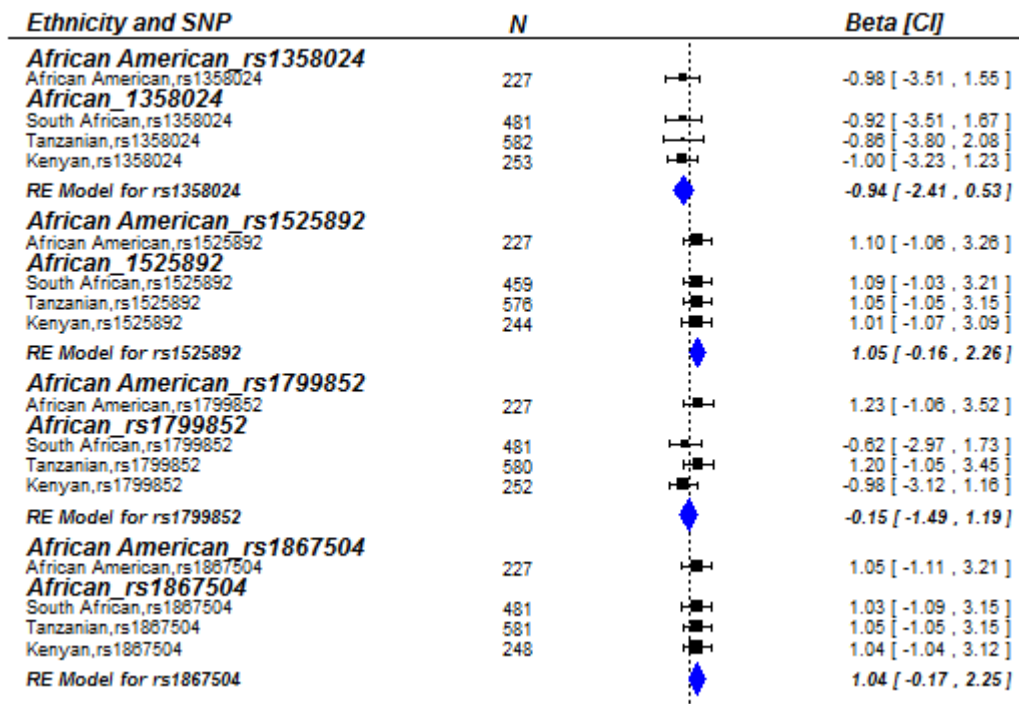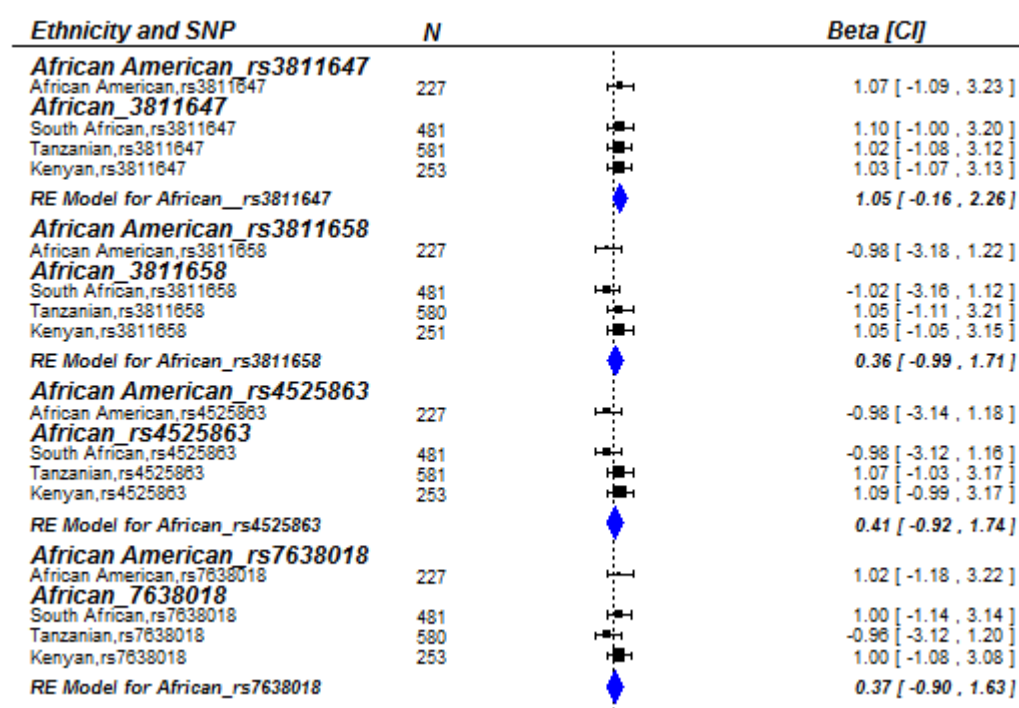
Supplementary Figure 1b:*OPRD1* SNPs and associations with serum ferritin (µg/L)

Values are Log$_{10}$ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A
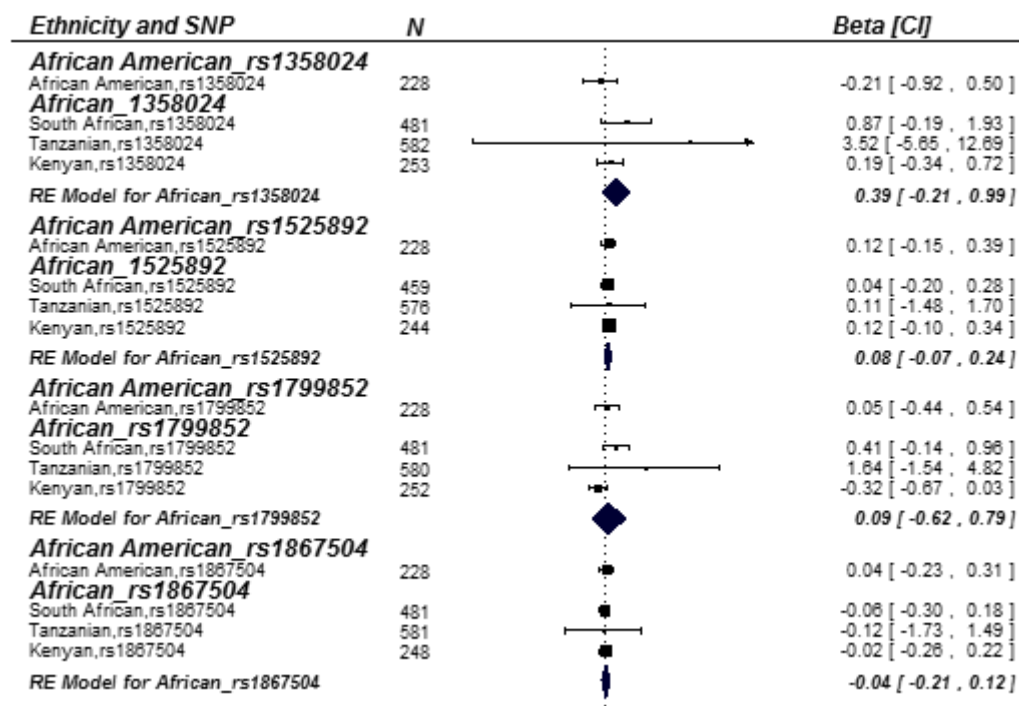
Supplementary Figure 2a: *SRPRB1* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = T

Supplementary Figure 2b: *SRPRB1* SNPs and associations with ferritin μg/L

Ferritin values are Log₁₀ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = T

Ethnicity and SNP	N	Beta [CI]
African American_rs3811647		
African American,rs3811647	228	0.10 [-0.19 , 0.39]
African_3811647		
South African,rs3811647	481	0.02 [-0.20 , 0.24]
Tanzanian,rs3811647	581	0.53 [-1.19 , 2.25]
Kenyan,rs3811647	253	0.08 [-0.17 , 0.33]
RE Model for African_rs3811647		*0.05 [-0.11 , 0.21]*
African American_rs3811658		
African American,rs3811658	228	0.08 [-0.23 , 0.39]
African_3811658		
South African,rs3811658	481	0.10 [0.02 , 0.18]
Tanzanian,rs3811658	580	1.57 [-0.37 , 3.51]
Kenyan,rs3811658	251	0.01 [-0.24 , 0.26]
RE Model for African_rs3811658		*0.09 [0.02 , 0.17]*
African American_rs4525863		
African American,rs4525863	228	-0.20 [-0.45 , 0.05]
African_rs4525863		
South African,rs4525863	481	-0.24 [-0.51 , 0.03]
Tanzanian,rs4525863	581	-0.68 [-2.25 , 0.89]
Kenyan,rs4525863	253	0.17 [-0.08 , 0.42]
RE Model for African_rs4525863		*-0.07 [-0.45 , 0.31]*
African American_rs7638018		
African American,rs7638018	228	0.18 [-0.13 , 0.49]
African_7638018		
South African,rs7638018	481	0.07 [-0.20 , 0.34]
Tanzanian,rs7638018	580	0.34 [-1.50 , 2.18]
Kenyan,rs7638018	253	0.06 [-0.18 , 0.30]
RE Model for African_rs7638018		*0.07 [-0.11 , 0.24]*

Supplementary Figure 3a: *TF* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = T for rs3811647, rs3811658, rs4525863, and G for rs7638018
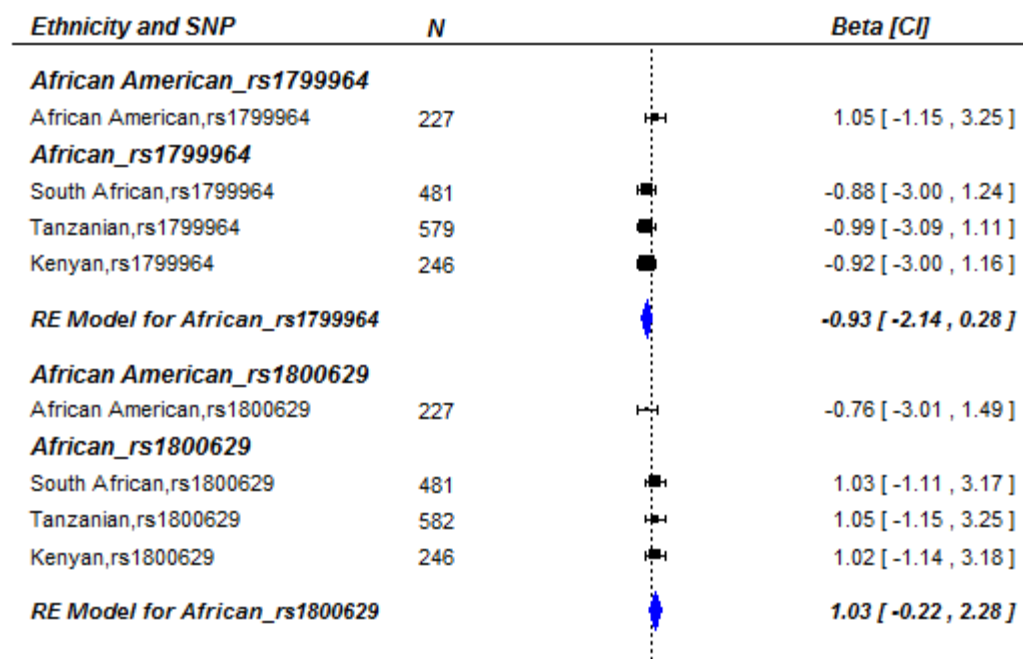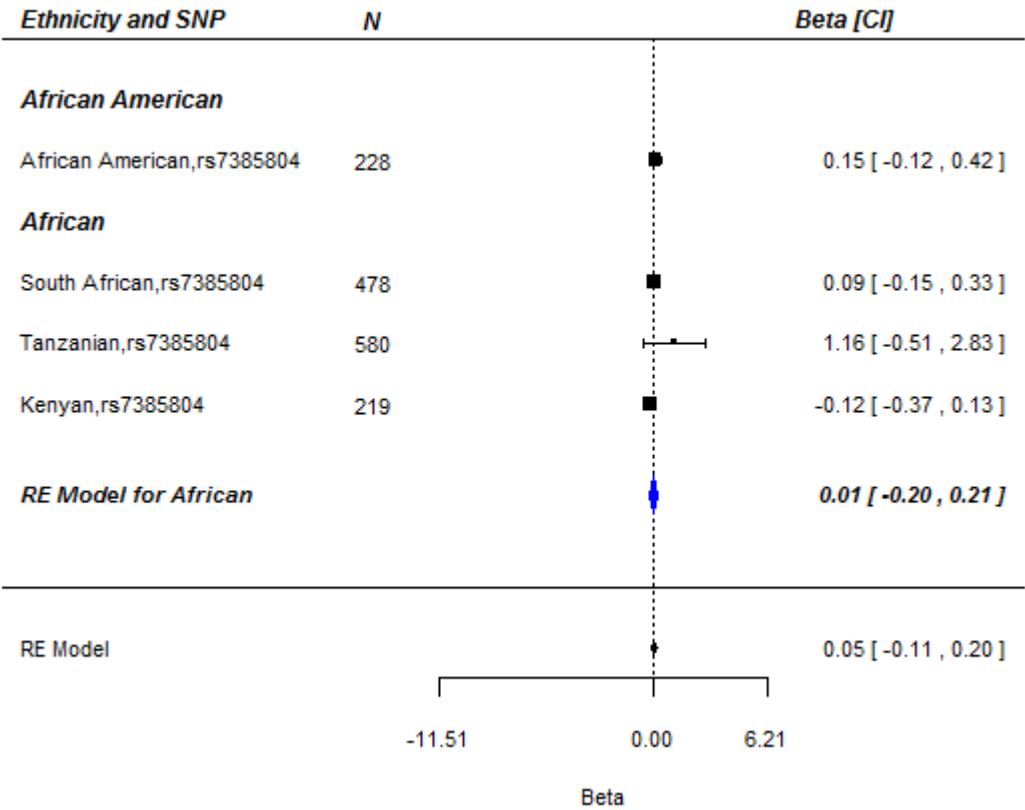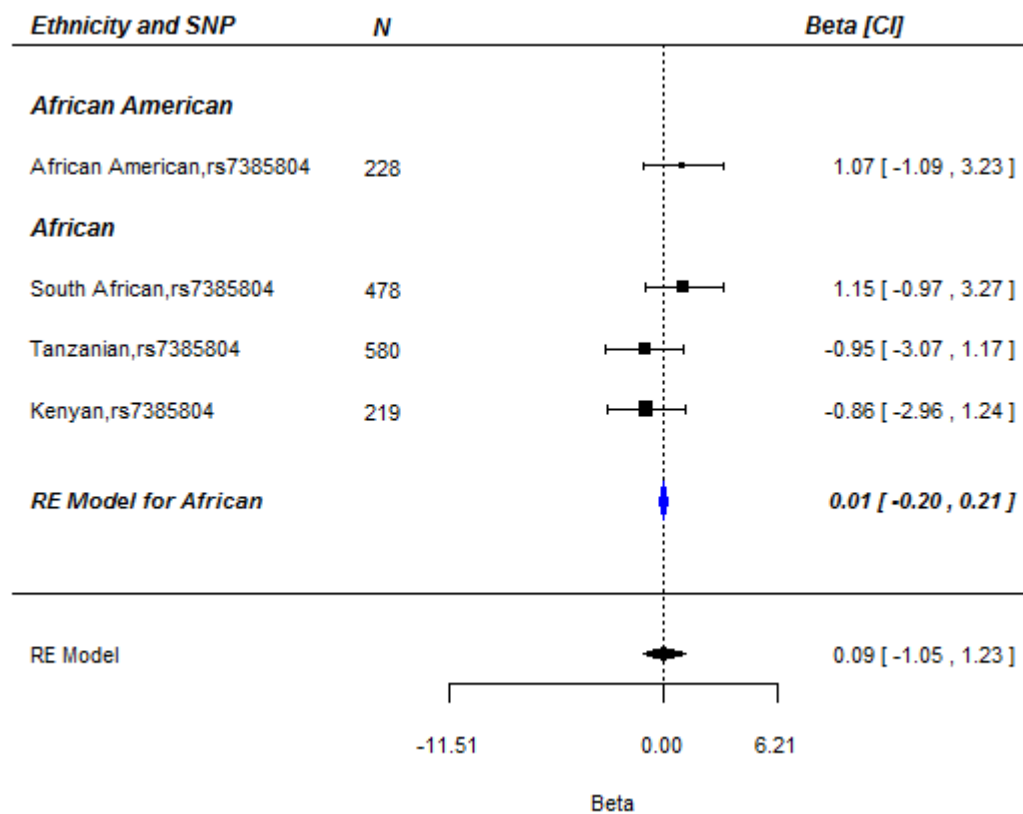
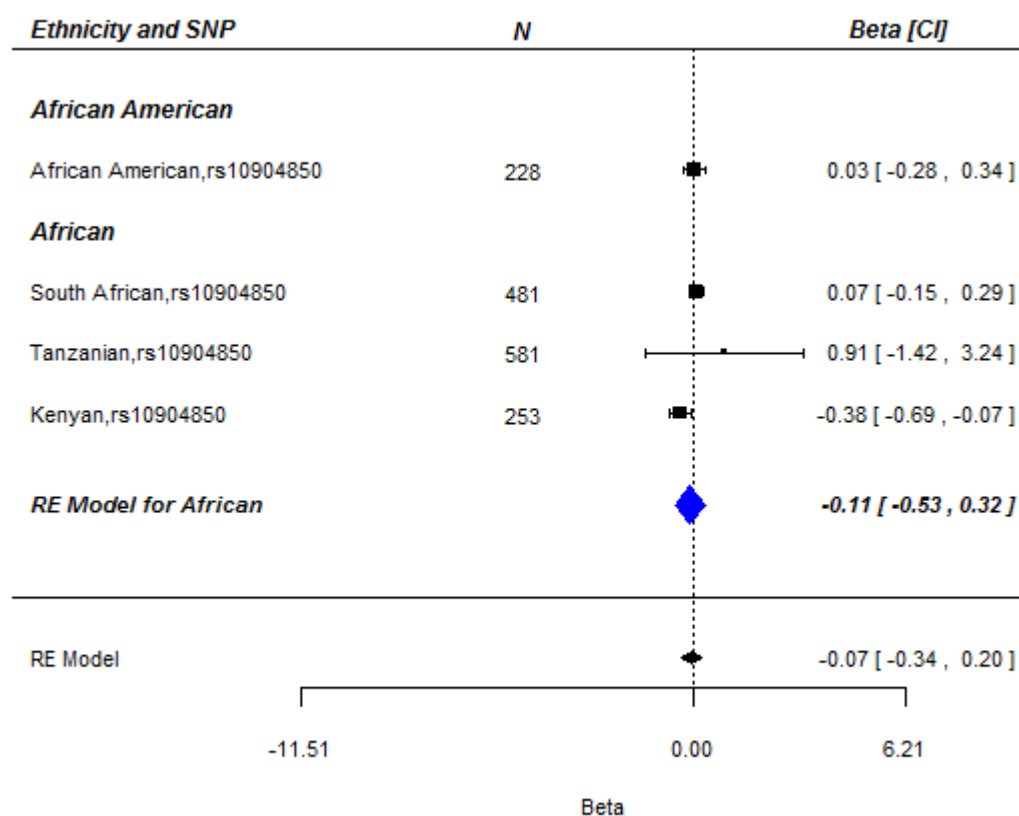**Supplementary Figure 3b**: *TF* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Red indicate summary estimates that were significant. Effect allele = A for rs1358024, rs1525892, rs1799852 and rs1867504

