haematologica
the hematology journal

Early Release Paper

CYBRD1 as a modifier gene that modulates iron phenotype in *HFE*p.c282y homozygous patients

by Sara Pelucchi, Raffaella Mariani, Stefano Calza, Anna Ludovica Fracanzani, Giulia Litta Modignani, Francesca Bertola, Fabiana Busti, Paola Trombini, Mirella Fraquelli, Gian Luca Forni, Domenico Girelli, Silvia Fargion, Claudia Specchia, and Alberto Piperno

Haematologica 2012 [Epub ahead of print]

Citation: Pelucchi S, Mariani R, Calza S, Fracanzani AL, Modignani GL, Bertola F, Busti F, Trombini P, Fraquelli M, Forni GL, Girelli D, Fargion S, Specchia C, and Piperno A. CYBRD1 as a modifier gene that modulates iron phenotype in HFEp.c282y homozygous patients. Haematologica. 2012; 97:xxx doi:10.3324/haematol.2012.062661

Haematologica (pISSN: 0390-6078, eISSN: 1592-8721, NLM ID: 0417435, www.haematologica.org) publishes peer-reviewed papers across all areas of experimental and clinical hematology. The journal is owned by the Ferrata Storti Foundation, a non-profit organization, and serves the scientific community with strict adherence to the principles of open access publishing (www.doaj.org). In addition, the journal makes every paper published immediately available in PubMed Central (PMC), the US National Institutes of Health (NIH) free digital archive of biomedical and life sciences journal literature.

DOI: 10.3324/haematol.2012.062661

CYBRD1 as a modifier gene that modulates iron phenotype in *HFE* p.c282y homozygous patients

Running title: Genetic modifiers of HFE-HH phenotype

Sara Pelucchi[1], Raffaella Mariani[2], Stefano Calza[3,4], Anna Ludovica Fracanzani[5], Giulia Litta Modignani[1], Francesca Bertola[6], Fabiana Busti[7], Paola Trombini[2], Mirella Fraquelli[8], Gian Luca Forni[9], Domenico Girelli[7], Silvia Fargion[5], Claudia Specchia[3], and Alberto Piperno[1,2,6]

[1]Department of Clinical Medicine and Prevention, University of Milano-Bicocca, Monza; [2]Centre for Diagnosis and Treatment of Hemochromatosis, S.Gerardo Hospital, Monza; [3]Department of Biomedical Sciences and Biotechnologies, University of Brescia, Brescia; [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; [5]Department of Medicine, IRCCS, Ospedale Maggiore Policlinico, University of Milano, Milano; [6]Consortium for Human Molecular Genetics, Monza; [7]Department of Medicine Policlinico GB Rossi, University of Verona,Verona; [8]Second Division of Gastroenterology, IRCCS, Ospedale Maggiore Policlinico, University of Milano, Milano, and [9]Centre of Microcitemia and Congenital Anemias, Galliera Hospital, Genova

Key words: hemochromatosis, SNP, iron, gene, ferritin, transferrin saturation

Funding

This work was supported by Progetto Quadro Regione Lombardia funding (CUP D41J10000390002) to AP. SP is supported by a research fellowship from University of Milano-Bicocca and by Associazione per lo Studio dell'Emocromatosi +Fe – ONLUS, Monza, Italy.

Correspondence

Alberto Piperno, MD, Department of Clinical Medicine and Prevention, Centre for Diagnosis and Treatment of Hemochromatosis S.Gerardo Hospital, via Pergolesi 33, 20900, Monza, Italy. Phone: international +39.039. 2339710. E-mail: alberto.piperno@unimib.it

DOI: 10.3324/haematol.2012.062661

Abstract

Background. Most patients with Hereditary Hemochromatosis are homozygous for the p.C282Y mutation in the *HFE* gene in Caucasian population. Penetrance and expression of Hemochromatosis largely differ in p.C282Y homozygous cases. Besides environmental factors, genetic factors might be implicated.

Design and Methods. In the present study, we analysed 50 candidate genes involved in iron metabolism and evaluated the association between 214 single nucleotide polymorphisms in these genes and three phenotypic outcomes of iron overload (serum ferritin, iron removed and transferrin saturation) in a large group of 296 Italian p.C282Y homozygous cases. Polymorphisms were tested for genetic association with each single outcome using linear regression models adjusted for age, sex and alcohol consumption.

Results. We found a series of 17 genetic variants located in different genes with possible additive effect on the studied outcomes. In order to evaluate if the selected polymorphisms could provide a predictive signature for adverse phenotype, we re-evaluated data by dividing patients in two extreme phenotype classes based on the three phenotypic outcomes. We found that only a small improvement in prediction can be achieved adding genetic information to clinical data. Among the selected polymorphisms, a significant association between rs3806562, located in the 5'UTR of *CYBRD1*, and transferrin saturation was observed. This variant belongs to the same haplotype block which contains the *CYBRD1* polymorphism rs884409, found to be associated with serum ferritin in another population of p.C282Y homozygotes, and able to modulate promoter activity. Luciferase assay indicates that rs3806562 has not a significant functional role, suggesting that it is a genetic marker linked to the putative genetic modifier rs884409.

Conclusions. While our results support the hypothesis that polymorphisms in genes regulating iron metabolism may modulate penetrance of *HFE*-HH, with emphasis on *CYBRD1*, they strengthen the notion that none of these polymorphisms alone is a major modifier of HH phenotype.

DOI: 10.3324/haematol.2012.062661

Introduction

Hereditary hemochromatosis (HH) is a heterogeneous disorder at both genetic and phenotypic level. *HFE*-HH is the most common form in Caucasian populations and homozygosity for p.C282Y mutation is the genotype most frequently associated with iron overload and related complications[1]. *HFE*-HH may lead to tissue iron accumulation and iron-related complications, but penetrance and expression varies greatly among p.C282Y homozygotes. Environmental and genetic factors have been implicated: blood loss, alcohol intake, coexistence of chronic hepatitis B and C, and non-alcoholic fatty liver disease (NAFLD) can influence clinical manifestations in humans[1,2]. Animal studies showed that genetic background modulates the expression of the disease[3] and identified several candidate modifiers regions in *HFE*-knock-out mice[4]. In humans, a high concordance of iron indices and/or iron-related disease has been found among related *HFE*-HH patients supporting the existence of genetic modifiers influencing phenotype expression[5,6]. Association studies between genetic markers and disease phenotype gave conflicting results[7-14]. More recently, candidate gene studies allowed to detect significant associations between SNPs in genes involved in iron metabolism and indices of iron overload in p.C282Y homozygotes. Milet *et al* focused on two biologically relevant gene categories: genes involved in non-*HFE* HH (*TFR2, HAMP,* and *SLC40A1*) and genes involved in the regulation of hepcidin expression (*BMP2, BMP4, HJV, SMAD1, 4* and *5,* and *IL6*). They found an association between the SNP rs235756 of the bone morphogenetic protein (*BMP2*) 2 gene and serum ferritin (SF) in a large French series of patients[15]. Constantine *et al* found the SNP rs884409 in *CYBRD1* as a possible novel modifier specific to *HFE*-HH, but were unable to confirm the association with the *BMP2* rs235756[16]. Moreover, genome-wide association studies performed in the general population in the last years showed associations between SNPs of Transmembrane protease, serine 6 (*TMPRSS6*) and serum iron[17,18] and transferrin saturation[19], suggesting relevant involvement of TMPRSS6 in control of iron homeostasis. In the present study, we evaluated the association between several SNPs in genes involved in iron metabolism and three phenotypic outcomes of iron overload (serum ferritin, iron

DOI: 10.3324/haematol.2012.062661

removed and transferrin saturation) in a group of Italian p.C282Y homozygous patients under the assumption that these SNPs may act as modifiers of their iron phenotype.

DOI: 10.3324/haematol.2012.062661

Design and Methods

Subjects. We enrolled in the study a group of 306 unrelated *HFE* cases. Patients were all p.C282Y homozygotes attending four Northern Italian centres (Milan, Verona, Genua and Monza). There was no selection based on disease severity. From the whole database of each Centre, patients were selected based on the following inclusion and exclusion criteria. Inclusion criteria were: p.C282Y homozygosity, availability of serum ferritin and transferrin saturation before iron depletion, good quality DNA, and information about age, sex and alcohol intake. Patients with previous history of regular blood donations were excluded from the study. Total iron removed was available in 211 patients. A group of 114 healthy controls were recruited among blood donors from the same geographic area of the patients only to further validate the quality of genotype data (calculation of Hardy-Weinberg equilibrium to check for bias and mistake in genotyping). Patients and controls gave their informed consent to the study. Regione Lombardia and University of Milano-Bicocca research fellowship committees approved the study.

