



HAL
open science

SNP and haplotype analysis reveals new HFE variants associated with iron overload trait

Yizhen Yang, Claude Ferec, Catherine Mura

► To cite this version:

Yizhen Yang, Claude Ferec, Catherine Mura. SNP and haplotype analysis reveals new HFE variants associated with iron overload trait. *Human Mutation*, 2011, 32 (4), 10.1002/humu.21461. hal-00613915

HAL Id: hal-00613915

<https://hal.science/hal-00613915>

Submitted on 8 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SNP and haplotype analysis reveals new HFE variants associated with iron overload trait

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0420.R2
Wiley - Manuscript type:	Mutation in Brief
Date Submitted by the Author:	22-Dec-2010
Complete List of Authors:	Yang, Yizhen; INSERM U613, Génétique moléculaire Ferec, Claude; INSERM U613, Génétique moléculaire Mura, Catherine; INSERM U613, Genetique
Key Words:	HFE, SNP, haplotype, iron overload, association study

SCHOLARONE™
Manuscripts

SNP and haplotype analysis reveals new *HFE* variants associated with iron overload trait



Yizhen Yang^{1,2}, Claude Férec^{1,2}, and Catherine Mura^{1,2}

¹INSERM U613 Génétique moléculaire et génétique épidémiologique, ²Université de Bretagne Occidentale, Faculté de Médecine et des Sciences de la Santé de Brest, 46 rue Félix Le Dantec, F-29275 Brest, France

*Correspondence to Catherine Mura, IEM CNRS 6218, 3 B rue de la ferrollerie, 45071 Orléans, France. Tel: +33 2 38257975; Fax: +33 2 38257979. E-mail: catherine.mura@cns-orleans.fr.

Short Title: *HFE* variants and iron overload trait.

Communicated by <Please don't enter>

ABSTRACT: Hereditary hemochromatosis is a common-recessive-autosomal disease characterized by progressive iron overload, and its prevalence correlates with c.845G>A (p. C282Y) mutation of the *HFE* gene. Two other variants c.187C>G and c.193A>T are associated with a mild iron overload phenotype. The correlation studies have revealed incompletely penetrance of the *HFE* mutations, as well as the lack of mutation on some chromosomes from patients. We screened for SNPs before examining allele and haplotype association with elevated iron parameters. We confirmed that the c.845G>A mutation is in complete linkage disequilibrium with a unique haplotype, whereas two haplotypes proved to account for 79.8 and 20.2% of the c.187G chromosomes whose only difference was the g.4694C>G variation. A greater prevalence of the g.4694G allele among patients' chromosomes, compared to controls, was observed. In addition, among non-mutant chromosomes the analyses revealed a risk haplotype and a protective haplotype, and the g.4694G and the c.1007-47A alleles were associated with a higher risk of elevated iron parameters. We determined that the g.4694C allele was located within a putative hypoxia-response element, protein binding was evidenced and was reduced with the g.4694C>G change. In addition, IVS4 was not spliced as well in the c.1007-47A allele compared to the c.1007-47G allele. ©2010 Wiley-Liss, Inc.

KEY WORDS: *HFE*; SNP; haplotype; iron overload; association study.

INTRODUCTION

HFE-linked hereditary hemochromatosis (OMIM 235200) is a common-autosomal-recessive iron homeostasis disorder affecting approximately 1:300 individuals of northern European descent [Edwards et al., 1988; Leggett et al., 1990; Olynyk et al., 1999; Adams, 2000]. This disease is characterized by hyperabsorption of dietary iron, which leads to progressive iron overload with increases in transferrin saturation and serum ferritin concentrations,

Received <date>; accepted revised manuscript <date>.

© 2010 WILEY-LISS, INC.

2 <First Author et al.>

and excess iron deposition in parenchymal cells of the liver, pancreas and heart liable to result in organ damage [Bothwell, 1995; Powell et al., 1999; Niederau et al., 1999]. The clinical onset usually appears after midlife, and the clinical outcomes and severity of the disease varies considerably. Prognosis depends on both an early diagnosis and iterative phlebotomies throughout life to prevent the occurrence of irreversible organ damage caused by iron overload [Adams et al., 1991; Niederau et al., 1996].

The *HFE* gene, located on the short arm of human chromosome 6, consists of 6 exons that encode a histocompatibility class I-like protein that associates with $\beta 2$ -microglobulin for normal processing and cell surface expression [Feder et al., 1996]. The HFE protein is predominantly expressed in tissues involved in iron storage, such as hepatocytes and macrophages [Holmström et al., 2003; Zhang et al., 2003], and its expression is transcriptionally regulated by hepatic C/EBP α and erythroid GATA-1 [Mura et al. 2004; Makui et al., 2005]. Functional HFE competes with holo-transferrin (holo-Tf) for binding to TfRC (OMIM 190010), and it interacts with TfR2 (OMIM 604720) independently of holo-Tf [Parkkila et al., 1997; Feder et al., 1998; Lebron et al., 1998; Gross et al., 1998; Chen et al., 2007]. Patients affected with *HFE*-linked hemochromatosis, *TfR2*-linked hemochromatosis, and *Hfe*^{-/-} mice, have inappropriately-low expression of the hepcidin peptide which regulates the ferroportin transmembrane iron exporter by inducing its internalization and degradation; both HFE and TfR2 are involved in a *HAMP* (OMIM 606464) regulatory pathway [Pigeon et al., 2001; Nicolas et al., 2002; Nemeth et al., 2004; Bridle et al., 2003; Nemeth et al., 2005; Wallace et al., 2009].

Most individuals affected by hemochromatosis (80-90%) are homozygous for the c.845G>A (p.C282Y) substitution. The prevalence of hemochromatosis correlates with the distribution of c.845A throughout the world, and linkage-disequilibrium studies have provided evidence of a founder effect consistent with a single ancestral occurrence of the c.845G>A mutation [Feder et al., 1996; Jazwinska et al., 1996; Carella et al., 1997; Mura et al., 1997]. The p.C282Y mutation disrupts a critical disulfide bond in the $\alpha 3$ domain, which impairs both its association with $\beta 2$ -microglobulin and its normal protein processing [Feder et al., 1997; Waheed et al., 1997]. However, variable expression of the disease and incomplete penetrance of the c.845A/c.845A genotype (asymptomatic homozygotes have been evaluated to 17.6%), both indicative of modulation of the severity of iron overload by secondary factors, have been observed in population screenings [Rhodes et al., 1997; Crawford et al., 1998; Mura et al., 2001; Beutler et al., 2002]. Non-c.845A patient chromosomes are characterized by a high prevalence of the c.187C>G (p.H63D) variant (40-70% vs. 14%-20% in non-c.845A patient and control chromosomes, respectively), and the c.193A>T (p.S65C) variant (7.2 and 2.49% in patient and control chromosomes with neither c.845G>A or c.187C>G variations) [Feder et al., 1996; Rish 1997; Beutler 1997; Bacon et al., 1999; Moirand et al., 1999; Mura et al., 1999]. Population studies have provided evidence for the very low penetrance of c.187G/c.187G, c.845A/c.187G, and c.845A/c.193T or c.187G/c.193T genotypes and their association with a milder form of iron overload compared to patients with the c.845A/c.845A genotype. A study of linkage disequilibrium by using external *HFE* microsatellites showed a similar pattern of haplotypes from c.187G chromosomes in patients and controls, and supports the occurrence of the c.187G variant on a single progenitor haplotype [Feder et al., 1996]. Aside from these variants, few *HFE* private nonsense and missense mutations in a compound heterozygous state with the c.845A mutation have been correlated with the hemochromatosis phenotype. In a few cases, hemochromatosis has been also correlated with mutations in the coding regions of other genes involved in iron homeostasis such as *TfR2*, ferroportin, hemojuvelin and hepcidin [Camaschella et al., 2000; Njajou et al., 2001; Papanikolaou et al. 2004], and digenic inheritance of *HFE* and one of the other mutated gene was also described [Merryweather-Clarke et al., 2003; Jacolot et al., 2004].