Ethnicity and SNP	N		Beta [CI]
African American_rs3811647			
African American,rs3811647	227		1.07 [-1.09 , 3.23]
African_3811647			
South African,rs3811647	481		1.10 [-1.00 , 3.20]
Tanzanian,rs3811647	581		1.02 [-1.08 , 3.12]
Kenyan,rs3811647	253		1.03 [-1.07 , 3.13]
RE Model for African__rs3811647			*1.05 [-0.16 , 2.26]*
African American_rs3811658			
African American,rs3811658	227		-0.98 [-3.18 , 1.22]
African_3811658			
South African,rs3811658	481		-1.02 [-3.16 , 1.12]
Tanzanian,rs3811658	580		1.05 [-1.11 , 3.21]
Kenyan,rs3811658	251		1.05 [-1.05 , 3.15]
RE Model for African_rs3811658			*0.36 [-0.99 , 1.71]*
African American_rs4525863			
African American,rs4525863	227		-0.98 [-3.14 , 1.18]
African_rs4525863			
South African,rs4525863	481		-0.98 [-3.12 , 1.16]
Tanzanian,rs4525863	581		1.07 [-1.03 , 3.17]
Kenyan,rs4525863	253		1.09 [-0.99 , 3.17]
RE Model for African_rs4525863			*0.41 [-0.92 , 1.74]*
African American_rs7638018			
African American,rs7638018	227		1.02 [-1.18 , 3.22]
African_7638018			
South African,rs7638018	481		1.00 [-1.14 , 3.14]
Tanzanian,rs7638018	580		-0.96 [-3.12 , 1.20]
Kenyan,rs7638018	253		1.00 [-1.08 , 3.08]
RE Model for African_rs7638018			*0.37 [-0.90 , 1.63]*

Supplementary Figure 4a: *TF* SNPs and associations with ferritin μg/L

Ferritin values are Log$_{10}$ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = T for rs3811647, rs3811658, rs4525863, and G for rs7638018
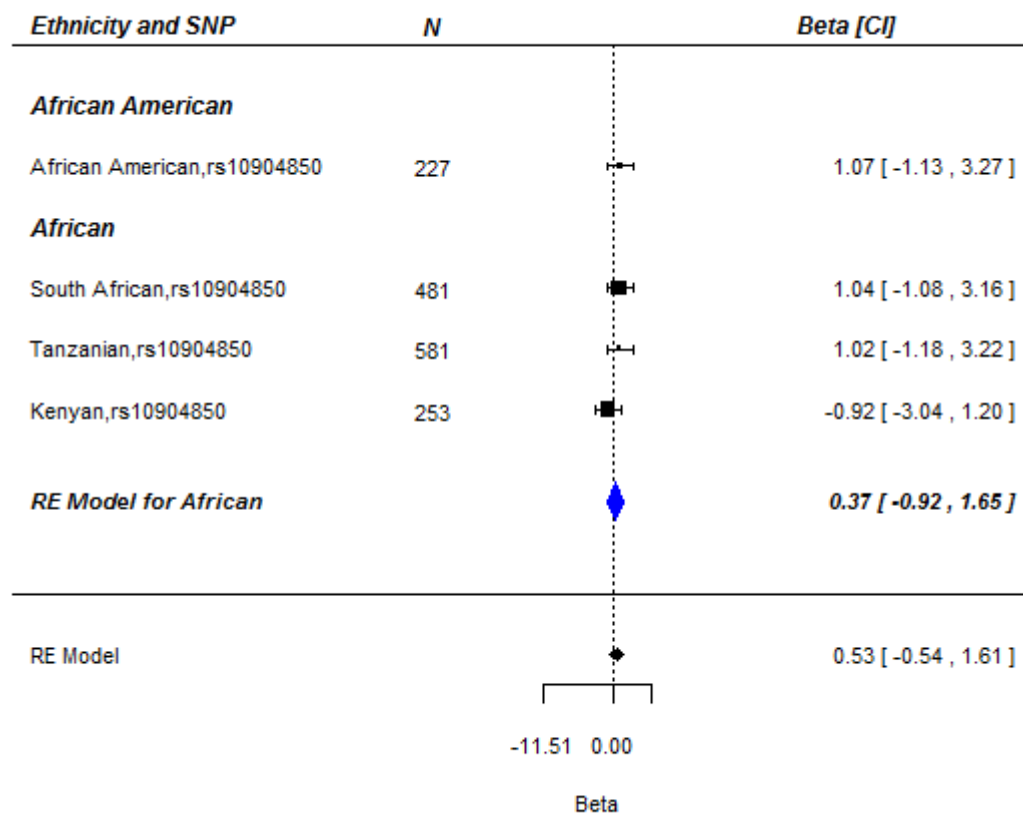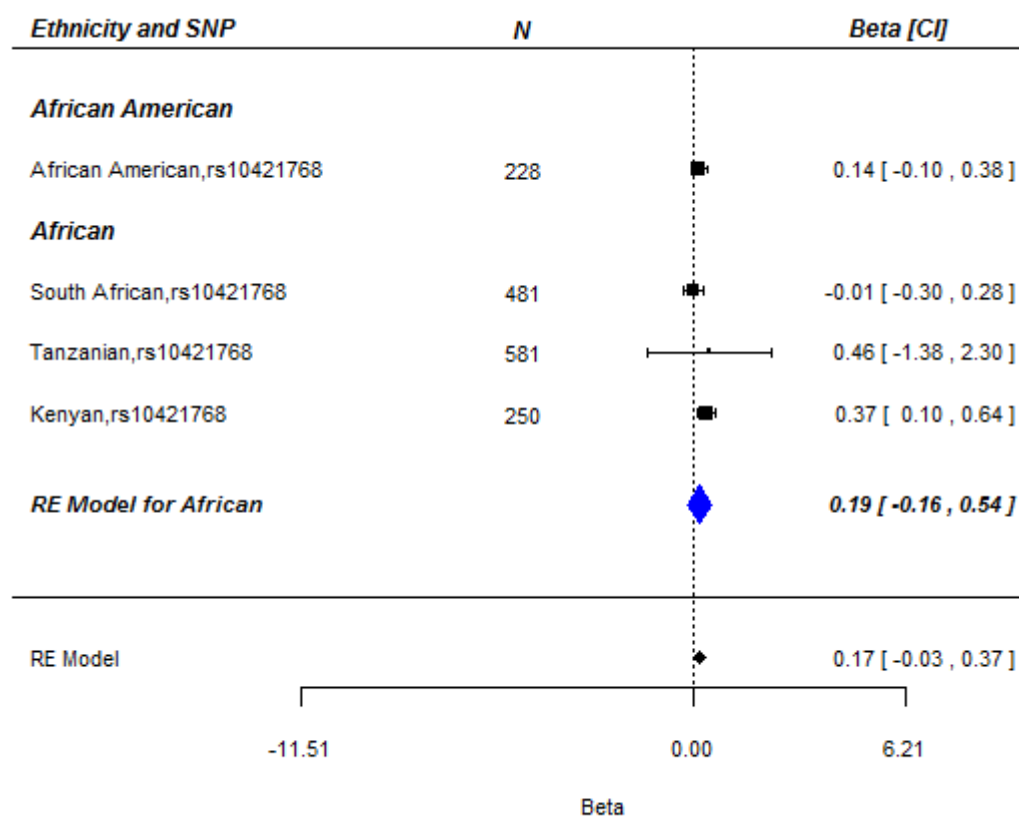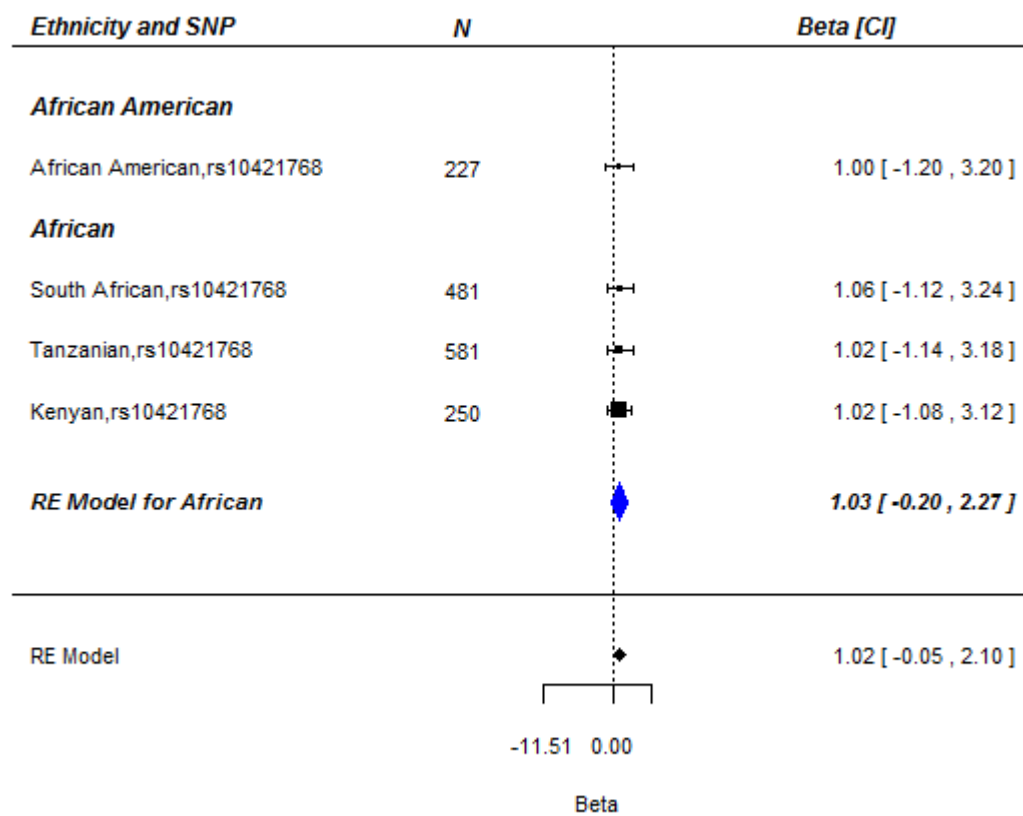
Supplementary Figure 4b: *TF* SNPs and associations with ferritin µg/L

Ferritin values are Log$_{10}$ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Red indicate summary estimates that were significant. Effect allele = A for rs1358024, rs1525892, rs1799852 and rs1867504

Ethnicity and SNP	N		Beta [CI]
African American_rs1799964			
African American,rs1799964	228		-0.05 [-0.38 , 0.28]
African_rs1799964			
South African,rs1799964	481		-0.16 [-0.41 , 0.09]
Tanzanian,rs1799964	579		0.65 [-1.09 , 2.39]
Kenyan,rs1799964	246		0.04 [-0.21 , 0.29]
RE Model for African_rs1799964			*-0.05 [-0.24 , 0.14]*
African American_rs1800629			
African American,rs1800629	228		0.18 [-0.21 , 0.57]
African_rs1800629			
South African,rs1800629	481		-0.02 [-0.29 , 0.25]
Tanzanian,rs1800629	582		2.02 [-0.47 , 4.51]
Kenyan,rs1800629	246		-0.03 [-0.38 , 0.32]
RE Model for African_rs1800629			*-0.01 [-0.22 , 0.21]*

Supplementary Figure 5a: *TNF α* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for s1800629 and G for rs1799964
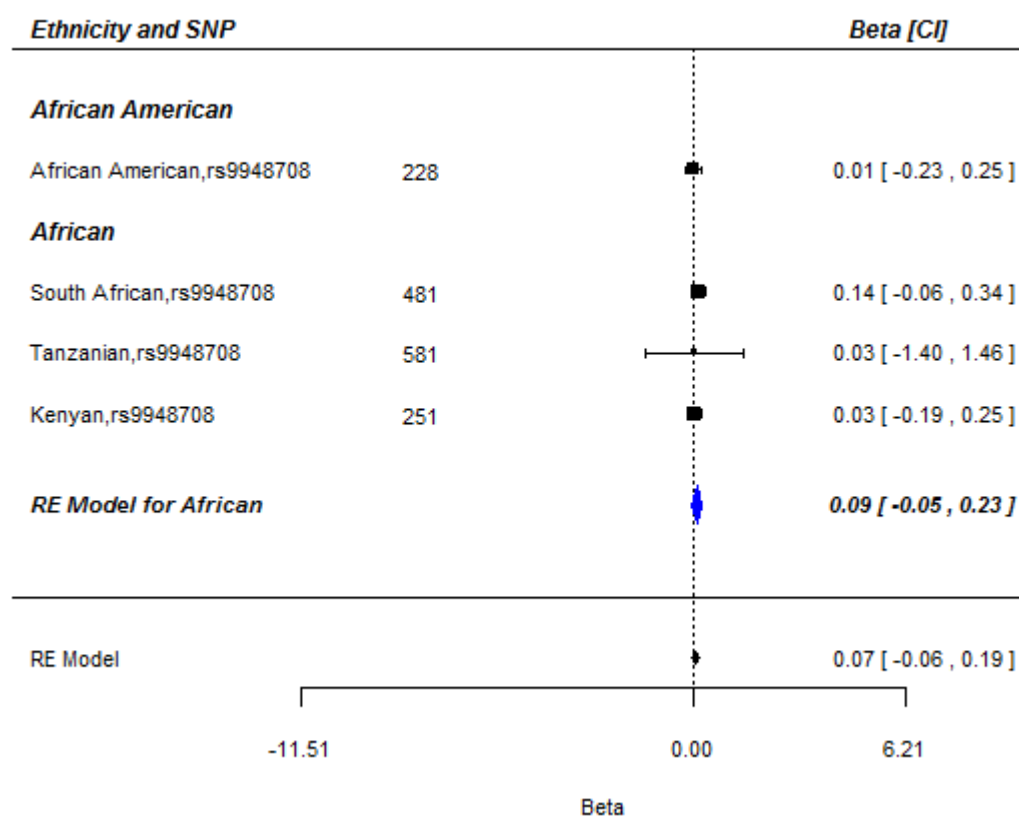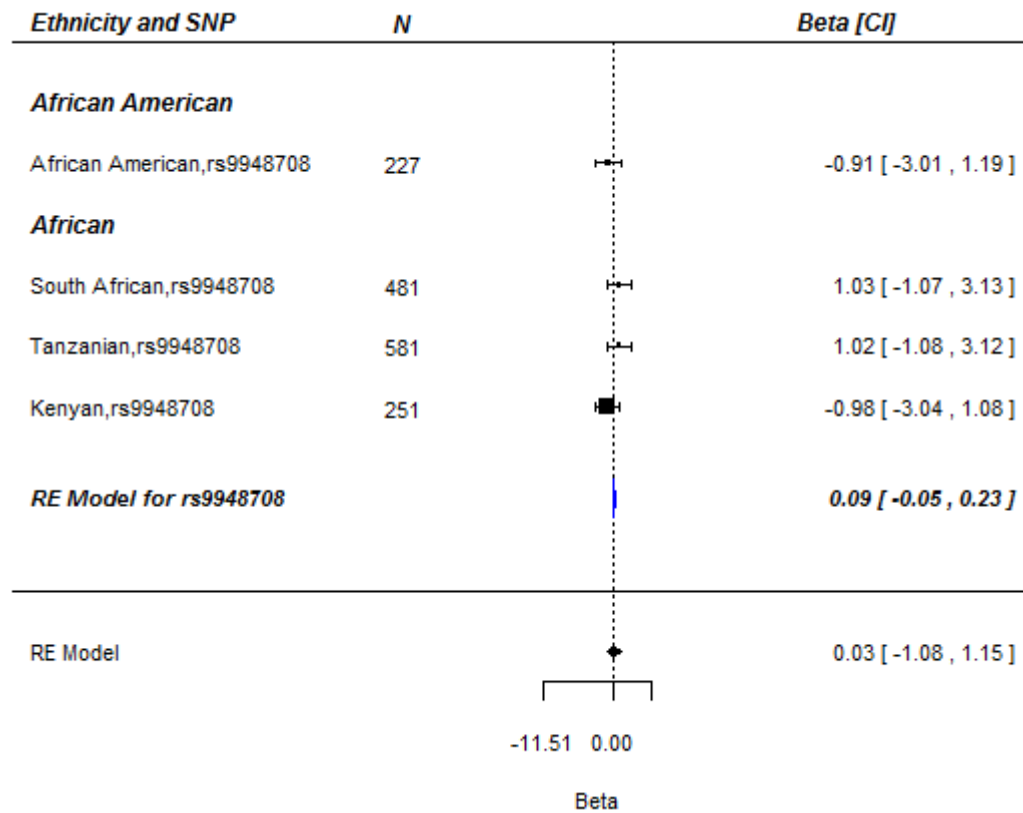
Supplementary Figure 5b: *TNF α* SNPs and associations with ferritin μg/L

Ferritin values are Log₁₀ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for s1800629 and G for rs1799964
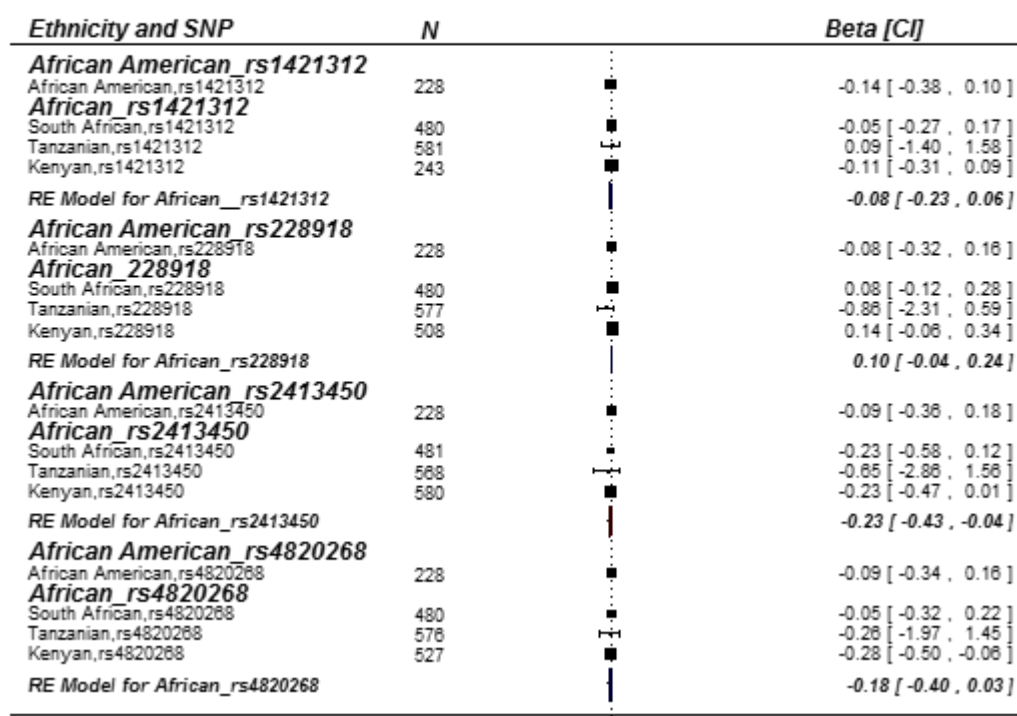
Supplementary Figure 6a: *TFR2* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = G for rs7385804
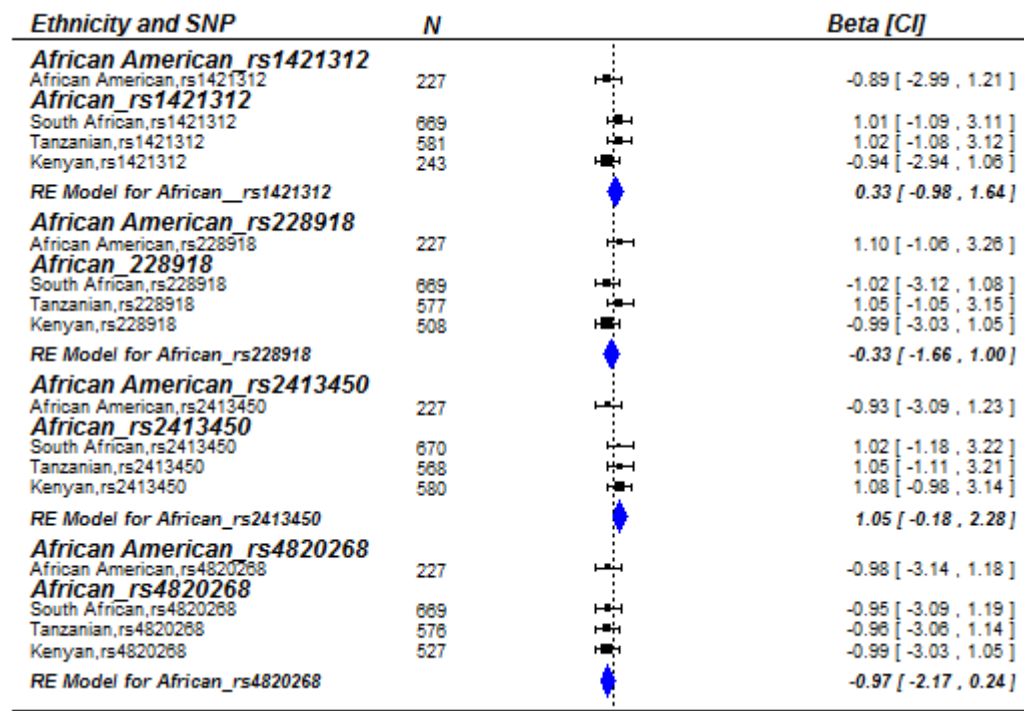
Supplementary Figure 6b: *TFR2* SNPs and associations with ferritin µg/L