Iron indices. Serum ferritin (SF) levels, total iron removed (IR) and % of transferrin saturation (TS) were used as markers of HH expression among patients. SF and TS were collected at time of diagnosis before phlebotomy therapy; IR was calculated based on the number of phlebotomies performed to achieve iron depletion as previously reported[20].

Extraction of DNA from blood. Blood samples for DNA extraction were collected in EDTA tube from all subjects. Genomic DNA was extracted from whole blood of each subject using the Wizard® Genomic DNA Purification kit (Promega, Madison, WI, USA), and stored at -20°C before use. DNA samples were adjusted to a 50ng/μl concentration.

SNPs selection. Two hundred and fourteen TagSNPs within 50 candidate genes were analysed. Two hundred and eleven were selected by Haploview Tagger[21] for the CEU population to be screened with a custom-designed 384-plex VeraCode GoldenGate genotyping assay on Illumina BeadXpress Reader platform (Illumina Inc. San Diego, CA, USA). Three (rs855791 in *TMPRSS6*,

DOI: 10.3324/haematol.2012.062661

rs884409 and rs3806566 in *CYBRD1*) were selected *a posteriori* based on previous results showing significant correlation with iron status in general population[18] and *HFE*-HH patients[16], respectively, and analysed by direct sequencing. Detailed description of SNPs selection criteria is reported in the Supplementary methods section, and the whole list of SNPs analysed in Supplementary Table 1.

SNPs genotyping. SNPs genotyping was performed using the GoldenGate Genotyping assay on an Illumina BeadXpress Reader platform according to the manufacturer's protocol. The GenomeStudio software was used to call genotypes (see Supplementary methods section). Illumina results were further validated by sequencing 100 samples for 12 SNPs on an ABI Prism 3130 Avant Automatic Sequencer (Applied Biosystems, Foster City, CA, USA).

Dual-Luciferase reporter assay. DNA fragments (415 bp) of the *CYBRD1* gene were obtained from genomic DNA of two individuals carrying TT and CC genotype of the SNP rs3806562, respectively, by PCR (forward primer: 5'-ggCTCgAgggCTggACCAgATCAAAgAA-3'; reverse primer: 5'-gggATATCgCCTgCCCTCTTTCCAACATT-3'). The PCR products, which does not include neither rs3806566 nor rs884409, were cloned into the pGL4.13 plasmid vector (Promega corp., Madison, WI, USA) upstream the firefly luciferase gene, by digestion with XhoI and EcoRV. Plasmid constructs were verified by direct sequencing. Plasmid DNAs were isolated by Pure Yield^{TM} Plasmid Miniprep System kit (Promega corp., Madison, WI, USA) for transfection. The recombinant plasmids were co-transfected with pGL4.74 plasmid (carrying hRluc luciferase reporter gene as expression control) into hepatoma carcinoma cells (Huh-7) by Lipofectamine 2000 (Life-technologies corp., Carlsbad, CA, USA). The pGL4.13 basic plasmid was also co-transfected with pGL4.74 as negative control. After incubation for 48 hr, cells were lysed and firefly and renilla luciferase activities were measured by Glomax Multi JR luminometer according to manufacturer's protocols (Promega corp., Madison, WI, USA). Each construct was tested in triplicate, and the transfection experiments were performed three times independently. Data were expressed as mean+standard deviation. Luciferase activities were compared by the Mann-Whitney test, using Prism 3.2 software (GraphPad Software, San Diego, CA, USA).

DOI: 10.3324/haematol.2012.062661

Statistical methods. Stringent quality control criteria were applied to all samples and genotype data. A per-SNP genotype rate threshold of 95% was used. Identity-by-state values were calculated for pairs of subjects to identify duplicates or possibly related subjects. For any pair with more than 95% identical SNP genotypes, the sample with the lower call rate has been removed from the analysis. SNPs with minor allele frequency less than 0.6%, have also been removed. To check for genotyping errors, Hardy-Weinberg equilibrium (HWE) for each SNP was tested among the controls with Pearson chi-squared test statistic. SNPs deviating from HWE were excluded from the analysis. Iron indices were analysed as continuously distributed outcomes and normalized using a loge transformation. A linear regression model was fitted to evaluate the effect on the three outcomes of age, sex and alcohol consumption (gr/day). Each SNP was tested for association with each single outcome by using separate linear regression models adjusted for age, sex and alcohol consumption. Genotypic association was considered and an additive genetic model was assumed. SNPs were ranked according to their uncorrected p-value and top ranked SNPs for each outcome were defined if $p<0.05$. The false discovery rate (FDR) as computed by the qvalue was applied to adjust for multiple comparisons, using a threshold of $FDR<0.2$. Interactions between pairs of top ranked SNPs with each outcome were evaluated by adding product terms to a multiple regression linear model, adjusted for age, sex and alcohol consumption. Multiple comparison adjustment was ignored while assessing significance of the interaction term. In order to evaluate if SNPs could provide a predictive signature for adverse phenotype, we re-evaluated data by dividing *HFE*-HH patients in two extreme iron-related phenotype classes. Identification of an extreme adverse phenotype, based on ferritin, iron removed and transferrin saturation values, was performed using a principal component analysis (PCA), based on correlation matrix, fitted on 209 patients with non-missing values in all three hematic parameters. The component retaining the highest proportion of variance (PCA first component) was used as a pseudo-marker and cut into tertiles, using the first and third one to define an extreme binary phenotype. Top ranked SNPs were included in classification procedures using four different algorithms; Support Vector Machine (SVM[22]),

DOI: 10.3324/haematol.2012.062661

Random Forest (RF[22]), Ridge Penalized Logistic Regression (PEN[22]) and K-nearest neighbours (KNN[22]). Classification performances of all algorithms were evaluated using the area under the ROC curve (AUC). The AUC for each procedure was computed using class probability i.e. the estimated probability of being a "case" (i.e. to belong to the highest tertile of the combined hematological parameters estimated via leave-one-out cross validation (LOOCV[22]). Class probability was estimated on the out of the bag sample to reduce the bias of evaluating a classification model on the same data used to build it. In order to estimate the best number of SNPs to use, the cross-validation procedure was repeated for varying number of SNPs. For every run of the LOOCV algorithm, all SNPs are ranked according to the size of the OR estimated by a logistic model accounting for SNPs itself, age, sex and alcohol consumption. Then a classification model was built for increasing number of SNPs, from a model with only one SNP (the most associated) to a model with all SNPs. The optimal number of SNP was then selected as the one with the highest AUC. All methods require some parameters tuning (e.g. the k value in KNN). This was performed with cross-validation using the whole set of SNPs. Tests for the equality of the AUC were performed based on the method by Delong et al[23] while model prediction improvement was evaluated as suggested by Pencina et al[24]. See Supplemental Methods for more information. The analyses were performed using R and the library GenABEL (genome-wide SNP association analysis R package version 1.7)[25].

DOI: 10.3324/haematol.2012.062661

Results

After imposing the quality control measures, 22 subjects (10 *HFE* patients and 12 healthy controls) and 10 SNPs were excluded from the analysis for low call rate and high frequency of SNPs that were identitical-by-state. In addition, sequence analysis of the SNPs rs3806566 and rs884409 in the 5' untranslated region of *CYBRD1*, which were recently found to be significantly associated with serum ferritin in *HFE*-HH patients[16], confirmed that they are in complete linkage with rs3806562, the SNP we previously selected and validated. For this reason, only the latter SNP was retained for statistical analysis. A total of 296 p.C282Y homozygous patients (220 men, 76 women) and 102 healthy controls (86 men, 16 women) for 202 SNPs were finally considered. None of these SNPs deviated from Hardy-Weinberg Equilibrium. Demographic data, alcohol intake, hemoglobin and iron indices of the patients are reported in Table 1. Age and sex were significantly associated with SF, IR and TS ($p<0.001$), while alcohol daily intake was significantly associated with SF and IR ($p<0.001$).

Seventeen SNPs resulted associated with the iron indices among patients (top ranked SNPs, uncorrected $p<0.05$). Their allele frequencies have been compared to the frequencies in CEU population (Table 2) and resulted similar with the exception of the HFE SNP.

For each single outcome, the top ranked associated SNPs with their location within the gene and the worst allele in terms of more severe phenotype have been reported in Table 3. SF was associated with rs12467409, rs9325886 and rs17804636, belonging to candidate genes *BMPR2, BMP9* and *SMAD8,* respectively. IR was associated with rs3178250, rs762642, rs12467409, rs11204215, rs1800702, rs149411, rs11684885 and rs3780474, belonging to candidate genes *BMP2, BMP4, BMPR2, BMP9, HFE, DMT1, HIF2A* and *IRP1,* respectively. TS was associated with rs4401458, rs2292915, rs701753, rs773050, rs701754, rs17554 and rs3806562, belonging to candidate genes *BMPR1B, NEO1, CP* and *CYBRD1,* respectively. The effect of the top ranked SNPs is expressed in terms of fold change induced on the estimated value of the outcome by adding a single worst allele and after adjusted for age, sex and alcohol consumption. After correction for multiple

DOI: 10.3324/haematol.2012.062661

comparison, a significant association adjusted for age, sex and alcohol consumption between a SNP located in *CYBRD1*, rs3806562, and TS was detected (uncorrected p<0.001, FDR=0.07). Observed mean TS levels were 0.93 among TT genotypes, 0.90 among TC genotype and 0.85 among CC genotypes.

