A novel MHC class I–like gene is mutated in patients with hereditary haemochromatosis

Article *in* Nature Genetics · September 1996

DOI: 10.1038/ng0896-399 · Source: PubMed

33 authors, including:

John N Feder
Bristol-Myers Squibb
63 PUBLICATIONS **7,837** CITATIONS
SEE PROFILE

Pete Lauer
University of California, Berkeley
12 PUBLICATIONS **3,650** CITATIONS
SEE PROFILE

Erin E Mcclelland
Middle Tennessee State University
43 PUBLICATIONS **3,784** CITATIONS
SEE PROFILE

Karen Brunke
Cardeus Pharmaceuticals, Inc.
11 PUBLICATIONS **3,458** CITATIONS
SEE PROFILE

A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis

J.N. Feder*, A. Gnirke*, W. Thomas*, Z. Tsuchihashi*, D.A. Ruddy, A. Basava, F. Dormishian, R. Domingo, Jr., M.C. Ellis, A. Fullan, L.M. Hinton, N.L. Jones, B.E. Kimmel, G.S. Kronmal, P. Lauer, V.K. Lee, D.B. Loeb, F.A. Mapa, E. McClelland, N.C. Meyer, G.A. Mintier, N. Moeller, T. Moore, E. Morikang, C.E. Prass, L. Quintana, S.M. Starnes, R.C. Schatzman, K.J. Brunke, D.T. Drayna, N.J. Risch[1], B.R. Bacon[2] & R.K. Wolff

Hereditary haemochromatosis (HH), which affects some 1 in 400 and has an estimated carrier frequency of 1 in 10 individuals of Northern European descent, results in multi-organ dysfunction caused by increased iron deposition, and is treatable if detected early. Using linkage-disequilibrium and full haplotype analysis, we have identified a 250-kilobase region more than 3 megabases telomeric of the major histocompatibility complex (MHC) that is identical-by-descent in 85% of patient chromosomes. Within this region, we have identified a gene related to the MHC class I family, termed *HLA-H*, containing two missense alterations. One of these is predicted to inactivate this class of proteins and was found homozygous in 83% of 178 patients. A role of this gene in haemochromatosis is supported by the frequency and nature of the major mutation and prior studies implicating MHC class I-like proteins in iron metabolism.

Mercator Genetics, Inc., 4040 Campbell Avenue, Menlo Park, California 94025, USA
[1]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA
[2]Department of Internal Medicine, Division of Gastroenterology and Hepatology, Saint Louis University School of Medicine, 3635 Vista at Grand Boulevard, St. Louis, Missouri 63110, USA

** These four authors contributed equally*

Correspondence should be addressed to R.K.W.

Hereditary haemochromatosis (HH) is an autosomal recessive disorder of iron metabolism wherein the body accumulates excess iron. It represents one of the most common inherited diseases among individuals of Northern European descent. Between 1 in 200 and 1 in 400 individuals have HH, leading to an estimated carrier frequency between 1 in 8 and 1 in 10 (refs 1,2). In HH patients, excess iron is deposited in a variety of organs leading to their failure, and resulting in serious illnesses including cirrhosis, hepatomas, diabetes, cardiomyopathy, arthritis and hypogonadotrophic hypogonadism[3–5]. Neither the precise physiologic mechanism of increased iron absorption nor the gene responsible has been described.

The symptoms of HH are often similar to those of other conditions, and the severe effects of the disease usually do not appear until after decades of progressive iron loading. Removal of excess iron by therapeutic phlebotomy decreases morbidity and mortality if instituted early in the course of the disease[6,7]. A combination of blood iron measurements and liver biopsy are used to identify individuals with HH, however, it would be desirable to provide a specific DNA-based test to aid in early detection in order to intervene and prevent organ damage.

The first observation indicating the location of the HH gene near the major histocompatibility complex (MHC) on chromosome 6p was the increased frequency of the *HLA-A3* allele in HH patients[8], a finding that also suggested a founder effect. Subsequent studies demonstrated linkage to the HLA locus, and positioned the HH gene within approximately 1–2 centiMorgans (cM) of the *HLA-A* gene[9,10]. Linkage-disequilibrium studies confirmed the existence of the founder effect[11,12], implying that a large percentage of the affected chromosomes carry the same mutation. However, further refinement of the location of this gene has been difficult. Some studies have implicated an approximately 400-kilobase (kb) candidate region within the MHC[11–13], supported by reports of recombinants within pedigrees that would place the gene centromeric to *HLA-F*[14,15]. However, other recombination events placed the gene telomeric to *HLA-F*[16,17], and more recent linkage disequilibrium studies localized the gene more than 1 megabase (Mb) telomeric to the MHC[18–23]. The search has been confounded by insufficient clone coverage, low numbers of markers, small numbers of recombinants observed in family studies and the large region of linkage disequilibrium where markers far apart appear to show comparable levels of linkage disequilibrium to HH, making fine localization of the HH locus difficult.

In order to identify the HH gene, we set out to characterize the entire HH candidate region, using a linkage-disequilibrium mapping approach with a combination of maximum allelic association and analysis of probable historic recombination events to narrow the candidate region. We now describe a mutation in a novel MHC class I-like gene within this region that accounts for 83% of our HH patients.

Clone coverage and physical mapping

Primary clone coverage of the genomic region telomeric of the MHC was obtained by assembling an overlapping set of yeast artificial chromosome (YAC) clones that cover the region between *D6S265*, a marker near *HLA-A*, and the more telomeric marker, *D6S276*. These boundaries were chosen because they exhibit lower

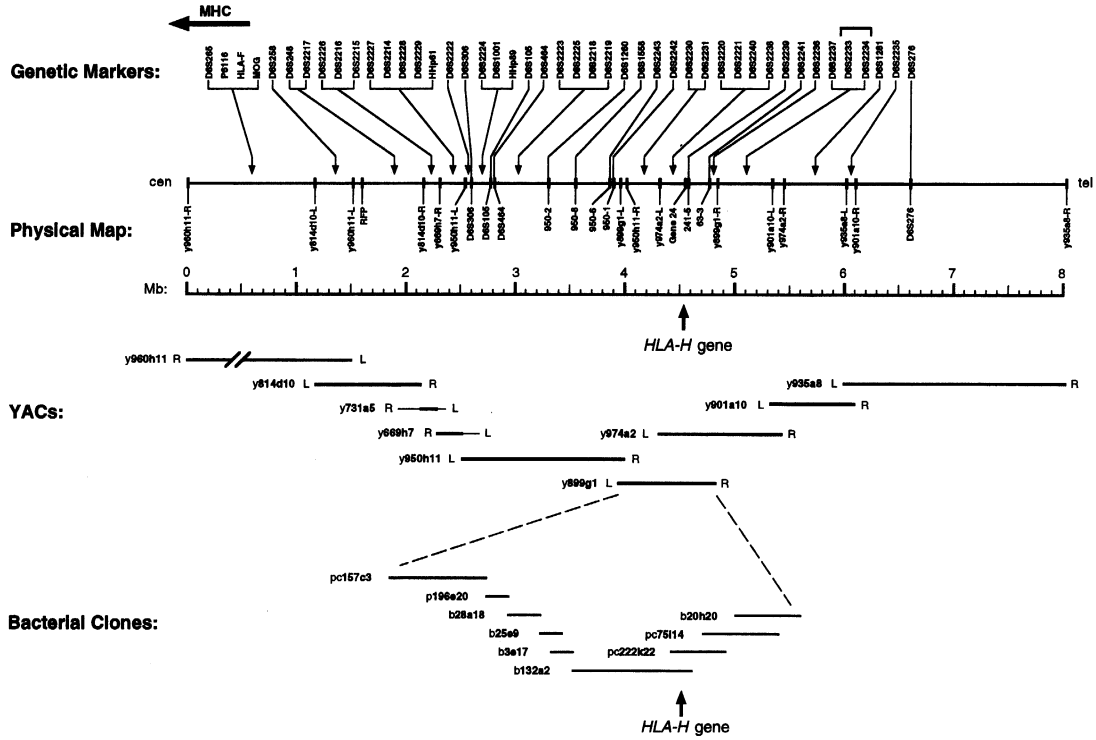

Fig. 1 Physical map of the hereditary haemochromatosis region on human chromosome 6p. The 45 genetic markers used are indicated at the top. The marker order was determined by STS-content mapping and, in some cases, sequencing. The order was unambiguous except for *D6S2233* and *D6S2234*. The left-pointing arrow denotes the distal end of the MHC. The physical map is based on the distance-calibrated YAC-contig shown below. The orientation relative to centromere (cen) and telomere (tel) is indicated. Vertical ticks denote the map positions of clone ends and other landmarks including genetic markers as determined by RARE-cleavage of YACs at adjacent restriction sites. Genetic markers that have not been localized precisely are assigned to their respective map intervals as indicated. YACs are represented as horizontal lines whose lengths are proportional to their size. Thin lines indicate non-contiguous parts of chimaeric YACs or, in the case of y731a5-L, a YAC end whose chromosomal origin has not been determined. L and R represent the centromeric and noncentromeric YAC-vector ends, respectively. HLA-F and MOG are not present in y960h11, indicating a deletion in this YAC. A bacterial clone contig covering approximately the same genomic region as YAC y899g1 is shown at the bottom. The bacterial clones are represented by lines, whose length reflect their respective STS content rather than their physical length. The arrow indicates the map position of the *HLA-H* gene described in the text. YACs, BACs, PACs and P1 clones are indicated by the prefices y, b, pc and p, respectively, followed by their respective library addresses.

allelic association with HH than intervening markers. Initial YAC contigs were seeded by screening the CEPH YAC library for the flanking markers, as well as the intervening sequence-tagged-sites (STSs) *HLA-F, D6S258, D6S306, D6S105* and *D6S464*. Additional YACs containing these STSs were identified in the CEPH/Généthon and MIT/Whitehead databases[24–26]. The three initial YAC contigs were expanded and merged into a single contig by bidirectional walking using STSs developed from the ends of YAC inserts.