Ferritin values are Log_{10} back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = G for rs7385804

Supplementary Figure 7a: *CUBN* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs10904850

Supplementary Figure 7b: *CUBN* SNPs and associations with ferritin µg/L

Ferritin values are Log10 back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs10904850

Supplementary Figure 8a: *HAMP* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates Effect allele = A for rs10421768

Supplementary Figure 8b: *HAMP* SNPs and associations with ferritin μg/L

Ferritin values are Log$_{10}$ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates Effect allele = A for rs10421768

Supplementary Figure 9a: *KIAA1468* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs9948708

Supplementary Figure 9b: *KIAA1468* SNPs and associations with ferritin μg/L

Ferritin values are Log$_{10}$ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs9948708

Ethnicity and SNP	N		Beta [CI]
African American_rs1421312			
African American,rs1421312	228		-0.14 [-0.38 , 0.10]
African_rs1421312			
South African,rs1421312	480		-0.05 [-0.27 , 0.17]
Tanzanian,rs1421312	581		0.09 [-1.40 , 1.58]
Kenyan,rs1421312	243		-0.11 [-0.31 , 0.09]
RE Model for African__rs1421312			*-0.08 [-0.23 , 0.06]*
African American_rs228918			
African American,rs228918	228		-0.08 [-0.32 , 0.16]
African_228918			
South African,rs228918	480		0.08 [-0.12 , 0.28]
Tanzanian,rs228918	577		-0.86 [-2.31 , 0.59]
Kenyan,rs228918	508		0.14 [-0.06 , 0.34]
RE Model for African_rs228918			*0.10 [-0.04 , 0.24]*
African American_rs2413450			
African American,rs2413450	228		-0.09 [-0.36 , 0.18]
African_rs2413450			
South African,rs2413450	481		-0.23 [-0.58 , 0.12]
Tanzanian,rs2413450	568		-0.65 [-2.86 , 1.56]
Kenyan,rs2413450	580		-0.23 [-0.47 , 0.01]
RE Model for African_rs2413450			*-0.23 [-0.43 , -0.04]*
African American_rs4820268			
African American,rs4820268	228		-0.09 [-0.34 , 0.16]
African_rs4820268			
South African,rs4820268	480		-0.05 [-0.32 , 0.22]
Tanzanian,rs4820268	576		-0.26 [-1.97 , 1.45]
Kenyan,rs4820268	527		-0.28 [-0.50 , -0.06]
RE Model for African_rs4820268			*-0.18 [-0.40 , 0.03]*

Supplementary Figure 10a: *TMPRSS6* SNPs and associations with haemoglobin (g/dL)

Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs2413450 and T for rs1421312, rs228918 and rs4820268

Supplementary Figure 10b: *TMPRSS6* SNPs and associations with ferritin µg/L

Ferritin values are Log₁₀ back transformed data. Black squares are point estimates for each study, horizontal lines are 95% CIs, black diamonds are summary estimates with the lateral tips of the open diamond indicating the standard errors for the summary estimates. Effect allele = A for rs2413450 and T for rs1421312, rs228918 and rs4820268

Chapter 5

Tumour necrosis factor allele variants and their association with the occurrence and severity of malaria in African children

Wanjiku N Gichohi-Wainaina
Alida Melse-Boonstra
Edith J Feskens
Ayse Y. Demir
Jacobien Veenemans
Hans Verhoef

Tumour necrosis factor allele variants and their association with the occurrence and severity of malaria in African children

Abstract

Background: Tumour necrosis factor-α (TNF-α) is central to the immune response to *Plasmodium* infection. Its plasma concentration is influenced by allele variants in the promoter region of the *TNF*-α gene.

Objectives: To assess the effect of *TNF*-α allele variants (*TNF*$_{-1031}$, *TNF*$_{-308}$) on 1) malaria rates in young Tanzanian children; 2) the severity of malaria as indicated by haemoglobin concentrations at the time of presentation with febrile episodes; 3) the association between *Plasmodium* infection and haemoglobin concentration in symptomless parasite carriers.

Methods: We used data from a placebo-controlled trial in which 612 Tanzanian children aged 6–60 months with height-for-age z-score≤−1.5 SD. Those with *Plasmodium* infection at baseline were treated with artemether-lumefantrine. An episode of malaria was pre-defined as current *Plasmodium* infection with an inflammatory response (axillary temperature ≥37.5°C or whole blood C-reactive protein concentration ≥8 mg/L) in children reported sick. Cox regression models used in the primary analysis accounted for multiple episodes per child.

Results: We genotyped 94.9% (581/612) children for *TNF*$_{-1031}$ (rs1799964) (*TNF*$_{-1031}$T>C); allele frequency was 0.39. Corresponding values for *TNF*$_{-308}$G>A (rs1800629) were 95.4% (584/612) and 0.17. Compared to their common genotype, malaria rates were increased in the *TNF*$_{-1031}$GG genotype (hazard ratio, HR [95%CI]: crude 1.41 [1.01–1.97], adjusted 1.31 [0.97–1.76]) but decreased in those with the *TNF*$_{-308}$AA genotype (crude HR: 0.13 [0.02–0.63], adjusted HR 0.16 [0.04–0.67]). These associations were weaker when analysing first episodes of malaria. We found no evidence that allele variants of *TNF*$_{-1031}$ and *TNF*$_{-308}$ affected haemoglobin concentration at first episode of malaria, or that they modified the association between *Plasmodium* infection and haemoglobin concentrations at baseline.

Conclusion: In this cohort of Tanzanian children, the *TNF*$_{-1031}$CC genotype was associated with increased rates of malarial episodes, whereas the *TNF*$_{-308}$ AA genotype was associated with decreased rates.

Introduction

Tumour necrosis factor-α (TNF-α) is central to the immune response to *Plasmodium* infection because of its pyrogenic properties and its key role in triggering the cascade of pro inflammatory cytokines that regulate immune cells. It is secreted predominantly by activated macrophages following exposure to *Plasmodium* antigens, and acts to suppress parasitaemia. Its plasma concentration is elevated during malarial episodes but also in symptomless parasite carriers [1]. In patients with malaria, a high TNF-α production capacity can result in faster clearance of fever and parasitaemia [2], but TNF-α acts as a two-edged sword: excessive production, as may occur during hyperparasitaemia, can contribute to pathological features of severe malaria [3, 4].

Allele variants in the promoter region of the *TNF-α* gene have been associated with changes in gene expression and plasma TNF-α concentration response to infection. Within this promoter region, the rs1799964 allele variant concerns a T>C change at the -1031 position relative to the transcription start site of the *TNF-α* gene, whilst the rs1800629 allele variant consists of a G>A change in the -308 position. In a case-control study in India, the $TNF_{1031}C$ allele was associated with elevated plasma TNF-α concentrations and increased susceptibility to severe falciparum malaria [5].

Homozygotes for the $TNF_{-308}A$ allele ($TNF_{-308}AA$) produce more TNF-α in response to infection than their heterozygote counterparts ($TNF_{-308}GA$) and those with the common genotype ($TNF_{-308}GG$) [6, 7]. Evidence that the $TNF_{-308}AA$ genotype is associated with an increased susceptibility to severe malaria is inconsistent [8], perhaps in part due to limitations in the case-control design of these studies.

There have been few cohort studies to longitudinally assess the effect of these *TNF-α* allele variants on malaria outcomes. Children with an elevated TNF-α response to *Plasmodium* infection are more likely to become ill than their peers with a lower TNF-α response capacity, so that the former may experience more frequent episodes of malaria. On the other hand, in children exposed to chronic or repeated *Plasmodium* infections, an increased TNF-α response associated with the $TNF_{-308}AA$ genotype may intensify a hepcidin-mediated block in iron absorption and consequently induce or exacerbate iron deficiency [9], which probably protects against malarial episodes [10, 11].

We aimed to assess the effect of the rs1799964 and rs1800629 allele variants on malaria rates in young Tanzanian children. Additionally, we aimed to investigate the effect of these variants on the severity of malaria as indicated by haemoglobin concentrations at

the time of presentation with febrile episodes. Lastly, we assessed the influence of these variants on the association between *Plasmodium* infection and haemoglobin concentration in symptomless parasite carriers.

Methods

Study area and population

We used data from a randomized trial that aimed to assess the effect of supplementation with zinc and other micronutrients on malaria rates. The trial was conducted between February 2008 and March 2009 in four rural villages in Handeni District, north-eastern Tanzania. Residents primarily belong to the Wazigua and Wabondei Bantu tribes, but settlement of migrant plantation workers has resulted in a mixture of tribes with different origins and much intermarriage. The area is mainly populated by poor farmer families involved in subsistence farming. Malaria transmission is intense and perennial, with nearly all infections being due to *P. falciparum* [12]. Apart from several local traditional healers the research dispensary was the only health facility in the area. The study was approved by the Ethical Review Committee of Wageningen University, The Netherlands and the National Health Research Ethics Review sub-Committee, Dar es Salaam, Tanzania. Informed consent was obtained from community leaders, local government officials and parents or guardians. Further details of this trial are described elsewhere [13].

Recruitment

All resident children were eligible for randomisation when aged 6–59 months and with a height-for-age z-score below –1.5 SD. Children with severe wasting (weight-for-age z-score < –3 SD), haemoglobin concentration < 70 g/L, signs of chronic illness, and those unlikely to remain permanently resident or comply with the supplementation for the duration of the trial, or whose parents or guardians declined consent, were excluded from the study. Venous blood samples were collected in EDTA tubes and centrifuged immediately. An aliquot of 90 µL erythrocyte sediment with the buffy coat was mixed with 90 µL phosphate-buffered saline and 180 µL of DNA stabilizing buffer (AS1; Qiagen, Hilden, Germany) and stored at 4°C for subsequent genotyping. Plasma samples were stored in liquid nitrogen (–196°C) in the field and at –80°C during transport and

subsequent storage until biochemical analysis in The Netherlands. Haemoglobin concentration was measured in an aliquot of whole blood by a haematology analyser (Sysmex KX21, Kobe, Japan). We detected *Plasmodium* infection in fresh blood by rapid dipstick test (CareStart, Access Bio, Monmouth Jct, USA). Children with a positive test result were treated immediately with artemether-lumefantrine. The location of the child's homestead was determined using a global positioning system. Further details about recruitment procedures are reported elsewhere [13].

Experimental intervention

Children were randomized within 6 strata defined by *Plasmodium* infection (binary) and age class (6–17 months, 18–35 months and 36–60 months) and randomly permuted blocks with size randomly selected of 4 or 8. They then received daily supplements with either zinc alone (10mg as gluconate), multi-nutrients without zinc, zinc combined with multi-nutrients or placebo. Supplements, in the form of powder in color-coded capsules, were contained in blister packs, and administered orally after suspending capsule contents in clean water or breast milk. Supplementation was performed by local community volunteers, who reported adherence daily to field staff at the research dispensary.

Follow-up and case detection

A clinical officer was on duty at the research clinic day and night. At recruitment, parents or guardians were requested to bring participating children to the dispensary immediately when detecting fever or any other illness during the intervention period. In samples collected at baseline and from sick children, the presence of parasite-specific lactate dehydrogenase (*P. falciparum* and other *Plasmodium* species) was detected by rapid tests (Monmouth Jct, USA) Axillary temperature was measured using an electronic thermometer and dipstick tests were administered for children with guardian-reported fever; for those with positive test results, we collected plasma samples and measured whole-blood C-reactive protein concentrations using a point-of-care test. Plasma was stored as described for the recruitment procedure. Artemether-lumefantrine (Novartis Pharma, Basel, Switzerland) was administered to any child with current *Plasmodium* infection upon enrolment, or with reported fever and a positive dipstick test result during the follow up period.

Laboratory procedures

For children who presented with malarial episodes, whole-blood concentrations of haemoglobin and C-reactive protein were measured using point-of-care tests (HemoCue, Ängelholm, Sweden and QuikRead, Orion Diagnostica, Espoo, Finland, respectively).Plasma concentrations of *P. falciparum*-specific histidine-rich protein-2 (HRP2) in samples collected during the first malaria episode were measured using a commercial enzyme-linked immunosorbent assay kit (Malaria Ag Celisa; Cellabs, Brookvale NSW, Australia). Plasma concentrations of C-reactive protein were measured (Meander Medical Centre, Amersfoort, The Netherlands) on a Beckman Coulter Unicel DxC880i system according to the manufactures' instructions. Genotypes were determined using Illumina's VeraCode™ GoldenGate Genotyping Assay on a BeadXpress™ platform. *TNFa* variant allele clustering was assessed visually to determine success of genotyping. Further quality control cut-offs were: a GenCall Score of >0.5 and a call rate of ≥0.95 [14].

Statistical analyses

Anthropometric indices were calculated using Epi Info software (version 3.3.2; http://www.cdc.gov/epiinfo). All analyses were performed using SPSS (v15·0 for Windows, SPSS, Chicago, IL), CIA (v2.1.2) [15] and STATA (v11; College Station, Tx). For each TNF-α genotype, we used Fisher's exact test to assess whether populations were in Hardy-Weinberg equilibrium. We calculated differences in baseline characteristics using the homozygote wildtype group as the reference for each variant using CIA (v2.1.2). Because variables of interest were normally distributed, we report means, SDs and 95% CIs.

Effect of TNF-α genotype on malaria rates The primary outcome, an episode of malaria, was pre-defined as a positive result for either a pLDH or a HRP2 dipstick test with any of the following: a) confirmed fever (axillary temperature ≥ 37.5°C as measured by electronic thermometer), or b) guardian-reported but unconfirmed 24-h history of fever in the presence of inflammation (whole blood C-reactive protein concentrations >8 mg/L), separated by at least 14 days from a previous malaria episode. We calculated incidence per *TNF*-α genotype, and incidence ratios based on time to first episodes, with wild type

homozygotes as reference group. In the primary analysis, we compared group rates using Cox regression with robust estimates of the standard error to account for multiple episodes within children. We explored to what extent supplementation with either zinc or multi-nutrients including iron influenced the magnitude of the effect of TNF-α genotype on malaria rates. We also examined to what extent adjustment for baseline factors that were a priori expected to be prognostic for malaria (*Plasmodium* infection status, distance between homestead and clinic [continuous variable], height-for-age z-score [continuous variable], mosquito net use [binary variable]) and experimental intervention influenced the estimated effect of genotypes. In this adjusted analysis, we included experimental intervention as a binary variable indicating pooled groups receiving multi-nutrients (with or without zinc) and pooled groups receiving no multi-nutrients (with or without zinc). We also used Kaplan-Meier analysis with Peto tests to assess effects of TNF-α genotypes on time-to-first episode of malaria.

Influence of TNF-α genotype on the association between P. falciparum infection and haemoglobin concentration at baseline

For each genotype, we indicated three classes by two dummy variables, resulting in two interaction terms per variant. Regression models were then used to assess the influence of genotypes on the associations between *P .falciparum* and haemoglobin concentration. We included main terms for *P. falciparum* infection and genotype dummies, and we adjusted as pre-planned for age class, mosquito net use, height-for-age z-score and distance between the child's homestead and the dispensary in regression analyses.

Influence of TNF-α genotype on the association between P. falciparum infection and haemoglobin concentration at first episode of malaria

For each genotype, we indicated three classes by two dummy variables. Influence of TNF-α genotype on the association between *P. falciparum* infection and haemoglobin concentration at first episode of malaria was then assessed by ANOVA and adjusted for age class, height-for-age z-score and distance between the child's homestead and the dispensary and treatment group.

Results

Of 612 children recruited in the original study, 581 (94.9%) and 584 (95.4%) respectively had DNA typed for TNF_{-1031} and TNF_{-308}, respectively. Only 3% (20/612) of children

failed to complete the trial. Minor allele frequencies were 0.39 (*TNF*$_{-1031}$) and 0.17 (*TNF*$_{-308}$), respectively. Regarding these *TNF-α* alleles, we found no evidence that the population was in Hardy-Weinberg disequilibrium (P=0.90 and 0.57 for *TNF*$_{-1031}$ and *TNF*$_{-308}$ respectively); thus there was no indication that were errors in genotyping of the two variants.

Baseline factors associated with TNF-α genotypes (Table 1)

Children with *TNF*$_{-1031}$C allele seemed to have larger weight-for-age z-scores than their peers with the common variant (0.14 SD and 0.41 SD in heterozygotes and homozygotes, respectively, compared to –0.18 SD in those with the common genotype). The *TNF*$_{-308}$A allele seemed to be associated with increased *Plasmodium* infection, but the evidence was weak (p-value: 0.09). We found no evidence of other factors being associated with the *TNF-α* genotypes investigated.