In vitro functional assay of rs3806562 did not show significant differences between constructs carrying the C or the T allele (CC: 1.26+0.28, TT: 2.11+1.48 Relative Luciferase Unit, p = NS).

We found two suggestive but not significant interactions between pairs of the 17 top ranked SNPs, one associated with total iron removed (rs149411 and rs762642, uncorrected p-value=0.027) and one associated to TS (rs773050 and rs4401458, uncorrected p=0.02). Figure 1 shows interaction plots of the estimated indices levels as a function of SNP's genotypes.

In order to evaluate if SNPs could provide a predictive signature for adverse phenotype, we re-evaluated data using a principal component analysis (PCA) as described in the Methods section. The PCA first component (hereafter called 'pseudo-marker') accounts for 62.4% of total variance of iron indices and reflects the overall iron load (*loadings*: SF=0.64, TS=0.43, IR=0.64). Therefore, high levels of the three indices result in high values of the pseudo-marker and viceversa. Based on the first and third tertiles of the pseudo-marker we cathegorize patients as controls (mild phenotype) and cases (adverse phenotype). So, the sample was reduced to 140 patients (70 controls and 70 cases). The average value of SF in the two groups was 619 μg/L in controls and 3718 μg/L in cases, with 85.5% of cases and no controls above 2000 μg/L and 91.3% of controls and only 1.5% of cases below 1000 μg/L. Similarly the average levels for IR were 3.73 g and 16.88 g in controls and cases, respectively, while for TS the average percentages were 67.2% in controls and 89.8% in cases. In order to evaluate if the top ranked SNPs could provide a predictive signature for adverse phenotype, we performed a classification procedure based on four different algorithms. For the classification algorithms we had to remove subjects with at least one missing value in any of the 17 selected SNPs. Overall we excluded 2 patients, and therefore performed the classification procedure on 69

DOI: 10.3324/haematol.2012.062661

controls and 69 cases. All models performed reasonably well with an AUC bigger than 65%. The best model was random forest (RF) with an AUC of 76.8% using 14 SNPs (KNN 67.6% with 6 SNPs; PEN 67.0% with 16 SNPs; SVM 71.9% with 15 SNPs). A predictive model based on logistic regression including only age, sex and alcohol consumption, was also fitted. The resulting AUC (70.0%) was lower than the best model with genetic effect, but the difference was not statistically significant (Test for Equality of AUC, p-value = 0.12). We then evaluated the change in predicted probability using the additional information from SNPs and we found that it increases the estimated probability of being an event among cases (sensitivity) of 2.97%, while reducing it among controls (specificity) of 3%. To combine the latter two quantities we computed the Integrated Discrimination Improvement (IDI) (REF) which was close to, but still did not reach statistical significance (IDI=0.061, p-value = 0.06). These results suggest that the 17 selected SNPs provide little additional predictive power for phenotype classification to the known clinical features.

DOI: 10.3324/haematol.2012.062661

Discussion

Homozygosity for the *HFE* p.C282Y mutation is necessary but not sufficient to develop disease phenotype in HFE-related HH. The present study shows that SNPs in several genes involved in iron metabolism may modulate the expression of the disease in p.C282Y homozygous patients. We analysed three different outcomes: transferrin saturation, serum ferritin and total iron removed. Transferrin saturation is not a quantitative index of iron overload, but might represent a qualitative index of alteration of iron homeostasis characterised by increased intestinal iron absorption and iron release from macrophages and storage cells[26]. Accordingly, very high transferrin saturation is usually found in the most severe forms of HH and in patients with ineffective erythropoiesis, both characterised by absent or very low level of hepcidin production and high iron absorption[27]. Serum ferritin is generally considered a good index of iron stores in HH[28] and in our series serum ferritin correlated significantly with the amount of iron removed ($r=0.613$, $p<0.0001$). However, serum ferritin can be influenced by hepatocellular necrosis, inflammation and alcohol intake, which may increase its concentration disproportionately to the amount of iron overload. Phlebotomy, with careful measurement of the amount of iron in the blood removed, is the most accurate means of measuring total body iron stores[29,30]. However, it was not available for all patients studied, this being a limitation when constituting a large cohort of patients.

Our sample is largely representative of the local population of p.C282Y homozygotes since it covered a diverse range of phenotypes, spanning from mild trough moderate to severe. It differed from the sample recruited in the study of Milet *et al,* which, although from a region in North-Western Europe, was not fully representative of the population of p.C282Y homozygotes from which it was drawn since it was, by the Authors admission, rich in individuals with serious symptoms[31]. In the present study we report a significant association between a variant in *CYBRD1*, rs3806562, and transferrin saturation. Moreover we suggested a series of 17 SNPs which could have a possible additive effect on the studied outcomes. The SNP rs3806562 of *CBRYD1* is located in 5'UTR of the gene and therefore likely to be functional. HapMap shows that this SNP is in linkage

DOI: 10.3324/haematol.2012.062661

disequilibrium with rs3806566 and rs884409 previously found to be associated with serum ferritin in Australian p.C282Y homozygotes[16]. So, we analysed SNPs rs3806566 and rs884409 in our cohort and we confirmed that the three SNPs are in complete linkage. However, luciferase assay did not show significance differences in promoter activity between different alleles of rs3806562. This suggests that SNP rs3806562 is a genetic marker, located in the 5'UTR of *CYBRD1*, linked to rs884409, a polymorphism able to modulate *CYBRD1* promoter activity[16]. A high *CYBRD1* activity might lead to increased amount of iron available for the Divalent Metal Transporter 1 (DMT1) at the epithelial intestinal mucosa and, in turn, to increased iron absorption and transferrin saturation. Differently to the Australian study we were not able to find correlation with quantitative indices of iron overload (SF and IR) and we have no clear explanation for this discrepancy. SF and IR have their own intrinsic limits (see above) and it is possible that acquired factors such dietary habits might contribute to modify the whole amount of body iron. However, our results support the hypothesis that *CYBRD1* could be a modifier gene of iron phenotype in *HFE*-HH.

Our results also suggest the involvement of genes of the BMPs in the modulation of iron overload in homozygotes p.C282Y. In particular, we found an association between three SNPs in *BMP9*, *SMAD8* and *BMPR2* and SF. In addition, IR associated with other polymorphisms present in *BMP2*, *BMP4*, *BMP9*, and *BMPR2* (same SNP associated with SF), and TS with another actor of this pathway: *BMPR1B*. Although we included rs235756 in *BMP2* (previously reported as genetic modifier in French patients with *HFE*-HH) in the analysis, we could not confirm the result in our series, and the rs3178250 in *BMP2* that we found associated with IR was not in linkage with rs235756. This could be due to difference in sample size between studies, but also to the inherent heterogeneity related to disease, as reported in other genetic studies of complex traits[32]. All these findings suggest that the expression of disease is not only related to the impairment of HFE function but also depends on the modulation exerted by the functional BMPs on the expression of hepatic hepcidin. BMP9 has been shown to be the most potent inducer of hepcidin *in vitro* and also in mice[33]. *In vitro* studies showed that also SMAD8 and receptors type I and type II (BMPR1A,

DOI: 10.3324/haematol.2012.062661

BMPR1B, and BMPR2) are hepcidin modulators[34], and that *Bmpr1a* is critically responsible for basal hepcidin expression and is required (together with *Acvr1*) for regulation of hepcidin in response to iron and BMP signalling in mice[35].