In the present study, we extended the analysis of *HFE* by using internal *HFE* SNPs, and derived haplotypes to determine whether particular alleles are associated with iron overload trait characterized by elevated serum iron parameters. Among non-mutant chromosomes, we discovered a protective haplotype and an at-risk haplotype with two of the five SNPs associated with elevated serum iron parameters; the latter being also in linkage disequilibrium with a subset of c.187G chromosomes. Functional studies of the putative pathogenic *HFE* SNPs showed that binding by the HIF-1 α transcription factor was affected by the g.4694G allele, which is located within the promoter region, and that IVS4 splicing was modified by the c.1007-47A allele.

PATIENTS AND METHODS

Subjects

The present study dealt with 265 unrelated patients with iron overload trait, and 55 healthy-unrelated controls, all from Brittany (France). All patients presented at least two elevated serum iron parameters (transferrin saturation above 60% in males and 50% in females, serum ferritin concentration above 400 μ g/L in males and 300 μ g/L in females, serum iron concentration above 20 μ M) [Mura et al., 1999], and all patients enrolled in the study were

1
2
3 treated by venesections. Control samples were obtained from healthy blood donors without abnormalities of serum
4 iron indices and without a clinical history of iron metabolism disorder. This study complied with French bioethics
5 regulations. Informed consent of patients was obtained before blood samples were taken.
6

7 **Reference sequences, nomenclature and primer sequences**

8 Entrez nucleotide database (<http://ncbi.nlm.nih.gov>) NG_008720.1:g, NM_000410.3:c, NP_000401.1 and
9 dbSNP ID were used as the reference sequences for *HFE* genomic, cDNA, and protein sequences. DNA sequence
10 variations are named according to the nomenclature recommendations from [http://www.hgvs.org/mutnomen/recs-](http://www.hgvs.org/mutnomen/recs-DNA.html)
11 DNA.html [den Dunnen et al., 2001; Ogino et al., 2007]. Primer sequences used in this study are listed in Supp.
12 Table S1.

13 **Genotyping**

14 Genomic DNA was extracted from peripheral white blood cells. *HFE* mutations c.845G>A (p.C282Y),
15 c.187C>G (p.H63D) and c.193A>T (p.S65C), and five SNPs (rs2794720, rs2071303, rs1800758, rs1800708,
16 rs1572982), were genotyped using PCR-RFLP for all patient and control individuals. PCR was carried out as
17 previously described [Mura et al., 1999], and restriction digests were performed according to the manufacturer's
18 specifications (New England Biolabs, Ipswich, MA, USA) using 10 μ L of PCR product and the appropriate
19 restriction enzymes (Supp. Table S2). Digested products were resolved on a 2% agarose gel. At least one
20 occurrence of each variant was confirmed by sequencing.
21

22 **Sequencing**

23 Chain terminator sequencing reactions were performed using the ABI BigDye Terminator v3.1 Cycle
24 Sequencing Kit (Applied Biosystems, Forster City, CA, USA) and analyzed on an ABI Prism 3130 Genetic
25 Analyzer (Applied Biosystems, Forster City, CA, USA) according to the manufacturer's specifications.
26

27 **Plasmid construction**

28 *HFE* minigenes were cloned into the pcDNA3.1(+) mammalian expression vector (Invitrogen Life
29 Technologies, Carlsbad, CA, USA). Plasmids containing *HFE* cDNA and IVS2 with c.187G>C, c.193A>T,
30 c.340+4T>C DNA variants were constructed as follows: the human *HFE* wild type cDNA was obtained by RT-
31 PCR using forward F2 and reverse R8 primers, and cloned into the pGEM-T plasmid (Promega, Madison, WI,
32 USA). Then, the 810-bp fragments amplified from the genomic DNA of individuals carrying either [c.187C;
33 c.340+4T], [c.187G; c.340+4C], [c.187C; c.340+4C], [c.193A; c.340+4T] or [c.193T; c.340+4C] using F4 and R5
34 primers were digested with *Hind*III and *Eco*RI to cleave the sequence at position g.8626 in exon 2 and position
35 g.9131 in exon 3, respectively. The resulting 505-bp digested product was inserted into the pGEM-T/*HFE* cDNA,
36 which was digested with *Hind*III and *Eco*RI to eliminate the 296-bp fragment. The clone containing the [c.187G;
37 c.340+4T] combination was obtained after substituting the *Nco*I fragment (*Nco*I cleaved at positions g.8700 in
38 exon 2 and g.11911 in exon 6) from the pGEM-T/*HFE* cDNA-IVS2 [c.187G; c.340+4C] clone with the 1.1-kb
39 *Nco*I fragment from the pGEM-T/*HFE* cDNA-IVS2 [c.187C; c.340+4T] clone. Plasmids containing the genomic
40 region encompassing exon 4 to exon 6, with the c.892+48G>A and c.1007-47G>A SNPs combinations, were
41 constructed by cloning the 1459-bp PCR product (obtained using the F7 and R8 primers and from genomic DNA
42 from individuals carrying either [c.892+48G; c.1007-47G], [c.892+48G; c.1007-47A], or [c.892+48A; c.1007-
43 47A] polymorphisms) into pGEM-T easy. All inserts were then excised from pGEM-T using *Eag*I, gel purified
44 based on size, and ligated into the pcDNA3.1 *Not*I-digested vector. Orientation of the inserts and the constructed
45 plasmids sequences were confirmed using automated DNA sequencing.

46 **Analysis of spliced transcripts**

47 *In vitro* transcription/splicing reactions were performed as described in Hicks et al. 2006. Briefly, pcDNA3.1
48 plasmid minigene constructs were linearized with *Pvu*I. One hundred nanograms of the digested products was
49 incubated for 2 h at 30°C with 30% nuclear extract (HeLaScribe Nuclear extract in vitro transcription system,
50 Promega, Madison, WI, USA); 0.4 mM each of rCTP, rUTP, and rATP, 16 μ M rGTP; 1 mM ATP; 20 mM
51 creatine phosphate; and 3 mM MgCl₂ in a 25- μ L reaction. After the addition of 1 u DNase I, reactions were
52 incubated for 15 min at 37°C. The reactions were stopped by the addition of 175 μ L of 0.3 M Tris-HCl pH 7.4,
53 0.3 M sodium acetate, 0.5% SDS, 2 mM EDTA, and 0.3 μ g/mL tRNA. The reaction products were extracted twice
54 with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated overnight at -70°C with ethanol. RNAs pellets
55 were resuspended in 20 μ L of RNase-free water. Two-step RT-PCR reactions were performed by standard method
56 using various sets of forward and reverse primers, and 5 μ L of total RNA. The products of semi-quantitative PCR
57 reactions (after 20-25 cycles, each consisting of 30 sec. at 95°C; 30 sec. at 57°C; and 1 min at 72°C), were
58
59
60

4 <First Author et al.>

analyzed on 2% agarose gel and gel images were digitally captured with a CCD camera and analyzed with the Bio-Profil Bio1D software imager program (Viber Lourmat, Marne-La-Vallée, France).