An STS-content map comprising 87 STSs and 50 YACs was constructed. Fourteen YACs were selected for analysis by RecA-assisted restriction endonuclease (RARE) cleavage[27, 28] to construct a map that indicates the extent of overlap among the clones and specifies precise map positions of reference landmarks. A minimal set of 9 YACs spanning approximately 8 Mb is depicted in Fig. 1. The physical distance between *D6S265* and *D6S276*, which are genetically less than 1 cM apart[29], is approximately 6 Mb, indicating that this region appears to have a low rate of recombination.

Parallel library screening using YAC-derived STSs and hybridization probes as well as serial chromosome walking were used to assemble a set of large-insert bacterial clones covering the central 3-Mb region extending from the right end of YAC y814d10 to the right end of YAC y899g1. Both YACs and bacterial clones were used to develop short tandem repeat polymorphic markers as well as single-base-pair substitution markers across the entire 6-Mb region (Fig. 1).

Genetic characterization of the HH region

Our approach for defining a minimal region within which to search for the HH gene employed two strategies — identification of the region of maximum linkage disequilibrium by pointwise analysis[30–32] and haplotype analysis to identify historic crossover events.

Linkage disequilibrium. Forty-five markers were used to genotype 101 HH patients from various parts of the United States and 64 controls (see Methods). At each marker, the allele with the highest excess frequency in the patient set when compared to the control set was defined as the ancestral allele, thereby allowing the reconstruction of the ancestral haplotype on which the common HH mutation occurred.

The *P*excess value, a measure of linkage disequilibrium[31, 33], was calculated at each marker and plotted against its respective physical location (Fig. 2*a*). Linkage disequilibrium was found to be high (*P*excess ≥ 0.5) across a 4-Mb region with association falling off at *D6S2217* (*P*excess = 0.44) and *D6S1281* (*P*excess = 0.19). The *P*excess increases towards the telomeric end of this interval, reaching a maximum value of 0.81 at

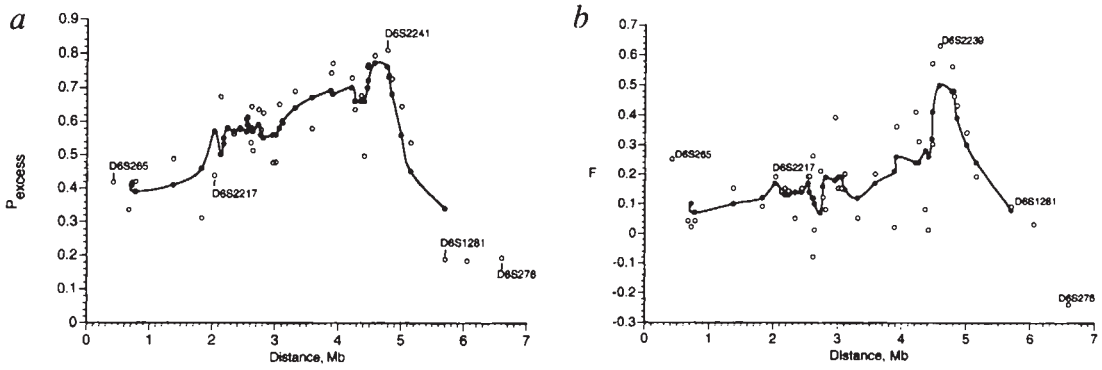

Fig. 2 Mapping the HH region by pointwise linkage-disequilibrium analysis. The marker order and distance scale is identical to that in Fig. 1 with HHp61 and HHp89 omitted. *a*, *P*excess values across the HH region. The linkage disequilibrium between the ancestral allele of polymorphic markers and the HH locus was measured as a *P*excess value which was calculated as described[33]. *b*, Deviation from Hardy-Weinberg equilibrium as measured by F, the proportionate increase of homozygotes over expected. $F = (H_o - H_e)/(1 - H_e)$, where H_o is the observed frequency of homozygotes and H_e is the expected frequency based on Hardy-Weinberg. Open circles represent the actual values. Filled circles represent a smoothed fitted curve based on a five-point moving average.

D6S2241. The peak of a rolling five-marker average of *P*excess values is between *D6S2240* and *D6S2233*, a region of approximately 600 kb (Fig. 2*a*).

The maximum *P*excess suggests that 80% of HH chromosomes carry the ancestral haplotype, implying that approximately 20% do not. If we assume that these remaining chromosomes carry distinct causative mutations in the same gene, because of the recessive nature of the disease, we would expect Hardy-Weinberg proportions for the genotypes AA, AX and XX, where A represents the ancestral mutation and X all other mutations combined. The same is true for marker alleles in linkage disequilibrium with the ancestral mutation. In examining the marker loci of Fig. 1 for Hardy-Weinberg proportions, we noticed a significant deficiency of heterozygotes AX for most of the marker loci (here A is the ancestral allele at the marker locus). While it is possible that other mutations in the HH gene may complement the ancestral mutation, leading to a normal phenotype in these individuals, the most likely explanation for the heterozygote deficiency is heterogeneity — that is, a subset of haemochromatosis is caused by (a) mutation(s) in this region, while other cases are due to either other genetic loci or non-genetic factors. In such circumstances, deviation from Hardy-Weinberg (as measured by homozygote excess) is also a sensitive indicator of the location of the ancestral mutation, because this deviation will diminish with decreasing linkage disequilibrium (see Methods).

To measure the degree of Hardy-Weinberg disequilibrium for each of the marker loci, we calculated the parameter F, which is the proportionate excess of homozygotes (see Methods). Conventionally, F is the inbreeding coefficient; but in this case it represents the proportionate increase in homozygotes due to heterogeneity. Thus, it is maximal at the disease locus itself. F is positive across most of the region but clearly peaks at *D6S2239*, implicating precisely the same location as indicated by *P*excess between *D6S2240* and *D6S2233* (Fig. 2*b*). We note that F is based only on genotype frequencies in the 101 patients and not on control allele frequencies. Thus, it gives statistically independent confirmation of the region implicated by *P*excess (Fig. 2*a*). Furthermore, the excess homozygosity in this region is quite substantial, exceeding 60%, reflecting significant heterogeneity. This finding has important implications for identifying mutations, as described below. However, while comparison of individual *P*excess and F values highlights the same 600-kb region as the most likely location of the HH gene, these two approaches do not define precise boundaries.

Haplotype analysis. As family studies have yielded almost no definition of the gene's precise location, we performed a haplotype analysis of chromosomes bearing the ancestral haplotype in order to identify probable historic recombination events on these chromosomes (identified by stretches of markers with the ancestral allele transitioning to stretches of markers with non-ancestral alleles). We inspected haplotypes on 46 HH chromosomes that had been separated in somatic-cell hybrid lines, allowing unambiguous marker phasing along the chromosome. Because of the previously detected heterogeneity, we performed this analysis in stages. In stage one, we only included individuals with the highest likelihood of being homozygous for the ancestral mutation; these are individuals who carry a substantial sequence of ancestral alleles (8 or more) on both chromosomes (Fig. 3*a*). Inspection of these chromosomes from 15 individuals revealed one region shared among all of them between markers *D6S2221–D6S2241*. This interval of about 400 kb contains three markers, *D6S2240*, *D6S2238* and *D6S2239*; the frequencies for the ancestral alleles for these three loci among controls are 11%, 27% and 41%, respectively. The full haplotype of these alleles together occurred on 2.3% of control chromosomes, which is consistent with the predicted mutant gene frequency. Thus, it is likely that this region is shared identical by descent with the original ancestral HH mutation rather than identical by state.