Table 1. Factors associated with rs1799964 and rs1800629 allele variants in the promoter region of the *TNF*α gene

Allele variant/factor	Wild type	Heterozygote	Difference (95% CI)	Homozygote mutant	Difference (95% CI)	P-value
rs1799964	**-1031TT**	**-1031TC**		**-1031CC**		
n	356	200		25		
Age (months)	33.4±15.5	31.9±16.1	–1.5 (–4.2,1.2)	28.06±14.9	–5.3 (–11.6, 0.99)	0.18
Sex (%male)	48.0% (171)	48.0% (96)	0.0% (–8.6%, 8.6%)	44.0% (11)	–4.0% (–22.1%, 15.6%)	0.93
Haemoglobin concentration, g/L	102.7±12.3	103.4±13.8	0.0 (–2.1,2.1)	101.5±10.4	–1.2 (–6.2, 3.8)	0.71
Anaemia, % (n) ¶	69.1% (246)	66.0% (132)	–3.1% (–11.3%, 4.9%)	76.0% (19)	6.9% (–13.1%, 20.4%)	0.50
Plasmodium infection, % (n)	43.8% (156)	43.0% (86)	–0.8% (–9.3%, 7.8%)	36.0% (9)	–7.8% (–24.4%, 12.3%)	0.75
Inflammation, % (n) †	32.3% (114)	32.0% (64)	0.0% (–7.9%, 8.2%)	32.0% (8)	–0.3% (–15.6%, 19.8%)	0.58
Mosquito net use, % (n) ††	32.9% (117)	32.0% (64)	–0.1% (–8.8%, 7.3%)	32.0% (8)	–0.9% (–16.5%, 19.3%)	0.99
Distance from homestead to dispensary, km **	3.5±2.1	3.5±2.2	0.0 (–0.4, 0.4)	3.8±2.6	0.3 (–0.6, 1.2)	0.90
Anthropometric indices						
Height-for-age z-score	–2.45±0.70	–2.38±0.69	0.07 (–0.05, 0.19)	–2.48±0.78	–0.03 (–0.32, 0.26)	0.48
Weight-for-height z-score	–0.18±0.81	–0.04±0.85	0.14 (0.00, 0.28)	0.23±1.00	0.41 (0.08, 0.75)	0.02
Weight-for-age z-score	–1.63±0.74	–1.52±0.75	0.11 (–0.02, 0.24)	–1.42±0.76	0.21 (–0.09, 0.51)	0.13
rs1800629	**-308GG**	**-308GA**		**-308AA**		
n	477	104		3		
Age	32.1±15.7	34.8±15.3	2.7 (–0.6, 6.0)	47.08±11.7	15.0 (–2.9, 32.8)	0.08
Sex (%male)	46.8% (223)	53.8% (56)	7% (–4%, 17%)	66.7% (2)	20% (–26%, 48%)	0.34
Haemoglobin concentration, g/L	102.3±12.8	105.2±12.2	2.9 (0.2, 5.6)	105.0±19.2	2.7 (–11.9, 17.3)	0.10
Anaemia, % (n) ¶	69.0% (329)	64.4% (67)	–4.5% (–14.9%, 5.0%)	66.7% (2)	–2.3% (–48.4%, 25.2%)	0.54
Plasmodium infection, % (n)	41.1% (196)	51.9% (54)	10.8% (0.3%, 21.1%)	66.7% (2)	25.6% (–20.5%, 53.1%)	0.09
Inflammation, % (n) †	31.0% (148)	33.7% (35)	2.6% (–6.8%, 12.9%)	33.3% (1)	2.3% (-25.2%, 48.4%)	0.23

Mosquito net use , % (n) ††	32.9% (157)	30.8% (32)	–2.1% (–11.3%, 8.1%)	33.3% (1)	0.4% (–27.1%, 46.5%)	0.95
Distance from homestead to dispensary, km **	3.6±2.2	3.6±2.2	0.0 (–0.5, 0.5)	3.5 ±1.0	–0.10 (–2.6, 2.4)	1.00
Anthropometric indices						
Height-for-age z-score	–2.42±0.69	–2.47±0.74	–0.05 (–0.20, 0.10)	–2.69±0.72	–0.27 (–1.06, 0.52)	0.61
Weight-for-height z-score	–0.11±0.84	–0.17±0.85	–0.06 (–0.24, 0.12)	0.07±0.78	0.04 (–0.92, 1.00)	0.73
Weight-for-age z-score	–1.57±0.75	–1.64±0.75	–0.07 (–0.23, 0.09)	–1.56±0.92	0.01 (–0.84, 0.86)	0.72

P-values for the associations between each factor and the allele variants investigated. Mean ± SD or % [n] unless indicated otherwise

% figures represent %with feature/%without feature (number with feature/number without feature).Difference is relative to the wild type group

¶ Haemoglobin concentration < 110 g/L. † Plasma C-reactive protein concentration ≥ 8 mg/L. †† Data missing for 11 children. ** As the crow flies, based measurements by global positioning system

*Effect of TNF-a genotype on malaria rates (**Table 2**)*

Overall, there were 1,511 malaria episodes recorded in 400 child-years of observation (incidence: 3.8/child-year). Of 581 children, 489 children (84%) experienced at least one malaria episode while recurrent episodes occurred in 406 (69%) children. Compared to the common genotype, malaria rates (all episodes) in homozygotes for the TNF_{-1031}C allele (TNF_{-1031}CC) were elevated by 41% (hazard ratio 1.41; crude analysis). In the analysis of first episodes, malaria rates also seemed elevated for this genotype (hazard ratio 1.17, Table 2), but the statistical evidence for such an association was much weaker. Only one episode of malaria occurred in homozygotes for the TNF_{-308} A allele (TNF_{-308}AA). This genotype was nonetheless associated with a decrease in malaria rates by 87% (hazard ratio 0.13; all episodes). Malaria rates were similarly decreased in the analysis of first episodes, but the statistical evidence was weak. Kaplan-Meier analysis did not indicate marked group-specific differences in malaria rates for specific time periods of follow-up for either of the two *TNF-a* genotypes investigated (**Figure 1**).

Table 2: Associations between *TNF*α genotype and rates of first episode malaria or all episodes of malaria

rs1799964	Homozygotes -1031TT		Heterozygotes -1031TC		Homozygote mutants -1031CC	
All episodes of malaria						
n		356		200		25
Incidence	2.93	(907/309.1)	2.92	(502/171.9)	4.15	(90/21.7)
Hazard ratio, crude	1.00	Reference	0.99	[0.85-1.14]	1.41	[1.01-1.97]
Hazard ratio, adjusted†	1.00	Reference	0.95	[0.82-1.09]	1.35	[1.01-1.80]
First episode of malaria						
Incidence	2.85	(296/104.0)	3.09	(168/54.4)	3.29	(20/6.07)
Incidence ratio	1.00	Reference	1.08	[0.89-1.31]	1.15	[0.68-1.76]
Hazard ratio, crude	1.00	Reference	1.08	[0.90-1.31]	1.17	[0.90-1.31]
Hazard ratio, adjusted†	1.00	Reference	1.07	[0.89-1.31]	1.11	[0.66-1.86]
rs1800629	**-308GG**		**-308GA**		**-308AA**	
All episodes of malaria						
n		477		104		3
Incidence	2.97	(1216/409.2)	3.08	(282/91.5)	0.36	(1/2.8)
Hazard ratio, crude	1.00	Reference	1.04	[0.87-1.24]	0.13	[0.02-0.63]
Hazard ratio, adjusted†	1.00	Reference	1.10	[0.93-1.29]	0.14	[0.04-0.48]
First episode of malaria						
Incidence	2.89	(392/135.6)	3.27	(91/27.8)	0.53	(1/1.9)
Incidence ratio	1.00	Reference	1.13	[0.90-1.44]	0.18	[0.00-0.98]
Hazard ratio, crude	1.00	Reference	1.07	[0.86-1.33]	0.27	[0.03-2.86]
Hazard ratio, adjusted†	1.00	Reference	1.14	[0.91-1.44]	0.33	[0.03-3.59]

Values between brackets indicate (episodes/child-years observed) or [95% CIs].
† Estimates adjusted for baseline *Plasmodium* infection status, distance between homestead and clinic (continuous variable), height-for-age z-score (continuous variable), mosquito net use and experimental intervention. There was no evidence of interactions between genotype and experimental intervention.
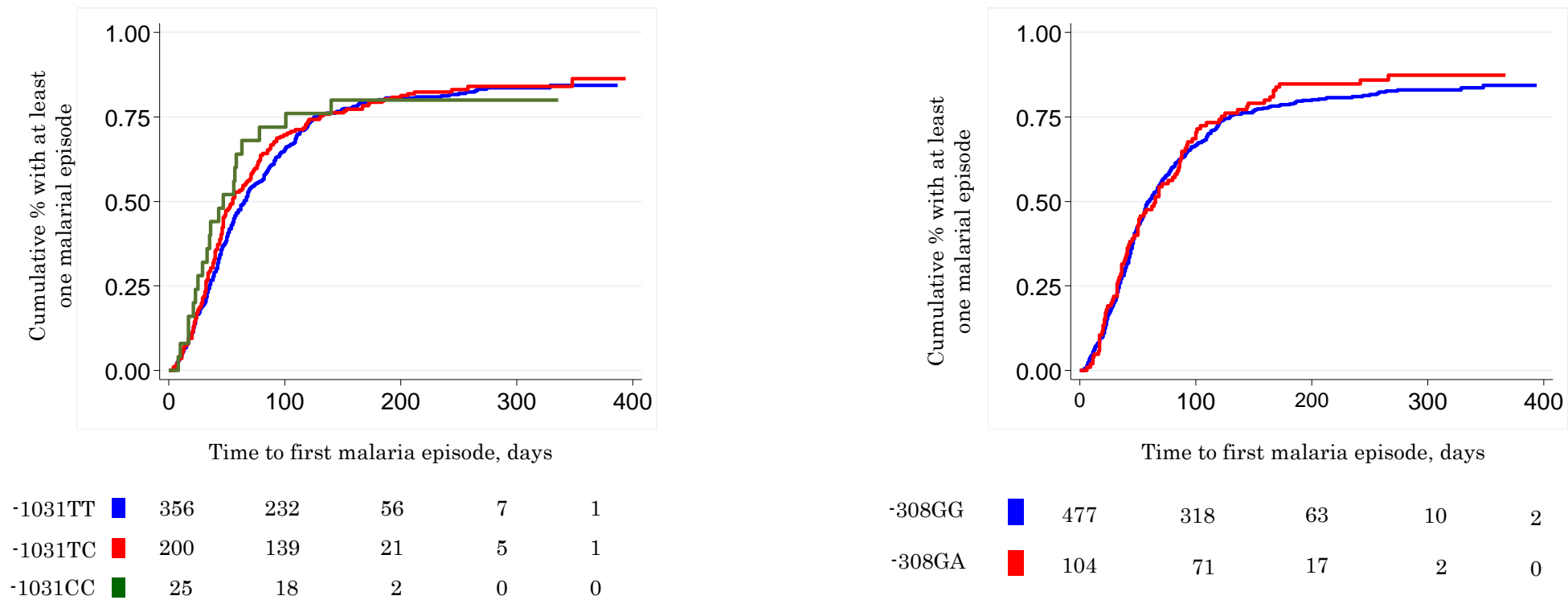
Figure 1. Effect of *TNF-α* genotype on time to first episode of malaria, as assessed by Kaplan-Meier analysis

Values indicate the number of children, by *TNF-α* genotype group, who had remained free of malaria at 0, 100, 200, 300 and 400 days of follow-up. For *TNF*$_{-1031}$ allele variant (*top panel*), pairwise group comparison with the common genotype as the reference group yielded p-values of 0.78 and 0.74 for heterozygotes *TNF*$_{-1031}$TC and homozygotes (*TNF*$_{-1031}$CC), respectively; for the *TNF*$_{-308}$ allele variant (*top panel*), corresponding p-values for heterozygotes (*TNF*$_{-308}$GA) and homozygotes (*TNF*$_{-308}$AA) were 0.54 and 0.31, respectively (Peto tests). For homozygotes of the *TNF*$_{-308}$A allele (*TNF*$_{-308}$AA), the curve is not shown because only one episode occurred at 400 days of follow-up.

Effect of TNF-α genotype on haemoglobin concentration at first malarial episode (**Figure 2**):

We found no evidence that haemoglobin concentrations at first episode of malaria varied by variants of TNF– 1031.
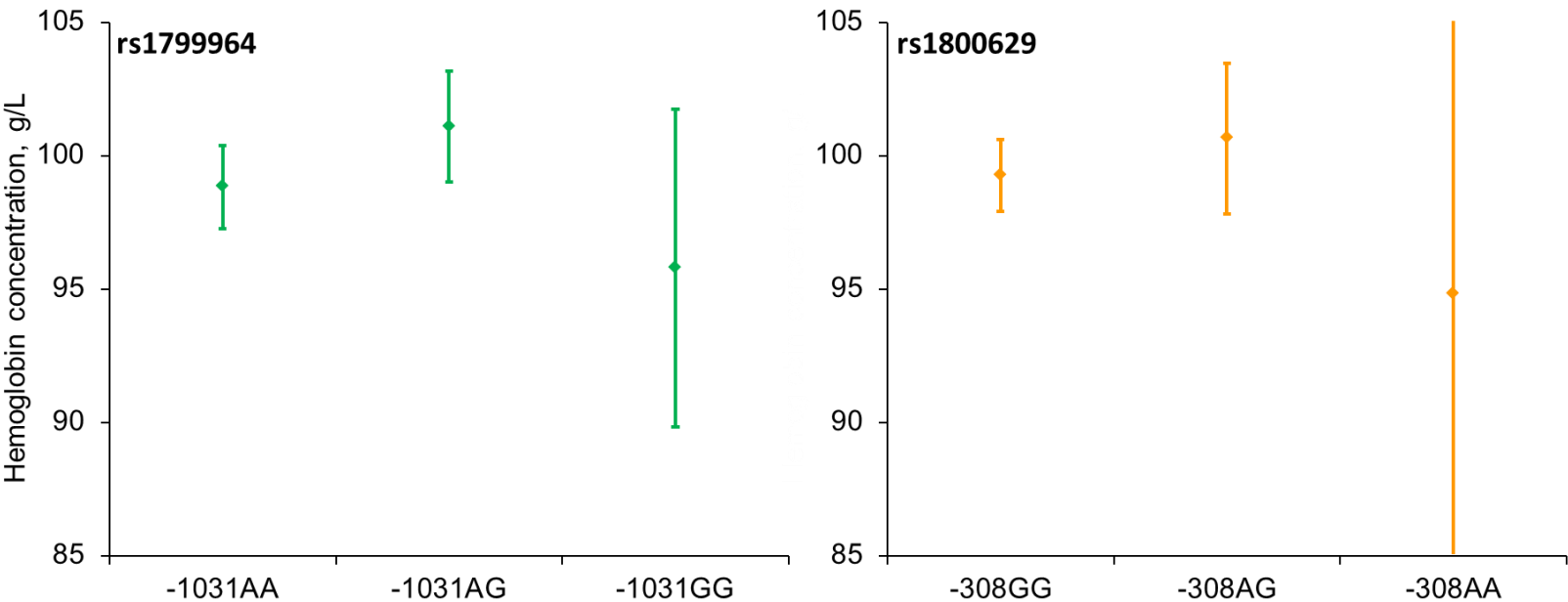
Figure 2. Association between *TNF-α* genotype and haemoglobin concentration at time of first episode of malaria

Left panel: *TNF*$_{-1031}$ (rs1799964) genotype; bottom panel: *TNF*$_{-308}$ (rs1800629) genotype. Genotypes are homozygous wildtype, heterozygous and homozygous mutant from left to right. Line bars indicate 95% CIs. Values on top of bars indicate differences in group means with corresponding 95% CIs as obtained by multivariate regression analysis. All values are adjusted for *Plasmodium* infection status, distance between homestead and clinic, height-for-age z-score, mosquito net use and experimental intervention. The 95% CI for haemoglobin concentration of the -308AA genotype is outside the indicated range; this 95% CI is: 68.0 to 121.6 g/L.

*Influence of TNF-α genotype on the association between Plasmodium infection and haemoglobin concentration at baseline (**Figure 3**)*
We found no evidence that heterozygosity for either $TNF_{-1031}C$ or $TNF_{-308}A$ modified the association between *Plasmodium* infection at baseline and haemoglobin concentrations at baseline (p-values or interaction: 0.47 and 0.57, respectively). Similarly, we found no evidence that this association was modified by homozygosity for either $TNF_{-1031}C$ (p-value 0.56)
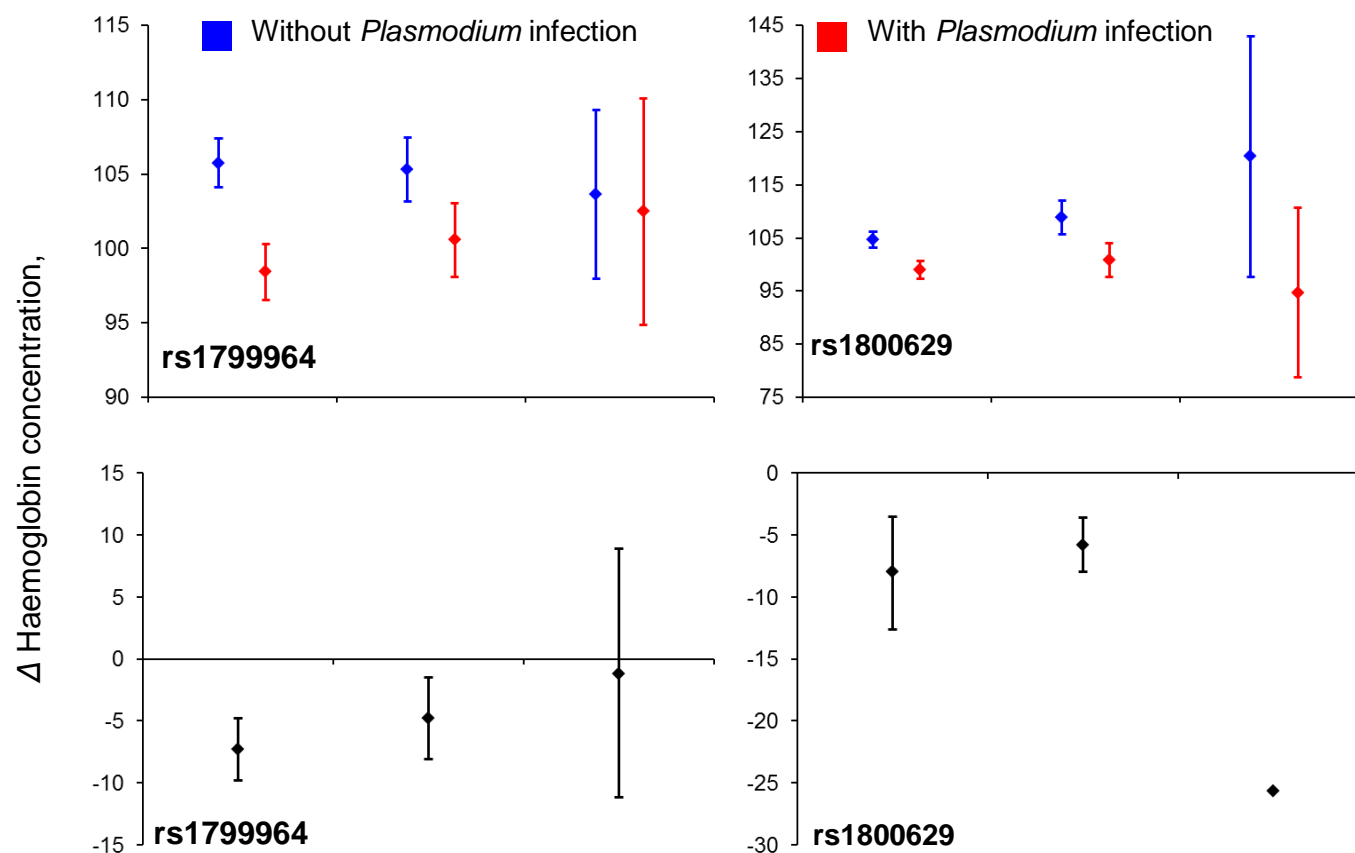
Figure 3. Association between *Plasmodium* infection and haemoglobin concentrations at baseline, by *TNF-α* genotype

Line bars indicate 95% CIs. No confidence interval indicated for $TNF_{-308}AA$ because there was only one case of *Plasmodium* infection. All values are adjusted for age class, mosquito net use, height-for-age z-score and distance from homestead.

Discussion

In our cohort of Tanzanian children, malaria rates due to *P. falciparum* were increased in homozygotes for the TNF_{-1031}C allele (TNF_{-1031}CC) and decreased in homozygotes for the TNF_{-308}A allele (TNF_{-308}AA). We found no evidence that allele variants of TNF_{-1031} and TNF_{-308} were associated with haemoglobin concentration at first episode of malaria, or that they modified the association between *Plasmodium* infection and haemoglobin concentrations at baseline.