In silico DNA analysis

The haplotypes based on the five SNPs, as well as c.845G>A, c.187C>G and c.193A>T mutations, were derived from all the patients and controls using the PLINK software (<http://pngu.mgh.harvard.edu/purcell/plink/>) [Purcell et al., 2007]. The Human Splicing Finder 2.4 (<http://www.umd.be/HSF/>) sequence analysis algorithm was used to predict the effects of intervening sequence variations on splicing signals [Desmet et al., 2009]. This algorithm defines consensus values to 5' and 3' splice sites, maximizing the difference between wild type active sites and mutant inactive sites. A normalized range scale from 0 to 100 is used for auxiliary sequences, enhancers, and silencers. TFSEARCH (<http://www.cbrc.jp/research/db/TFSEARCH.html>) and MatInspector (<http://www.genomatix.de/cgi-bin/eldorado/main.pl>) programs were used to search for putative transcription factor binding sites.

Electrophoretic mobility shift assay

Sense and antisense single-stranded oligonucleotides from position g.4677 to position g.4710, including either the g.4694C or g.4694G variant (sense 5'-TTATGAAGGCTGTGGAAC/GGTGTTTCAGTAGGATC-3'), were labeled using the Biotin 3'-end DNA labeling kit (Pierce, Thermo Scientific, Essex, UK) according to the manufacturer's specifications. Gel shift assays were carried out using the Light Shift Chemiluminescent EMSA kit (Pierce, Thermo Scientific, Essex, UK). Briefly, 5 µg of nuclear extract, in 10 mM Tris pH7.5/50 mM KCl/1 mM DTT reaction buffer and 1 µg of poly(dI-dC) were pre-incubated for 10 min at room temperature. After the addition of labeled probe (20 fmol), the reaction mixture was incubated at room temperature for 30 min in a final volume of 20 µL. For competition experiments, 25- to 100-fold excess unlabeled oligonucleotides were incubated with the reaction mixture for 20 min before the addition of labeled probe. Then, the complexes were separated on a 5% polyacrylamide gel (29:1) in 0.5 x TBE, then chemiluminescence with streptavidin-horseradish peroxidase conjugate and substrate was used for detection of western blots. The signals were analyzed by densitometry on a GelDoc 2000 apparatus (Bio-Rad, Hercules, CA, USA).

Statistical analysis

The two-tailed Student's *t*-test was used to compare the mean age and mean value of serum iron parameters between groups of individuals. Deviation from Hardy-Weinberg equilibrium (HWE) of the genotyped SNPs (Fisher's exact test), haplotype and allele associations analysis with elevated serum iron parameters, (Fisher's exact test, Pearson's Chi square, and odds ratio), as well as linkage disequilibrium (D') calculations were performed using PLINK software (<http://pngu.mgh.harvard.edu/purcell/plink/>) [Purcell, et al., 2007]. A *p*-value less than 0.05 was considered statistically significant.

RESULTS**Mutation genotyping**

In this study, patients, that all had elevated serum iron parameters and were treated by venesections, were divided into two groups. The first group, consisting of patients (n=195) either homozygous or compound heterozygous for the c.845G>A (p.C282Y), c.187C>G (p.H63D) or c.193A>T (p.S65C) mutations within the *HFE* coding sequence, was used to derived haplotypes. The alleles were distributed within this group as follows: c.845A/c.845A (n=128), c.187G/c.187G (n=34), c.845A/c.187G (n=19), c.845A/c.193T (n=10), c.187G/c.193T (n=4) (mean of iron removed 8.76 ± 5.68 g for c.845A/c.845A and 3.92 ± 1.65 g for the others). The second group consisted of patients (n=70) lacking mutations within the *HFE* coding region on at least one of their chromosomes (mean of iron removed 2.97 ± 1.17 g) and either c.845A heterozygotes (n=19), c.187G heterozygotes (n=15), or c.193T heterozygotes (n=2); thirty-four individuals from this group had no mutations within *HFE* coding region. For all patients lacking at least a mutation, the six *HFE* exons and intron-exons boundaries regions were sequenced to verify there were no additional mutations. The age, gender, and serum iron parameters of the patients (according to their genotype for the c.845G>A, c.187C>G and c.193A>T mutations) are listed in Table 1. There were no significant differences observed for the age of diagnosis of iron overload trait between the genotypic groups, but c.845A/c.845A individuals were diagnosed at a younger age than the c.187G/wt, c.193T/wt and wt/wt individuals ($p < 10^{-3}$, Student's *t*-test). Transferrin saturation and serum ferritin levels at diagnosis indicated that iron overload was milder in patients that were not homozygous for the c.845A mutation than in patients homozygous for the c.845A mutation ($p < 0.05$, Student's *t*-test).

Table 1. Patient features for the C282Y, H63D and S65C HFE substitutions.

Features	Patients (n=195)				Patients (n=70)		
	c.845A/c.845A (n=128)	c.187G/c.187G (n=34)	c.845A/c.187G (n=19)	c.845A/c.193T c.187G/c.193T (n=14)	c.845A /wt (n=19)	c.187G /wt or c.193T /wt (n=17)	wt/wt (n=34)
Age (years)							
mean \pm SD	46.0 \pm 11.7	48.8 \pm 12.98	48.0 \pm 12.2	48.4 \pm 16.3	48.8 \pm 12.9	54.7 \pm 13.4	52.9 \pm 12.2
range	[25-70]	[32-74]	[23-66]	[23-73]	[27-65]	[22-77]	[35-65]
		p=0.23	p=0.15	p=0.41	p=0.14	p=1.10 ⁻³	p=1.10 ⁻⁴
Gender (n)							
male	103 (80.5%)	28 (82.4%)	17 (89.5%)	11 (78.7%)	19 (100%)	14 (82.4%)	32 (94.1%)
female	25 (19.5%)	6 (17.6%)	2 (10.5%)	3 (21.4%)	0 (0%)	3 (17.6%)	2 (5.9%)
Iron parameters							
TS (%) mean \pm SD	78.8 \pm 15.1	52.6 \pm 11.7 p=10 ⁻⁵	55.4 \pm 22.8 p=10 ⁻⁸	55.6 \pm 22.8 p=0.05	53.1 \pm 11.6 p=10 ⁻¹³	48.2 \pm 20.3 p=10 ⁻⁷	46.4 \pm 16.3 p=10 ⁻¹⁶
SF (μ g/mL) mean [range]	1419 [116-8890]	778 [350-1500] p=10 ⁻⁷	688 [138-1890] p=10 ⁻¹¹	488 [170-784] p=10 ⁻⁴	707 [319-2145] p=10 ⁻⁷	960 [318-4000] p=10 ⁻³	854 [346-1920] p=10 ⁻⁷

The mean age, with standard deviation (SD), age range, gender, transferrin saturation (TS) and serum ferritin (SF) values at diagnosis. P values were calculated with Student's *t*-test to compare age and iron parameters, respectively, between homozygotes and nonhomozygotes for the c.845A substitution.

Allele and haplotype frequencies analysis

SNPs and derived haplotypes were analyzed to gain more insight into a possible association between the specific alleles found within the patients (and therefore overrepresented in chromosomes of patients) and the iron status. Five previously reported SNPs located within the *HFE* locus [Beutler and West, 1997; Totaro et al., 1997], one located within the promoter region and four located within intronic regions near splice boundaries, were screened by PCR-RFLP in patients and controls (Supp. Table S2). The minor allele frequency of all SNPs was greater than 5% in both controls and patients, with the exception of c.893-44T>C. Moreover, all the SNPs in the controls were within the expected Hardy-Weinberg proportions ($p > 0.01$, Fisher's exact test; no calculation for c.893-44T>C). The haplotypes based on the five SNPs, as well as c.845G>A, c.187C>G and c.193A>T mutations, were derived from all the patients and controls using the PLINK software. The analysis revealed that all of the 289 c.845A-carrier chromosomes investigated from the two series of patients and controls were in complete linkage disequilibrium with the [g.4694C; c.340+4T; c.892+48G; c.893-44T; c.1007-47G] haplotype, subsequently noted as [C; T; G; T; G], confirming the unique occurrence of this mutation. In regards to the c.187G mutation, this analysis showed that 79.8% out of a total of 110 carrier chromosomes were in linkage disequilibrium with the [C; C; G; T; A] haplotype and 20.2% with the [G; C; G; T; A] haplotype. A discrepancy tended to emerge, though not statistically significant: the g.4694G allele appeared to be more common among patients than in controls. The g.4694G allele accounted for 21.4% of the c.187G patient chromosomes and 11.35%, of the c.187G carrier-control chromosomes. Interestingly, the twelve c.193T carrier chromosomes were in complete linkage disequilibrium with the [C; C; G; T; A] haplotype. This indicates that the c.187G and c.193T carrier chromosomes share an identical 3' region of the *HFE* [c.340+4C; c.892+48G; c.893-44T; c.1007-47A] ancestral haplotype.