In stage two, we examined additional individuals likely to be homozygous for the ancestral HH mutation, but with less certainty than for those in Fig. 3*a*. These were individuals whose chromosomes both have a minimum of 5 contiguous loci with ancestral alleles that overlap the 400-kb shared region (Fig. 3*a*). The chromosomes for these three individuals are given in Fig. 3*b*. Four of the six chromosomes overlap completely with the 400-kb region. The two remaining chromosomes both showed ancestral recombinations at *D6S2238*, which thereby represents the centromeric flanking marker. Each of these

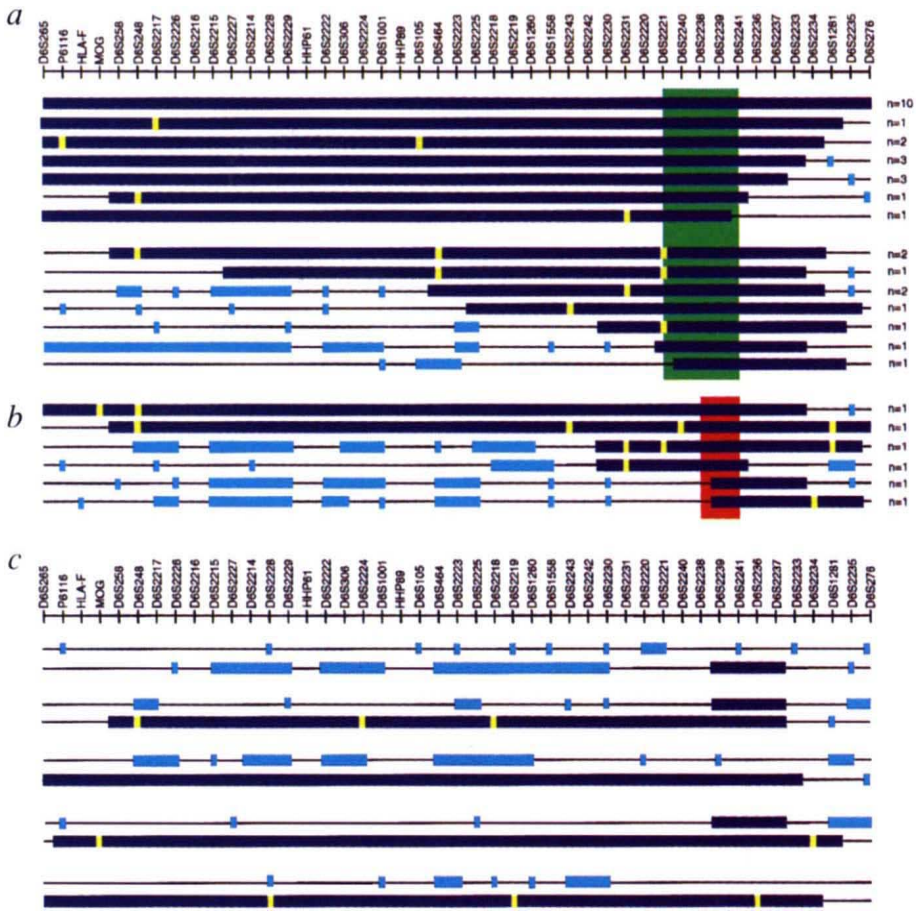

Fig. 3 Haplotype analysis of patient chromosomes. Haplotypes were determined by genotyping patient chromosomes that were separated in somatic-cell hybrid lines. Ancestral alleles that are interpreted as identical by descent are drawn in dark blue. Ancestral alleles believed to be identical by state rather than by descent are indicated by light blue. The presence of a non-ancestral allele is indicated as a black line, unless it was interpreted as being introduced by a mutation, rather than a recombination event, in which case it is drawn in yellow. n indicates the number of chromosomes in each category out of 46 total. The number and position of "yellow" and "light blue" markers varied among the members of each category. Only one representative example from each class is shown. *a*, Haplotypes from patients with the highest likelihood of being homozygous for the ancestral allele were used to define conservative telomeric and centromeric boundaries of the candidate region, depicted in green. This represents the only region in common among all 30 chromosomes. *b*, Chromosomes from individuals carrying at least five contiguous ancestral alleles on both chromosomes overlapping the region defined by the chromosomes in (*a*). The region in common among these chromosomes defines a more narrow candidate region (indicated in red). *c*, Remaining chromosomes from 5 individuals not used in this analysis due to uncertainty as to their homozygous status, shown as homologous pairs.

two chromosomes contains at least five consecutive ancestral alleles. The telomeric flanking marker remains *D6S2241*, based on a chromosome with 37 consecutive ancestral alleles. The interval from *D6S2238* to *D6S2241* is about 250 kb. Although this interval is defined by a small number of chromosomes, the large number of consecutive ancestral alleles makes it highly likely that these chromosomes contain the ancestral mutation.

We did not use the remaining five individuals for fine localization because of uncertainty regarding their homozygosity for the ancestral mutation (Fig. 3*c*). In four of the five cases, one of the chromosomes appears to have a substantial patch of ancestral alleles, while the other chromosome does not. Three of these latter chromosomes have a smaller patch of alleles (4) overlapping the region of interest; however, we did not use them to narrow the region further. Three chromosomes contained no ancestral segment in the region. The region defined by this analysis is consistent with inferred crossovers on an additional 156 unseparated (unphased) chromosomes (data not shown). Taken together, all of our genetic analyses converge on the same region, with haplotype analysis defining a 250-kb candidate region between *D6S2238* and *D6S2241* as the likely location of the HH gene.

Genes within the minimally defined region

We identified potential genes within the 250-kb candidate region by cDNA selection[34], exon trapping[35] and genomic DNA sequencing. cDNA selection was carried out using YAC y899g1 DNA, while the exon-trapping and DNA sequencing utilized the bacterial clones b132a2, pc222k22, pc75l14 and b20h20, spanning the minimal region. We sequenced products from cDNA selection and exon trapping, and localized expressed sequence tags (ESTs) generated from the cDNA selection products within the bacterial clone contig. Selected cDNA fragments and trapped exons derived from the minimal region were used as hybridization probes on testis, liver and brain cDNA libraries to obtain full-length cDNA clones.

The resulting clones were sequenced and evaluated for nucleic acid and protein similarities by BLAST, revealing three novel genes with amino-acid similarities to a 52-kD Ro/SSA ribonucleoprotein[36] ($P = 1.4$ e–115), a sodium-phosphate transporter[37] ($P = 1.5$e–146) and *HLA-A2* (ref. 38) ($P = 8.7$ e–71), respectively. The latter was termed cDNA 24. All three cDNAs contained long open reading frames, poly $(A)^+$ addition signals and poly $(A)^+$ tails, and all were found by all three methods. We undertook complete genomic sequencing of the 250-kb region as a final measure to identify all of the genes. Analysis of this sequence identified the previously discovered three novel genes, as well as 12 histone genes. (As histone genes are not polyadenylated, and have a single exon structure, cDNA selection and exon trapping are not optimal methods to identify them.) In total, 15 genes were discovered within the region. Our comprehensive approach to gene identification, including complete genomic sequencing of the candidate region, makes it unlikely that additional genes remain unidentified.

Mutation analysis in candidate genes

We first set out to identify the putative ancestral HH mutation in genes from the 250-kb region using the following criterion. The ancestral mutation must be common in patients (approaching 80% or higher, as estimated by *P*excess values) and rare in controls (approximately 5%, the estimated allele frequency in the population). To accomplish this, we analysed all 15 genes identified between *D6S2238* and *D6S2241* for sequence variation by comparing two patients who were homozygous for the ancestral haplotype to two control samples. We used either DNA sequencing of RT-PCR fragments amplified from lymphoblastoid RNA, or DNA sequencing of genomic PCR products, to analyse the entire coding region of these genes. Eighteen DNA sequence polymorphisms were found, 15 of which were silent changes that did not affect the amino acid coding of these genes (data not shown).

Two of the 15 genes contained base differences predicted to result in amino acid alterations. The first two changes were in a histone H1 gene: a valine/leucine polymorphism at amino acid 14 (the leucine codon was present on 4/4 ancestral chromosomes and 3/4 control chromosomes) and a glutamine/lysine variant at amino acid 178 (lysine was found on 4/4 ancestral chromosomes and 2/4 control chromosomes). The leucine and lysine alleles were present on 61% and 32% of a larger set of control chromosomes ($n = 28$), respectively. These high frequencies suggest that neither variant is consistent with the ancestral HH mutation.

A cysteine mutation in cDNA 24

The only nucleotide change consistent with the ancestral HH mutation occurs in the MHC class I-like gene, cDNA 24: a G to A transition at nucleotide 845 of the open reading frame that results in a cysteine-to-tyrosine substitution at amino acid 282. All four ancestral patient chromosomes contained a tyrosine codon while all 4 control chromosomes contained a cysteine. This Cys282Tyr missense mutation occurred at a highly conserved residue involved in intramolecular disulphide bridging in MHC class I proteins[39,40], and could therefore disrupt the structure and function of this protein.