Besides genes of the BMP/SMAD pathways and *CYBRD1*, other genes involved in iron homeostasis emerged from our study. TS associated with SNPs in genes involved in iron release from storage cells (*Cp*). Previous studies showed the existence of complex interaction between Cp and HFE in transgenic mice, suggesting Cp as a modifier gene with protective effect of *HFE*-HH. IR correlated with SNPs in *DMT1*, *HIF2A*, *IRP1*, and *HFE* itself. Previous studies showed that *DMT1* is over-expressed in *HFE*-HH[36] and that genetic loss of DMT1 modulates iron overload in *Hfe* knockout mice[37]. Although the great majority of our patients carry identical *HFE* SNPs according to the observation that c.845G>A (p.C282Y) mutation is in complete linkage disequilibrium with a unique haplotype[38], we found that one *HFE* SNPs (rs1800702) was associated with variable IR. This finding is quite unexpected because it indirectly suggests that p.C282Y mutation does not completely abolish HFE function. Although this hypothesis needs to be validated, it is to be noted that patients carrying null *HFE* mutations or *Hfe* knockout mice have more severe iron phenotype than their counterparts carrying p.C282Y or p.C282Y hortologue mutation[3]. Among the SNPs emerging from the analysis, some (rs9325886, rs17804636, rs3178250, rs11204215, rs1800702, rs149411, rs3806562) were in regulatory 5'and 3' UTR and one in the coding region (rs701753), suggesting that all currently described genes in these pathways might be candidates as modifier genes in homozygotes p.C282Y and opening the way to functional studies to confirm the effects of these variants on gene expression. We were also able to suggest some interactive effects on TS and IR of couple of different SNPs. This result should be taken with caution because interaction analysis was not corrected for multiple comparisons. However, the interactive effect on IR of SNPs in *BMP4* and *DMT1* which are involved in hepcidin regulation and intestinal iron absorption, respectively, is intriguing because it suggests that two different pathways regulating iron status might cooperate in modulating iron overload in p.C282Y homozygotes.

DOI: 10.3324/haematol.2012.062661

In order to evaluate if the selected SNPs could provide a predictive signature for adverse phenotype, we re-evaluated our data by dividing our *HFE*-HH patients in two extreme iron-related phenotype classes. As extensively described in Methods section and in supplementary methods, we used the Principal component analysis (PCA) to extract relevant information from the set of data including all the three iron indices, serum ferritin, iron removed and transferrin saturation. We compared the predictive performance, as measured by a cross-validated AUC, between various algorithms considering variable number of SNPs with a model with only clinical characteristics and we found that only a small, but still not significant improvement in prediction can be achieved adding genetic information. This might due either to an inappropriate definition of the phenotype or, most likely, to the strong association between the binary phenotype and clinical variables, namely alcohol consumption.

We did not observe associations with SNPs of *TF* and *TMPRSS6* which have emerged as modulators of some indices of iron status in general population in recent GWAS[17]. Transferrin heritability ranged from 0.2 to 0.5 is different isolated populations in Italy[39] and it seems reasonable to hypothesise that genetic factors might influence transferrin saturation in HH by modulating transferrin levels, thus increasing iron deposition and storage[19]. Decreased serum transferrin level is a common observation in HH and it is generally considered secondary to hepatocellular iron overload[40], and our results seem to confirm it. We tested 9 SNPs of *TMPRSS6* including the common SNPs (rs4820268 and rs855791) associated with serum iron and transferrin saturation in general population[17,18]. Recent studies suggested that rs855791 is a *TMPRSS6* functional variant able to modulate hepcidin production[41], and that genetic loss of *Tmprss6* in *Hfe* knock-out mice reduced systemic iron overload by increasing Bmp/Smad signalling in an *Hfe*-independent manner[42]. Although these findings suggest that natural genetic variation in the human ortholog *TMPRSS6* might modify the clinical penetrance of *HFE*-associated Hereditary Hemochromatosis, our results indicate that the SNPs studied had not enough power to modify iron phenotype in our series.

DOI: 10.3324/haematol.2012.062661

In conclusion, the present study suggests that SNPs in genes regulating iron metabolism may modulate penetrance of *HFE*-HH. These results support the role of BMPs as possible modifiers of *HFE*-HH phenotype and further expand these observations on a larger number of genes involved in iron absorption and release, with emphasis on the 5' UTR region in *CYBRD1*. Our results also strengthen the notion that none of these polymorphisms alone is a major modifier of HH phenotype, suggesting that iron phenotype in HH is the result of a complex interaction between a major gene defect, genetic background and environmental factors (alcohol intake in particular) supporting the idea that *HFE*-HH is a multifactorial disease. From a practical point of view the identification of all these factors including one or more risk SNPs might add information on patients' susceptibility to fully penetrant *HFE*-HH and would help in modulating the clinical approach better defining follow-up and therapeutic approaches.

DOI: 10.3324/haematol.2012.062661

Authorship and Disclosures

SP: design of the study, genotyping, analysis and interpretation of data, manuscript writing;
RM: patients enrolment, collection and assembly of phenotypic data, analysis and interpretation, manuscript writing; SC: statistical data analysis and interpretation; ALF: patients enrolment and collection of phenotypic data; GLM: genotyping and manuscript writing; FB: genotyping; FB: patients enrolment and assembly of phenotypic data; PT: patients enrolment, collection and assembly of phenotypic data; MF: patients enrolment assembly of phenotypic data; GLF: provision of study patients assembly of phenotypic data; DG: patients enrolment assembly of phenotypic data; SF: patients enrolment assembly of phenotypic data; CS: statistical data analysis, interpretation and manuscript writing; AP: conception and design, data analysis and interpretation, manuscript writing, final approval of manuscript.

DOI: 10.3324/haematol.2012.062661

References

1. Pietrangelo A. Hereditary hemochromatosis: pathogenesis, diagnosis, and treatment. Gastroenterology. 2010;139(2):393-408, e1-2.
2. Deugnier Y, Brissot P, Loreal O. Iron and the liver: update 2008. J Hepatol. 2008;48 Suppl 1:S113-23.
3. Fleming RE, Holden CC, Tomatsu S, Waheed A, Brunt EM, Britton RS, et al. Mouse strain differences determine severity of iron accumulation in Hfe knockout model of hereditary hemochromatosis. Proc Natl Acad Sci U S A. 2001;98(5):2707-11.
4. Bensaid M, Fruchon S, Mazeres C, Bahram S, Roth MP, Coppin H. Multigenic control of hepatic iron loading in a murine model of hemochromatosis. Gastroenterology. 2004;126(5):1400-8.
5. Crawford DH, Halliday JW, Summers KM, Bourke MJ, Powell LW. Concordance of iron storage in siblings with genetic hemochromatosis: evidence for a predominantly genetic effect on iron storage. Hepatology. 1993;17(5):833-7.
6. Whiting PW, Fletcher LM, Dixon JK, Gochee P, Powell LW, Crawford DH. Concordance of iron indices in homozygote and heterozygote sibling pairs in hemochromatosis families: implications for family screening. J Hepatol. 2002;37(3):309-14.
7. Distante S, Elmberg M, Foss Haug KB, Ovstebo R, Berg JP, Kierulf P, et al. Tumour necrosis factor alpha and its promoter polymorphisms' role in the phenotypic expression of hemochromatosis. Scand J Gastroenterol. 2003;38(8):871-7.
8. Mura C, Raguenes O, Ferec C. HFE mutations analysis in 711 hemochromatosis probands: evidence for S65C implication in mild form of hemochromatosis. Blood. 1999;93(8):2502-5.
9. van Dijk BA, Kemna EH, Tjalsma H, Klaver SM, Wiegerinck ET, Goossens JP, et al. Effect of the new HJV-L165X mutation on penetrance of HFE. Blood. 2007;109(12):5525-6.
10. Lee PL, Gelbart T, West C, Halloran C, Felitti V, Beutler E. A study of genes that may modulate the expression of hereditary hemochromatosis: transferrin receptor-1, ferroportin, ceruloplasmin, ferritin light and heavy chains, iron regulatory proteins (IRP)-1 and -2, and hepcidin. Blood Cells Mol Dis. 2001;27(5):783-802.
11. Jacolot S, Le Gac G, Scotet V, Quere I, Mura C, Ferec C. HAMP as a modifier gene that increases the phenotypic expression of the HFE pC282Y homozygous genotype. Blood. 2004;103(7):2835-40.
12. Krayenbuehl PA, Maly FE, Hersberger M, Wiesli P, Himmelmann A, Eid K, et al. Tumor necrosis factor-alpha -308G>A allelic variant modulates iron accumulation in patients with hereditary hemochromatosis. Clin Chem. 2006;52(8):1552-8.
13. Tolosano E, Fagoonee S, Garuti C, Valli L, Andrews NC, Altruda F, et al. Haptoglobin modifies the hemochromatosis phenotype in mice. Blood. 2005;105(8):3353-5.
14. Gouya L, Muzeau F, Robreau AM, Letteron P, Couchi E, Lyoumi S, et al. Genetic study of variation in normal mouse iron homeostasis reveals ceruloplasmin as an HFE-hemochromatosis modifier gene. Gastroenterology. 2007;132(2):679-86.
15. Milet J, Le Gac G, Scotet V, Gourlaouen I, Theze C, Mosser J, et al. A common SNP near BMP2 is associated with severity of the iron burden in HFE p.C282Y homozygous patients: a follow-up study. Blood Cells Mol Dis. 2010;44(1):34-7.
16. Constantine CC, Anderson GJ, Vulpe CD, McLaren CE, Bahlo M, Yeap HL, et al. A novel association between a SNP in CYBRD1 and serum ferritin levels in a cohort study of HFE hereditary haemochromatosis. Br J Haematol. 2009;147(1):140-9.
17. Tanaka T, Roy CN, Yao W, Matteini A, Semba RD, Arking D, et al. A genome-wide association analysis of serum iron concentrations. Blood. 2010;115(1):94-6.