SNP and haplotypes from controls (n=42) and the group of patients with no mutations on at least one of their chromosomes (n=70) were analyzed. In this group of patients the c.845A carrier chromosomes, associated with [C; T; G; T; G] haplotype, tended to be at risk for iron overload trait (with frequency of 15.67% in patients versus 7.3% in healthy controls). Table 2 lists the SNP allele frequency distribution from the group of patients that lack mutations on at least one of their chromosomes and the control group, excluding the c.845A mutation-carrier chromosomes. The g.4694C (position: 5'UTR-467) allele is negatively associated with iron overload trait (Fisher's exact test, $p = 0.025$; Chi square, 5.135 $p = 0.021$; OR 0.5 [0.279-0.905]), and a total of 20 homozygous for g.4694G allele have been found in patients that do not carry any mutation within the coding sequence of *HFE* (n=34), whereas only 2 homozygous have been found in controls (n=42). The c.1007-47G allele had moderate negative association with iron overload trait (Fisher's exact test, $p = 0.058$; Chi square, 3.864 $p = 0.049$; OR 0.56 [0.31-1]). The frequency of other SNP alleles was not significantly different in patients compared to controls. Interestingly, the c.187G and c.193T alleles were both in complete linkage disequilibrium with the c.1007-47A allele, but c.1007-47A was also found on some other non-mutant chromosomes. In addition, as previously noted the g.4694G

6 <First Author et al.>

Table 2. Distribution of *HFE* SNPs in controls and heterozygous patients, excluding the C282Y mutation, or non-carrier of a mutation. Allelic association test results using Fisher's test, chi square and OR. n= number of chromosomes studied.

SNP	Allele	Allele count (frequency)		Allele association (p value Fisher's test)	χ^2 p value	OR [95% CI]
		Patients (n=108)	Controls (n=84)			
rs2794720 -467C>G	C*	35 (0.3241)	41 (0.4881)	0.025	5.315	0.50 [0.279-0.905]
	G	85 (0.7872)	43 (0.5119)		0.021	
rs1799945 c.187C>G	C	0.1296	0.1548	0.67	0.247	0.8134 [0.359-1.838]
	G*				0.6192	
rs1800730 c.193A>T	T*	0.0185	0.0238	1	0.064	0.7736 [0.1067-5.609]
	A				0.799	
rs2071303 IVS2 +4T>C	T	86 (0.7979)	58 (0.6905)	1	0.006	1.025 [0.5538 -1.897]
	C*	22 (0.3148)	26 (0.3095)		0.9375	
rs1800758 IVS4+48G>A	G	84 (0.7766)	74 (0.8801)	0.17	1.984	1.786 [0.7912-4.033]
	A*	24 (0.194)	10 (0.119)		0.159	
rs1800708 IVS4-44T>C	T	105 (0.9680)	83 (0.9881)	1	0.1344	1.566 [0.1396-17.57]
	C*	3 (0.018)	1 (0.0119)		0.7139	
rs1572982 IVS5-47G>A	G*	45(0.4167)	47 (0.5595)	0.058	3.864	0.5625 [0.3161-1]
	A	63 (0.5833)	37 (0.4405)		0.049	

Single asterisk indicates the risk allele of each SNP. Abbreviations: SNP=single nucleotide polymorphism; OR= odds ratio; CI=confidence interval. *P* values are calculated by comparing chromosomes from patients to controls.

Table 3. *HFE* haplotype analysis of controls and patients lacking mutations on at least one of their chromosomes.

SNP								Haplotype frequency		
rs2794720 g.4694C>G	rs1799945 c.187C>G	rs1800730 c.193A>T	rs2071303 c.340 +4T>C	rs1800562 c.845G>A	rs1800758 c.892 +48G>A	rs1800708 c.893 -44T>C	rs1572982 c.1007 -47G>A	Patients (n=128)	Controls (n=84)	Fisher p value
C	C	A	T	A	G	T	G	20 (0.1567)	6.17 (0.073)	0.0857
C	G	A	C	G	G	T	A	12 (0.093)	11.1 (0.132)	0.5002
G	G	A	C	G	G	T	A	1.39 (0.011)	1.2 (0.014)	1
C	C	T	C	G	G	T	A	2.1 (0.016)	2 (0.024)	1
G	C	A	T	G	G	T	G	33.6 (0.263)	24.8 (0.296)	0.5182
C	C	A	T	G	G	T	G	14 (0.108)	17.3 (0.206)	0.0399
C	C	A	C	G	G	T	A	6.1 (0.047)	3 (0.036)	1
G	C	A	C	G	G	T	A	13.1 (0.103)	8.2 (0.098)	1
G	C	A	T	G	A	T	A	18.6 (0.145)	9.67 (0.115)	0.5295
G	C	A	T	G	G	T	A	7 (0.054)	0.22 (0.002)	0.083

All haplotypes used had a frequency greater than 1%.

allele was more prevalent on c.187G-carrier chromosomes from patients than in those of controls. The c.893-44C allele was only observed in chromosomes from three patients with different haplotypes, but all with the g.4694G variant. Moreover, the c.340+4T>C change in IVS2 showed frequencies of 0.726 and 0.310 in patients and controls, respectively, but the c.340+4C allele was in complete linkage disequilibrium with the c.187G ($D' = 0.94$)

and c.193T mutations, thus with the [c.892+48G; c.893-44T; c.1007-47A] *HFE* haplotype. In addition, in this group of patients the [G; T; G; T; A] within non-mutant chromosomes was also identified as an at risk haplotype (5.4% in patients versus 0.2% in healthy controls), whereas the [C; T; G; T; G] haplotype (with frequency of 10.8% in patients versus 20.6% in controls), proved to be a protective haplotype (Table 3). This association study suggests that the g.4694G, c.340+4C, and c.1007-47A alleles could play a role in the genetic etiology of iron overload in the population of being investigated.

Functional analysis of sequence variations

Examination of the human *HFE* promoter sequence using the TFSEARCH and MatInspector softwares shows that the g.4693 to g.4697 sequence, including nucleotide g.4694C (5'UTR-467C), encompasses a putative consensus hypoxia-response element (HRE), which is defined by the sequence [A/G]CGTG. HRE binds to the HIF-1 α transcription factor, and the g.4694G nucleotide abolishes this binding site. To further characterize the g.4694C>G (5'UTR-467C>G) variation, located within the putative HRE, we tested for binding by the nuclear protein factor. Electrophoretic mobility shift assays (EMSA) of labeled-synthetic double-stranded oligonucleotides from both variants, after incubation with nuclear extract, provided evidence of the formation of protein-DNA complex with the sequence containing either the g.4694C or G variants (Fig. 1). The interaction of the protein was abolished through competition experiments using an excess of an unlabeled homologous double-stranded oligonucleotide. Competition experiments using the g.4694G variant oligonucleotide sequence was two-fold less efficient to abolish protein binding to the g.4694C variant oligonucleotide than the g.4694C unlabeled oligonucleotide.