Using an allele-specific oligonucleotide-ligation assay on our total of 178 patients, we detected the Cys282Tyr mutation on 85% of all HH chromosomes. In contrast, only 3.2% ($n = 10$) of the control chromosomes ($n = 310$) carried the mutation — a carrier frequency of 10/155 = 6.4%. We found 148 of our 178 patients are homozygous for this mutation, 9 are heterozygotes and 21 carry only the normal allele. These numbers are extremely discrepant from Hardy-Weinberg equilibrium ($\chi^2 = 112.3$, $P < 0.0001$). This is corroborating evidence for heterogeneity among our haemochromatosis patients, with 83% of cases related to Cys282Tyr homozygosity.

All chromosomes with a Cys282Tyr mutation also contained the ancestral haplotype (including all chromosomes in Fig. 3*a* and *b*). Of the 10 chromosomes in Fig. 3*c*, five also carried this mutation. One individual was homozygous for this mutation while one other was homozygous for the normal allele. The correlation between the presence of the Cys282Tyr mutation and the ancestral haplotype was also observed on 3 of 128 control chromosomes.

We next considered more carefully the number of heterozygotes (AX) and homozygotes (XX) at this locus. According to a simple heterogeneity model, where A (= Cys282Tyr) represents the only causative mutation at this locus and the remaining cases are unlinked, the ratio of AX to XX genotypes among patients will be $R = 2pq/q2$, where p is the frequency of the mutation A in the population and $q = 1-p$. R reduces to $2p/q = 2p/1-p$. As we estimated that $p = 0.032$ from our control chromosomes, we derive an expected ratio R of 0.064/0.968 = 0.066. Thus, of a total of 30 non-AA genotypes, we would expect $(0.066/1.066) \times 30 \approx 2$ to be heterozygotes (AX) and the remaining 28 to be homozygotes (XX). These expected numbers also differ significantly from the observed (9 and 21, respectively; $\chi^2 = 26.2$, $P < 0.001$). Thus, there are too many heterozygotes for this model to be accurate, suggesting that some of these heterozygotes carry a second HH mutation.

A second missense variant in cDNA 24

We therefore searched for additional mutations within the MHC class I-like gene on the non-ancestral chromosomes present in the nine Cys282Tyr heterozygotes. We identified a C to G change in exon 2 that results in a histidine to aspartic acid substitution at position 63. This His63Asp variant was present in 8 of the 9 (89%) non-ancestral chromosomes, representing a significant enrichment over the 17% frequency ($n = 51$ out of 308) observed in control chromosomes ($P < 10^{-4}$, Fisher exact test). One patient was homozygous for the His63Asp variant.

Although the His63Asp variant is frequent in Cys282Tyr heterozygous patients, its high frequency on control chromosomes and the relative absence of patients homozygous for this allele raises the possibility that this variant is not directly causative of disease, but in linkage disequilibrium with a rare disease-causing mutation. We therefore examined linkage disequilibrium on chromosomes carrying the His63Asp variant in patients as well as controls. Linkage disequilibrium was not observed for marker loci beyond the candidate region described for ancestral disease chromosomes. The only markers showing significant linkage disequilibrium span *D6S2221* and *D6S2237* (Table 1), with *P*excess values from 0.5–1.0 in both patients and controls. All control chromosomes were phased as were 2 of the 8 patient chromosomes. For the remaining 6 patient chromosomes, alleles were inferred based on the assumption that ancestral alleles were present on the second chromosome. The His63Asp variant appears to have occurred on a single progenitor haplotype common in both patients and controls on which subsequent mutations at markers and/or recombinations have occurred.

The pattern of haplotypes (Table 1) appears similar between the patients and controls, arguing against the patient chromosomes bearing a rare mutation carried by only a subset of these chromosomes. Further evidence against this hypothesis is the fact that multiple versions of this haplotype appear in both patients and controls, an observation inconsistent with a single rare mutation being carried on patient chromosomes. The fact that significant disequilibrium was observed on these chromosomes in patients (that is, 8 out of 9 heterozygotes for the primary ancestral chromosome) in the same region as implicated by the ancestral homozygotes provides addi-

Table 1 Haplotypes of Chromosomes Carrying the His63Asp Allele

D6S2243	D6S2242	D6S230	D6S2231	D6S2220	D6S2221	D6S2240	D6S2236	D6S2239	D6S2241	D6S2236	D6S2237	D6S2233	D6S2234	D6S1281	D6S2235	D6S276
153	146	142	171	159	208	109	107	108	167	141	131	153	109	198	167	210
153	146	144	171	161	208	109	107	108	167	141	131	151	103	200	167	210
151	142	142	161	161	206	109	107	108	167	141	131	147	105	200	167	224
153	146	142	171	159	208	109	107	108	167	141	133	151	109	200	167	224
151	142	144	177	161	208	109	113	108	167	141	131	155	109	104	163	206
153	154	144	171	167	208	109	107	110	167	139	131	149	111	192	169	218
151	142	144	177	167	208	109	107	110	167	139	131	149	107	192	167	210
151	142	144	177	167	210	109	107	110	167	139	131	155	109	192	169	220
163	144	144		159	206	109	107	108	167	141	131	153	109	192	163	226
151	152	140	173	161	210	109	107	108	167	141	131	147	103	192	167	206
		142	171	169	208	109	107	108	167	141	131	151	103	200	167	210
153	148	142	159	159	208	109	107	108	167	141	131	153	109	200		206
155	142	144	159	159	208	109	107	108	167	141	129	153	109	192	163	212
151	142	146	175	169	208	109	107	108	167	141		155	109	200	167	198
151	152	144	171	161	212	109	107	108	167	141	113	153	105	192	167	224
151	154	144	171	169	208	109	107	108	167	141	133	147	109	200	163	210
163	150	142	173	169	208	109	107	108	169	139	117	153	109	200	163	222
155	148	140	159	159	208	109	107	108	171	157	133	145	105	188	163	222
165	142	140	159	163	206	109	107	108	169	161	113	149	103	204	167	224
165	142	140	159	173	204	109	107	108	169	161	113	149	103	208	163	198
153	144	140	159	159	208	109	107	108		141	131	157	103	176	169	222
		140	173	169	206	109	105	108	167	141	131	143	105	196	167	218
161	144	144	173	169	206	109	106	108	167	141		151	111	200	167	210
153	146	144	171	169	204	109	113	108	167	157	113	153	103	200	167	220
153	144		171	161	208	109	105	108	169	141	131	151	105	200	163	214
151	142	146	177	167	210	109	107	110	167	139	131	149	107	204	169	222
	142	144	177	167	210	109	107	110	167	139	131	149		192	163	206
153	142	140	177	159	210	109	105	110	169	139	131	149	107	196	163	

Numbers denote the respective allele sizes in bp. The top 8 rows represent the His63Asp-containing chromosomes from Cys282Tyr/His63Asp compound heterozygous HH patients. The bottom 20 rows represent His63Asp-containing control chromosomes. The shared seven-marker haplotype, or portions thereof, is underlined. Blanks indicate loci where the phase could not be determined.

tional evidence that this region harbours the HH locus.

We next analysed the 21 patients who did not possess the Cys282Tyr mutation on either chromosome. Our previous detection of heterogeneity has significant implications for what to expect on these chromosomes. If the heterogeneity were due to these individuals being affected through an entirely unrelated mechanism, we would not expect to find additional mutations on these chromosomes. Out of the 42 chromosomes studied, 9 carried the His63Asp variant (21%), a frequency similar to control chromosomes. Further analysis of the other non-ancestral chromosomes by sequencing of individual exons and exon/intron boundaries did not identify any additional mutations nor any evidence for linkage disequilibrium, consistent with the scenario that a subset of our cases are unrelated to chromosome 6p. Nonetheless, 87% of our patients are either Cys282Tyr homozygotes or Cys282Tyr/His63Asp compound heterozygotes, providing compelling evidence for this gene being a strong candidate for the HH gene. We have termed this gene *HLA-H* due to its similarity to HLA class I genes.

HLA-H similarity to MHC class I

cDNA 24 is a 2.7-kb clone with an open reading frame of 1,029 bases encoding a predicted protein of 343 amino acids. A consensus polyadenylation signal occurs 15 bases upstream of the poly $(A)^+$ tail. Database comparisons show that the HLA-H protein is most similar to MHC class I molecules including HLA-A2[38] ($P = 8.7e{-}71$) and non-classical class-I-like molecules such as HLA-G[41] ($P = 3.3\ e{-}27$). We also observed similarity with the human Fc receptor[42]. Aligning the amino acid sequence to several types of MHC class I proteins reveals all the structural hallmarks of an MHC class I molecule including signal sequence, peptide-binding region ($\alpha 1$ and $\alpha 2$ domains), an immunoglobulin-like domain ($\alpha 3$), a transmembrane region and a small cytoplasmic portion (Fig. 4). One of the most important conserved structural features of MHC class I molecules in HLA-H are the four cysteine residues that form disulphide bridges in the $\alpha 2$ and $\alpha 3$ domains (Fig. 4; see below). The correct conformation of the $\alpha 3$ domain is required for non-covalent interaction with β_2-microglobulin (β_2M)[43] and correct cell-surface presentation. It is one of these conserved cysteines that is altered in the Cys282Tyr mutation. Mutating the cysteine at the reciprocal position (Cys 203) in the disulphide bond of the mouse H-$2L^D$ eliminates intracellular transport of the protein from the endoplasmic reticulum to the plasma membrane[44]. By analogy, the Cys282Tyr mutation may have the same consequences.

