Blood Cells, Molecules and Diseases 57 (2016) 50–53

Contents lists available at ScienceDirect

Blood Cells, Molecules and Diseases

journal homepage: www.elsevier.com/locate/bcmd

# Genome-wide association study of serum iron phenotypes in premenopausal women of European descent

Daniel L. Koller [a,*], Erik A. Imel [b], Dongbing Lai [a], Leah R. Padgett [b], Dena Acton [b], Amie Gray [b], Munro Peacock [b], Michael J. Econs [a,b], Tatiana Foroud [a]

[a] Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[b] Department of Medicine, Indiana University School of Medicine, Indianapolis, IN 46202, USA

ARTICLE INFO

*Article history:*
Submitted 11 September 2015
Accepted 2 December 2015
Available online 3 December 2015

Editor: Mohandas Narla

*Keywords:*
Iron
Premenopausal
GWAS
Transferrin
Hemochromatosis

ABSTRACT

A genome-wide association study was performed on 1130 premenopausal women to detect common variants associated with three serum iron-related phenotypes. Total iron binding capacity was strongly associated ($p = 10^{-14}$) with variants in and near the *TF* gene (transferrin), the serum iron transporting protein, and with variants in *HFE* ($p = 4 \times 10^{-7}$), which encodes the human hemochromatosis gene. Association was also detected between percent iron saturation ($p = 10^{-8}$) and variants in the chromosome 6 region containing both *HFE* and *SLC17A2*, which encodes a phosphate transport protein. No significant associations were detected with serum iron, but variants in *HFE* were suggestive ($p = 10^{-6}$). Our results corroborate prior studies in older subjects and demonstrate that the association of these genetic variants with iron phenotypes can be detected in premenopausal women.



## 1. Introduction

Iron deficiency anemia is a common disease, particularly among women of reproductive age due to menstrual blood loss [1,2]. Iron overload is relatively uncommon, but causes toxicity in a variety of cell types and tissues [3]. Iron balance is regulated at the level of the enterocytes and absorbed iron is bound to serum transferrin for transport to the tissues. Serum iron concentration, along with the unsaturated iron binding capacity of transferrin, reflects to a large extent the state of the iron cycle in an individual. Consequently, clinical measurement of serum iron, total iron binding capacity (TIBC) and percent iron saturation may be used in assessing critical parameters indicating iron status [4]. Substantial portions of the variation in iron phenotypes among individuals have been attributed to genetic factors in several populations [5,6]. Importantly, Benyamin [7], Tanaka [8], McLaren [9] and Li [10] have conducted genome-wide association studies (GWAS) of serum iron and related measures to detect common genetic polymorphisms contributing to the observed variation in these measures. They found significant evidence of association between these iron-related phenotypes and single nucleotide polymorphisms (SNPs) in genes known to be involved in iron metabolism (hemochromatosis, *HFE*; transferrin, *TF*; transmembrane protease serine 6, *TRMPSS6*), as well as at other loci.

These GWAS have included few or no premenopausal women, either because of concerns regarding the influence of menstrual blood loss on serum iron levels, or because younger women were absent from the cohorts. Thus, it is unclear at which age the genetic polymorphisms contributing to variation in serum iron-related phenotypes have their effect [11]. The aim of this study was to provide data regarding genes that influence iron metabolism in healthy premenopausal women; and further, by comparing these results with those from published studies, to provide information on the period of life in which genetic variants may have their primary effect. To address these aims, we performed GWAS studies of serum iron, total iron binding capacity (TIBC) and percent iron saturation in a cohort of 1130 healthy, premenopausal European-American women.

## 2. Material and methods

### 2.1. Subjects

Full sibships consisting of healthy European-American premenopausal women were recruited for a genetic study of bone mineral density in Indiana, along with additional unrelated women or half-sibships meeting the same criteria [12]. DNA was isolated from whole blood, and an additional blood sample was collected to obtain serum at the baseline visit (age 20–45) or during a second visit 4–5 years later (for

* Corresponding author at: Indiana University School of Medicine, Health Information and Translational Sciences Building (HS 4000), 410 W. 10th Street, Indianapolis, IN 46202-3022, USA.
*E-mail address:* dkoller@iu.edu (D.L. Koller).

http://dx.doi.org/10.1016/j.bcmd.2015.12.002
1079-9796/

**Table 1**
Iron phenotypes and sample demographics.

| Variable | n | Mean | SD | Min | Max |
|---|---|---|---|---|---|
| Age, years | 1130 | 34.3 | 7.7 | 20 | 50 |
| Weight, kg | 1130 | 71.6 | 16.9 | 39.6 | 133.9 |
| TIBC, mcg/dl | 1122 | 371.7 | 56.0 | 220.5 | 577.3 |
| Iron, mcg/dl | 1130 | 97.5[b] | 42.6 | 14.9 | 292.6 |
| ln(iron)[a] | 1130 | 4.5 | 0.5 | 2.7 | 5.7 |
| Iron saturation, % | 1122 | 26.6 | 11.7 | 3.5 | 75.3 |
| UIBC, mcg/dl | 1122 | 274.1 | 66.1 | 85.0 | 483.7 |

[a] Summary statistics for iron levels after natural-log transformation, as used in the statistical analyses presented. Untransformed values (Iron, mcg/dl) in the previous row of the Table are included for reference.
[b] 102 (9.0%) of the study subjects were iron deficient, as defined by serum iron levels below the normal range (<50 mcg/dl).

subjects yet to reach menopause by that time) [13]. Studies were performed at the General Clinical Research Center of the Indiana University School of Medicine. Informed, written consent (Indiana University IRB: 8502-23) was obtained from all participants.