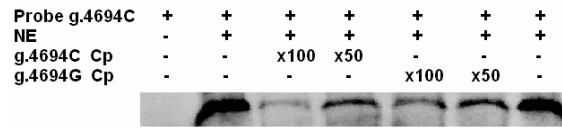


Figure 1. Electrophoretic mobility shift assay. Biotin-labeled double-stranded oligonucleotides were incubated with 5 μ g of nuclear extract (NE) and 1 μ g of poly(dI-dC) in a final volume of 20 μ L. For competition experiments (Cp) a 50- or 100-fold excess of g.4694C or g.4694G unlabeled oligonucleotide was incubated with the reaction mixture for 20 min before addition of the labeled probe. Complexes were then separated on 5% polyacrylamide gel (29:1) in 0.5 x TBE by electrophoresis. The mobility results showed the formation of protein-DNA complex for the g.4694C variant sequence after incubation with nuclear extract compared to incubation without nuclear extract. The complexes that formed with the g.4694C variant were abolished through competition experiments with the nuclear extracts that were preincubated with an excess of unlabeled homologous oligonucleotide. This demonstrates the specificity of the interaction between the nuclear proteins and the double-stranded oligonucleotide sequence. Quantitation of the binding efficiency by CCD camera scanning from 4 independent experiments obtained from competition experiments showed that the g.4694G variant oligonucleotide sequence was two-fold less efficient to abolish protein binding to the g.4694C variant oligonucleotide than the g.4694C unlabeled oligonucleotide.

To determine whether intronic sequence variations located within the splice site region (c.340+4T>C), or within a short distance of a splice site (c.892+48G>A, c.893-44T>C and c.1007-47G>A), could interfere with correct pre-mRNA splicing, sequence variations were submitted to Human Splicing Finder 2.4 (<http://www.umd.be/HSF/>) to predict any alteration in splicing in addition to splice factor binding sites that result in misregulated splicing. c.340+4T>C gives a Δ Consensus Value of +9.71% exceeded the 7% for +4 position of the 5' splice site, a threshold value considered by the software to be indicative of a significant impact on splicing. RT-PCR products, analyzed for size, from *HFE* minigene expression plasmids harboring a combination of variants, were used as substrates for *in vitro* RNA synthesis and splicing in HeLa cell nuclear extracts. Spliced products revealed identical transcripts from the [c.187C; c.340+4C], [c.187C; c.340+4T], [c.187G; c.340+4C] and [c.187G; c.340+4T] minigenes (obtained with F3, F5, R3 and R4 primers), and were equally efficient in producing the expected size product of the normal IVS4 spliced mRNA form (data not shown). Splicing efficiency of the [c.193A; c.340+4T] and the [c.193T; c.340+4C] minigenes were comparable (data not shown). According to the Human Splicing Finder 2.4 software, the c.1007-47G>A variant disrupts an IIE silencer motif GATGGTGCT, and an hnRNP A GATGGT site (68.33 to 0), both of which could interfere with splicing. Minigene spliced products from the [c.892+48A; c.893-44T; c.1007-47A] and [c.892+48G; c.893-44T; c.1007-47A] haplotypes differ from the [c.892+48G; c.893-44T; c.1007-47G] haplotype (Supp. Fig. 1). There was a reduction of full-length spliced transcripts, and the amount of the unspliced IVS4 variant was increased 2- to 4-fold. On the other hand, the unspliced IVS5 transcript was unchanged in the haplotype carrying the c.1007-47A allele compared to the haplotype carrying the c.1007-47G allele.

DISCUSSION

To gain more insight into *HFE* sequence variations, we analyzed a large number of patients (all from Brittany) affected with elevated serum iron parameters through use of internal *HFE* SNPs (g.4694G>C; c.340+4T>C; c.892+48G>A; c.893-44T>C; c.1007-47G>A) and derived haplotypes. The haplotype analysis confirmed that c.845G>A (p.C282Y) mutation-carrier chromosomes are in complete linkage disequilibrium with a unique [C; T; G; T; G] haplotype, which is quite common and accounts for about 20.6% of non-mutant chromosomes. The results of this study, in agreement with previous linkage disequilibrium data obtained by using external *HFE* gene microsatellites [Feder et al., 1996], argues that the c.845G>A mutation occurred once on a progenitor haplotype, and spread throughout the Caucasian population.

HFE c.187C>G (p.H63D), and, to a lesser extent, c.193A>T (p.S65C) are variants associated with non-c.845A chromosomes from hemochromatosis patients. Both variants have a very low penetrance and mild expressivity for iron overload. Here, haplotype analysis revealed that the [C; C; G; T; A] and [G; C; G; T; A] haplotypes account for 79.8% and 20.2%, respectively, of c.187G-carrier chromosomes; these haplotypes only diverge for the g.4694G>C variation. This may suggest that the c.187G variant occurred once in the ancestral [C; C; G; T; A] haplotype, and that a subsequent divergence appeared at the g.4694C>G (5'UTR-467C>G) position. In this respect, g.4694C>G divergence could have arisen either by a mutational event or through recombination, but the lack of recombination observed in the *HFE* gene region, and the similar pattern of haplotypes defined with external microsatellites observed in the c.187G chromosomes, argues against the latter hypothesis [Feder et al., 1996]. Another assumption is that the c.187G mutation would have occurred twice on ancestral haplotypes, which account for 3.6% and 9.8%, respectively, of non-carrier chromosomes in controls. Interestingly, previous studies have shown that in populations of European descent, c.187G is associated with the [c.340+4C; c.893-44T; c.1007-47A] haplotype. On the other hand, in Sri Lanka, where the allele frequency is about 10%, c.187G is associated with three haplotype backgrounds [c.340+4T; c.893-44T; c.1007-47G], [c.340+4T; c.893-44T; c.1007-47A] and [c.340+4C; c.893-44T; c.1007-47G]. These findings suggest that the c.187G mutation may have arisen more than once in Sri Lanka, independently from the mutation in Europe [Rochette et al., 1999]. Furthermore, despite its high prevalence in European populations, the low penetrance of the c.187G variant, together with the low prevalence of the disease among the Sri Lankan (Asian) population, casts doubt on the risk of developing the disease induced by this variant. This study showed that in Caucasian population the c.187G chromosomes, as well as the c.193T chromosomes, are in complete linkage disequilibrium with the c.1007-47A variant, whereas the g.4694G variant is in linkage disequilibrium with a subset of c.187G chromosomes that are increased in patients compared to controls. In addition, through comparisons with control subjects, the present study provides evidence of a significant association of the g.4694G variant with mild iron overload trait in non- c.845A, - c.187G or - c.193T patient chromosomes, as well as a weak association of the c.1007-47A variant. The results also show a risk of iron overload trait linked to the [g.4694G; c.340+4T; c.892+48G; c.893-44T; c.1007-47A] haplotype, and a protective role of the [g.4694C; c.340+4T; c.892+48G; c.893-44T; c.1007-47G] haplotype for non-mutant chromosomes differing by g.4694C>G and c.1007-47G>A variations, confirming the potential role of g.4694G and c.1007-47A in iron overload. g.4694C>G is located within a putative HRE binding site for the hypoxia-inducible transcription factor, HIF-1 α , found in the *HFE* 5'-flanking region. The g.4694G allele was proved to reduce protein binding. It is worth noting that HIF-1 α constitutes a link between the physiology of hypoxic response and the control of iron availability; in particular, its binding within the *hepcidin* sequence promoter is involved in the downregulation of *hepcidin* transcription [Peyssonnaud et al., 2007], HIF-1 could thus be implicated in *HFE* transcriptional regulation. In addition, the c.1007-47A allele causes a reduction of IVS4 normal splicing compared to c.1007-47G. An analysis using the Human Splicing Finder software revealed that this change would abolish a IIE silencer motif and an hnRNPA site, and in doing so could have an impact on splicing. Therefore, the association of both g.4694C>G and c.1007G>A changes in sequence leading to a reduced response to HIF-1 transcription factor and/or to incorrect splicing of the transcript, would result in lower levels of intact mRNA and a reduction of functional HFE protein. The low penetrance of the c.187G mutation could thus be related to the increased risk for developing iron overload trait from a subset of c.187G chromosomes associated with the [g.4694G; c.340+4C; c.892+48G; c.893-44T; c.1007-47A] haplotype. The g.4694C>G variant and to a lesser extent c.1007G>A are significantly defined at risk for mild iron overload trait, consequently the screening of the g.4694C>G and c.1007G>A variations would bring additional refinement to the correlation between iron overload trait and *HFE* genotype.