The associations between *TNF-α* allele variants and malaria rates were generally more pronounced in the analysis of all episodes than in the analysis of first episodes (**Table 2**). Thus the question arises how such differences should be interpreted. Some children experience malaria more frequently than others. This between-individual variation may be due to differences in exposure to infectious mosquito bites, immunity (and thus age) and therapy-seeking behaviour. In the presence of such heterogeneity in susceptibility to malaria, 'high risk' individuals tend to become sick more rapidly than 'low risk' individuals, and once they have experienced an episode they are no longer considered in a time-to-first-event analysis. Thus, with the passing of follow-up time, only the 'low risk' individuals remain in the risk set and, accordingly, the event rate observed will decrease. In cohort studies such as ours, this decrease is generally more pronounced in groups with high overall event rates, which have 'high-risk' individuals dropping out more efficiently. Thus in a time-to-first-event analysis, group contrasts can become underestimated and effects on malaria may appear decrease with time. For this reason, we believe that analysis of all episodes more validly reflects associations between *TNF-α* allele variants and malaria rates than analysis of first episodes. In addition, the former better reflects effects on the total population burden of disease.

There was no evidence that the association between TNF_{-1031}CC genotype and malaria rates were confounded by *Plasmodium* infection status, distance between homestead and clinic, height-for-age z-score and mosquito net use factors assessed at baseline, or by the experimental intervention. Similarly, we found no evidence for such confounding on the association between TNF_{-308}AA and malaria rates. These associations should nonetheless be interpreted with caution, because the statistical evidence was weak in the analysis of first episodes, even though the direction of these associations was similar. We also have no good explanation why the TNF_{-1031}CC genotype was associated with increased malaria rates whilst the TNF_{-308}AA genotype was associated with a decrease

in these rates, even though both genotypes have been reported to be associated with elevated plasma TNF-α concentration [5–7].

Our findings add to the growing but conflicting evidence that allele variants of the *TNF-α* gene are associated with susceptibility to malaria [8, 16, 17]. Several factors can explain apparent discrepancies in results obtained thus far in various studies. First, studies varied in design. Most studies compared hospitalised cases of severe malaria with hospital- or community-based controls [8, 16, 18, 19]. Such studies can be implemented quickly and at a relatively low cost, but they are inherently vulnerable to selection bias if the distribution of *TNF-α* allele variants differs between cases and controls in the absence of a true association between these allele variants and malaria. As a consequence, for studies that recruited controls from a hospital setting, the strength of the association between *TNF-α* allele variants and malaria will be underestimated if these controls suffered from a condition that is associated with *TNF-α* genotype. This is plausible, because *TNF-α* genotypes have also been associated with a variety of disorders other than malaria [20-23]. A better approach is to select controls that are representative of the population that produced the cases. Thus Clark et al. [8] selected as controls cord blood samples obtained from birth clinics in the same hospital from which they recruited cases of severe malaria. Even so, bias may still have occurred in that study if *TNF-α* genotypes varied in frequency of delivery in hospital, or in the frequency of care-seeking behavior and hospital admission for severe malaria. Such differentials may have occurred, for example, if the geographical distribution of *TNF-α* genotypes in relation to the hospital is uneven, or if *TNF-α* genotypes vary in the risk of birth complications. In addition, it is difficult to rule out that the results of *TNF-α* genotyping performed by Clark et al. (2009) in cord blood may have been affected by maternal DNA originating from contamination of cord blood at birth by maternal cells or by contact with maternal blood or tissue [24]. In our cohort study, bias may have occurred if there was differential loss to follow-up such that the risk of being lost to follow-up was related both to malaria and *TNF-α* genotype, or if there were group differentials in the detection of malarial episodes. However, loss to follow-up was very low (3%), genotyping was successful for almost all (95%) children, and we believe to have captured virtually all malaria cases [25].

Second, the relationship between *TNF-α* genotype and susceptibility to malaria may depend on population-specific factors. For example, as noted in our introduction, *TNF-α* allele variants that are associated with an enhanced TNF-α response to chronic or

repeated infections can possibly protect against newly acquired *Plasmodium* infections through an altered iron metabolism. Such protection is conceivably more effective in children who already have marginal iron status than in their iron-replete peers, and in children who are frequently exposed to infection.

Third, differences in results may be due to haplotype heterogeneity between populations studied. It has been suggested that the causal allele variants regulating *TNF-α* responses may actually be located some distance downstream of the *TNF-α* gene, and that *TNF-α* allele variants considered are in sufficient linkage disequilibrium with the causal variant in some populations, but not in others [8].

In conclusion, in pre-school Tanzanian children living in an area of intense transmission, the TNF_{-1031}CC genotype was associated with increased rates of malarial episodes, whereas the TNF_{-308} AA genotype was associated with decreased rates. Further studies are needed to confirm these findings.

Acknowledgements

References

1. Kurtzhals JA, Addae MM, Akanmori BD, Dunyo S, Koram KA, Appawu MA, Nkrumah FK, Hviid L: **Anaemia caused by asymptomatic Plasmodium falciparum infection in semi-immune African schoolchildren.** *Trans R Soc Trop Med Hyg* 1999, **93:**623-627.

2. Mordmuller BG, Metzger WG, Juillard P, Brinkman BM, Verweij CL, Grau GE, Kremsner PG: **Tumor necrosis factor in Plasmodium falciparum malaria: high plasma level is associated with fever, but high production capacity is associated with rapid fever clearance.** *Eur Cytokine Netw* 1997, **8:**29-35.

3. Kwiatkowski D: **Tumour necrosis factor, fever and fatality in falciparum malaria.** *Immunol Lett* 1990, **25:**213-216.

4. Richards AL: **Tumour necrosis factor and associated cytokines in the host's response to malaria.** *Int J Parasitol* 1997, **27:**1251-1263.

5.   Sinha S, Mishra SK, Sharma S, Patibandla PK, Mallick PK, Sharma SK, Mohanty S, Pati SS, Mishra SK, Ramteke BK, et al: **Polymorphisms of TNF-enhancer and gene for FcgammaRIIa correlate with the severity of falciparum malaria in the ethnically diverse Indian population.** *Malar J* 2008, **7:**13.

6.   Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW: **Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation.** *Proc Natl Acad Sci U S A* 1997, **94:**3195-3199.

7.   Natividad A, Hanchard N, Holland MJ, Mahdi OS, Diakite M, Rockett K, Jallow O, Joof HM, Kwiatkowski DP, Mabey DC, Bailey RL: **Genetic variation at the TNF locus and the risk of severe sequelae of ocular Chlamydia trachomatis infection in Gambians.** *Genes Immun* 2007, **8:**288-295.

8.   Clark TG, Diakite M, Auburn S, Campino S, Fry AE, Green A, Richardson A, Small K, Teo YY, Wilson J, et al: **Tumor necrosis factor and lymphotoxin-alpha polymorphisms and severe malaria in African populations.** *J Infect Dis* 2009, **199:**569-575.

9.   Atkinson SH, Rockett KA, Morgan G, Bejon PA, Sirugo G, O'Connell MA, Hanchard N, Kwiatkowski DP, Prentice AM: **Tumor necrosis factor SNP haplotypes are associated with iron deficiency anemia in West African children.** *Blood* 2008, **112:**4276-4283.

10.   Gwamaka M, Kurtis JD, Sorensen BE, Holte S, Morrison R, Mutabingwa TK, Fried M, Duffy PE: **Iron deficiency protects against severe Plasmodium falciparum malaria and death in young children.** *Clin Infect Dis* 2012, **54:**1137-1144.

11.   Prentice AM, Cox SE: **Iron and malaria interactions: research needs from basic science to global policy.** *Adv Nutr* 2012, **3:**583-591.

12.   Ellman R, Maxwell C, Finch R, Shayo D: **Malaria and anaemia at different altitudes in the Muheza district of Tanzania: childhood morbidity in relation to level of exposure to infection.** *Ann Trop Med Parasitol* 1998, **92:**741-753.

13.   Veenemans J, Milligan P, Prentice AM, Schouten LR, Inja N, van der Heijden AC, de Boer LC, Jansen EJ, Koopmans AE, Enthoven WT, et al: **Effect of supplementation with zinc and other micronutrients on malaria in Tanzanian children: a randomised trial.** *PLoS Med* 2011, **8:**e1001125.

14. Ritchie ME, Liu R, Carvalho BS, Irizarry RA: **Comparing genotyping algorithms for Illumina's Infinium whole-genome SNP BeadChips.** *BMC Bioinformatics* 2011, **12**:68.

15. G Altman DM, Trevor N Bryant and Martin J Gardner: *Statistics with Confidence.* 2nd edn: BMJ Books; 2000.

16. McGuire W, Knight JC, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D: **Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles.** *J Infect Dis* 1999, **179**:287-290.

17. Gimenez F, Barraud de Lagerie S, Fernandez C, Pino P, Mazier D: **Tumor necrosis factor alpha in the pathogenesis of cerebral malaria.** *Cell Mol Life Sci* 2003, **60**:1623-1635.

18. Randall LM, Kenangalem E, Lampah DA, Tjitra E, Mwaikambo ED, Handojo T, Piera KA, Zhao ZZ, de Labastida Rivera F, Zhou Y, et al: **A study of the TNF/LTA/LTB locus and susceptibility to severe malaria in highland papuan children and adults.** *Malar J* 2010, **9**:302.

19. Church J, Maitland K: **Invasive bacterial co-infection in African children with Plasmodium falciparum malaria: a systematic review.** *BMC Med* 2014, **12**:31.

20. Wajant H: **The role of TNF in cancer.** *Results Probl Cell Differ* 2009, **49**:1-15.

21. Daniele G, Guardado Mendoza R, Winnier D, Fiorentino TV, Pengou Z, Cornell J, Andreozzi F, Jenkinson C, Cersosimo E, Federici M, et al: **The inflammatory status score including IL-6, TNF-alpha, osteopontin, fractalkine, MCP-1 and adiponectin underlies whole-body insulin resistance and hyperglycemia in type 2 diabetes mellitus.** *Acta Diabetol* 2014, **51**:123-131.

22. Mootoo A, Stylianou E, Arias MA, Reljic R: **TNF-alpha in tuberculosis: a cytokine with a split personality.** *Inflamm Allergy Drug Targets* 2009, **8**:53-62.

23. Elahi MM, Asotra K, Matata BM, Mastana SS: **Tumor necrosis factor alpha -308 gene locus promoter polymorphism: an analysis of association with health and disease.** *Biochim Biophys Acta* 2009, **1792**:163-172.

24. Hall JM, Lingenfelter P, Adams SL, Lasser D, Hansen JA, Bean MA: **Detection of maternal cells in human umbilical cord blood using fluorescence in situ hybridization.** *Blood* 1995, **86**:2829-2832.

25. Veenemans J, Schouten LR, Ottenhof MJ, Mank TG, Uges DR, Mbugi EV, Demir AY, Kraaijenhagen RJ, Savelkoul HF, Verhoef H: **Effect of preventive**

supplementation with zinc and other micronutrients on non-malarial morbidity in Tanzanian pre-school children: a randomized trial. *PLoS One* 2012, **7**:e41630.

Chapter 6

General Discussion

General Discussion

Although a large portion of the variation in iron status undoubtedly results from non-genetic factors, the role of genetic factors in influencing iron absorption, uptake and transport, and therefore eventual iron status, has recently come to the fore. The studies herein were therefore conducted to unravel the genetics of iron status in populations of African ancestry.

With the success of genome-wide association (GWA) studies, many common genetic variants underlying iron metabolism have been identified. These genetic loci could refine the biological understanding of iron metabolism and in so doing, contribute to strategies to address iron deficiency. Furthermore, this would also impact on the development of appropriate biomarkers to assess iron status. A major drawback of current genetic association studies is that they have been limited to populations of European ancestry. This is a problem as European populations contain only a subset of the human genetic variation present globally. Additionally, populations vary in terms of their allele frequencies, biological adaptations and other properties that may affect the detectability and importance of risk variants associated with iron status across populations.

Before genetic markers associated with iron status are utilised in African populations at the public health level, they must be rigorously assessed using among others, association analyses. The research presented in this thesis is a step towards understanding the genetics of iron status in African populations using mainly the candidate gene approach. As presented in the general introduction, this research focussed on several research questions.

The **primary research question** was therefore:

Do the single nucleotide polymorphisms previously identified as being associated with iron status in non-African populations present with similar associations in African populations?

The **secondary research questions** were:

1.   Are the minor allele frequencies of single nucleotide polymorphisms associated with iron status in African populations similar to those of non-African populations?

2.   How do the single nucleotide polymorphisms associated with iron status influence disease severity in the African population?

In this chapter, we present the main findings determined in the studies presented herein and highlight some of the methodological considerations that may have impacted on their internal and external validity. Finally, we offer insights into future perspectives on investigating the genetics of iron status in African populations.

6.1 Main Findings

In this section, we discuss the contribution of our studies to the objectives previously stated, as well as to recent findings from other studies. The main findings are summarized in Table 6.1.

A large number of GWA studies point to the importance of TMPRSS6 variants in iron metabolism in populations of European ancestry but no data is available from populations of African ancestry

Before the start of this project, genetic epidemiological research into iron status markers or haematological measures focussed mainly on populations of European ancestry, with limited studies from populations of Asian ancestry [1-4]. These studies resulted in the conclusive identification of *TMPRSS6* variants as being associated with iron status measures. In order to shed light on the contradictory findings from the different GWA studies and to evaluate the role of ethnicity as an explanatory factor, we performed a systematic review with meta-analysis of loci in the *TMPRSS6* gene as identified in cohorts of various ancestry [5]. We observed that risk alleles in the key *TMPRSS6* loci such as rs855791 and rs4820268 are associated with lower Hb and serum ferritin concentrations across both European and Asian ancestry populations. We also observed that there were differences in the minor allele frequencies of these variants across ethnicities investigated. Since there are currently no GWA studies on iron status parameters in African populations, genetic association studies conducted in African American populations can be leveraged to identify regions where African ancestral alleles are associated with iron status. At the time that we conducted the study selection for the review, only one study included information from a cohort of African ancestry. Since then, results from five GWA studies that were either conducted in African Americans only or included African Americans as part of cohorts studied have been reported [6-10]. In a paediatric cohort of 7943 African American children, no *TMPRSS6* variant reached genome wide significance in association testing with 8 haematological traits. This same study consisted of 6234 Caucasian children in which the variant

rs855791 was significantly associated with mean corpuscular haemoglobin as well as mean corpuscular volume *[7]*. In four other studies *[6, 8-10]* conducted among African American adults, there was no association of the *TMPRSS6* loci with measures of iron status. These findings suggest that the genetic architecture of haematological traits may differ by ethnicity. This first study constructed the foundation for conducting the next studies described in **Chapters 3 and 4**.

Our candidate gene association study in a cohort of South African women generally did not confirm findings from studies on the genetics of iron status in populations of European ancestry. Furthermore, the minor alleles of single nucleotide polymorphisms that were investigated were more common in populations of European ancestry compared to our South African cohort.

The study described in Chapter 3 is the first extensive candidate gene association study on the genetics of iron status conducted in an African population. The candidate genes selected for genotyping were based on information from GWA studies conducted in populations of European ancestry *[1-4]*. We were generally unable to uncover genetic loci associated with iron status with convincing statistical evidence ($p<0.05$). Only two single nucleotide polymorphisms (SNPs) in *TF* (rs3811647 and rs1799852) were significantly associated with various iron status measures. These associations have previously been observed in European ancestry populations *[11]*. We also observed that generally the minor allele frequencies (MAFs) of investigated SNPs were lower than those of the populations of European ancestry included in the 1000 Genomes Project. Investigations on gene-gene interactions revealed that the chromosome 6 SNP allele combination (AAA) consisting of rs1799964 and rs1800629 both in *TNF α* and rs2071592 in the *NFKBIL1* was associated with higher odds for low serum ferritin concentrations. The *TNF α* and *NFKBIL1* loci are both involved in the inflammatory response *[12]*. Elevated serum and tissue levels of *TNF* are found in inflammatory and infectious conditions [13, 14] and serum levels correlate with the severity of infections *[15, 16]*. Nuclear Factor-kappa B (NF-κB) which is encoded by NFKBIL-1, on the other hand is required for the transcriptional activation of the *TNF* α gene *[17]*.

Our observation corroborated a previous finding in which a similar allele combination consisting of both rs1799964 and rs1800629 as well as other loci was associated with an increased risk of iron deficiency in a cohort of West African children *[12]*. Though further investigations are required, this confirms that the regulation of iron status through

inflammatory pathways is an area necessitating of attention in African populations. One implication of this finding is that among other strategies, treatment of infectious diseases in African populations may be important in lowering the iron deficiency burden in African populations compared to other populations. Additionally, this finding suggests that assessment of iron status using serum ferritin (SF) may be less useful in Africa This is because the use of serum ferritin is limited due to confounding by inflammation [18].

*Further candidate gene association studies in various cohorts of African ancestry pointed to TMPRSS6 variants as possibly being important in iron metabolism in African cohorts (**Chapter 4**), however, only few single nucleotide polymorphisms were associated with iron status measures.*

We performed a candidate gene association study involving various cohorts of African ancestry (Kenyan, Tanzanian and South African cohorts). Thereafter, meta-analyses of the genetic association results from these cohorts was conducted and compared with an African American cohort. The results of the meta-analyses point to the need to further investigate the influence of *TMPRSS6* and *TF* loci on iron status in African populations. We additionally observed that the candidate SNPs genotyped were more common in European ancestry populations. This study further contributed to the evidence that genetic variants associated with iron status measures in cohorts of European ancestry do not necessarily translate into similar associations in cohorts of African ancestry. Further population specific studies are therefore required to uncover the genetic architecture of iron metabolism in African populations.

*It is vital to explore how information from genetic association studies may be useful in the African context. In **Chapter 5**, a study was performed to explore how single nucleotide polymorphisms within the tumour necrosis factor alpha (TNF α) gene (previously associated with iron status) influence malaria severity in a cohort of Tanzanian children.*

Iron is essential for the survival of the malaria parasite [19]. The relationship between TNF α (an inflammatory cytokine) and iron status, and by extension malaria severity, has been a subject of recent investigations. *TNF α* has previously been identified as a modifier of body iron status [17]. On the other hand, during *P. falciparum* malarial infection, the cytokine TNF α has been described as both protective and pathogenic [20]. It has further been reported that the SNPs (i.e. rs1799964, rs1799724, rs1800750,

rs1800629 and rs361525) in the promoter region of the TNF α gene have different associations with malaria in various populations *[21-23]*.

We investigated the association of two SNPs in the *TNF a* gene *TNF*$_{-1031}$ (rs1799964) and *TNF*$_{-308}$ (rs1800629) with malaria incidence and severity. It was observed that compared to their common allele variants, malaria rates were increased in those with the *TNF*$_{-1031}$CC genotype by 30-40%, but decreased in those with the *TNF*$_{-308}$AA genotype by 84-87%. We found no evidence that either allele variant affected haemoglobin concentration at first episode of malaria, or that they modified the association between *Plasmodium* infection and haemoglobin concentrations at baseline. In this study, we concluded that replication studies and mechanistic studies are required to further validate our findings. Results from this study further underscore the need to treat malaria in endemic areas in order to alleviate the burden of iron deficiency.

