

## Commentary

# Haemochromatosis: a gene at last?

Simon *et al*<sup>1</sup> in 1976 first reported the association between the leucocyte antigens HLA-A3 and HLA-B14 and genetic haemochromatosis (GH). Now, 20 years later, Feder *et al*<sup>2</sup> report the cloning of what is the strongest candidate gene for this disorder to have been identified so far. Formal proof from functional studies is now required to confirm that mutations in this gene are responsible for causing haemochromatosis. The question now arises as to why this gene has proved so elusive.

(J Med Genet 1997;34:148-151)

Keywords: haemochromatosis; genetic mapping.

### Background

Although GH may be perceived by many to be of limited clinical importance,<sup>3</sup> prevalence rates for primary iron overload, as determined by measurement of serum iron indices and by liver biopsy, suggest that GH is the commonest single gene disorder in caucasian populations, affecting as many as 1 in 300 people.<sup>4</sup> Progressive accumulation of dietary iron results in an age related clinical presentation, with secondary end organ damage such as cirrhosis, diabetes, and cardiomyopathy apparent in those patients between the ages of 40 and 60.<sup>3</sup> Diagnosis is also compounded by the non-specific symptoms resulting from iron overload, such that serum iron measurements are not automatically considered, which would otherwise confirm the diagnosis of GH. Early detection and treatment by repeated venesection leads to normal life expectancy.<sup>5,6</sup> Until now there has been no reliable, cheap, and easy method for detecting people with GH in the early asymptomatic phase of the disease before the onset of iron overload. The possibility of screening the adult population for GH using a DNA based assay is appealing.

There was no evidence to implicate iron storage (ferritin), transport (transferrin, transferrin receptor), or control (iron responsive) proteins or their genes as candidate genes for GH as these map to regions outside the region of linkage to the major histocompatibility complex (MHC).

### Genetics

Major contributions to the genetics of haemochromatosis have come from the group in Rennes. They were the first to show linkage to the MHC and subsequently proposed a founder effect as many of the affected chromosomes carried the HLA-A3 allele.<sup>7</sup> They positioned the haemochromatosis gene (HFE) within 1 cM of the HLA-A gene. HLA-A is at

the telomeric end of the MHC which maps to 6p21.3 (fig 1). Progress in defining the critical region was slow because of the lack of informative chromosomal translocations and recombinations, and the lack of ordered polymorphic markers telomeric to HLA-A. Analyses of markers around HLA-A and HLA-F (250 kb telomeric of HLA-A) was originally interpreted as suggesting that the haemochromatosis gene lay within this region.<sup>8,9</sup> A number of new genes were identified from this region, but none has proved to be the haemochromatosis gene.<sup>10</sup>

### Fine mapping

The critical interval for the haemochromatosis gene expanded substantially with the analysis of the microsatellite D6S105,<sup>11</sup> which mapped 1-2 cM telomeric to HLA-A.<sup>12,13</sup> The marker D6S105 is now known to be 2.5 Mb telomeric of HLA-A.<sup>14,15</sup> An informative recombination supported this more telomeric position for the haemochromatosis gene by firstly separating the gene from HLA-F<sup>16</sup> and more recently placing this recombinational breakpoint telomeric to the Ret-Finger Protein locus (that is, more than 1 Mb telomeric of HLA-A and within 1.5 Mb of D6S105).<sup>17</sup> Lack of recombination in this region hindered the physical ordering of microsatellite markers which were known to map here. The position of the gene relative to D6S105 was unknown and the telomeric limit was considered to be the marker D6S299 which mapped 2 cM telomeric of D6S105.<sup>18</sup> Application of a new microsatellite marker D6S1260<sup>19</sup> suggested that the gene mapped telomeric to D6S105 and possibly telomeric to D6S1260. This work confounded those who still believed that the haemochromatosis gene would be found closer to HLA-A than to D6S105 as it moved the map location of the gene at least 3 Mb telomeric of the original candidate search area.