The primary structure of HLA-H resembles the highly polymorphic antigen-presenting HLA-A protein, yet it displays a low degree of polymorphism similar to HLA-G[41]. However, it differs from HLA-G in that of the four critically spaced tyrosine residues in HLA-G that are important for peptide binding[45], HLA-H retains only two, those

Fig. 4 Alignment of the predicted HLA-H protein sequence with HLA-A2, HLA-G and the human neonatal Fc receptor (hFcRn). Gaps were introduced to optimize alignment. Black boxes indicate residues identical to HLA-H, while grey boxes indicate amino acids of similar chemical character. Invariant cysteine residues are marked with an asterisk (*). Cys282Tyr and His63Asp are marked with their respective amino acid changes. Filled circles above the sequence alignment denote the conserved tyrosine residues of HLA-A2 known to function in peptide binding. The proline conserved with the human neonatal Fc receptor and believed to affect the structure of the peptide groove is indicated by a triangle (▲).

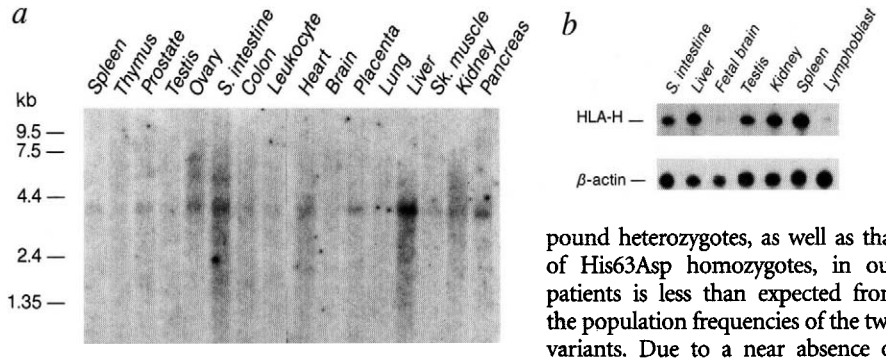

Fig. 5 Expression levels of the *HLA-H* gene. *a*, Northern blot containing poly $(A)^+$ mRNAs from the tissues indicated was hybridized with cDNA 24 as a probe. One major transcript of approximately 4 kb was detected in all tissues except brain. The autoradiography was performed for 2 weeks. *b*, RNase-protection analysis of gene expression performed on total RNA from the tissues indicated. Fetal brain and lymphoblast show considerably less steady-state expression. β-actin-protected products were used as internal controls for the amount of RNA loaded in each lane. Autoradiography was 24 hours and 3 hours for the upper and lower panel, respectively.

equivalent to residues 7 and 159 of HLA-G and HLA-A2 (Fig. 4). In this respect, HLA-H is more similar to the Fc receptor which also retains only two of the four tyrosine residues[46]. An additional similiarity between HLA-H and the Fc receptor is proline 188; this residue introduces a kink in the α2 helix, causing a closure of one side of the peptide-binding cleft[46].

HLA-H gene expression

Comparison of the cDNA 24 sequence with the corresponding genomic sequence indicates that *HLA-H* is comprised of seven exons encompassing 12 kb of DNA. The seventh exon is completely non-coding and contains one poly $(A)^+$ addition signal.

We assayed the level of mRNA by northern blot analysis and RNase protection. Tissue northern blots show that the *HLA-H* gene was widely expressed, albeit at low levels. One major transcript, approximately 4 kb, was detected in all tissues tested except for brain (Fig. 5*a*). RNase protection revealed that the steady-state levels of the transcript vary only slightly from tissue to tissue with the exception of fetal brain and lymphoblast where the levels are considerably less (Fig. 5*b*). Thus, *HLA-H* is expressed in most, if not all, of the tissues believed to be affected by HH.

Discussion

We have used a positional cloning approach to identify a compelling candidate gene for hereditary haemochromatosis. Although we cannot prove formally that the two alterations we have described are not in linkage disequilibrium with another as yet unidentified causative locus, several lines of evidence argue strongly against this. 1) A 250-kb region was highlighted by four different genetic approaches; 2) We identified most if not all of the genes within this region; 3) All of these genes were screened for mutation by DNA sequencing; 4) Only one gene, *HLA-H*, shows a mutation with characteristics consistent with causing disease; 5) This mutation is predicted to have deleterious effects on the function of HLA-H; 6) Independent *in vivo* evidence implicates MHC class I molecules and their cognate partners in iron overload (see below).

Genetics of HH. The Cys282Tyr mutation is present on 85% of the chromosomes in our patients, but only 3.2% of control chromosomes, consistent with frequencies expected for the ancestral HH mutation. Eighty-three percent of our HH patients are homozygous for this mutation. An additional 4% of patients are compound heterozygotes for the Cys282Tyr mutation and a second variant, His63Asp. Interestingly, the frequency of compound heterozygotes, as well as that of His63Asp homozygotes, in our patients is less than expected from the population frequencies of the two variants. Due to a near absence of His63Asp homozygotes among our patients, we conclude that homozygosity for this allele does not increase risk for HH. If we assume that the second variant, His63Asp, increases risk in association with Cys282Tyr, then we can estimate the relative penetrance for the compound heterozygote, Cys282Tyr/His63Asp. In the general population there is a frequency ratio of $(0.032)^2/2(0.032)(0.17) = 1/10.6$ of Cys282Tyr homozygotes to compound heterozygotes. However, in our patients the observed ratio is $148/9 = 16.44$. Therefore, we estimate the relative penetrance of the compound heterozygote versus the Cys282Tyr homozygote to be $1/10.6(16.44) = 0.0057$. Thus, these compound heterozygotes would be at much lower risk than Cys282Tyr homozygotes and hence might present clinically less often. We note, however, that the evidence for causality of the second variant is considerably less than for the Cys282Tyr mutation.

In view of these observations, do patients who do not carry two copies of the Cys282Tyr mutation present a clinically distinct picture? We blindly examined all patients who met the clinical criteria for HH (see Methods), but could find no differences among these groups, suggesting there is no obvious clinical–genetic correlation to the underlying heterogeneity. We cannot exclude the possibility that His63Asp is a polymorphism in linkage disequilibrium with an undiscovered mutation. However, the His63Asp-containing haplotypes in patients and controls look similar, suggesting that such a hypothetical second mutation would have similar penetrance and population frequency as the His63Asp variant.

The relatively high frequency of the ancestral HH mutation suggests an old mutation whereas the large region of linkage disequilibrium is consistent with a more recent mutational event. The most likely explanation for widespread linkage disequilibrium is suppression of recombination in this region, a notion supported by our mapping results that place markers within 1 cM more than 6 Mb apart. Consequently, we had to analyse a large physical region including large portions that had not been surveyed before. The region showing increased linkage disequilibrium and containing the *HLA-H* gene lies more than 500 kb distal to the regions previously considered[47]. The immediate region surrounding *HLA-H* is not represented in the public YAC database[48], probably hampering the search for the HH gene.

The possibility of heterogeneity in HH has been raised before. For example, Simon *et al.*[8] described a family with five affected siblings who carried all four possible HLA haplotypes from their parents, none of which carried the associated HLA-A3 allele. HH is likely not linked to chromosome 6p in this family. Given the small proportion of families believed to be unlinked,

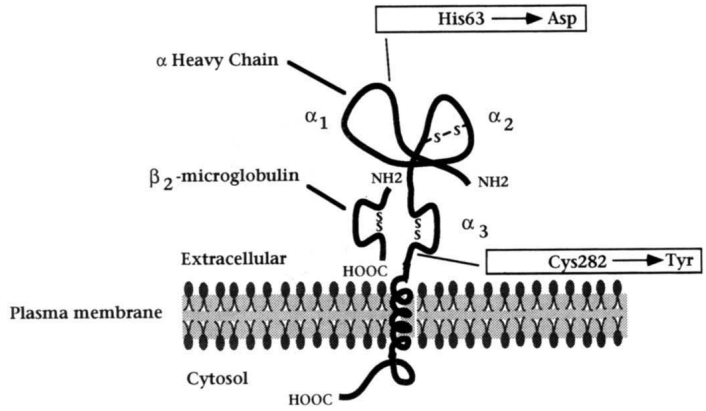

Fig. 6 Hypothetical model of the HLA-H protein based upon its homology with MHC class I molecules. The HLA-H protein is a single polypeptide with three extracellular domains which would be analogous to the α1, α2 and α3 domains of other MHC class I proteins. In contrast to other members of the MHC class I family, the α1 and α2 domains in the HLA-H protein are non-polymorphic. β_2-microglobulin is a separate protein and would interact with the HLA-H gene product in a non-covalent manner in the α3 homologous region. In addition, the HLA-H protein contains a membrane spanning region and a short cytoplasmic tail. The approximate locations of Cys282Tyr and His63Asp are indicated.

power to detect heterogeneity would be low. Our ability to detect heterogeneity relied upon finding marker loci in stronger disequilibrium with disease, as markers further away (for example *HLA-F*) showed much less evidence for deviation from Hardy-Weinberg.