Table 6.1: Main findings of the studies described in this thesis

Chapter	Study design	Study population	Objectives	Main Findings
2	Systematic review with meta analyses	Several European, Asian and mixed ancestry cohorts of varied ages	Investigate potential inter-ethnic differences in associations of genetic variants within the transmembrane protease, serine 6 (*TMPRSS6*) gene with iron status indicators	• Differences in minor allele frequencies (MAF) of 8 *TMPRSS6* SNPs (rs855791, rs4820268, rs2111833, rs1421312, rs228921, rs228918, rs228919 and rs575620) across investigated cohorts were observed; with the MAF of rs855791 considerably higher in Asian populations than in Caucasians (0.55 vs 0.42) • The A allele of rs855791 was associated with lower haemoglobin and ferritin concentrations and increased serum transferrin receptor and transferrin concentrations in all populations. Similar associations for the G allele in rs4820268 made. • No data from African cohorts

| 3 | Candidate gene association study | 686 South African women (mean age: 50 years) | Compare frequencies of minor alleles identified in populations of European ancestry to those of a South African cohort of women

Investigate associations of reported SNPs with iron status in a female black South African cohort

Investigate gene-gene interactions that influence iron status | • MAFs were lower than those reported for European ancestry populations in the 1000 Genomes Project

• Two SNPs in *TF* were associated with iron status: rs1799852 with decreased serum ferritin and body iron and increased serum transferrin receptor concentrations; rs3811647 with serum transferrin receptor and body iron (U-shaped association)

• Chromosome 6 allele combination (AAA) consisting of rs1799964 and rs1800629, both in TNF α and rs2071592 in NFKBIL1, was associated with higher odds for low serum ferritin concentrations

• The chromosome 22 allele combination (GG) of rs228918 and rs228921 in the TMPRSS6 gene was associated with lower odds for increased sTfR concentrations |
| 4 | Candidate gene association study | 686 South African women (mean age: 50 y)

628 Kenyan children (mean age: 9 y)

609 Tanzanian children (mean age: 3 y)

228 African American men and women (mean age: 67 y) | Compare minor allele frequencies between several cohorts of African ancestry

Compare combined associations of 20 SNPs with markers of iron status between the African cohorts and an African American cohort

Compare combined associations of 20 SNPs with markers of iron status between the African ancestry cohorts (including African American cohort) with findings from European ancestry cohorts | • Minor allele frequencies of candidate SNPs were generally lower in the African ancestry cohorts as compared to Europeans in the 1000 Genomes Project

• Associations of two TMPRSS6 single nucleotide polymorphisms, rs2413450 and rs4820268, with lower haemoglobin concentrations replicated in combined associations of all African cohorts, as well as African cohorts together with the African American cohort.

• rs3811658 in TF associated with higher haemoglobin concentrations, rs1525892 in TF and rs228918 in TMPRSS6 with increased ferritin concentrations in combined associations of all African cohorts as well as African cohorts together with the African American cohort. These associations have previously been observed in European cohorts, although direction of association is not reported. |

| 5 | Candidate gene association study | 581 Tanzanian children (mean age: 3 y) | Assess the effect of SNPs at the *TNF α* locus rs1799964 (TNF–1031) and rs1800629 (TNF–308) allele variants on malaria incidence. | • | Compared to their common genotype, malaria rates were increased in the TNF_{-1031}GG genotype (hazard ratio, HR [95%CI]: crude 1.41 [1.01-1.97], adjusted 1.31 [0.97-1.76]) but decreased in those with the TNF_{308}AA genotype (crude HR: 0.13 [0.02-0.63], adjusted HR 0.16 [0.04-0.67]) |
| | | | Investigate the effect of these variants on the severity of malaria as indicated by haemoglobin concentrations at the time of presentation with febrile episodes as well as at baseline | • | No evidence that allele variant of TNF_{-1031} or TNF_{-308} affected haemoglobin concentration at first episode of malaria, or that they modified the association between *Plasmodium* infection and haemoglobin concentrations at baseline. |

6.2 Methodological considerations

We have already outlined some of the methodological aspects that may have impacted on results from the specific studies reported in Chapters 2, 3 and 4. In this section, we elaborate further on the methods used, from study design to data analysis, which could possibly impact on both the internal and external validity of our studies.

The discovery of associations in an observational study is a fundamental first step in understanding disease aetiology [24]. We attempted to replicate findings of GWA studies from European ancestry populations in various African cohorts using the candidate gene association method and only found a few associations. It is possible that most of the previous observations were not replicated for the following reasons:

Heterogeneity between our study populations and European cohorts

Population heterogeneity across studies may lead to differences in associations. There exists in Africa today a rich diversity of environments (diet, geography, climate, pathogen exposure and many other factors) that likely act in concert to modulate iron status in ways that may overlap but may also be different than in non-African populations. It has been shown that including a non-confounding covariate associated with a quantitative trait is beneficial as it may explain some of the variability in the outcome, thus reducing noise and increasing power to detect genetic associations [25]. To address the effects of differences due to environments and factors specific to cohorts, we adjusted analyses for factors that could potentially influence iron status such as age, gender, iron intake, BMI and menopausal status.

In terms of diet, the intake of animal foods are the most important dietary determinant of iron status of a population [26]. It is thus necessary to obtain information on dietary iron intake when conducting studies on the genetics of iron status. We included haem iron intake as a proportion of total dietary iron intake as a covariate in association analyses described in Chapter 3. In Chapter 4, total iron intake was included as a covariate in association analyses involving the African American cohort. There was however only a negligible difference in effect estimates between the crude model and the model adjusted for iron intake in this second case. We therefore did not adjust for iron intake in associations involving the African American cohort. We additionally did not adjust for iron intake in any of the cohorts involved in the study described in Chapter 4.

The finding that adjusting for iron intake had no impact on effect estimates within the African American cohort probably means that there are more important factors that determine iron status within this population. One explanation may be the effect of comorbidities, for example, chronic kidney disease that is increased among African Americans. However it has been observed that haemoglobin levels remain lower in African Americans than in white persons even after taking into account disease status, behavioural risk factors, nutritional intake, and iron status *[27, 28]*. This may point to the influence of polymorphisms retained in the genome that were previously useful in adaptation to environments of high inflammation found in Africa. For example, a recent study has reported a SNP in the Glucose-6-phosphate dehydrogenase (*G6PD)* gene to be significantly associated with ferritin in an African American cohort *[6]*. Glucose-6-phosphate dehydrogenase (G6PD) deficiency due to polymorphisms in the *G6PD* is associated with protection from severe malaria, and potentially uncomplicated malaria phenotypes in African populations *[29]*. The genetics of iron status does therefore require further research in African American populations. Information on dietary iron intake was not available for the Tanzanian and Kenyan cohorts, implying that effect estimates obtained from analyses in these cohorts may be somewhat less precise.

We additionally adjusted for age and gender in all analyses, and for inflammation for analyses involving ferritin concentrations. While dietary iron remains the foremost determinant of iron status, scientific evidence suggests that adiposity may also be a determinant of iron status *[30]*. An inverse association between adiposity and iron status has been observed *[31-33]*. Whereas studies have shown that adiposity might increase the risk of iron deficiency, adiposity has additionally been shown to induce elevation of ferritin concentrations *[34]*. This has been attributed to the fact that ferritin is an acute-phase protein that may be elevated by the low-grade inflammation that occurs when adipose tissue is enlarged *[32]*. Menstrual blood loss has also been observed to affect iron status *[35, 36]*. When BMI and menopausal status were included as covariates in association analyses for the South African cohort there were negligible differences between the crude and adjusted models, indicating that these two factors may not be crucial in explaining the lack of associations between genetic factors and iron status in this cohort of women.

Since the cohorts included in this thesis were not specifically designed to investigate the genetics of iron status, not all covariates of interest may have been assessed. This

especially relates to information on dietary iron intake as well as information on G6PD deficiencies and any thalassemias present. Since such important covariates are often not taken into account, association results from GWA studies may lack validity. Our studies therefore provide more accurate effect estimates than previously reported estimates as we took into account some of the covariates that explain iron status.

Difference in minor allele frequencies and impact on sample sizes required

In Chapters 3 and 4 we observed that the SNPs investigated in this thesis were generally less common in our cohorts compared to European ancestry populations. The statistical power to detect a significant association depends on the effect size of the association and the frequency of the allele of interest *[37]*. SNPs with low allele frequencies would need to have large effect sizes in order to be detected in a candidate-gene study with small numbers. The sample sizes in Chapters 2 and 3 were therefore too small for estimation of small genetic effects with high precision (e.g. 780 individuals are required to conduct association analysis for a SNP explaining 1% of the variation in the outcome measure, with $\alpha = 0.05$ and power 80%). Lack of power can be addressed by the combination of multiple studies by way of meta-analyses. Such studies have confirmed many associations that were unclear from individual study reports *[38-40]*. To overcome the sample size issue, we attempted to combine results from various African cohorts using meta-analytic methods in Chapter 4. In general, we were unable to replicate previously observed associations from populations of European ancestry apart from 2 loci in the *TMPRSS6* gene (rs241450 and rs4820268) as well as rs3811658 in the *TF* gene. In Chapter 3, we have not observed associations with any *TMPRSS6* variants probably due to the small sample size. As information from more genetic studies become available in Africa, meta-analytic methods as a means to increase power will continue to be a useful tool in identification of SNPs associated with low iron status. One of the SNPs that was significantly associated with serum ferritin in chapter 3, rs3811647 is in linkage disequilibrium (LD) with rs3811658. This further emphasizes the potential importance of the *TF* locus in determining iron status.

Difference in linkage disequilibrium structure between European ancestry populations and African cohorts

It has previously been observed that European tag SNPs performed poorly in the African population, capturing only an average of 35% of SNP variation *[41]*. This could be due to differences in LD structure between the European ancestry populations and the African population. A genetic marker that is targeted in an association study may not be a disease-causing locus itself, but may be linked to a causal one, such that passage of the marker from one generation to the next is correlated with passage of the disease-causing mutation. The resulting linkage disequilibrium (LD) then yields a significant association between the marker and the disease itself. The degree of linkage disequilibrium (LD) may be less in individuals of recent African descent because of the greater variation in the genome within this group *[42]*. We did not examine linkage patterns between our cohorts so as to compare with those of European ancestry cohorts. Determining linkage disequilibrium in a pilot sample could have been useful to optimize SNP selection and could have provided information about possible haplotypes for analysis. This presents an area for further study.

6.2.1 Implications of findings: Does information on genetics of iron status have practical applications?

One of the questions from public health practitioners would be the manner by which information on the genetics of iron status could be utilized to influence population health. Genetic variation that confers subtle differences in nutrient requirements may exert greater effect in certain subgroups or populations and contribute to iron deficiency in certain environmental contexts. Much genetic variation associated with iron status remains undiscovered, especially in African populations that have existed in unique and diverse nutritional environments for many generations where adaptive alleles may have expanded. In suggesting a way forward, it is necessary to first state what practical implications our results may have especially concerning potential benefits for public health.

We have observed that the *TMPRSS6* as well as *TF* loci are important insofar as iron status in African populations is concerned. The *TMPRSS6* SNPs, rs241350 and rs4820268, are quite common with MAFs of approximately 15% and 25% and reduction

of Hb of approximately 0.20 g/dL per copy of each minor allele. Although the reduction in Hb is not large, further research on the *TMPRSS6* locus and its influence on iron status is required. This is because variants in this locus have consistently been associated with iron status measures. Additionally, a significant proportion of the population harbours the minor alleles of these variants. We also demonstrated that rs1525892 in *TF with a* MAF of approximately 25% increased serum ferritin concentrations by approximately 1 µg/L per copy of each minor allele in African populations. Other polymorphisms in *TF* (rs17999852 and rs3811647) were also associated with iron status measures in the population of black South African women. Since the effect sizes of these variants were small and the fact that the polymorphisms were not present in large proportions of the population, the application of these findings for public health may be limited at present. However, this does not preclude the need for further studies to investigate these loci. One of the examples of the manner in which genetics can be applied to influence public health is the case of the hereditary hemochromatosis. Hereditary hemochromatosis (HH) due to mutations in the *HFE* gene are associated with iron overload, which is a strong risk factor for several non-communicable diseases *[43, 44]*. The *HFE* gene Cys282Tyr polymorphism has already affected government policy, with Denmark and Sweden stopping their iron fortification policies, in part because of a potential risk to persons at risk of hereditary hemochromatosis *[45, 46]*. In the case of iron deficiency in African populations, the discovery of loci that predispose populations or sub groups to iron deficiency could lead to a refinement of iron fortification or supplementation policies. In terms of having a further impact on strategies that address iron deficiency, we have also observed findings that support current efforts to treat malaria in malaria endemic areas as a means to alleviate iron deficiency. We found that the TNF_{-1031} CC (MAF 22%) is associated with increased malaria rates. This SNP has previously been identified as a risk factor for iron deficiency *[12]*. It therefore follows that treatment of malaria in an area where this SNP is common would substantially reduce prevalence of iron deficiency.

Genetics may also redefine how iron adequacy or inadequacy is defined. Recommended dietary allowance (RDA) for a particular nutrient is defined based on development of deficiency diseases, but with greater understanding of the genetics of iron status, it may be possible to formulate more appropriate RDAs to prevent iron deficiency.

Finally, investigating the genetics of iron status may also aid in development of reliable biomarkers to define iron deficiency. Currently, commonly used biomarkers of iron

status, sTfR and serum ferritin, lack sensitivity and specificity as they are influenced by inflammation *[47]*. Novel biomarkers may be used in future to identify population groups that would benefit most from strategies to address iron deficiency. For example, recently, a study has found that the expression of the iron hormone hepcidin distinguishes different types of anaemia in African children *[48]*. Therefore, a better understanding of the molecular regulation of iron metabolism may lead to the discovery of new biomarkers.

6.3 Future directions: which way forward?

The increasing availability of high throughput genotyping technologies makes it plausible that genetic association studies will continue and will in the future provide useful information on the genetics of iron status. However, because we have shown that many of the loci previously associated with iron status in populations of European ancestry are not replicated in African populations, it may be necessary to conduct GWA studies within the latter to discover novel loci associated with iron status. After the GWA studies have been conducted, candidate gene studies can be used to validate findings as well as explore the interactions between genes with factors such as age, gender, and other demographic characteristics.

To more specifically identify genetic markers associated with iron deficiency, fine mapping studies are required. Fine mapping is the process of searching a region identified by GWA studies for possible risk alleles, and they can be used to precisely characterise risk alleles for certain conditions. An example of this is the fine mapping of associations of the fat mass and obesity associated (*FTO*) gene locus with BMI and diabetes *[49]*. Knowledge about exact location of risk alleles for iron deficiency can preclude the need for investigations based on marker loci, thus reducing the error and variability in association testing.

Animal models present a useful option to elucidate action of polymorphisms identified from fine mapping studies. With animal studies, it is possible to control iron intake and environmental factors rigorously in the different groups under study. Knockout mice are particularly important for studying the role of reported genes whose functions have not been determined. Knockout mouse models have been extensively used to study the regulation of *TMPRSS6* and its activity *[50, 51]*. However, these models have not yet been used to study the response to dietary iron at different concentrations. Animal

models are advantageous in that dose response questions can be assessed in ways that are difficult in human studies. Establishment of exposure challenge-response data is critical to a better understanding of the mechanism, and ultimate relevance, to iron deficiency.

Finally, several options exist to conduct human studies to assess action of risk alleles in the presence of high or low iron diets. In Africa specifically, the potential to conduct intervention studies investigating the genetics of iron status is largely untapped. One option is to conduct targeted intervention studies in which individuals with or without risk alleles are involved. This would provide a clearer picture of the manner in which individual factors such as age and gender, as well as diet, interact with genes to influence eventual iron status. Another option is to incorporate collection of DNA samples as part of biological material collected from participants within the numerous micronutrient intervention studies that are ongoing. This information may help to explain the heterogeneity in response to iron interventions. DNA material may also be collected immediately after interventions to determine gene expression. Such functional studies will assist to confirm the role of putative polymorphisms in iron deficiency.

The studies described in this thesis only provide a first attempt towards the understanding of the genetic regulation of iron status in African populations. The potential underlying the discovery of novel variants associated with iron deficiency would offer opportunities towards combating iron deficiency in African populations. This thesis higlights the need for population-specific systems biological approaches to unravel the complex interplay of genes and the environment in iron deficiency development.

References

1. Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, Hottenga JJ, Henders AK, Campbell MJ, Wallace L, et al: **Common variants in *TMPRSS6* are associated with iron status and erythrocyte volume.** *Nat Genet* 2009, **41:**1173-1175.

2. Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L, et al: **Genome-wide association study identifies variants in *TMPRSS6* associated with hemoglobin levels.** *Nat Genet* 2009, **41:**1170-1172.

3.   Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, Chen MH, Kottgen A, Glazer NL, Dehghan A, et al: **Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium.** *Nat Genet* 2009, **41:**1191-1198.

4.   Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, Walston JD, Fried LP, Singleton A, Guralnik J, et al: **A genome-wide association analysis of serum iron concentrations.** *Blood* 2010, **115:**94-96.

5.   Gichohi-Wainaina WN, Towers GW, Swinkels DW, Zimmermann MB, Feskens EJ, Melse-Boonstra A: **Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (***TMPRSS6***) gene associated with iron status indicators: a systematic review with meta-analyses.** *Genes Nutr* 2015, **10:**442.

6.   Li J, Lange LA, Duan Q, Lu Y, Singleton AB, Zonderman AB, Evans MK, Li Y, Taylor HA, Willis MS, 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* 2014.

7.   Li J, Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, Mentch FD, Guo Y, Kim C, Xia Q, et al: **GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children.** *Hum Mol Genet* 2013, **22:**1457-1464.

8.   Chen Z, Tang H, Qayyum R, Schick UM, Nalls MA, Handsaker R, Li J, Lu Y, Yanek LR, Keating B, et al: **Genome-wide association analysis of red blood cell traits in African Americans: the COGENT Network.** *Hum Mol Genet* 2013, **22:**2529-2538.

9.   Ding K, de Andrade M, Manolio TA, Crawford DC, Rasmussen-Torvik LJ, Ritchie MD, Denny JC, Masys DR, Jouni H, Pachecho JA, et al: **Genetic variants that confer resistance to malaria are associated with red blood cell traits in African-Americans: an electronic medical record-based genome-wide association study.** *G3 (Bethesda)* 2013, **3:**1061-1068.

10.   Lo KS, Wilson JG, Lange LA, Folsom AR, Galarneau G, Ganesh SK, Grant SF, Keating BJ, McCarroll SA, Mohler ER, 3rd, et al: **Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans.** *Hum Genet* 2011, **129:**307-317.

11.   Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, Peltonen L, Martin NG, Montgomery GW, Whitfield JB, Visscher PM: **Variants in TF and**

HFE explain approximately 40% of genetic variation in serum-transferrin levels. *Am J Hum Genet* 2009, **84**:60-65.

12.   Atkinson SH, Rockett KA, Morgan G, Bejon PA, Sirugo G, O'Connell MA, Hanchard N, Kwiatkowski DP, Prentice AM: **Tumor necrosis factor SNP haplotypes are associated with iron deficiency anemia in West African children.** *Blood* 2008, **112**:4276-4283.

13.   Robak T, Gladalska A, Stepien H: **The tumour necrosis factor family of receptors/ligands in the serum of patients with rheumatoid arthritis.** *Eur Cytokine Netw* 1998, **9**:145-154.

14.   Nurnberger W, Platonov A, Stannigel H, Beloborodov VB, Michelmann I, von Kries R, Burdach S, Gobel U: **Definition of a new score for severity of generalized Neisseria meningitidis infection.** *Eur J Pediatr* 1995, **154**:896-900.

15.   Waage A, Halstensen A, Espevik T: **Association between tumour necrosis factor in serum and fatal outcome in patients with meningococcal disease.** *Lancet* 1987, **1**:355-357.

16.   Kwiatkowski D, Hill AV, Sambou I, Twumasi P, Castracane J, Manogue KR, Cerami A, Brewster DR, Greenwood BM: **TNF concentration in fatal cerebral, non-fatal cerebral, and uncomplicated Plasmodium falciparum malaria.** *Lancet* 1990, **336**:1201-1204.