### 2.2. Iron measures

Iron and unsaturated iron binding capacity (UIBC) were measured on stored serum samples collected in subjects after an overnight fast using a spectrophotometric assay (Randox Laboratories; Antrim, Northern Ireland, United Kingdom). TIBC was calculated as the sum of serum iron and UIBC. Iron saturation was calculated as serum iron/TIBC and expressed as a percentage.

### 2.3. Genotyping and imputation

Genotyping was performed on the Illumina Human610Quadv1_B BeadChips (Illumina, San Diego, CA, USA) by the Center for Inherited Disease Research (CIDR) using the Illumina Infinium II assay protocol. This array contains 592,532 markers. Genotype calls and quality control filters were applied as described previously [12]. Briefly, DNA samples with a 98% SNP call rate and genotypes consistent with reported family structures were included. A principal component analysis was performed in Eigenstrat [14] to cluster these samples with HapMap reference samples (CEU, YRI, CHB, and JPT). Only samples consistent with European ancestry were retained. Subsequently, individual SNP metrics were computed. SNPs having a call rate less than 95%, estimated minor allele frequency (MAF) less than 0.03 or significant deviation ($p < 0.0001$) from the Hardy Weinberg equilibrium were removed from further analysis. The genotype dataset taken forward for imputation contained 539,566 SNPs.

The IMPUTE2 software [15,16] was used to generate genotype data for the 1000 Genomes Phase I SNP map in chromosomal regions with genome-wide significant evidence of association ($p < 5 \times 10^{-8}$) to one or more of the iron phenotypes. The integrated variant set for samples of European descent was used as the source of reference haplotypes. The original genotype call was not overwritten for the set of SNPs genotyped on the Illumina array. A large subset of rare and low-frequency variants (MAF < 0.03) or those of lower imputation certainty (IMPUTE2 info score < 0.3) were excluded from association analyses.

### 2.4. Statistical analysis

For each of the phenotypic measures, distributions were evaluated and transformations applied as necessary to remove skewness or kurtosis. Pearson correlation coefficients (r) between the measures were computed using SASv9.1 (SAS Institute, Cary, NC, USA), and a subset of phenotypes selected for GWAS analyses to minimize the number of tests performed.

For each genotyped SNP in the filtered dataset, association tests were performed separately for each of the selected phenotypes. These models were fitted with R package GWAFv2.1, using a mixed model to account for relatedness within our sample [17]. Genome-wide results for each phenotype were evaluated for evidence of inflation of the association statistics. Genomic control as implemented in METAL [18] was applied as needed to correct for cryptic admixture or phenotypic non-normality. Imputed SNPs in chromosomal regions with evidence of association to one or more phenotypes were tested for association in the same manner.

**Table 2**
Pearson correlation coefficients (r) for the assayed or calculated iron-related phenotypes (n = 1130).

| Variable | TIBC | Iron | Iron saturation |
|---|---|---|---|
| Iron | 0.08 | | |
| Iron saturation | −0.22 | 0.90 | |
| UIBC | 0.77 | −0.55 | −0.79 |

## 3. Results

Evaluation of the iron phenotype distributions showed that only iron concentration was not normally distributed (skewness = 0.71); this was corrected by a natural-logarithm transformation (skewness = 0.11). Summary statistics for the iron phenotypes and subject demographics are provided in Table 1, and Pearson correlation coefficients (r) between the four iron phenotypes are shown in Table 2. Serum samples from study visits 4–5 years after baseline were included for 7 premenopausal subjects among the subset of our GWAS cohort included in the current report, when serum from the baseline visit was unavailable. Table 1 shows that demographic parameters are unchanged from the previous publication. Pairwise values of |r| greater than 0.5 with all other phenotypes were observed for UIBC; therefore, TIBC, iron, and iron saturation were selected as outcome measures for quantitative GWAS analyses.

Results for the genome screen for TIBC are shown in Fig. S1A (Supplementary material). There was evidence of inflation of the genome-wide association statistics ($\lambda = 1.18$) and genomic control was applied. Evidence of association for TIBC far surpassing the typical threshold for genome-wide significance ($5 \times 10^{-8}$) was found on chromosome 3 (133.48 Mb, build 37 coordinates), with the most significant results found for a group of relatively common SNPs (MAF = 0.33) in the region of the transferrin (*TF*) gene (Fig. 1A). SNPs at or above the genome-wide significance level spanned the full length of the *TF* gene, extending from the region 5′ of *TF*, to the 5′ region of the neighboring gene, *SRPRB8*.

Results of the association analysis of the iron saturation phenotype are shown in Fig. S1B. Inflation of the full set of test statistics was minimal ($\lambda = 1.05$) and corresponding corrections were not applied. Genome-wide significant evidence for association ($p = 2 \times 10^{-8}$) was observed for iron saturation near the major histocompatibility complex (MHC) region on chromosome 6, with SNPs of modest allele frequency (MAF = 0.06–0.08). These included the nonsynonymous variant rs1800562 (C282Y) in the hereditary hemochromatosis (*HFE*) locus (Fig. 1B). SNPs in this region also approached the significance threshold with the other iron phenotypes as well ($p \le 1 \times 10^{-6}$; Tables S1A–S1C, Supplementary material).

Results of the GWAS for iron level are shown in Fig. S1C. As for iron saturation, genomic control correction was not necessary ($\lambda = 1.04$). None of the genotyped or imputed SNPs demonstrated genome-wide evidence of association.

## 4. Discussion and conclusions

Our GWAS, in a sample of healthy premenopausal women for three iron phenotypes, detected highly significant associations with several loci previously reported in samples of older subjects. The strongest