As no positional candidate for the disease had been identified and the cellular phenotype was unknown, positional cloning was the most profitable route to identifying the gene for haemochromatosis. For this approach to prove successful the region containing the gene needed to be cloned and mapped in detail. One of the main stumbling blocks in identifying the haemochromatosis gene has been that the further it mapped from the MHC the less was known about markers and YAC clones for the region. Feder *et al*,<sup>2</sup> from the biotechnology company Mercator, began by generating a YAC contig that spanned the region from HLA-A to D6S276 (a microsatellite marker

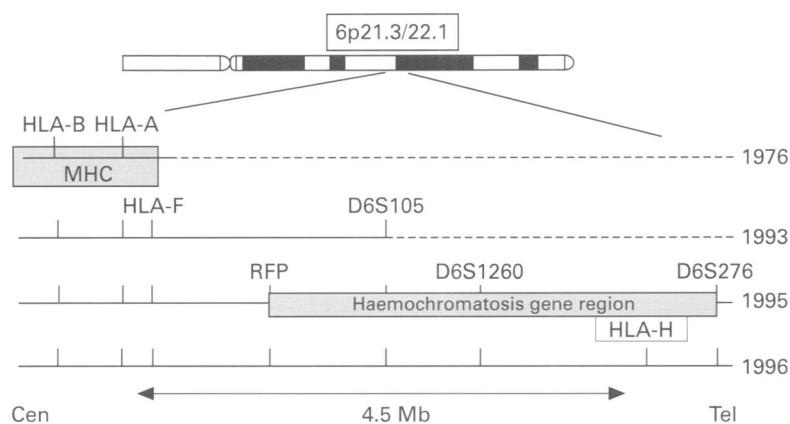


Figure 1 A historical perspective on the position of the gene responsible for GH.

known to map telomeric to D6S1260 and close to the microsatellite D6S299) in order to cover the whole candidate region. The genetic distance between HLA-A and D6S276 is 1-2 cM but the physical distance is 6 Mb, confirming earlier observations that this region of chromosome 6 appears to be subject to low rates of recombination. Of the YACs in the public database mapping to the region, it is worth noting that none contains the candidate gene. The major YAC for the region (947f6) spans the gene locus and is now recognised as being deleted for at least the 250 kb studied by Mercator. Comparing maps, the YAC 947f6 probably contains an internal deletion of the order of 1 Mb.

### Finding the gene

Having generated a YAC contig for the region from HLA-A to D6S276 the next step was to use these YACs together with bacterial clones (BACs, PACs, and cosmids) from the region to develop short tandem repeat polymorphic markers across the region for linkage disequilibrium studies.<sup>2</sup> This approach had previously been used in the identification of the genes for diastrophic dysplasia and myoclonic epilepsy.<sup>20-22</sup> The region had to be saturated with ordered markers. Until this point few groups had specifically isolated new markers with which to refine the position of the gene. Mercator isolated a total of 29 new microsatellites, more than doubling the number of informative markers in the region. With these they were able to identify a peak of linkage disequilibrium. The maximum value for  $p_{\text{excess}}$ , a measure of linkage disequilibrium, was 0.81, suggesting that over 80% of disease bearing chromosomes carry a common mutation. The microsatellite giving this maximum value, D6S2239, mapped over 1 Mb telomeric of D6S1260, the marker which had previously been shown to be closest to the gene<sup>19</sup> and had given a  $p_{\text{excess}}$  of 0.74. Combining this approach with that of deviation from the Hardy-Weinberg equilibrium and detailed analysis of ancestral haplotypes, the location of the gene was restricted to a 250 kb candidate region. In the absence of family material, hybrid cell lines containing only one copy of chromosome 6 were generated from patients, permitting haplotype analysis.

Standard approaches such as exon trapping and cDNA selection were then applied to the region of interest. To ensure comprehensive identification of all genes, complete genomic sequencing of the region was undertaken. This region was gene rich as had been expected from studies around the MHC.<sup>23</sup> Three novel genes with amino acid sequence similarities to Ro/SSA ribonucleoprotein, a sodium phosphate transporter, and to HLA-A2 were found as well as 12 histone genes. In order to identify the haemochromatosis gene the coding sequence and intron/exon boundaries of all these genes were analysed. A single base substitution in the MHC class I-like gene was compatible with this being the causative mutation in haemochromatosis.

Did this mutation fulfil other criteria required of the haemochromatosis gene? Analysis of DNA from all the other patients showed that this mutation indeed occurred on 85% of all haemochromatosis bearing chromosomes and in 3.2% of controls, which concurred with earlier predictions of carrier frequency.

### The mutations

The mutation found was a G to A transition at nucleotide 845 of the open reading frame and results in a cysteine to tyrosine substitution at amino acid 282. The frequency of this mutation in patients is compelling evidence that this class I-like gene is indeed the haemochromatosis gene or one very closely linked to it. A second variant was found elsewhere in the molecule, a C to G change in exon 2 resulting in a histidine to aspartic acid change at position 63.