17.   Sharp P: **SNPs linking TNF with anemia.** *Blood* 2008, **112**:3923-3924.

18.   Lynch S: **The rationale for selecting and standardizing iron status indicators. In:World Health Organization. Report: Priorities in the assessment of vitamin A and iron status in populations, Panama City,Panama, 15–17 September 2010. Geneva, World Health Organization.** 2012.

19.   Clark MA, Goheen MM, Cerami C: **Influence of host iron status on Plasmodium falciparum infection.** *Front Pharmacol* 2014, **5**:84.

20.   Angulo I, Fresno M: **Cytokines in the pathogenesis of and protection against malaria.** *Clin Diagn Lab Immunol* 2002, **9**:1145-1152.

21.   McGuire W, Knight JC, Hill AV, Allsopp CE, Greenwood BM, Kwiatkowski D: **Severe malarial anemia and cerebral malaria are associated with different tumor necrosis factor promoter alleles.** *J Infect Dis* 1999, **179**:287-290.

22.   Ubalee R, Suzuki F, Kikuchi M, Tasanor O, Wattanagoon Y, Ruangweerayut R, Na-Bangchang K, Karbwang J, Kimura A, Itoh K, et al: **Strong association of a

tumor necrosis factor-alpha promoter allele with cerebral malaria in Myanmar. *Tissue Antigens* 2001, **58**:407-410.

23.   Hananantachai H, Patarapotikul J, Ohashi J, Naka I, Krudsood S, Looareesuwan S, Tokunaga K: **Significant association between TNF-alpha (TNF) promoter allele (-1031C, -863C, and -857C) and cerebral malaria in Thailand.** *Tissue Antigens* 2007, **69**:277-280.

24.   Lewis CM, Knight J: **Introduction to genetic association studies.** *Cold Spring Harb Protoc* 2012, **2012**:297-306.

25.   Mefford J, Witte JS: **The Covariate's Dilemma.** *PLoS Genet* 2012, **8**:e1003096.

26.   Reddy MB, Hurrell RF, Cook JD: **Meat consumption in a varied diet marginally influences nonheme iron absorption in normal individuals.** *J Nutr* 2006, **136**:576-581.

27.   Johnson-Spear MA, Yip R: **Hemoglobin difference between black and white women with comparable iron status: justification for race-specific anemia criteria.** *Am J Clin Nutr* 1994, **60**:117-121.

28.   Pan WH, Habicht JP: **The non-iron-deficiency-related difference in hemoglobin concentration distribution between blacks and whites and between men and women.** *Am J Epidemiol* 1991, **134**:1410-1416.

29.   Maiga B, Dolo A, Campino S, Sepulveda N, Corran P, Rockett KA, Troye-Blomberg M, Doumbo OK, Clark TG: **Glucose-6-phosphate dehydrogenase polymorphisms and susceptibility to mild malaria in Dogon and Fulani, Mali.** *Malar J* 2014, **13**:270.

30.   Seltzer CC, Mayer J: **Serum iron and iron-binding capacity in adolescents. ii. comparison of obese and nonobese subjects.** *Am J Clin Nutr* 1963, **13**:354-361.

31.   Cepeda-Lopez AC, Osendarp SJ, Melse-Boonstra A, Aeberli I, Gonzalez-Salazar F, Feskens E, Villalpando S, Zimmermann MB: **Sharply higher rates of iron deficiency in obese Mexican women and children are predicted by obesity-related inflammation rather than by differences in dietary iron intake.** *Am J Clin Nutr* 2011, **93**:975-983.

32.   Tussing-Humphreys LM, Liang H, Nemeth E, Freels S, Braunschweig CA: **Excess adiposity, inflammation, and iron-deficiency in female adolescents.** *J Am Diet Assoc* 2009, **109**:297-302.

33. Cepeda-Lopez AC, Aeberli I, Zimmermann MB: **Does obesity increase risk for iron deficiency? A review of the literature and the potential mechanisms.** *Int J Vitam Nutr Res* 2010, **80**:263-270.

34. Zafon C, Lecube A, Simo R: **Iron in obesity. An ancient micronutrient for a modern disease.** *Obes Rev* 2010, **11**:322-328.

35. Milman N, Clausen J, Byg KE: **Iron status in 268 Danish women aged 18-30 years: influence of menstruation, contraceptive method, and iron supplementation.** *Ann Hematol* 1998, **77**:13-19.

36. Harvey LJ, Armah CN, Dainty JR, Foxall RJ, John Lewis D, Langford NJ, Fairweather-Tait SJ: **Impact of menstrual blood loss and diet on iron deficiency among women in the UK.** *Br J Nutr* 2005, **94**:557-564.

37. Tabor HK, Risch NJ, Myers RM: **Candidate-gene approaches for studying complex genetic traits: practical considerations.** *Nat Rev Genet* 2002, **3**:391-397.

38. Altshuler D, Hirschhorn JN, Klannemark M, Lindgren CM, Vohl MC, Nemesh J, Lane CR, Schaffner SF, Bolk S, Brewer C, et al: **The common PPARgamma Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes.** *Nat Genet* 2000, **26**:76-80.

39. Ioannidis JP, Ntzani EE, Trikalinos TA: **'Racial' differences in genetic effects for complex diseases.** *Nat Genet* 2004, **36**:1312-1318.

40. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN: **Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.** *Nat Genet* 2003, **33**:177-182.

41. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D: **Efficiency and power in genetic association studies.** *Nat Genet* 2005, **37**:1217-1223.

42. Tishkoff SA, Kidd KK: **Implications of biogeography of human populations for 'race' and medicine.** *Nat Genet* 2004, **36**:S21-27.

43. Agudo A, Bonet C, Sala N, Munoz X, Aranda N, Fonseca-Nunes A, Clavel-Chapelon F, Boutron-Ruault MC, Vineis P, Panico S, et al: **Hemochromatosis (HFE) gene mutations and risk of gastric cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) study.** *Carcinogenesis* 2013, **34**:1244-1250.

44. Zhao Z, Li C, Hu M, Li J, Liu R: **Plasma ferritin levels, HFE polymorphisms, and risk of pancreatic cancer among Chinese Han population.** *Tumour Biol* 2014, **35**:7629-7633.

45.   Milman N, Byg KE, Ovesen L, Kirchhoff M, Jurgensen KS: **Iron status in Danish women, 1984-1994: a cohort comparison of changes in iron stores and the prevalence of iron deficiency and iron overload.** *Eur J Haematol* 2003, **71**:51-61.

46.   Olsson KS, Vaisanen M, Konar J, Bruce A: **The effect of withdrawal of food iron fortification in Sweden as studied with phlebotomy in subjects with genetic hemochromatosis.** *Eur J Clin Nutr* 1997, **51**:782-786.

47.   Zimmermann MB: **Methods to assess iron and iodine status.** *Br J Nutr* 2008, **99 Suppl 3**:S2-9.

48.   Pasricha SR, Atkinson SH, Armitage AE, Khandwala S, Veenemans J, Cox SE, Eddowes LA, Hayes T, Doherty CP, Demir AY, et al: **Expression of the iron hormone hepcidin distinguishes different types of anemia in African children.** *Sci Transl Med* 2014, **6**:235re233.

49.   Akiyama K, Takeuchi F, Isono M, Chakrawarthy S, Nguyen QN, Wen W, Yamamoto K, Katsuya T, Kasturiratne A, Pham ST, et al: **Systematic fine-mapping of association with BMI and type 2 diabetes at the FTO locus by integrating results from multiple ethnic groups.** *PLoS One* 2014, **9**:e101329.

50.   Maurer E, Gutschow M, Stirnberg M: **Matriptase-2 (*TMPRSS6*) is directly up-regulated by hypoxia inducible factor-1: identification of a hypoxia-responsive element in the *TMPRSS6* promoter region.** *Biol Chem* 2012, **393**:535-540.

51.   Meynard D, Vaja V, Sun CC, Corradini E, Chen S, Lopez-Otin C, Grgurevic L, Hong CC, Stirnberg M, Gutschow M, et al: **Regulation of *TMPRSS6* by BMP6 and iron in human cells and mice.** *Blood* 2011, **118**:747-756.

Summary

Anaemia has been associated with increased mortality and morbidity as well as with decreased physical functioning and quality of life. On average, globally, 50% of anaemia is assumed to be attributable to iron deficiency. There exists in Africa today a rich diversity of environments (diet, geography, climate, pathogen exposure and many other factors) that likely act in concert to influence the development of iron deficiency. Of interest is the manner in which genetic factors interact with environmental factors such as diet to increase susceptibility to iron deficiency. Understanding the genetics of iron status is key to optimizing strategies to address iron deficiency as well as discovering appropriate biomarkers to monitor iron status in populations.

With the success of genome-wide association (GWA) studies, many common genetic variants underlying iron status have been uncovered. These genetic loci could refine the biological understanding of iron metabolism and in so doing, contribute to strategies to address iron deficiency. Furthermore, this would also impact on the development of appropriate biomarkers to assess iron status. As European populations contain only a subset of the human genetic variation present globally, this limits transferability of association results to African populations. Additionally, African populations when compared to European populations vary in terms of their allele frequencies and biological adaptations which are all factors that may affect the detectability and importance of risk variants associated with iron status. Before markers associated with iron status are utilised in African populations at the public health level, they must be rigorously assessed using among others, association analyses.

In **Chapter 2,** we performed a systematic review with meta-analysis of loci in the *TMPRSS6* gene that have previously been associated with iron status measures in European ancestry cohorts. This study was performed in order to get a more precise evaluation of identified *TMPRSS6* single nucleotide polymorphism associations with iron status in cohorts of differing continental ancestry. We observed that risk alleles of the key *TMPRSS6* variants such as rs855791 and rs4820268 are associated with lower Hb and serum ferritin concentrations across both European and Asian ancestry populations. We also observed that there were differences in the minor allele frequencies of these variants across ethnicities investigated. At the time that we conducted the study selection for the review, only one study included information from a cohort of African ancestry i.e. African Americans. There was no information from African populations.

In **Chapter 3,** we report the results of the first extensive candidate gene association study on the genetics of iron status conducted in a population of black South African women. The candidate genes selected for genotyping were based on information from GWA studies conducted in populations of European ancestry. We were generally unable to uncover genetic loci associated with iron status measures with convincing statistical evidence (p<0.05). Minor allele frequencies (MAF) of the investigated SNPs from the female participants in the PURE population were lower compared to those of males and females of European ancestry populations in the 1000 genomes project. In the TF gene, the single nucleotide polymorphism (SNP) rs1799852 was associated with decreased serum ferritin (p=0.01) and body iron (p=0.03) and increased serum transferrin receptor (sTfR) concentrations (P=0.004), while rs3811647 was associated with transferrin receptor and body iron (both P=0.03). The chromosome 6 SNP allele combination (AAA) consisting of rs1799964 and rs1800629 both in *TNF α* and rs2071592 in *NFKBIL1* was associated with higher odds of low serum ferritin concentrations (serum ferritin<15µg/L; OR:1.86 (95%-CI, 1.23-2.79)). The chromosome 22 SNP allele combination (GG) consisting of rs228918 and rs228921 in the *TMPRSS6* gene was associated with a lower odds of increased sTfR concentrations (sTfR>8.3mg/L; OR:0.79 (95%-CI,0.63-0.98).

In **Chapter 4** we investigated single nucleotide polymorphisms associated with iron status in several cohorts of African ancestry. This was a candidate gene association study and thereafter, meta-analyses of the genetic association results from the individual African cohorts (Kenyan, Tanzanian and South African) was undertaken to compare these results to association results of an African American cohort. We also combined all results from the four cohorts and compared this to previous association findings from cohorts of European ancestry. We successfully replicated reported significant associations with lowered haemoglobin concentrations for two variants in *TMPRSS6* (rs2413450 and rs4820268) and with increased haemoglobin concentrations for one variant in *TF* (rs3811658). When ferritin was considered as an outcome measure, we replicated associations with increased ferritin concentrations in two loci namely rs228918 in *TMPRSS6* and rs1525892 in *TF* (P<0.05 for all significant associations). No other significant associations with either concentrations of haemoglobin or serum ferritin were observed.

In **Chapter 5** a study to explore the association of two SNPs in the *TNF α* gene namely *TNF-1031* (rs1799964) and *TNF-308* (rs1800629) with malaria incidence and severity in a

cohort of Tanzanian pre-school children was conducted. Our main finding was that compared to their common allele variants, malaria rates were increased in those with the *TNF*$_{-1031}$ CC genotype by 30-40%. We determined no evidence that either allele variant affected haemoglobin concentration at first episode of malaria, or that they modified the association between *Plasmodium* infection and haemoglobin concentrations at baseline.

Finally, the main findings and the methodological aspects that may have impacted on internal and external validity of our findings are discussed in **Chapter 6**. The public health implications and recommendations, and directions for future research are also presented.

To our knowledge the work presented herein presents the first extensive investigation of the genetics of iron status in African populations. In conclusion, we observed that most of the associations previously identified in populations of European ancestry do not replicate in African ancestry populations. Additionally minor allele frequencies of single nucleotide polymorphisms associated with iron status in European ancestry cohorts are higher than those of African ancestry populations. This finding may indicate that novel loci may be responsible for the heritability of iron status in African populations. The observation that a common *TNF α* variant *TNF*$_{-1031}$ (rs1799964) increased malaria rates indicates that to alleviate iron deficiency in malaria endemic areas- prevention and treatment is necessary. This is because malaria is known to negatively impact on iron status. Based on our observations, further investigations are necessary to uncover genetic variants associated with iron status in African populations. Several important factors should be taken into consideration in the design and execution of future association studies such as sample size, and collection of all information that could impact on iron status such as dietary iron intake, age, and measures of inflammation. Furthermore mechanistic studies would be useful to further ascertain action of identified variants. Single nucleotide polymorphisms that influence iron status will continue to help advance our understanding of iron metabolism and homeostasis. Besides contributing towards the development of biomarkers that are devoid of the shortfalls of current iron status biomarkers, unravelling the genetics of iron status is promising in efforts aimed at optimizing strategies to address iron deficiency.

Samenvatting

Bloedarmoede is geassocieerd met verhoogde mortaliteit, morbiditeit en met een verminderd fysiek functioneren en kwaliteit van leven. Wereldwijd wordt aangenomen dat ongeveer 50% van de bloedarmoede komt door een tekort aan ijzer. Vandaag de dag zijn er in Afrika een veel factoren (dieet, geografische ligging, klimaat, blootstelling aan ziektekiemen en vele andere factoren), die wellicht in combinatie met elkaar de ontwikkeling van ijzertekort beïnvloeden. De wijze waarop genetische factoren samen hangen met milieufactoren, zoals dieet, is van belang in de ontwikkeling en vatbaarheid van een ijzertekort. Inzicht in de genetica van ijzerstatus is de sleutel tot het optimaliseren van strategieën om ijzertekort te verminderen, evenals het ontdekken van geschikte biomarkers om de ijzerstatus in populaties te monitoren.

Met succesvolle genoombrede associatie (GWA) studies zijn veel gemeenschappelijke genetische varianten die met ijzerstatus verbonden zijn ontdekt. Met deze genetische varianten is de kennis van het ijzer metabolisme vergroot en daarnaast heeft het bijgedragen aan de ontwikkeling van strategieën om ijzertekort aan te pakken. Bovendien heeft dit ook invloed op de ontwikkelen van geschikte biomarkers voor het monitoren van de ijzerstatus. Omdat de Europese bevolking slechts over een klein deel van de wereldwijd beschikbare menselijke genetische variatie beschikt, wordt de generaliseerbaarheid van resultaten naar de Afrikaanse bevolking beperkt. Daarnaast varieert de Afrikaanse populatie in vergelijking met Europese populaties qua allel frequenties en biologische aanpassingen, allemaal factoren die invloed hebben op de detecteerbaarheid en het belang van risico varianten die verbonden zijn met ijzerstatus. Voordat biomarkers die geassocieerd zijn met ijzerstatus worden ingezet op volksgezondheidsniveau voor de Afrikaanse populatie, moeten ze eerst grondig onderzocht worden met onder andere associatie studies.

In **hoofdstuk 2** hebben we een systematisch review inclusief meta-analyse uitgevoerd met loci in het TMPRSS6 gen waar in eerdere studies, in cohorten met een populatie van Europese afkomst, associaties zijn gevonden met ijzerstatus. Deze studie werd uitgevoerd om een nauwkeuriger beeld te krijgen van de eerder geïdentificeerde associatie van het TMPRSS6 single nucleotide polymorfisme met ijzerstatus in cohorten van verschillende continenten.

We vonden dat risico-allelen van de belangrijkste TMPRSS6 varianten zoals rs855791 en rs4820268 een associatie vertoonden met lagere Hb en serum ferritine concentraties voor zowel populaties van Europese als Aziatische afkomst. We hebben ook geconstateerd dat er verschillen waren in de minor allel frequenties van deze varianten voor verschillende

etnische groepen. Op het moment dat we de studies selecteerden voor inclusie in de review, werd slechts één onderzoek gevonden met informatie over een cohort van Afrikaanse afkomst, in dit geval van Afro-Amerikanen. Er was geen informatie beschikbaar over Afrikaanse bevolkingen.

In **hoofdstuk 3** beschrijven we de resultaten van de eerste grote kandidaat-gen associatie studie over de genetica van ijzerstatus, uitgevoerd onder zwarte Zuid-Afrikaanse vrouwen. De onderzochte selectie van kandidaat-genen voor genotypering waren gebaseerd op informatie van GWA studies in populaties van Europese afkomst. Wij hebben geen overtuigend statistisch bewijs gevonden (p<0,05) dat specifieke genetische loci geassocieerd zijn met ijzerstatus. Minor allel frequenties (MAF) van de onderzochte SNP's van de vrouwelijke deelnemers van de PURE populatie waren lager dan die van mannen en vrouwen van Europese kom af uit het 1000 genome project. In het TF-gen, werd een associatie gevonden tussen de enkel-nucleotide polymorfie (SNP) rs1799852 en een verlaagd serum ferritine (p=0,01) en lichaamsijzer (p=0,03) en verhoogde serum transferrine receptor (sTfR) concentraties (P=0,004), terwijl rs3811647 was geassocieerd met de transferrine receptor en lichaamsijzer (beide P=0,03). De chromosoom 6 SNP allel combinatie (AAA) bestaande uit rs1799964 en rs1800629, beide in TNF α en rs2071592 in NFKBIL1, waren geassocieerd met een verhoogde kans op lage serum ferritine concentraties (serum ferritine <15 ug/L, OR: 1,86 (95%BI, 1,23 -2,79)). De chromosoom 22 SNP allel combinatie (GG), bestaande uit rs228918 en rs228921 in het TMPRSS6 gen, werd geassocieerd met een lagere kans op verhoogde concentraties sTfR (sTfR> 8.3mg L, OR: 0,79 (95%BI, 0,63-0,98).