Our analysis featured the use of deviation from Hardy-Weinberg equilibrium (parameter F) to refine the location of the HH locus by linkage disequilibrium mapping, perhaps for the first time. This technique does not require the use of external controls, but is based entirely on patients, providing a statistically independent line of evidence for localizing the HH gene. Can this approach be generalized to map other diseases? Success probably depends on the expected amount of Hardy-Weinberg disequilibrium at the disease locus. This deviation depends on the degree of non-independence of alleles inherited by an affected individual from the two parents. For example, in a homogeneous recessive disease this approach would not be feasible. For a dominant disease, however, it could be quite practical, based on expected heterozygote excess. Similarly, for a heterogeneous recessive disorder as described here, it can also be quite powerful.

Potential biological functions. HLA-H represents a novel MHC class I-like molecule and retains significant features of that family, including the number and spacing of the cysteine residues involved in disulphide bridges. In particular, residues 225 and 282 are equivalent to those in MHC class I molecules at positions 203 and 259 (Fig. 6). This disulphide bridge is required for the non-covalent association of β_2-microglobulin with MHC class I molecules and proper cell-surface presentation[44]. By analogy, the Cys282Tyr mutation in HLA-H would disrupt the formation of the disulphide bridge and likely eliminate cell-surface presentation. The second variant, His63Asp, is located on the loop between the third and fourth β strands of the peptide-binding domain[49]. This variant occurs in an extracellular portion of the molecule, and might disrupt a ligand-receptor interaction.

Independent support for a defective MHC class I-like gene causing HH comes from studies of β_2M-deficient mice[50,51]. These mice display a progressive hepatic iron overload similar to that observed in human HH. As β_2M is known to associate with MHC class I proteins, it has been hypothesized that HH may be caused by a class I-like molecule. The HLA-H protein fulfills this requirement (Fig. 6). In addition, a striking phenotype of the β_2M-deficient mice is the absence of $CD4^-$ $CD8^+$ T cells[52–54]. Interestingly, some HH patients display abnormally low $CD4^-$ $CD8^+$ T cell counts as well[55].

Despite its MHC class I-like characteristics, HLA-H differs from the antigen-presenting members of this family. Only two of four Tyr residues in the peptide-binding region are conserved. In addition, the presence of a proline residue at position 188, similar to that found in the Fc receptor and other non-peptide binding MHC homologues[56,57], also suggests that the protein is not involved in antigen presentation but rather may internalize and/or recycle ligand via receptor-mediated pathways.

How an aberrant MHC class I-like protein could contribute to iron overloading is not obvious. However, any model must take into consideration the major site of iron absorption, the small intestine, and the organs of deposition, such as the liver. The primary defect in intestinal iron absorption of HH patients is probably the transfer of mucosal iron from the basolateral surface of the enterocyte to the plasma[58]. We envision three potential mechanisms for the involvement of the HLA-H protein in iron metabolism. First, the molecule may behave as a receptor for an iron-binding ligand: as the Cys282Tyr mutation appears to inactivate the gene product, this protein may normally act in a negative fashion, limiting or controlling the iron-absorption process. A second possibility is a role in signal transduction, whereby the HLA-H protein senses plasma iron levels and regulates the appropriate genes or gene products which control the rate of iron transfer. Involvement of MHC class I molecules in signal transduction has been reported in the insulin and EGF-receptor systems[59,60]. Lastly, this protein may act more indirectly through association with components of the immune system which influence iron metabolism. The disruption of any of these processes could result in progressive iron loading as seen in HH.

While the arguments for the candidacy of *HLA-H* as the HH gene are compelling, formal proof must await functional studies, which should be facilitated by the similarity of HLA-H to a known class of proteins. The discovery of this novel gene may have important clinical implications in that it contains an easily detectable mutation that identified 83% of patients in our sample with a common but widely underdiagnosed disease for which there is effective available therapy.

Methods

Human subjects and diagnosis. 178 patients diagnosed with haemochromatosis using standard criteria[61] (see below) were collected from 32 different clinical centres across the US. Clinical diagnosis was established by the absence of a known cause of secondary iron overload, and the presence of at least two of the following criteria: a) Hepatic iron concentration (HIC) > 4,500 μg/g dry weight; b) calculated hepatic iron index[5,62] (HII > 2.0); c) increased stainable iron in the liver (Perls' Prussian blue stain, grade 3+ or 4+); and d) > 4 g of iron removed by quantitative phlebotomy (> 16 units of blood)[61]. All patients had a liver biopsy performed. Blood samples were obtained with informed consent, using a protocol approved by the Committee for Investigation in Humans at Saint Louis University School of Medicine. Initially,

101 patient samples were used in genotyping studies across the entire region. An additional 77 patients were used for the determination of mutation frequencies. Control subjects were 64 Caucasian members of the grandparental generation in the CEPH collection. Additional control samples for the determination of mutation frequencies were 92 random Caucasian individuals. Genomic DNA was prepared as described[63]. Immortalized cell lines were created according to standard techniques. Hybrid cell lines containing chromosome 6 were produced as described[64].

Clone coverage and physical mapping. YACs were identified by PCR-based screening of plates 613–984 of the CEPH YAC library (purchased from Research Genetics), or by searching the CEPH/Généthon and MIT/Whitehead databases[24–26]. At least 4 different clone isolates were analysed from each positive library address, and the largest YAC containing the STS of interest was studied further. YAC-insert ends were isolated using a ligation-mediated PCR method[65]. PCR products were sequenced for the development of YAC-end specific PCR and RARE cleavage assays[27, 28]. The chromosomal origin of YAC-end-specific STSs was determined using the NIGMS somatic-cell-hybrid-mapping panel 2 (Coriell Institute, Camden, NJ). Large-insert bacterial clones were identified by PCR-based or hybridization-based screening of commercially available human BAC[66] (Research Genetics), PAC[67] or P1[68] (Genome Systems) libraries. Insert ends of bacterial clones were isolated by a vector-specific modification of the ligation-mediated PCR method[65] or a conventional digestion/re-ligation strategy.

Genotyping. Newly developed polymorphic markers (*D6S2214* through *D6S2243*) have been deposited with the Genome Data Base (GDB). These (CA)-repeat markers were identified through direct sequencing of entire cosmid clones, or sequencing of (CA)-containing subclones from P1, BAC or YAC clones. The subclones were identified by hybridization essentially as described[69]. The markers *D6S248, GATA–P6116, D6S265, HLA–F*[70], *MOG, D6S258, D6S306, D6S1001, D6S105, D6S464, D6S1260, D6S1558, D6S1281* and *D6S276* are available from GDB and were used as described. All markers were assayed by standard procedures[71]. The single-basepair-substitution markers HHp61 and HHp89 were identified by sequence comparison between cosmids from an affected individual who was homozygous for the ancestral haplotype and P1 clones from a random individual.

Fine localization by Hardy-Weinberg disequilibrium. We performed fine localization of the disease locus using deviation from Hardy-Weinberg equilibrium among the patients. This analysis utilizes linkage disequilibrium to localize the disease gene, analogous to the approach using *P*excess. However, this analysis uses genotype information from patients only, whereas the *P*excess analysis uses allele frequencies from patients and controls. Hence, this analysis provides statistically independent evidence for the location of the HH gene.

The theoretical justification for using homozygote excess to localize the disease gene is as follows. Let A represent the disease locus, with alleles A and a, which have frequencies of p and q = 1–p, respectively. Let F be the excess homozygosity over expected among patients at locus A. F can be given by formula $(H_o–H_e)/(1–H_e)$, where H_o is the observed frequency of homozygotes and H_e is the expected frequency assuming Hardy-Weinberg equilibrium. The value of F at locus A depends on the precise mode of inheritance and allele frequencies. For example, F is large for a heterogeneous recessive disease, but 0 for a homogeneous recessive disease.