ACKNOWLEDGMENTS

This work was supported by Institut National de la Santé et de la Recherche Médicale.

REFERENCES

- Adams PC, Speechly M, Kertesz AE. 1991. Long-term survival analysis in hereditary hemochromatosis. *Gastroenterol* 101:368-372.
- Adams PC. 2000. Population screening for hemochromatosis. *Gut* 46:301-303.
- Bacon BR, Powell LW, Adams PC, Kresina TF, Hoofnagle JH. 1999. Molecular medicine and hemochromatosis: at the cross-roads. *Gastroenterol* 116:193-207.
- Beutler E, Felitti VJ, Koziol JA, Ho NJ, Gelbert T. 2002. Penetrance of 845G→A (C282Y) hereditary haemochromatosis mutation in the USA. *Lancet* 359:211-218.
- Beutler E, West C. 1997. New Diallelic Markers in the HLA Region of Chromosome 6. *Blood Cells Mol Dis* 23:219-229.
- Beutler E. 1997. The significance of the 187G (H63D) mutation in hemochromatosis. *Am J Hum Genet* 61:762.
- Bothwell TH. 1995. Overview and mechanisms of iron regulation. *Nutr Rev* 53:237-245.
- Bridle KR, Frazer DM, Wilkins SJ, Dixon JL, Purdie DM, Crawford DHG, Subramaniam VN, Powell LW, Anderson GJ, Ramm GA. 2003. Disrupted hepcidin regulation in *HFE*-associated haemochromatosis and the liver as a regulator of body iron homeostasis. *Lancet* 361:669-673.
- Camaschella C, Roetto A, Cali A, De Gobbi M, Garozzo G, Carella M, Majorano N, Totaro A, Gasparini P. 2000. The gene *TFR2* is mutated in a new type of haemochromatosis mapping to 7q22. *Nat Genet* 25:14-15.
- Carella M, D'Ambrosio L, Totaro A, Grifa A, Valentino MA, Piperno A, Girelli D, Roetto A, Franco B, Gasparini P, Camaschella C. 1997. Mutation analysis of the HLA-H gene in Italian hemochromatosis patients. *Am J Hum Genet* 60:828-832.
- Chen J, Chloupková M, Gao J, Chapman-Arvedson TL, Enns CA. 2007. HFE modulates transferrin receptor 2 levels in hepatoma cells via interactions that differ from transferrin receptor 1-HFE interactions. *J Biol Chem* 282:36862-36870.
- Crawford DHG, Jazwinska EC, Cullen LM, Powell LW. 1998. Expression of HLA-linked hemochromatosis in subjects homozygous or heterozygous for the C282Y mutation. *Gastroenterol* 114:1003-1008.
- den Dunnen JT, Antonarakis SE. 2001. Nomenclature for the description of human sequence variations. *Hum Genet* 109:121-124.
- Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 37:e67 doi :101093/nar/gkp215.
- Edwards CQ, Griffen LM, Goldgar D, Drummond C, Skolnick MH, Kushner JP. 1988. Prevalence of hemochromatosis among 11065 presumably healthy blood donors. *N Engl J Med* 318:1355-1362.
- Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, Dormishian F, Domingo R, Ellis MC, Fullan A, Hinton LM, Jones NL, Kimmel BE, Kronmal GS, Lauer P, Lee VK, Loeb DB, Mapa FA, McClelland E, Meyer NC, Mintier GA, Moeller N, Moore T, Morikang E, Prass CE, Quintana L, Starnes SM, Schatzman RC, Brunke KJ, Drayna DT, Risch NJ, Bacon BR, Wolff RK. 1996. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 13:399-408.
- Feder JN, Penny DM, Irrinki A, Lee VK, Lebron JA, Watson N, Tsuchihashi Z, Sigal E, Bjorkman PJ, Schatzman RC. 1998. The hemochromatosis gene product complexes with the transferrin receptor and lowers its affinity for ligand binding. *Proc Natl Acad Sci USA* 95:1472-1477.
- Feder JN, Tsuchihashi Z, Irrinki A, Lee VK, Mapa FA, Morikang E, Prass CE, Starnes SM, Wolff RK, Parkkila S, Sly WS, Schatzman RC. 1997. The hemochromatosis founder mutation in HLA-H disrupts β 2-microglobulin interaction and cell surface expression. *J Biol Chem* 272:14025-14028.

10 <First Author et al.>

- 1
2
3 Gross CN, Irrinki A, Feder JN, Enns CA. 1998. Co-trafficking of HFE a nonclassical major histocompatibility
4 complex class I protein with the transferrin receptor implies a role in intracellular iron regulation. *J Biol Chem*
5 273:22068-22074.
6
7 Hicks MJ, Yang CR, Kotlajich MV, Hertel KJ. 2006. Linking splicing to Pol II transcription stabilizes pre-mRNAs
8 and influences splicing patterns. *PLOS Biology* 6:943-951.
9
10 Holmström P, Dzikaite V, Hultcrantz R, Melefors O, Eckes K, Stål P, Kinnman N, Smedsrød B, Gåfvæls M,
11 Eggertsen G. 2003. Structure and liver cell expression pattern of the HFE gene in the rat. *J Hepatol* 39:308-314
12
13 Jacolot S, Le Gac G, Scotet V, Quere I, Mura C, Ferec C. 2004. HAMP as a modifier gene that increases the
14 phenotypic expression of the HFE pC282Y homozygous genotype. *Blood* 103:2835-2840.
15
16 Jazwinska EC, Cullen LM, Busfield F, Pyper WR, Webb SI, Powell LW, Morris CP, Walsh TP. 1996.
17 Haemochromatosis and HLA-H. *Nat Genet* 14:249-251.
18
19 Lebron JA, Bennett MJ, Vaughn DE, Chirino AJ, Snow PM, Mintier GA, Feder JN, Bjorkman PJ. 1998. Crystal
20 structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor.
21 *Cell* 93:111-123.
22
23 Leggett BA, Halliday JW, Brown NN, Bryant S, Powell LW. 1990. Prevalence of hemochromatosis amongst
24 asymptomatic Australians. *Br J Hematol* 74:525-530.
25
26 Makui H, Soares RJ, Jiang W, Constante M, Santos MM. 2005. Contribution of Hfe expression in macrophages to
27 the regulation of hepatic hepcidin levels and iron loading. *Blood* 106:2189-2195.
28
29 Merryweather-Clarke AT, Cadet E, Bomford A, Capron D, Viprakasit V, Miller A, McHugh PJ, Chapman RW,
30 Pointon JJ, Wimhurst VL, Livesey KJ, Tanphaichitr V, Rochette J, Robson KJ. 2003. Digenic inheritance of
31 mutations in HAMP and HFE results in different types of haemochromatosis. *Hum Mol Genet* 12:2241-2247.
32
33 Moirand R, Jouanolle AM, Brissot P, Le Gall JY, Davis V, Deugnier Y. 1999. Phenotypic expression of HFE
34 mutations: a French study of 1110 unrelated iron-overloaded patients and relatives. *Gastroenterol* 116:372-377.
35
36 Mura C, Le Gac G, Jacolot S, Férec C. 2004. Transcriptional regulation of the human HFE gene indicates high
37 liver expression and erythropoiesis coregulation. *FASEB Journal* 18:1922-1924.
38
39 Mura C, Le Gac G, Scotet V, Raguénes O, Mercier AY, Férec C. 2001. Variation of phenotypic expression in
40 C282Y homozygous haemochromatosis probands and sibpairs. *J Med Genet* 38:632-636.
41
42 Mura C, Nousbaum JB, Verger P, Moalic MT, Raguénes O, Mercier AY, Ferec C. 1997. Phenotype-genotype
43 correlation in haemochromatosis patients. *Hum Genet* 101:271-276.
44
45 Mura C, Raguénes O, Férec C. 1999. HFE mutations analysis in 711 hemochromatosis probands: evidence for
46 S65C implication in mild form of hemochromatosis. *Blood* 93:2502-2505.
47
48 Nemeth E, Roetto A, Garozzo G, Ganz T, Camaschella C. 2005. Hepcidin is decreased in Tfr2 hemochromatosis.
49 *Blood* 105:1803-1806.
50
51 Nemeth E, Tuttle MS, Powelson J, Vaughn MB, Donovan A, Mc Vey Ward D, Ganz T, Kaplan J. 2004. Hepcidin
52 regulates cellular iron efflux by binding to ferroportin and inducing its internalization. *Science* 306:2090-2093.
53
54 Nicolas G, Chauvet C, Viatte L, Danan JL, Bigard X, Devaux I, Beaumont C, Kahn A, Vaulont S. 2002. The gene
55 encoding the iron regulatory peptide hepcidin is regulated by anemia hypoxia and inflammation. *J Clin Invest*
56 110:1037-1044.
57
58 Niederau C, Erhardt A, Haussinger D, Strohmeyer G. 1999. Hemochromatosis and the liver. *J Hepatol* 30:6-11.
59
60 Niederau C, Fisher R, Purschel A, Stremmel W, Haussinger D, Strohmeyer G. 1996. Long-term survival in
patients with hereditary hemochromatosis. *Gastroenterol* 110:1107-1119.
Najjou OT, Vaessen N, Joosse M, Berhuis B, van Dongen JWF, Breuning MH, Snijders PJLM, Rutten WPF,
Sandkuijl LA, Oostra BA, van Duijn CM, Heutink P. 2001. A mutation in SLC11A3 is associated with
autosomal dominant hemochromatosis. *Nat Genet* 28:213-214.