In **hoofdstuk 4** onderzochten we enkel-nucleotide polymorfismen die geassocieerd zijn met ijzerstatus in verschillende Afrikaanse cohorten. Dit was een kandidaat-gen associatie studie en daarna werden meta-analyses van de resultaten van de genetische associaties van de individuele Afrikaanse cohorten (uit Kenia, Tanzania en Zuid-Afrika) vergelijken met de resultaten van een Afro-Amerikaans cohort. We hebben ook alle resultaten uit de vier cohorten gecombineerd en vergeleken met eerdere bevindingen uit Europese. Wij hebben met succes associaties gereproduceerd voor verlaagde hemoglobine concentraties met twee varianten in TMPRSS6 (rs2413450 en rs4820268) en voor een verhoogde hemoglobine concentratie met één variant in TF (rs3811658). Wanneer ferritine als uitkomstmaat werd genomen, hebben we associaties gereproduceerd voor verhoogde ferritine concentraties met twee loci namelijk rs228918 in TMPRSS6 en rs1525892 TF (P<0,05 voor alle significante associaties). Verder werden geen significante

associaties met hemoglobine of serum ferritine concentraties waargenomen. In **hoofdstuk 5** is een studie beschreven die de associatie verkent tussen twee SNPs in het TNF α-gen namelijk TNF-1031 (rs1799964) en TNF-308 (rs1800629) en het optreden en de ernst van malaria aanvallen in een cohort van Tanzaniaanse kinderen. Onze belangrijkste bevinding was dat in vergelijking met hun gemeenschappelijke allel varianten, een verhoging van 30-40% van malaria bij kinderen met het TNF-1031 CC genotype werd gevonden. We vonden geen bewijs dat een van beide allel varianten het hemoglobinegehalte beïnvloed bij de eerste malaria aanval, of dat de associatie tussen *Plasmodium* infectie en hemoglobine concentratie veranderde bij aanvang van de studie. Tot slot, worden de belangrijkste bevindingen en de methodologische aspecten die mogelijk invloed hebben gehad op de interne en externe validiteit van onze bevindingen besproken in **hoofdstuk 6**. Daarnaast worden ook de implicaties en adviezen voor de volksgezondheid, en aanbevelingen voor toekomstig onderzoek beschreven.

Voor zover wij weten is het werk dat hier gepresenteerd word het eerste uitgebreide onderzoek naar de genetica van ijzerstatus in de Afrikaanse populatie. We concluderen dat de meeste associaties die eerder geïdentificeerd zijn in Europese populaties niet zijn gevonden in Afrikaanse populaties. Daarnaast zijn minor allel frequenties van enkel-nucleotide polymorfismen geassocieerd met ijzer-status in Europese cohorten hoger dan in Afrikaanse populaties. Deze bevinding duidt er op dat nog niet ontdekte enkel-nucleotide polymorfismen de ijzerstatus in de Afrikaanse populatie zouden kunnen beïnvloeden. De waarneming dat bij een veel voorkomende TNF-α variant, TNF-1031 (rs1799964), malaria vaker voorkwam geeft aan dat om ijzertekort te verminderen in malaria endemische gebieden- preventie en behandeling noodzakelijk is. Te meer omdat het bekend is dat malaria een negatieve invloed heeft op de ijzerstatus. Op basis van onze observaties, is nader onderzoek nodig om genetische varianten geassocieerd met ijzerstatus in de Afrikaanse populatie te onderzoeken. Een aantal belangrijke factoren moeten in aanmerking worden genomen in het ontwerp en de uitvoering van toekomstige associatie studies zoals de steekproefgrootte, en het verzamelen van alle belangrijke gegevens die van invloed kunnen zijn op ijzerstatus zoals ijzer uit de voeding, leeftijd, en het vaststellen van de aanwezigheid van ontstekingen. Verdere mechanistische studies zouden nuttig zijn om de specifieke actie van de geïdentificeerde varianten nader vast te stellen. Enkel-nucleotide polymorfismen die ijzerstatus beïnvloeden zullen helpen tot een beter begrip van ijzer metabolisme en homeostase. Naast een bijdrage aan de ontwikkeling van biomarkers die niet beïnvloed worden door

de tekortkomingen van de huidige ijzerstatus biomarkers, is het ontrafelen van de genetica van ijzerstatus veelbelovend voor de inspanningen die gericht zijn op het optimaliseren van strategieën om ijzertekort te verminderen.

Acknowledgements

Writing the acknowledgements section has been every bit as tasking as getting the thesis chapters ready. This is because so many minds and hands influenced the final product. While some offered the academic support required, others offered the equally important social support. I therefore must confess that the list of people mentioned in gratitude in this section is by no means exhaustive. My hope is that all who contributed, those mentioned here and otherwise, will feel my appreciation and share in the success of knowing that the work presented herein adds to the limited body of work on the genetics of iron status in African populations, thus seeking to contribute positively to combating iron deficiency within the African continent.

The caterpillar goes from being one of the most ordinary creatures to being a beautiful butterfly. It no longer has to crawl on the ground; its perspective changes because as a butterfly it can now fly. The story of this PhD journey is one of metamorphosis from a caterpillar to a butterfly.

The Wageningen leg of my academic research journey began in 2007 when I first arrived to undertake an MSc in Nutrition and Health. I undertook my MSc thesis under the direction of **Dr. Alida Melse** and we have worked together since. I would like to express my deep gratitude to you Alida, for believing in me. My metamorphosis began with you: you have given me wings to fly and work independently as a scientist, and prepared me for the journey that is my career. In particular, I thank you for your patient guidance, enthusiastic encouragement and useful critiques of this research work.

I thank my promotor **Prof. Dr. Michael Zimmermann,** for the conversations we had whenever you visited the Division of Human Nutrition at Wageningen. Your comments in the planning stage as well as in the preparation of manuscripts and the thesis helped me to think more critically of my work.

I also thank my promotor **Prof Edith Feskens,** for statistical advice and assistance in keeping my progress on schedule. You injected life into my PhD work and were enthusiastic at every meeting. Your easy-going, yet focussed nature kept me fervent about my work. You helped me push my boundaries (especially in teaching myself new statistical software) and to go beyond what I thought I was capable of. Under you I have learned that if I put my mind to it, I can do it.

As my work evolved, I worked with several co-authors to prepare manuscripts. I thank you **Dr. Hans Verhoef**, **Prof. Dorine Swinkels** and **Dr. Toshiko Tanaka**. I have learnt a lot from you. Your mentorship as we prepared the manuscripts was invaluable.

My appreciation is also extended to **Dr. Mark Boekschoeten and Jan Harryvan** of the nutrigenomics group at Wageningen for your help in orienting me to the lab work involved in my study. Additionally I would like to thank **Richard Crooijmans** together with the staff at the **Animal Breeding Group (ABG) of Wageningen University**. You did all my genotyping work, answered my questions, as well as facilitated Lab visits. Thank you for a job well done, for your availability and your patience.

The administrative tasks that came with this PhD could not have been managed without the support of the late **Lidwien Van der Heijden**, **Lous Duym**, **Gea Brussen**, **Cornelia Van Bree**, **Jacqueline Verhoef-te Brake, Jasmijn Mater** and **Karen Zweers,** all of the Division of Human Nutrition, Wageningen. Thank you all for all your assistance in arranging my visas, flights, and medical insurance. I also thank **Riekie Janssen** and **Eric van Munster** for promptly arranging all my financial matters.

To my fellow PhDs' at the division, past and present - **Catherine, Razak, Yarra, Akwilina, Sophie, Martin, Laura, Elise, Por, Nidhi, Ana Carla, Phyllis, Nicole,** and **Rina** - thank you for the times and discussions with each one of you in the division at Wageningen, which enriched my PhD experience. I especially thank **Catherine, Akwilina, Sophie, and Phyllis** for the many meals and conversations that I enjoyed sharing with you and especially that you gave me the much needed escape from the studies. **Elise and Gloria**, I appreciate your kindness during my stay with you at Kibwezi.

As this was a collaborative project together with the Centre of Excellence for Nutrition (CEN) in North West University, Potchefstroom South Africa, I owe every success to the team at the CEN as well. I thank **Prof. Johann Jerling and Prof. Marius Smuts,** who together with the Wageningen team, secured the funding required for my stay in South Africa. Thank you, **Prof. Johann,** for your open door policy that made my stay at the CEN comfortable. I thank you, **Prof. Smuts,** for your being so approachable and always available to sort out any issue concerning my stay.

I also thank you, **Prof. Salome,** for working tirelessly together with us to ensure we secured funding for the PURE project work. Without your efforts, we would not have managed to work on the data from this study.

My gratitude is also extended to **Dr. Wayne Towers,** my local supervisor in South Africa. From you I learnt how to be meticulous as a scientist - to dot my i's and cross my t's. I will definitely carry this lesson with me. In addition, thank you for your time and effort in arranging a home for me to stay at in South Africa, and for going out of your way to ensure I was comfortable.

To **Tinashe Chikowore** and **Herman Myburg**, who helped me in sample preparation; thank you. You put in the extra hours required to meet my deadline for sample shipment. I also thank **Bianca Swanepoel** for taking time to orient me on the lab equipment use.

I thank **Ronel Benson**, **Henrëtte Claasen** and **Sanet Vermeulen** for making all the administrative arrangements to get me to South Africa, as well as ensuring my stay was comfortable.

I thank **Dr. Cornelie Nienaber-Rousseau** and husband **Jacques Rousseau**, for your unparalleled kindness and generosity. The three of us had many long philosophical discussions about life, my perceptions of South Africa and Africa in general, and most of all on research. You entertained my many questions and insatiable curiosity. Above all, thank you for opening your home to me, and Cornelie, for giving me ideas on how to tackle various practical issues in my work. I only hope that one day I can repay your acts of kindness.

Dr. Namukolo Covic - what an awesome, wise and patient lady! Every young African woman scientist needs a mentor like you - you are the best! I appreciate that you always had a listening ear and gave the most excellent advice. I hope that as a thank you to your warmth and kindness, this PhD thesis does you proud.

To the CEN lunch table girls **Jennifer Akosy, Olusula Sotunde, Christiana Berti, Jeanine Baumgartner, Marilize Richter** and **Marie Nel** - what great laughs we had at the lunch table! Thank you for making my days in Potchefstroom interesting. Oh and the surprise goodbye party that you all colluded to arrange, I will never forget how special

that made me feel. Thank you! **Wisdom Dube,** you were a mine of information on survival for the international student in Potchefstroom, and for that I thank you.

Throughout my PhD journey as well as stay in Wageningen, spiritual nourishment has been my rock. In particular I thank the Redeemed Christian Church of God, Amazing Grace Parish Church in Wageningen for a spiritual home in Wageningen. I thank **Pastors Farai and Busi Maphosa, Pastor Adesuwa, Deji and Olayinka Bamidele, Sunday and Maureen Makama, Elton and Ednah Zvinavashe, Peter and Kumba Morsay-** you all opened your hearts and homes to me. In you I found comfort and as we shared in His word, rest. I also thank the members of the Praise and Worship team with whom I sang together all these years: for *"I will sing unto the LORD, because he hath dealt bountifully with me (Psalm 13:6)"*.

To the Kenyan community in Wageningen, parties with you were always heart-warming. I shall cherish the memories.

To my friends far and wide – **Grace, Juddy, Chibole, Liza, Susan, Njeri ("Mama Viktoria"), Shiku Karue, Cyrus, Rita** and **Tabitha** - thank you for encouragement and keeping our friendship alive and for some of you even when we were separated by great distances. Your visits and encouraging messages energised me to trudge on. Susan thank you for always checking if I was still sane and making me laugh. Grace, such a pity you could not see your "little scholar" graduate today. Thank you for encouraging me and being my friend all these years. You are special. I particularly thank you **Juddy** and **Njeri** for booking flights and sacrificing your time to be there for me on my graduation.

I thank my home institution - **Jomo Kenyatta University of Agriculture and Technology,** as well as my colleagues at the Department of Food Science for supporting me during my studies. Special thanks to you **Irene,** for all the "little" errands you ran for me. I do not think you know how much it meant to me as your assistance allowed me to concentrate on my work and gave me peace; I am eternally grateful.

I cannot forget to thank my paranymphs, **Laura Trijsburg** and **Mwikali Kioko**. To **Laura**, thank you for your friendship and for all the time and energy you put into my defense preparations. I particularly cannot forget the invites to your home for delicious meals and for letting me know that if I needed help getting my way around the Dutch systems, you would be there to help. I hope I can do the same for you whenever you visit Kenya.

Mwikali, since we met in 2006 you have been a great source of inspiration. You have encouraged me, helped me keep things in perspective, and always taught me to focus on the main thing. I am pleased that you are beside me during my defence and graduation.

I wish to thank my parents **Charles and Lucy Gichohi,** for your support and encouragement throughout my study. You have made me the woman I am today. Thank you, mum, for all the thoughtful and practical things you did during the writing phase of this thesis. Most important how every week you brought to our home a basketful of fruits and vegetables to ensure I did not fall sick as I finalised my thesis. Baba, you have always been my sounding board and the one who thought highly of my abilities. I faced many challenges during this PhD, and yours was the shoulder I cried on and your thoughts offered much needed clarity and direction. You taught me the value of hard work and reminded me that hard work always pays. Well, it has paid off ! I thank you for engaging me in discussions about my work and making useful contributions - you have enriched the final product. For your love, encouragement and wise words, I dedicate this thesis to you.

To my siblings, **Muraya, Muhoro, Wambugu and Watetu,** I thank you for believing in me and cheering me on tirelessly. I couldn't have asked for a better family.

Special gratitude goes to my cousins, **Nyawira** and **Shiku,** who have supported me during this PhD. Nyawira I hold you dear to my heart. You have made the journey from the US, to be there for me on this day. Having family present during my graduation means the world to me. Thank you!

To the entire **Muiruri family**, thank you for keeping Martin company in my absence.

Last, but certainly not the least, I thank the man of my life, the mate of my soul - my husband **Martin**. We did it! Thank you for making my dream our dream. You were always there to support me - staying up late with me and taking care of our home. You never complained about the numerous goodbyes and trips to the airport, as well as the constant packing and unpacking. I dedicate this thesis to you and hope I have done you proud. Thank you!

Above all I give glory and honour to God!

About the author

Wanjiku N Gichohi was born on 4th June 1982 in Nairobi, Kenya. In 2006, she graduated from the department of Food science, Jomo Kenyatta University of Agriculture and Technology with a BSc. degree in Food science and nutrition (First class honours) .

She then worked for a Caritas Internationalis, a non-governmental organization, as a programme officer. Her main role was on co-ordination of multi stakeholder activities with other NGOs' involved in women empowerment activities. In 2007, she was employed as a member of the teaching staff at Jomo Kenyatta University of Agriculture and Technology. In 2007, she was granted a scholarship by nestle nutrition institute (NNI), Lausanne, Switzerland, to study for an MSc degree in nutrition and health at Wageningen University, The Netherlands. In 2009, she received her MSc degree in nutrition and health (specialization-public health nutrition) with her main research topic in the area of iron bioavailability. Her MSC thesis ignited an interest to delve deeper into the area of iron deficiency.

In 2010 she was appointed as a PhD fellow on a collaborative research project between the division of human nutrition, Wageningen university and the centre for excellence of Nutrition, Northwest University, Potchefstroom Campus. Wanjikus' main focus in the thesis was on understanding the genetics of iron status in African populations. She joined the PhD tour to Mexico and South West USA in 2010. In 2012, she was selected to join the African Nutrition Leadership Programme (ANLP) and was part of the ANLP junior faculty in 2013. During the PhD period she also attended several courses and conferences. Wanjiku looks forward to a continued contribution to improving the health of mothers and children.

List of Publications

Publications in peer reviewed journals

Wanjiku N Gichohi-Wainaina, G, Wayne Towers, Dorine W. Swinkels, Michael B. Zimmermann, Edith J. Feskens, Alida Melse-Boonstra. Inter-ethnic differences in genetic variants within the transmembrane protease, serine 6 (*TMPRSS6*) gene associated with iron status indicators: a systematic review with meta-analyses. *Genes and nutrition. 2015 Jan;10(1):442. doi: 10.1007/s12263-014-0442-2*

Wanjiku N Gichohi-Wainaina, Alida Melse-Boonstra, Dorine W Swinkels, Michael B Zimmermann, Edith J Feskens, G Wayne Towers. Common variants and haplotypes in the *TF, TNF α* and *TMPRSS6* genes are associated with iron status in a female black South African population. *Journal of nutrition (In press)*

Submitted publications

Wanjiku N Gichohi-Wainaina , Toshiko Tanaka , G Wayne Towers , Hans Verhoef , Jacobien Veenemans , Elise F. Talsma, Jan Harryvan , Mark Boekschoten , Edith J Feskens, Alida Melse-Boonstra. Associations between common variants in iron-related genes with haematological traits in populations of African ancestry

Wanjiku N Gichohi-Wainaina, Alida Melse-Boonstra, Edith J Feskens, Ayse Y Demir, Jacobien Veenemans, Hans Verhoef. Tumour necrosis factor allele variants and their association with the occurrence and severity of malaria in African children: a longitudinal study

Conference abstracts

Alida Melse-Boonstra, Wanjiku Gichohi, Elise F. Talsma, Jan Harryvan, Mark Boekschoten, Wayne Towers and Michael B. Zimmermann. Single nucleotide polymorphisms in the *TMPRSS6* gene and iron status indicator in Kenyan school children. Bio iron conference; University College London UK; April 2013 (Poster presentation)

W Gichohi-Wainaina, W Gordon Towers, D Swinkels, MB Zimmermann, EJ Feskens, A Melse-Boonstra. Genetic variants in the TMPRSS6 gene and iron status indicators: A systematic review. Nutrition congress Africa, Bloemfontein, South Africa, September 2014 (Oral presentation)

Wanjiku Gichohi-Wainaina, Alida Melse-Boonstra, Dorine Swinkels, Michael B Zimmermann, Herman Myburgh, Tinashe Chikowore, Edith J Feskens, G Wayne Towers. Gene-environment interactions of specific *TF* gene variants in relation to iron status in black South Africans, International Congress of Nutrition (IUNS), Granada, Spain; September 2013 (Poster presentation)

Overview of completed training activities

Activity	Organizers and Location	Year
Discipline specific Courses		
Polymorphisms and responsiveness to diet	Nutrigenomics Organization (NUGO)	2010
Principles of genetic epidemiology	NIHES, Rotterdam, The Netherlands	2011
Genome wide association analysis	NIHES, Rotterdam, The Netherlands	2011
Discipline specific conferences, workshops and seminars		
Genetics in child cohort studies	Rotterdam, The Netherlands	2010
Nutrition congress Africa	Bloemfontein, South Africa	2012
IUNS 20th International Congress of Nutrition	Granada, Spain	2013
Micronutrient Forum	Addis Ababa	2014
Iodine provision during the first 1000 days of life	Division of Human nutrition, Wageningen University, The Netherlands	2014
Nutritional iron, anaemia and infectious diseases	Division of Human nutrition, Wageningen University, The Netherlands/ Cell biology and immunology group	2014
General courses		
1st Master class in Linear and Logistic Regression	Wageningen University, The Netherlands	2010
Working with endnote	Wageningen University, Netherlands	2010
Project and time management	Wageningen University, The Netherlands	2010
Regression analysis	Wageningen University, Netherlands	2011

PhD week	Graduate School VLAG (Advanced Studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences)	2011
Good Clinical Practice	Centre of Excellence for Nutrition (CEN), North West University, Potchefstroom, South Africa	2012
Africa Nutrition Leadership programme (ANLP), participant	Centre of Excellence for Nutrition (CEN), North West University, Potchefstroom, South Africa	2012
International course on evidence based nutrition	Institute of Tropical Medicine, Antwerp, Belgium	2013
Optional courses and activities		
PhD tour Mexico and South West USA	UC Davis, San Francisco, Berkley (USA), Universidad de Monterrrey, Mexico, Unilever Mexico, National Institute of Public Health, Mexico	2010
Nutritional Genomics and Genetics	Wageningen University, The Netherlands	2011
Journal club	Centre of Excellence for Nutrition (CEN), North West University, Potchefstroom, South Africa	2012-2013
Rapporteur-High level forum on micronutrient deficiencies	Division of Human nutrition, Wageningen University, The Netherlands	2014
Staff Seminars/student presentations 2010/2014	Division of Human nutrition, Wageningen University, The Netherlands	2010-2014
NL for NIL malnutrition! The Lancet Nutrition Series 2013: What's their message – what's our action?	Ministry of Foreign Affairs, The Hague, The Netherlands	2013