Let B represent a linked locus with alleles B (associated with A) and b, whose frequencies are r and s = 1–r, respectively. For this analysis, b can represent all non-associated alleles combined. Let D represent the usual linkage disequilibrium parameter, defined by P(AB)–pr, where P(AB) is the frequency of haplotype AB. Then it is straightforward to show that the excess homozygosity at locus B, F_B, is given by $F_B = (D^2/pqrs)F = \Delta^2F$, where F, again, is the excess homozygosity at disease locus A, and Δ is simply the correlation coefficient in the 2 × 2 table of haplotype frequencies for loci A and B. Thus, F_B is proportional to Δ^2, which is a function of $(1–\theta)^{2g}$, where g is the number of generations since the founding of the mutation. Therefore, F_B decays approximately at a rate of $2g\theta$ with recombination distance θ between A and B. Thus, F_B can be used as a measure of distance from marker locus B to the disease locus A, similar to *P*excess.

Analysis of expressed sequences. cDNA selection experiments were carried out on cDNA from fetal brain, intestine and liver poly $(A)^+$ RNA (Clontech) and gel-purified YAC DNA as described[72], except that the resulting cDNA selection products were cloned using uracil DNA glycosylase (Gibco-BRL) and a version of pSP72 (Promega) constructed to accept PCR fragments containing uracil residues. Exon-trapping was carried out using a modification of Gibco-BRL's exon-trapping kit. 72 clones were sequenced for each trapped genomic bacterial clone, and sequence comparisons performed using BLAST (NCBI). cDNA libraries were obtained from Gibco-BRL and screened by hybridization using standard procedures. Northern blots were purchased from Clontech. RNase protection assays were carried out according to the kit manufacturer's (Ambion) protocol. Run-off transcripts used for RNase protection were made from a genomic clone containing the 120-bp exon 5.

Mutation detection. Mutation detection was performed by fluorescent sequencing of either genomic PCR or RT-PCR products. Poly $(A)^+$ RNA for RT-PCR was prepared from lymphoblastoid cell lines using Fasttrack (Invitrogen), and first-strand cDNA was synthesized by SuperScript reverse transcriptase (Gibco-BRL) primed with a mixture of random primers and oligo (dT). PCR products were gel-purified using Gelase (Epicentre) before being sequenced.

Oligonucleotide ligation assay (OLA). Genotyping by OLA was performed on PCR-amplified genomic DNA samples as described[73]. The PCR primers for amplification of the Cys282Tyr locus were 5'–TGGCAAGGGTAAACAGATCC and 5'–CTCAGGCACTCCTCTCAACC. The oligonucleotide specific for OLA detection of the normal and mutant alleles were 5'–biotin–GGAAGAGCAGAGATATACGTG and 5'–biotin–GGAAGAGCAGAGATATACGTA, respectively. The common 3' oligonucleotide was 5'–phosphate–CCAGGTGGAGCACCCAGG–3'–digoxigenin. The corresponding OLA oligonucleotides used to discriminate between normal and His63Asp alleles were 5'–ACATGGTTAAGGCCTGTTGC (PCR), 5'–GCCACATCTGGCTTGAAATT (PCR), 5'–biotin–AGCTGTTCGTGTTCTATGATC (normal allele), 5'–biotin–AGCTGTTCGTGTTCTATGATG (His63Asp allele) and 5'–phosphate–ATGAGAGTCGCCGTGTGGA–3'–digoxigenin (common 3'–oligonucleotide).

Genomic sequencing. Sequencing of 3-kb plasmid subclones from large-insert bacterial clones was performed using a modification of the transposon-insertion method[74]. The minimal overlapping 3-kb clone path was determined by comparison of end sequences using a strategy similar to those described[75,76].

GenBank accession number. The *HLA-H* cDNA and deduced protein sequence have been deposited in GenBank (U60319).

Acknowledgements

This work would not have been possible without the collaboration of numerous patients and their physicians. We thank E. Avecilla for excellent technical assistance; D. Burris and L. Kendig for help in setting up a patient database; C. Harmon, G. Hartzell and J. Tisdall for computer support; P. Bjorkman for helpful discussions; and D. Cox, R. Myers and E. Sigal for continuous support and guidance.

Received 3 June; accepted 10 July 1996.