DOI: 10.3324/haematol.2012.062661

18. Benyamin B, Ferreira MA, Willemsen G, Gordon S, Middelberg RP, McEvoy BP, et al. Common variants in TMPRSS6 are associated with iron status and erythrocyte volume. Nat Genet. 2009;41(11):1173-5.
19. Benyamin B, McRae AF, Zhu G, Gordon S, Henders AK, Palotie A, et al. Variants in TF and HFE explain approximately 40% of genetic variation in serum-transferrin levels. Am J Hum Genet. 2009;84(1):60-5.
20. Piperno A. Classification and diagnosis of iron overload. Haematologica. 1998;83(5):447-55.
21. 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(11):1217-23.
22. Hastie T TR, Friedman J. The Elements of Statistical Learning, 2nd ed. In: Springer, ed, 2009.
23. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44(3):837-45.
24. Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. Stat Med. 2008;27(2):157-72; discussion 207-12.
25. Team RDC. R: A language and environment for statistical computing., 2012.
26. Mariani R, Pelucchi S, Arosio C, Coletti S, Pozzi M, Paolini V, et al. Genetic and metabolic factors are associated with increased hepatic iron stores in a selected population of p.Cys282Tyr heterozygotes. Blood Cells Mol Dis. 2010;44(3):159-63.
27. Pietrangelo A, Caleffi A, Corradini E. Non-HFE hepatic iron overload. Semin Liver Dis. 2011;31(3):302-18.
28. Camaschella C, Poggiali E. Towards explaining "unexplained hyperferritinemia". Haematologica. 2009;94(3):307-9.
29. Worwood M. Laboratory determination of iron status. In: Brock JH HJ, Pippard MJ, Powell LW, ed. Iron metabolism in health and disease: W.B. Saunders, 1994:449-76.
30. Piperno A, Arosio C, Fargion S, Roetto A, Nicoli C, Girelli D, et al. The ancestral hemochromatosis haplotype is associated with a severe phenotype expression in Italian patients. Hepatology. 1996;24(1):43-6.
31. Milet J, Dehais V, Bourgain C, Jouanolle AM, Mosser A, Perrin M, et al. Common variants in the BMP2, BMP4, and HJV genes of the hepcidin regulation pathway modulate HFE hemochromatosis penetrance. Am J Hum Genet. 2007;81(4):799-807.
32. Chanock SJ, Manolio T, Boehnke M, Boerwinkle E, Hunter DJ, Thomas G, et al. Replicating genotype-phenotype associations. Nature. 2007;447(7145):655-60.
33. Truksa J, Peng H, Lee P, Beutler E. Bone morphogenetic proteins 2, 4, and 9 stimulate murine hepcidin 1 expression independently of Hfe, transferrin receptor 2 (Tfr2), and IL-6. Proc Natl Acad Sci U S A. 2006;103(27):10289-93.
34. Babitt JL, Huang FW, Xia Y, Sidis Y, Andrews NC, Lin HY. Modulation of bone morphogenetic protein signaling in vivo regulates systemic iron balance. J Clin Invest. 2007;117(7):1933-9.
35. Steinbicker AU, Bartnikas TB, Lohmeyer LK, Leyton P, Mayeur C, Kao SM, et al. Perturbation of hepcidin expression by BMP type I receptor deletion induces iron overload in mice. Blood. 2011;118(15):4224-30.
36. Griffiths WJ, Sly WS, Cox TM. Intestinal iron uptake determined by divalent metal transporter is enhanced in HFE-deficient mice with hemochromatosis. Gastroenterology. 2001;120(6):1420-9.
37. Levy JE, Montross LK, Andrews NC. Genes that modify the hemochromatosis phenotype in mice. J Clin Invest. 2000;105(9):1209-16.

DOI: 10.3324/haematol.2012.062661

38. Yang Y, Ferec C, Mura C. SNP and haplotype analysis reveals new HFE variants associated with iron overload trait. Hum Mutat. 2011;32(4):E2104-17.
39. Traglia M, Sala C, Masciullo C, Cverhova V, Lori F, Pistis G, et al. Heritability and demographic analyses in the large isolated population of Val Borbera suggest advantages in mapping complex traits genes. PLoS One. 2009;4(10):e7554.
40. Morton AG, Tavill AS. The role of iron in the regulation of hepatic transferrin synthesis. Br J Haematol. 1977;36(3):383-94.
41. Nai A, Pagani A, Silvestri L, Campostrini N, Corbella M, Girelli D, et al. TMPRSS6 rs855791 modulates hepcidin transcription in vitro and serum hepcidin levels in normal individuals. Blood. 2011;118(16):4459-62.
42. Finberg KE, Whittlesey RL, Andrews NC. Tmprss6 is a genetic modifier of the Hfe-hemochromatosis phenotype in mice. Blood. 2011;117(17):4590-9.

DOI: 10.3324/haematol.2012.062661

Table 1. Median and (range) of age, alcohol intake, hemoglobin and serum iron indices in p.C282Y homozygotes patients by sex.

	Missing (%)	All patients (N=296)	Males (N=220)	Females (N=76)
Age (yrs)	0	45.5 (11-77)	43 (11-76)	56 (21-77)
Alcohol intake (g/day)	0	10 (0-250)	10 (0-250)	5 (0-60)
Hemoglobin (g/dL)	22.6	14.8 (9.0-18.7)	15.0 (9.0-18.7)	13 (9.4-16.9)
Transferrin saturation (%)	2.4	85 (30-100)	86 (41-100)	80 (30-100)
Serum ferritin (μg/L)	0	1060 (31-13136)	1209 (32-13136)	552.5 (31-5089)
Iron removed (g)	28.7	7 (0.5-41)	8 (1.6-41)	4.4 (0.5-25)

DOI: 10.3324/haematol.2012.062661

Table 2. Allelic frequencies in p.C282Y homozygotes patients and in the CEU population of the 17 SNPs associated with the outcomes (top ranked SNPs, uncorrected p<0.05).

SNPs	CEU allelic frequencies	Allelic frequencies in p.C282Y homozygous patients
rs17554 (G/A)	G: 0.55; A: 0.45	G: 0.67; A: 0.33
rs149411 (C/T)	C:0.58; T: 0.42	C:0.63; T: 0.37
rs701753 (A/T)	A: 0.92 ; T: 0.08	A: 0.94 ; T: 0.06
rs701754 (A/T)	A: 0.85 ; T: 0.15	A: 0.84 ; T: 0.16
rs762642 (T/G)	T: 0.61; G: 0.39	T: 0.56; G: 0.44
rs1800702 (C/G)*	C: 0.61; G:0.39	C: 0.01; G:0.99
rs2292915 (C/T)	C: 0.66; T: 0.34	C: 0.65; T: 0.35
rs3178250 (T/C)	T: 0.81; C: 0.19	T: 0.72; C: 0.28
rs3780474 (A/C)	A: 0.64; C:0.36	A: 0.64; C:0.36
rs3806562 (T/C)	T: 0.85; C: 0.15	T: 0.84; C: 0.16
rs4401458 (T/C)	T: 0.52; C: 0.48	T: 0.46; C: 0.54
rs9325886 (C/A)	C: 0.93; A: 0.07	C: 0.94; A: 0.06
rs11204215 (A/C)	A: 0.87; C: 0.13	A: 0.83; C: 0.17
rs11684885 (G/T)	G: 0.66; T: 0.34	G: 0.62; T: 0.38
rs12467409 (G/T)	G: 0.86; T: 0.14	G: 0.87; T: 0.13
rs17804636 (A/G)	A: 0.92; G: 0.08	A: 0.93; G: 0.07
rs773050 (A/G)	A: 0.94; G: 0.06	A: 0.94; G: 0.06

*HFE SNP

DOI: 10.3324/haematol.2012.062661

Table 3. SNPs associated with serum ferritin, iron removed and transferrin saturation variability.