- 1
2
3 Ogino S, Gulley MI, den Dunnen JT, Wilson RB. 2007. Standard mutation nomenclature in molecular diagnostics:
4 practical and educational challenges. *J Mol Diagn* 9:1-6.
5
6 Olynyk JK, Cullen DJ, Aquilia S, Rossi E, Summerville L, Powell LW. 1999. A population-based study of the
7 clinical expression of the hemochromatosis gene. *N Engl J Med* 341:718-724.
8
9 Papanikolaou G, Samuels ME, Ludwig EH, MacDonald MLE, Franchini PL, Dubé MP, Andres L, MacFarlane J,
10 Sakellaropoulos N, Politou M, Nemeth E, Thompson J, Risler JK, Zaborowska C, Babakaiff R, Radomski CC,
11 Pape TD, Davidas O, Christakis J, Brissot P, Lockitch G, Ganz T, Hayden MR, Goldberg YP. 2004. Mutations
12 in *HFE2* cause iron overload in chromosome 1q-linked juvenile hemochromatosis. *Nat Genet* 36:77-82.
13
14 Parkkila S, Waheed A, Britton RS, Bacon BR, Zhou XY, Tomatsu S, Fleming RE, Sly WS. 1997. Association of
15 the transferrin receptor in human placenta with HFE the protein defective in hereditary hemochromatosis. *Proc
Natl Acad Sci USA* 94:13198-13202.
16
17 Peyssonnaud C, Zinkernagel AS, Schuepbach RA, Rankin E, Vaulont S, Haase VH, Niset V, Johnson RS. 2007.
18 Regulation of iron homeostasis by the hypoxia-inducible transcription factors (HIFs). *J Clin Invest* 117:1926-
19 1932.
20
21 Pigeon C, Ilyn G, Courselaud B, Leroye P, Turlin B, Brissot P, Loréal O. 2001. A new mouse-liver specific gene
22 encoding a protein homologous to human antimicrobial peptide hepcidin is overexpressed during iron overload.
23 *J Biol Chem* 276:7811-7819.
24
25 Powell LW, Subramaniam VN, Yapp TR. 1999. Haemochromatosis in the new millenium. *J Hepatol* 32:48-62.
26
27 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ,
28 Sham PC. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses.
29 *Am J Hum Genet* 81:559-575.
30
31 Rhodes DA, Raha-Chowdhury R, Cox TM, Trowsdale J. 1997. Homozygosity for the predominant Cys282Tyr
32 mutation and absence of disease expression in hereditary haemochromatosis. *J Med Genet* 34:761-764.
33
34 Rish N. 1997. Haemochromatosis HFE and genetic complexity. *Nat Genet* 17:375.
35
36 Rochette J, Pointon JJ, Fisher CA, Perera G, Arambepola M, Arichchi DS, De Silva S, Vandwalle JL, Monti JP,
37 Old JM, Merryweather-Clarke AT, Weatherall DJ, Robson KJ. 1999. Multicentric origin of hemochromatosis
38 gene (HFE) mutations. *Am J Hum Genet* 64:1056-1062.
39
40 Totaro A, Grifa A, Carella M, D'Ambrosio L, Valentino M, Roth MP, Borot N, Coppin H, Roetto A, Camaschella
41 C, Gasparini P. 1997. Hereditary hemochromatosis: a HpaI polymorphism within the HLA-H gene. *Mol Cell
Probe* 11:229-230.
42
43 Waheed A, Parkkila S, Zhou XY, Tomatsu S, Tsuchihashi Z, Feder JN, Schatzman RC, Britton RS, Bacon BR, Sly
44 WS. 1997. Hereditary hemochromatosis: effects of C282Y and H63D mutations on association with β 2-
45 microglobulin intracellular processing and cell surface expression of the HFE protein in COS-7 cells. *Proc Natl
Acad Sci USA* 94:12384-12389.
46
47 Wallace DF, Summerville L, Crampton EM, Frazer DM, Anderson GI, Subramaniam VN. 2009. Combined
48 deletion of hfe and transferrin receptor 2 in mice leads to marked dysregulation of hepcidin and iron overload.
49 *hepatol* 50:1992-2000.
50
51 Zhang AS, Davies PS, Carlson HL, Enns CA. 2003. Mechanisms of HFE-induced regulation of iron homeostasis:
52 Insights from the W81A HFE mutation. *Proc Natl Acad Sci USA* 100:9500-9505.
53
54
55
56
57
58
59
60

12 <First Author et al.>

Supp. Table S1. Name, location, and nucleotide positions within the *HFE* gene, and sequences of the oligonucleotide primers used in this study for both cloning experiments and to detect polymorphisms.