1. Dadone, M.M., Kushner, J.P., Edwards, C.Q., Bishop, D.T. & Skolnick, M.H. Hereditary hemochromatosis: analysis of laboratory expression of the disease by genotype in 18 pedigrees. *Am. J. Clin. Pathol.* **78**, 196–207 (1982).
2. Edwards, C.Q. *et al.* Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *N. Engl. J. Med.* **318**, 1355–1362 (1988).
3. McKusick, V.A. Mendelian inheritance in man (The Johns Hopkins University Press, Baltimore, 1994).
4. Bothwell, T.H., Charlton, R.W. & Motulski, A.G. in The metabolic and molecular bases of inherited disease (eds. Scriver, C.R., Beaudet, A.L., Sly, W.S. & Valle, D.) 2237–2269 (McGraw-Hill, New York, 1995).
5. Bacon, B.R. & Tavill, A.S. in Hepatology. A textbook of liver disease (eds. Zakim, D. & Boyer, T.D.) 1439–1472 (W.B. Saunders, Philadelphia, 1996).
6. Bomford, A. & Williams, R. Long-term results of venesection therapy in idiopathic haemochromatosis. *Quart. J. Med.* **45**, 611–623 (1976).
7. Milder, M.S., Cook, J.D., Stray, S. & Finch, C.A. Idiopathic hemochromatosis, an interim report. *Medicine* **59**, 34–49 (1980).
8. Simon, M., Bourel, M., Fauchet, R. & Genetet, B. Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis. *Gut* **17**, 332–334 (1976).
9. Jazwinska, E.C., Lee, S.C., Webb, S.I., Halliday, J.W. & Powell, L.W. Localization of the hemochromatosis gene close to D6S105. *Am. J. Hum. Genet.* **53**, 347–352 (1993).
10. Lalouel, J.M. *et al.* Genetic analysis of idiopathic hemochromatosis using both qualitative (disease status) and quantitative (serum iron) information. *Am. J. Hum. Genet.* **37**, 700–718 (1985).
11. Boretto, J. *et al.* Anonymous markers located on chromosome 6 in the HLA-A class I region: allelic distribution in genetic haemochromatosis. *Hum. Genet.* **89**, 33–36 (1992).
12. Yaouanq, J. *et al.* Anonymous marker loci within 400 kb of HLA-A generate haplotypes in linkage disequilibrium with the hemochromatosis gene (HFE). *Am. J. Hum. Genet.* **54**, 252–263 (1994).
13. Totaro, A. *et al.* New polymorphisms and markers in the HLA class I region: relevance to hereditary hemochromatosis (HFE). *Hum. Genet.* **95**, 429–434 (1995).
14. Edwards, C.Q., Griffen, L.M., Dadone, M.M., Skolnick, M.H. & Kushner, J.P. Mapping the locus for hereditary hemochromatosis: localization between HLA-B and HLA-A. *Am. J. Hum. Genet.* **38**, 805–811 (1986).
15. Gasparini, P. *et al.* Linkage analysis of 6p21 polymorphic markers and the hereditary hemochromatosis: localization of the gene centromeric to HLA-F. *Hum. Mol. Genet.* **2**, 571–576 (1993).
16. Powell, L.W. *et al.* Expression of hemochromatosis in homozygous subjects. *Gastroenterology* **98**, 1625–1632 (1990).
17. Calandro, L.M., Baer, D.M. & Sensabaugh, G.F. Characterization of a recombinant that locates the hereditary hemochromatosis gene telomeric to HLA-F. *Hum. Genet.* **96**, 339–342 (1995).
18. Jazwinska, E.C. *et al.* Haplotype analysis in australian hemochromatosis patients: evidence for a predominant ancestral haplotype exclusively associated with hemochromatosis. *Am. J. Hum. Genet.* **56**, 428–433 (1995).
19. Raha-Chowdhury, R. *et al.* New polymorphic microsatellite markers place the haemochromatosis gene telomeric to D6S105. *Hum. Mol. Genet.* **4**, 1869–1874 (1995).
20. Stone, C. *et al.* Isolation of CA dinucleotide repeats close to D6S105; linkage disequilibrium with haemochromatosis. *Hum. Mol. Genet.* **3**, 2043–2046 (1994).
21. Seese, N.K. *et al.* Localization of the hemochromatosis gene: linkage disequilibrium analysis using an American patient collection. *Blood Cells Mol. Dis.* **22**, 36–46 (1996).
22. Worwood, M. *et al.* Alleles at D6S265 and D6S105 define a haemochromatosis-specific genotype. *Brit. J. Haematol.* **86**, 863–866 (1994).
23. Gandon, G. *et al.* Linkage disequilibrium and extended haplotypes in the HLA-A to D6S105 region: implications for mapping the hemochromatosis gene (HFE). *Hum. Genet.* **97**, 103–113 (1996).
24. Cohen, D., Chumakov, I. & Weissenbach, J. A first-generation physical map of the human genome. *Nature* **366**, 698–701 (1993).
25. Chumakov, I.M. *et al.* A YAC contig map of the human genome. *Nature* **377** Suppl., 175–183 (1995).
26. Hudson, T. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954 (1995).
27. Ferrin, L.J. & Camerini-Otero, R.D. Selective cleavage of human DNA: recA-assisted restriction endonuclease (RARE) cleavage. *Science* **254**, 1494–1497 (1991).
28. Gnirke, A., Iadonato, S.P., Kwok, P.-Y. & Olson, M.V. Physical calibration of yeast artificial chromosome contig maps by recA-assisted restriction endonuclase (RARE) cleavage. *Genomics* **24**, 199–210 (1994).
29. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
30. Kerem, B.-T. *et al.* Identification of the cystic fibrosis gene: genetic analysis. *Science* **245**, 1073–1080 (1989).
31. Hästbacka, J. *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* **78**, 1073–1087 (1994).
32. Lehesjoki, A.-E. *et al.* Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum. Mol. Genet.* **2**, 1229–1234 (1993).
33. Bengtsson, B.O. & Thomson, G. Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* **18**, 356–363 (1981).
34. Lovett, M., Kere, J. & Hinton, L.M. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc. Natl. Acad. Sci. USA* **88**, 9628–9632 (1991).
35. Church, D.M. *et al.* Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**, 98–105 (1994).
36. Itoh, K., Itoh, Y. & Frank, M.B. Protein heterogeneity in the human Ro/SSA ribonucleoproteins: the 52- and 60-kD Ro/SSA autoantigens are encoded by separate genes. *J. Clin. Invest.* **87**, 177–186 (1991).
37. Chong, S.S., Kristjansson, K., Zoghbi, H.Y. & Hughes, M.R. Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3–p23. *Genomics* **18**, 355–359 (1993).
38. Selavakumar, A. *et al.* A novel subtype of A2 (A*0217) isolated from the South American Indian B-cell line AMALA. *Tissue Ant.* **45**, 343–347 (1995).
39. Orr, H.T., Lopez de Castro, J.A., Parham, P., Ploegh, H.L. & Strominger, J.L. Comparison of amino acid sequences of two human histocompatibility antigens, HLA-A2 and HLA-B7: location of putative alloantigenic sites. *Proc. Natl. Acad. Sci. USA* **76**, 4395–4399 (1979).
40. Hood, L., Kronenberg, M. & Hunkapiller, T. T cell antigen receptors and the immunoglobulin supergene family. *Cell* **40**, 225–229 (1985).
41. Stroynowski, I. Molecules related to class-I major histocompatibility complex antigens. *Ann. Rev. Immunol.* **8**, 501–530 (1990).
42. Simister, N.E. & Mostov, K.E. An Fc receptor structurally related to MHC class I antigens. *Nature* **337**, 184–187 (1989).
43. Bjorkman, P.J. & Parham, P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Ann. Rev. Biochem.* **59**, 253–288 (1990).
44. Miyazaki, J.-I., Apella, E. & Ozato, K. Intracellular transport blockade caused by disruption of the disulfide bridge in the third external domain of major histocompatibility complex class I antigen. *Proc. Natl. Acad. Sci. USA* **83**, 757–761 (1986).
45. Madden, D.R., Gorga, J.C., Strominger, J.L. & Wiley, D.C. The three-dimensional structure of HLA-B27 at 2.1 Å resolution suggests a general mechanism for tight peptide binding to MHC. *Cell* **70**, 1035–1048 (1992).
46. Burmeister, W.P., Gastinel, L.N., Simister, N.E., Blum, M.L. & Bjorkman, P.J. Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor. *Nature* **372**, 336–343 (1994).
47. Burt, M.J., Smit, D.J., Pyper, W.R., Powell, L.W. & Jazwinska, E.C. A 4.5-Megabase YAC contig and physical map over the haemochromatosis gene region. *Genomics* **33**, 153–158 (1996).
48. Hudson, T. *et al.* An STS-based map of the human genome. *Science* **270**, 1945–1954, with supplementary data from the Whitehead Institute/MIT Center for Genome Research, human genetic mapping project, data release 10 (May 1996) (1995).
49. Bjorkman, P.J. *et al.* Structure of the human class I histocompatibility antigen, HLA-A2. *Nature* **329**, 506–512 (1987).
50. De Sousa, M. *et al.* Iron overload in β_2-microglobulin-deficient mice. *Immunol. Lett.* **39**, 105–111 (1994).
51. Rothenberg, B.E. & Voland, J.R. β_2 knockout mice develop parenchymal iron overload: a putative role for class I genes of the major histocompatibility complex in iron metabolism. *Proc. Natl. Acad. Sci. USA* **93**, 1529–1534 (1996).
52. Zijlstra, M. *et al.* β_2-microglobulin deficient mice lack CD4$^-$8$^+$ cytolytic T cells. *Nature* **344**, 742–746 (1990).
53. Koller, B.H., Marrack, P., Kappler, J.W. & Smithies, O. Normal development of mice deficient in β_2M, MHC class I proteins, and CD8$^+$ T cells. *Science* **248**, 1227–1230 (1990).
54. Raulet, D.H. MHC class I-deficient mice. *Adv. Immun.* **55**, 381–421 (1994).
55. Arosa, F.A. *et al.* Decreased CD8–p56lck activity in peripheral blood T-lymphocytes from patients with hereditary hemochromatosis. *Scand. J. Immunol.* **39**, 426–432 (1994).
56. Story, C.M., Mikulska, J. & Simister, N.E. A major histocompatibility complex class I-like Fc receptor cloned from human placenta. *J. Exp. Med.* **180**, 2377–2381 (1994).
57. Ueyama, H., Deng, H.X. & Ohkubo, I. Molecular cloning and chromosomal assignment of the gene for human Zn-α_2-glycoprotein. *Biochemistry* **32**, 12968–12976 (1993).
58. McLaren, G.D., Nathanson, M.H., Jacobs, A., Trevett, D. & Thomson, W. Regulation of intestinal iron absorption and mucosal iron kinetics in hereditary hemochromatosis. *J. Lab. Clin. Med.* **117**, 390–401 (1991).
59. Schreiber, A.B., Schlessinger, J. & Edidin, M. Interaction between major histocompatibility complex antigens and epidermal growth factor receptors on human cells. *J. Cell Biol.* **98**, 725–731 (1984).
60. Verland, S. *et al.* Specific molecular interaction between the insulin receptor and a D product of MHC class 1. *J. Immun.* **143**, 945–951 (1989).
61. Crawford, D.H.G. *et al.* Evidence that the ancestral haplotype in Australian hemochromatosis patients may be associated with a common mutation in the gene. *Am. J. Hum. Genet.* **57**, 362–367 (1995).
62. Bassett, M.L., Halliday, J.W. & Powell, L.W. Value of hepatic iron measurements in early hemochromatosis and determination of the critical iron level associated with fibrosis. *Hepatology* **6**, 24–29 (1986).
63. Wolff, R.K. *et al.* Analysis of chromosome 22 deletions in neurofibromatosis type 2-related tumors. *Am. J. Hum. Genet.* **51**, 478–485 (1992).
64. Jackson, C.L. in *Current protocols in human genetics* (eds Dracopoli, N.C. *et al.*) (J. Wiley and Sons, New York, 1994).
65. Riley, J. *et al.* A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18**, 2887–2890 (1990).
66. Shizuya, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**, 8794–8797 (1992).
67. Ioannou, P.A. *et al.* A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84–89 (1994).
68. Shepherd, N.S. *et al.* Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl. Acad. Sci. USA* **91**, 2629–2633 (1994).
69. Hudson, T.J. in *Current protocols in human genetics* (eds Dracopoli, N.C. *et al.*) (J. Wiley and Sons, New York, 1994).
70. Fullan, A. & Thomas, W. A polymorphic dinucleotide repeat at the human HLA-F locus. *Hum. Mol. Genet.* **3**, 2266 (1994).
71. Weber, J.L. & May, P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
72. Morgan, J.G., Dolganov, G.M., Robbins, S.E., Hinton, L.M. & Lovett, M. The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucl. Acids Res.* **20**, 5173–5179 (1992).
73. Delahunty, C., Ankener, W., Deng, Q., Eng, J. & Nickerson, D.A. Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay. *Am. J. Hum. Genet.* **58**, 1239–1246 (1996).
74. Strathmann, M. *et al.* Transposon-facilitated DNA sequencing. *Proc. Natl. Acad. Sci. USA* **88**, 1247–1250 (1991).
75. Chen, E.Y., Schlessinger, D. & Kere, J. Ordered shotgun sequencing, a strategy for integrated mapping and sequencing of YAC clones. *Genomics* **17**, 651–656 (1993).
76. Roach, J.C., Boysen, C., Wang, K. & Hood, L. Pairwise end sequencing: a unified approach to genomic mapping and sequencing. *Genomics* **26**, 345–353 (1995).