Outcome	SNPs	Gene symbol	Location	Worst allele	N	Fold change* (95% CI)	p-value°
Serum Ferritin	rs12467409 (G/T)	*BMPR2*	Intron	T	293	1.26 (1.02-1.58)	0.036
	rs9325886 (C/A)	*BMP9*	3'UTR	C	296	1.39 (1.04-1.86)	0.026
	rs17804636 (A/G)	*SMAD8*	3'UTR	G	296	1.43 (1.06-1.92)	0.019
Iron Removed	rs3178250 (T/C)	*BMP2*	3'UTR	T	211	1.19 (1.02 -1.37)	0.023
	rs762642 (T/G)	*BMP4*	Intron	G	210	1.16 (1.02 -1.33)	0.030
	rs12467409 (G/T)	*BMPR2*	Intron	T	208	1.30 (1.03 -1.64)	0.026
	rs11204215 (A/C)	*BMP9*	5'UTR	A	211	1.22 (1.02 -1.46)	0.028
	rs1800702 (C/G)	*HFE*	5'UTR	C	211	1.56 (1.01-2.38)	0.043
	rs149411 (C/T)	*DMT1*	3'UTR	T	211	1.18 (1.02 -1.35)	0.025
	rs11684885 (G/T)	*HIF2A*	Intron	A	211	1.16 (1.01 - 1.33)	0.041
	rs3780474 (A/C)	*IRP1*	Intron	C	211	1.21 (1.06-1.38)	0.005
Transferrin saturation	rs4401458 (T/C)	*BMPR1B*	Intron	T	289	1.23 (1.02 - 1.49)	0.031
	rs2292915 (C/T)	*NEO1*	Intron	C	287	1.21 (1.00-1.47)	0.046
	rs701753 (A/T)	*CP*	Coding	A	288	1.56 (1.06-2.30)	0.023
	rs773050 (A/G)	*CP*	Intron	G	289	1.45 (1.002-2.09)	0.048
	rs701754 (A/T)	*CP*	Intron	A	288	1.32 (1.04-1.67)	0.022
	rs17554 (G/A)	*CYBRD1*	Intron	G	289	1.23 (1.02-1.48)	0.034
	rs3806562 (T/C)	*CYBRD1*	5'UTR	T	288	1.54 (1.21-1.96)	<0.001

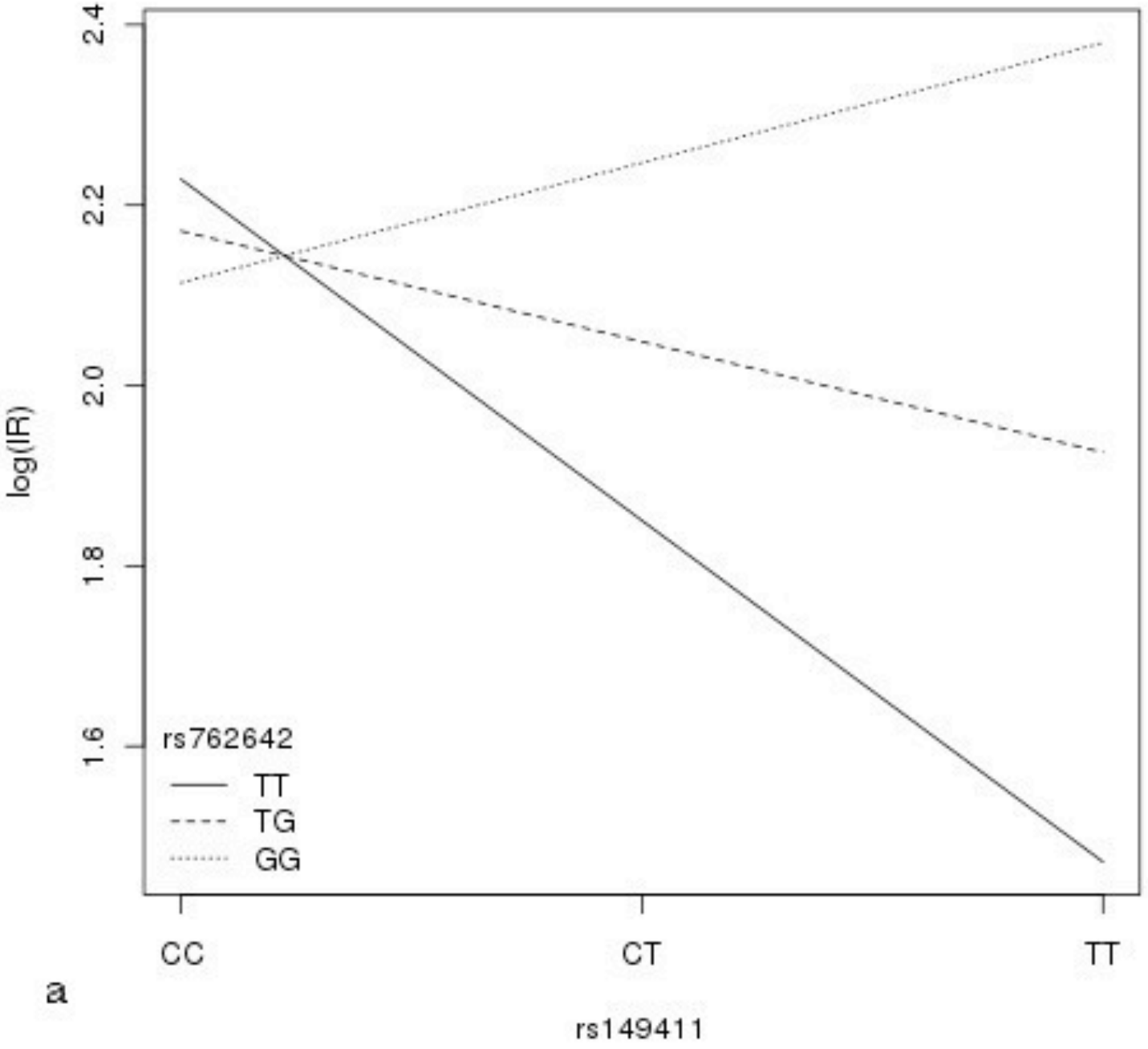
*Adjusted for age, gender and alcohol intake in a multiple linear regression model; ° p-value uncorrected for multiple testing

DOI: 10.3324/haematol.2012.062661

LEGEND

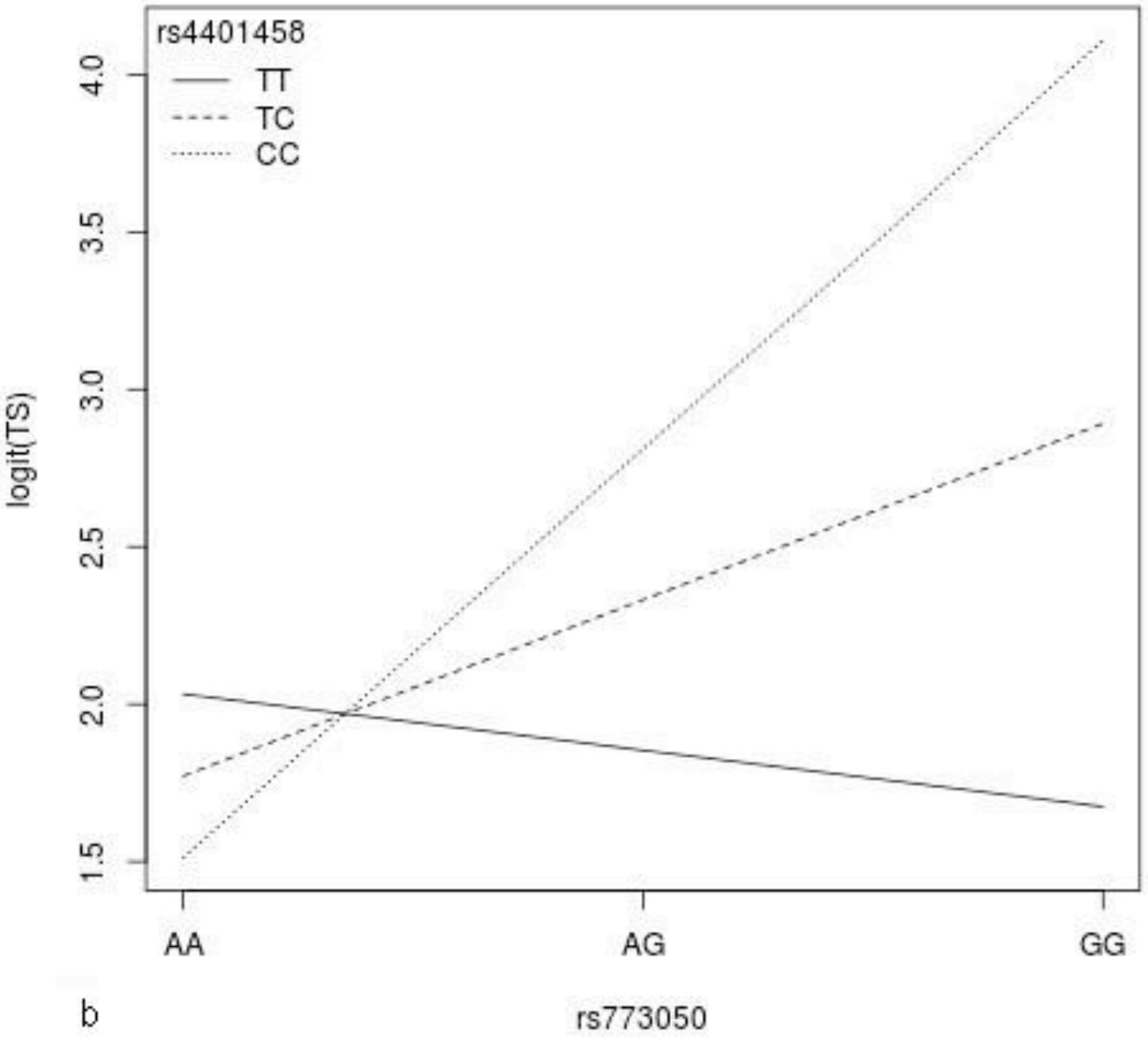
Figure 1. Boxplots of the estimated total iron removed (IR) levels (log scale) according to rs149411 (*DMT1*) and rs762642 (*BMP4*) genotypes (a) and of the estimated transferrin saturation (TS) levels (logit scale) according to rs773050 (*CP*) and rs4401458 (*BMPR1B*) genotypes (b).