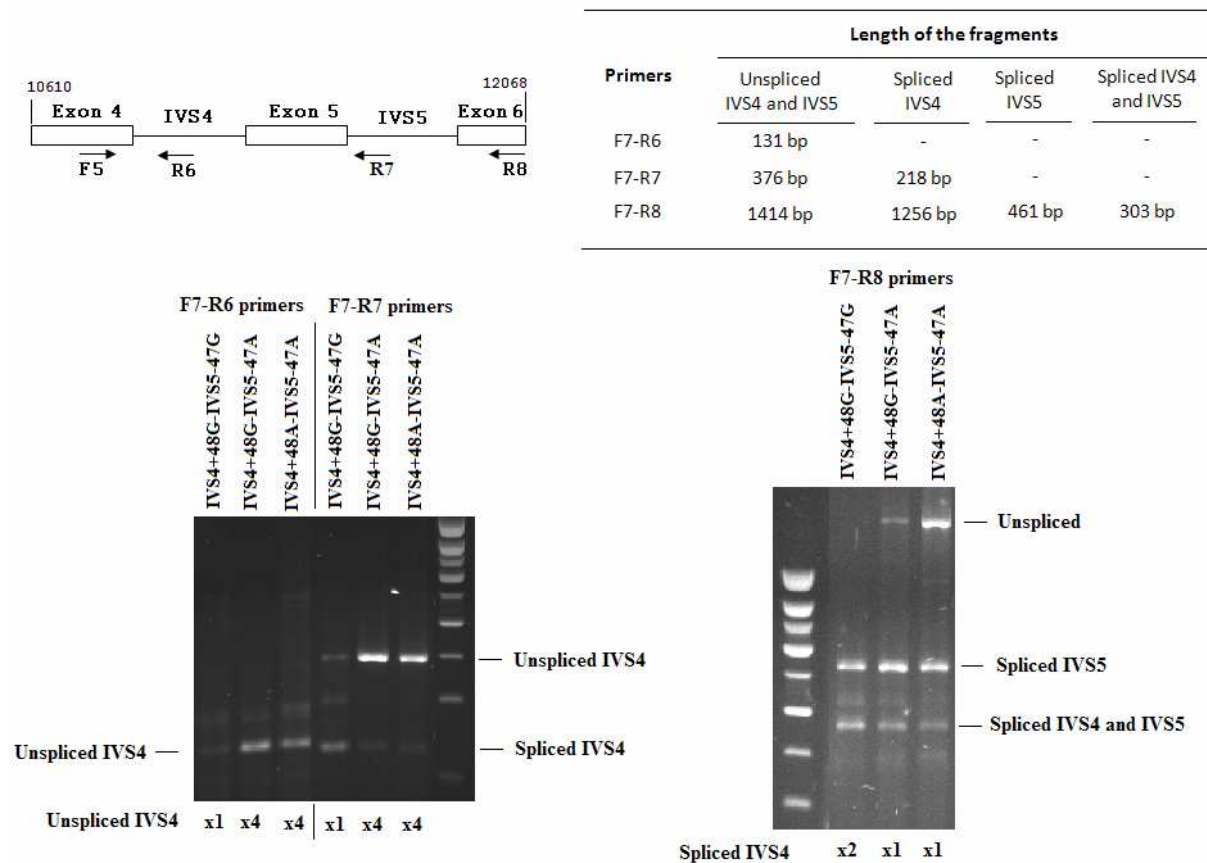
Name	Location	Position	Sequence
F1	5'UTR	Forward 4469-4489	5'-GAATTTAGGTGGGCCAGTGG-3'
F2	Exon 1	Forward 5060-5080	5'-GGCAATAGCTGTAGGGTGACT-3'
		Forward	
F3	Exon1/Exon2	5224-5232/8560-8570	5'-GCTTGCTGCGTTCACACTCT-3'
F4	IVS1	Forward 8529-8548	5'-ACATGGTTAAGGCCTGTTGC-3'
R1	IVS1	Reverse 5398-5417	5'-CTATCTGCAGTTCAGTGGTA-3'
F5	Exon 2	Forward 8719-8739	5'-TTTCAAGCCAGATGTGGCTGC-3'
R2	Exon 2	Reverse 8717-8736	5'-GCCACATCTGGCTTGAAATT-3'
R3	IVS2	Reverse 8849-8868	5'-AAAGCTGTGACAACCTCAGG-3'
R4	Exon 3	Reverse 9079-9098	5'-AGTAGCCCTCGGTACTGTTG-3'
F6	IVS3	Forward 10367-10386	5'-ATCCCTCTCCTCATCTTC-3'
R5	IVS3	Reverse 9320-9339	5'-GTATAGGGGCAGAAAGTGTGT-3'
F7	Exon 4	Forward 10610-10629	5'-GGAAGAGCAGAGATATACGT-3'
R6	IVS4	Reverse 10728-10746	5'-TACCTCCTCAGGCACTCCT-3'
R7	Exon 5	Reverse 10876-10897	5'-GACGACAAAAACAGCAATTCCA-3'
R8	Exon 6	Reverse 12049-12068	5'-AGTTCGTCAGGCAATTTCTA-3'

Reference sequence NG_008720.1:g was used for sequence primer location.

Supp. Table S2. DNA sequence variations of *HFE* and the experimental conditions used for the analysis.

SNP ID	Sequence variation	Location	Primer	Restriction enzyme	Allele fragments
rs2794720	g.4694C>G	5'UTR -467	Forward F1 Reverse R1	<i>Afl</i> III	C: 221 + 728 bp G: 949 bp
rs1799945	c.187 g.8671C>G p.H63D	exon 2	Forward F4 Reverse R2	<i>Bcl</i> I	C: 138 bp + 70 bp G: 208 bp
rs1800730	c.193 g.8677A>T p.S65C	exon2		<i>Hinf</i> I	T: 147 bp + 61 bp A: 208 bp
rs2071303	g.8828T>C c.340+4T>C	IVS2+4	Forward F4 Reverse R5	<i>Rsa</i> I	T: 820 bp C: 297 + 513 bp
rs1800562	c.845 g.10633G>A p.C282Y	Exon 4	Forward F7 Reverse R6	<i>Rsa</i> I	G: 137 bp A: 20 bp + 117 bp
rs1800758	g.10728G>A c.892+48G>A	IVS4+48	Forward F6	<i>Mse</i> I	G :1702 bp A: 357 + 1345 bp
rs1800708	g.10795T>C c.893-44T>C	IVS4-44	Reverse R8	<i>Sau</i> 96I	T: 1702 bp C: 422 + 1280 bp
rs1572982	g.11859G>A c.1007-47G>A	IVS5-47		<i>Ban</i> I	G: 1489 + 213 bp A: 1702 bp

Entrez database (<http://www.ncbi.nlm.nih.gov>) dbSNP ID numbers, reference sequence NG_008720.1:g and NM_00410.3:c for, respectively, nucleotide variations, genomic DNA, and cDNA positions within *HFE*. Names of the primers used for PCR, enzymes, and expected allele fragment sizes in bp. IVS= intervening sequence (intron).



Supp. Figure 1. mRNA splicing assay. To carry out *in vitro* transcription/splicing studies a set of *HFE* minigenes that contained partial genomic sequences from exons 4 to 6 including various combinations of c.892+48G>A; c.893-44T>C; c.1007-47G>A within the pcDNA3.1 mammalian expression vector were constructed. Spliced mRNAs were revealed by RT-PCR using a set of primers (F7 R6 R7 R8) as indicated. The exons and introns (IVS) included in the minigenes are given the first and last nucleotide are numbered according to the NG_008720.1:g database and SNPs according to the NM_000410.3:c database. Forward (F) and reverse (R) primer locations and sequences used are listed in Supp. Table S1. The expected fragment lengths (bp) are listed. Quantitation of RT-PCR products by CCD camera scanning from 3 independent experiments indicated a 4-fold increase in unspliced IVS4 mRNA from the IVS4+48G-IVS5-47A or IVS4+48A-IVS5-47A minigenes compared to the IVS4+48G-IVS5-47G minigene with F7 R6 and F7 R7 primers and a 2-fold increase in spliced IVS4 mRNA from the IVS4+48G-IVS5-47G minigene compared to the IVS4+48G-IVS5-47A or IVS4+48A-IVS5-47A minigenes with F7 R8 primers.