DOI: 10.3324/haematol.2012.062661



a

DOI: 10.3324/haematol.2012.062661



b

DOI: 10.3324/haematol.2012.062661

Supplementary Methods

SNP selection

Illumina assigns designability ranks and SNP scores. They are items strictly linked since they both give the same information about the ability to design a successful assay. Illumina SNP scores range from 0 to 1.1 and designability rank is represented by scores 0, 0.5 or 1. SNP_Score < 0.4 corresponding to designability rank of 0, gives a low success rate with consequent high risk to perform OPA (Oligo Pull Assay). SNP_Score from 0.4 to <0.6 corresponds to designability rank of 0.5 and gives a moderate success rate with consequent moderate risk to perform OPA. SNP_Score from 0.6 to 1.1 = designability rank of 1 gives a high success rate with consequent low risk to perform OPA. All the SNPs we analyzed had Illumina SNP scores ≥1, designability rank ≥1, no failure code, validation class=3 and GoldenGate® validation. Validation class data and validation bin are additional items to consider; they are numeric and textual representation of genotyping reaction feasibility. GoldenGate® validation bin SNPs have a validation class of 3 and SNP_scores of 1.1. Two-hit validated SNPs have a validation class of 2 (it means that both alleles of the SNP have been read by two different methods and in two independent populations) and SNP_scores from 0 to 1. Not validated SNPs have a validation class of 1 and SNP_scores of 0. We analysed 133 GoldenGate® and 78 two-hit validated class SNPs, with 3 and 2 validation bin, respectively. SNP scores were 1.1 for GoldenGate® ones and between 0.668 and 0.997 for Two-hit validated SNPs (mean 0.886).

A set of non-redundant tag SNPs was identified for each region, so that all the SNPs with a minor-allele frequency (MAF) ≥0.05 in the database, have a pairwise r^2≥0.80 (www.hapmap.org - hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap3r2_B36/#search). Tagging was performed using the algorithm implemented in Tagger[21]. The linkage disequilibrium (LD) blocks were determined using data from HapMap data release #28, on National Center for Biotechnology (NCBI) B36 assembly, dbSNP b126.

Among all the TagSNPs selected from Haploview we retained only SNPs belonging to coding, intronic and 5′ and 3′ untranslated regions with similar proportions in order to cover the full gene.

DOI: 10.3324/haematol.2012.062661

SNP genotyping

For laboratory quality assurance, we qualified SNPs that had Illumina GenCall_10 scores ≥0.4 and call rates ≥88%. So, we excluded SNPs with GenCall_10 scores below 0.40 and/or call rates below 88%. VeraCode Raw data generated from the genotyping were analyzed by the GenomeStudio software to define, for each SNP, the called genotypes into the three different cluster area. Data were then processed in order to infer all SNP genotypes *via a* genotyping cluster. All genotypes were manually checked and re-scored if any errors in calling homozygous or heterozygous clusters were evident. Samples falling out of these cluster area corresponding to the different genotypes were failed. Four duplicate samples were genotyped for all assays for quality control with 100% reproducibility. All SNPs showed high-genotyping quality: the genotyping call rate for the studied SNPs was in the range of 99-100%.

Statistical Methods

Principal component analysis

Principal component analysis (PCA) is one of the most used and valuable application of linear algebra, being a simple, non-parametric method of dimensionality reduction, that is of extracting relevant information from a complex set of data.

The procedure allows to extract from a number *p* of correlated variables as much as *p* new factors derived as linear combination of the original variables. The biggest advantage is in that one or few of the PCA factors accounts for a great proportion of the total variance (hence information) of the *p* variables. Let us consider only two variables with approximately the same variance and reasonably high correlation. Then let us plot them in a scatter-plot. The PCA would draw two new *orthogonal* axis, one passing along the direction with higher variance in the cloud of point and the second being perpendicular to the first one. These axis are the new coordinates for the derived PCA factors, and can be simply computed as a linear combination of the original variables. In the ideal setting of two variables with very high correlation, the new first component (axis) would account for the big majority of information (variability) contained in the data, and could be used as a surrogated of the two variables combined. The same idea can be applied to any number *p* of variables.

The coefficients used to compute the different linear combinations are traditionally called *loadings*. The proportion of variance explained or accounted for by the *p* new axis is an indication of the information loss by the dimensionality reduction.

DOI: 10.3324/haematol.2012.062661

Classification

A classifier is defined as a model describing the specific classification algorithm, like Support Vector Machine (SVM), K-nearest neighbours (KNN) etc. Every classifier was built iteratively on all samples but a single one (the so called "*out of the bag*" sample). Specifically, for every classification algorithm we fitted 138 different models (where 138 is the overall number of patients used), each one built on 137 samples excluding each sample in turn. In every single iteration we used the model fitted on 137 samples to predict the status (case/control, i.e. high/low tertile of SF, IR and TS combined values) of the excluded (*out of the bag*) sample based on the covariate in the model (sex, age, alcohol and a given number of SNPs). In iteration one we would then exclude sample one and build the classifier on the remaining 137 samples; such classifier is then used to predict the status for sample one. The same for sample two and so on.

DOI: 10.3324/haematol.2012.062661

Online Supplementary Table 1. 214 SNPs analyzed in 50 genes.

SNPs	Chromosome	Gene name	Location	Function
rs4693924	4	ABCG2	Intron	Superfamily of ATP-binding cassette (ABC) transporters
rs2054576	4	ABCG2	Intron	
rs2725256	4	ABCG2	Intron	
rs4148155	4	ABCG2	Intron	
rs2622624	4	ABCG2	Intron	
rs1901531	15	B2M	Intron	HFE pathway
rs1801621	11	BEST1	3'UTR	Carrier calcium-activated chloride-ion channels
rs6077060	20	BMP2	5' UTR	BMPs pathway
rs235768	20	BMP2	Coding	
rs3178250	20	BMP2	3'UTR	
rs6054512	20	BMP2	3'UTR	
rs173107	20	BMP2	3'UTR	
rs235756	20	BMP2	3'UTR	
rs910141	20	BMP2	3'UTR	
rs17563	14	BMP4	Coding	BMPs pathway
rs762642	14	BMP4	Intron	
rs4901474	14	BMP4	3'UTR	
rs3812163	6	BMP6	5UTR	BMPs pathway
rs6910759	6	BMP6	Intron	
rs267201	6	BMP6	Intron	
rs1225934	6	BMP6	Intron	
rs1044104	6	BMP6	3'UTR	
rs9325886	10	BMP9	3'UTR	BMPs pathway
rs9971293	10	BMP9	Intron	
rs11204215	10	BMP9	5'UTR	
rs3905377	10	BMPR1A	5'UTR	BMPs pathway
rs2354353	10	BMPR1A	Intron	
rs2883420	10	BMPR1A	Intron	
rs10887666	10	BMPR1A	Intron	
rs7091555	10	BMPR1A	Intron	
rs7074064	10	BMPR1A	Intron	
rs4401458	4	BMPR1B	Intron	BMPs pathway
rs7661049	4	BMPR1B	Intron	
rs6815044	4	BMPR1B	Intron	
rs9997720	4	BMPR1B	Intron	
rs3821964	4	BMPR1B	Intron	
rs3796443	4	BMPR1B	Intron	
rs11097457	4	BMPR1B	3'UTR	
rs13010656	2	BMPR2	Intron	BMPs pathway
rs6751210	2	BMPR2	Intron	
rs7575056	2	BMPR2	Intron	
rs12467409	2	BMPR2	Intron	
rs1048829	2	BMPR2	3'UTR	
rs1053709	3	CP	Coding	Iron reductase
rs701754	3	CP	Intron	

DOI: 10.3324/haematol.2012.062661

rs773050	3	CP	Intron	
rs701753	3	CP	Coding	
rs701748	3	CP	5'UTR	
rs3806562	2	CYBRD1	5'UTR	Iron reductase
rs3806566	2	CYBRD1	5'UTR	
rs884409	2	CYBRD1	5'UTR	
rs960748	2	CYBRD1	Intron	
rs17554	2	CYBRD1	Intron	
rs10455	2	CYBRD1	Coding	
rs2542938	2	CYBRD1	3'UTR	
rs1435166	1	EGLN1	Intron	Hypoxia pathway
rs2486742	1	EGLN1	Intron	
rs1538664	1	EGLN1	Intron	
rs7544596	1	EGLN1	5'UTR	
rs3736329	19	EGLN2	Intron	Hypoxia pathway
rs10405596	19	EGLN2	3'UTR	
rs1680710	14	EGLN3	3'UTR	Hypoxia pathway
rs1680694	14	EGLN3	Intron	
rs1047881	1	FLVCR1	5'UTR	Heme pathway
rs12756625	1	FLVCR1	Intron	
rs12125982	1	FLVCR1	Intron	
rs10779594	1	FLVCR1	Intron	
rs1390501	1	FLVCR1	Intron	
rs3207090	1	FLVCR1	Coding	
rs4932178	15	FURIN	5'UTR	Hepcidin cleavage
rs6227	15	FURIN	3'UTR	
rs12459782	19	GDF15	5'UTR	Growth factor
rs1059519	19	GDF15	Coding	
rs1227731	19	GDF15	Intron	
rs16982345	19	GDF15	3'UTR	
rs8101249	19	GDF15	3'UTR	
rs10405246	19	USF2	Intron	Transcription factor
rs1882694	19	USF2	3'UTR	
rs8101606	19	HAMP	Intron	Hepcidin pathway
rs7251432	19	HAMP	Intron	
rs12971321	19	HAMP	3'UTR	
rs1264218	X	HEPH	Intron	Iron oxidase
rs5919024	X	HEPH	Intron	
rs1800702	6	HFE	5'UTR	Hepcidin regulator
rs2794719	6	HFE	Intron	
rs9366637	6	HFE	Intron	
rs2858996	6	HFE	Intron	
rs707889	6	HFE	Intron	
rs16827043	1	HFE2	5'UTR	Hepcidin regulator
rs7536827	1	HFE2	5'UTR	
rs1535921	1	HFE2	3'UTR	
rs2301106	14	HIF1A	Intron	Hypoxia pathway
rs12434438	14	HIF1A	Intron	
rs10873142	14	HIF1A	Intron	
rs2301113	14	HIF1A	Intron	
rs2057482	14	HIF1A	3'UTR	

DOI: 10.3324/haematol.2012.062661

rs11684885	2	HIF2A	Intron	Hypoxia pathway
rs11689011	2	HIF2A	Intron	
rs6756667	2	HIF2A	Intron	
rs1374748	2	HIF2A	Intron	
rs7571218	2	HIF2A	Intron	
rs13424253	2	HIF2A	3'UTR	
rs9924964	16	HP	5'UTR	Heme pathway
rs2856836	2	IL1A	3'UTR	Inflammation
rs3783546	2	IL1A	Intron	
rs1800587	2	IL1A	5'UTR	
rs1878319	2	IL1A	5'UTR	
rs2069832	7	IL6	Intron	Inflammation
rs2069849	7	IL6	Coding	
rs4601580	1	IL6R	Intron	Inflammation
rs7518199	1	IL6R	Intron	
rs4553185	1	IL6R	Intron	
rs4845625	1	IL6R	Intron	
rs4129267	1	IL6R	Intron	
rs11265618	1	IL6R	Intron	
rs4072391	1	IL6R	3'UTR	
rs4297112	9	IRP1	Intron	Cell iron regulation
rs7874815	9	IRP1	Intron	
rs10970971	9	IRP1	Intron	
rs10813813	9	IRP1	Intron	
rs3780474	9	IRP1	Intron	
rs4878497	9	IRP1	Intron	
rs10813816	9	IRP1	Intron	
rs10970978	9	IRP1	Intron	
rs7042042	9	IRP1	3'UTR	
rs17483548	15	IRP2	5'UTR	Cell iron regulation
rs12916396	15	IRP2	Intron	
rs2938674	15	IRP2	Intron	
rs13180	15	IRP2	Coding	
rs2292116	15	IRP2	Intron	
rs16969906	15	IRP2	3'UTR	
rs3814526	9	LCN2	5'UTR	Iron Carrier
rs721183	8	MFRN1	5'UTR	Mitochondrial iron transporter
rs4872154	8	MFRN1	Intron	
rs1047384	8	MFRN1	Coding	
rs922516	15	NEO1	Intron	HJV pathway
rs1979409	15	NEO1	Intron	
rs3736510	15	NEO1	Coding	
rs2292915	15	NEO1	Intron	
rs1878940	15	NEO1	3'UTR	
rs739439	17	SARM1	3'UTR	Inflammation
rs149411	12	DMT1	3'UTR	Iron absorption
rs161047	12	DMT1	Intron	
rs445520	12	DMT1	Intron	
rs364627	12	DMT1	Intron	
rs870843	3	SLC25A38	Intron	Mitochondrial carrier

DOI: 10.3324/haematol.2012.062661

rs6890	3	SLC25A38	3'UTR	
rs2352262	2	SLC40A1	3'UTR	Iron exporter
rs2304704	2	SLC40A1	Coding	
rs4145237	2	SLC40A1	Intron	
rs1439812	2	SLC40A1	Intron	
rs3811621	2	SLC40A1	5'UTR	
rs6537355	4	SMAD1	5'UTR	Signal transduction
rs2289737	4	SMAD1	Intron	
rs714195	4	SMAD1	Intron	
rs11724813	4	SMAD1	Intron	
rs12457540	18	SMAD4	Intron	Signal transduction
rs2276163	18	SMAD4	Intron	
rs8084630	18	SMAD4	Intron	
rs8096092	18	SMAD4	Intron	
rs948588	18	SMAD4	Intron	
rs9304407	18	SMAD4	Intron	
rs6596289	5	SMAD5	Intron	Signal transduction
rs13179769	5	SMAD5	Intron	
rs10068371	5	SMAD5	Intron	
rs10515478	5	SMAD5	Intron	
rs7031	5	SMAD5	3'UTR	
rs17804636	13	SMAD8	3'UTR	Signal transduction
rs7993661	13	SMAD8	Intron	
rs9547689	13	SMAD8	Intron	
rs9576129	13	SMAD8	5'UTR	
rs1053005	17	STAT3	3'UTR	Inflammation
rs3744483	17	STAT3	3'UTR	
rs8074524	17	STAT3	Intron	
rs6503695	17	STAT3	Intron	
rs1026916	17	STAT3	Intron	
rs17405722	17	STAT3	5'UTR	
rs838082	2	STEAP3	Intron	Iron reductase
rs1867749	2	STEAP3	Intron	
rs3731603	2	STEAP3	3'UTR	
rs1530561	2	STEAP3	3'UTR	
rs8177178	3	TF	5'UTR	Iron transport
rs8177213	3	TF	Intron	
rs8177240	3	TF	Intron	
rs3811647	3	TF	Intron	
rs1525889	3	TF	Intron	
rs10247962	7	TFR2	Intron	Hepcidin regulator
rs2075674	7	TFR2	Coding	
rs7457868	7	TFR2	Intron	
rs4727457	7	TFR2	Intron	
rs4434553	7	TFR2	3'UTR	
rs1052897	7	TFR2	3'UTR	
rs6772320	3	TFRC	3'UTR	Iron uptake
rs3326	3	TFRC	Intron	
rs3827556	3	TFRC	Intron	
rs3817672	3	TFRC	Coding	
rs4820268	22	TMPRSS6	Coding	HJV pathway

DOI: 10.3324/haematol.2012.062661

rs2543519	22	TMPRSS6	Intron	
rs2179229	22	TMPRSS6	Intron	
rs2235323	22	TMPRSS6	Intron	
rs2235324	22	TMPRSS6	Coding	
rs2743824	22	TMPRSS6	Intron	
rs732755	22	TMPRSS6	Intron	
rs855791	22	TMPRSS6	Coding	
rs228910	22	TMPRSS6	5'UTR	
rs779805	3	VHL	5'UTR	
rs1642742	3	VHL	3'UTR	Hypoxia pathway
rs17610448	3	VHL	3'UTR	
rs12544577	8	ZIP14	5'UTR	
rs11136017	8	ZIP14	Intron	
rs2280521	8	ZIP14	Intron	Metal ion
rs12545575	8	ZIP14	Intron	transporter
rs10101909	8	ZIP14	Intron	
rs12679702	8	ZIP14	3'UTR	