This gene has been termed HLA-H supposedly as it is similar to HLA class I genes.<sup>2</sup> There is already an HLA-H gene (OMIM 142925) which maps between HLA-A and HLA-G so this nomenclature is unacceptable. Hopefully confirmation of this gene as being that responsible for haemochromatosis will lead to it being renamed HFE or HLA-HFE; HFE was assigned to the haemochromatosis gene before its identification.

Twenty one of the 178 patients studied by Mercator lacked the Cys282Tyr mutation on either chromosome and carried the His63Asp variant at a frequency of 21%, which was similar to the control chromosomes. Despite further analysis by sequencing the individual exons and exon/intron boundaries of the HLA-H gene, neither additional mutations nor any evidence of linkage disequilibrium were found, suggesting that a subset of patients either had haemochromatosis through a second locus unrelated to chromosome 6 or were sporadic cases.

This class I-like gene is certainly the best candidate to date. What is the function of this molecule in iron metabolism? This is not going to be an easy question to answer. The gene appears to be expressed at a low level in a number of tissues but functional studies are required to confirm that this is indeed the haemochromatosis gene.

The high degree of sequence similarity with other class I molecules suggests structural

homology. The crystal structures of several class I molecules have now been determined and extrapolation to HLA-H shows some interesting features. Class I molecules interact non-covalently with  $\beta_2$ -microglobulin. The groove formed by the  $\alpha_1$  and  $\alpha_2$  domains is responsible for peptide binding in class I molecules. This molecule differs from class I proteins in that two of the four tyrosine residues forming the classical class I groove are missing and there is a proline side chain blocking one end of the groove. These features are found in a related molecule, the Fc receptor (hFcRn), which forms a heterodimer with  $\beta_2$ -microglobulin and lacks a peptide binding groove interacting with an altogether larger molecule, IgG. The His63Asp mutation is predicted to occur in the  $\alpha_1$  domain and is predicted to be located on the loop between the third and fourth  $\beta$  strands of the peptide binding domain. There are four conserved cysteine residues in class I molecules and the second two form a disulphide bridge giving rise to the Ig like  $\alpha_3$  domain which interacts with  $\beta_2$ -microglobulin. The common Cys282Tyr mutation loses one of these conserved residues and is predicted to inactivate the protein by preventing correct folding of the  $\alpha_3$  domain and hence interfering with its interaction with  $\beta_2$ -microglobulin, if it does interact with this protein.  $\beta_2$ -microglobulin deficient mice display a progressive iron overload,<sup>24</sup> although the mechanism by which this occurs has not been defined. These mice have been proposed as an animal model for GH.<sup>25</sup>

Various hypotheses are put forward by Feder *et al*<sup>2</sup> to explain how this MHC class I-like molecule could contribute to iron loading. These include HLA-H acting as an iron binding ligand, a role in signal transduction by sensing plasma iron levels, or an indirect mechanism through association with components of the immune system which might influence iron metabolism. None is any more than speculative.

### Implications

What is important from this work is that the Cys282Tyr mutation is by far the best marker for chromosomes bearing the haemochromatosis gene. The clinical implications are that there is now a straightforward way of identifying patients at risk for this common and widely underdiagnosed genetic disease. Treatment by venesection is safe and of proven benefit in preventing development of the disease before the secondary end organ damage. The development of a test to screen the adult population for haemochromatosis will prove a major advance in public health care.

It is instructive to note that once a tight linkage has been established by academic research, biotechnology companies have the resources rapidly to identify the disease causing gene.

### Why has the gene been so elusive?

The initial tight linkage to the MHC was a major breakthrough in the days before informative DNA markers. What had not been

predicted was the extent of the linkage disequilibrium telomeric to the MHC. Lack of chromosomal recombinations and breakpoints forced a reliance on linkage disequilibrium studies. There is clearly suppression of recombination in this region. The accepted genome wide conversion factor assumes that 1 Mb is equivalent to approximately 1 cM. However, for this region of chromosome 6, markers that appear to be 1 cM apart are separated by a physical distance of 6 Mb. This has also meant that although there appeared to be a number of Généthon markers that mapped to the region these were not all physically close to each other. This was also confounded by the fact that the region containing this HLA-H gene was greatly under represented in the YAC libraries and absent in the public YAC map databases.

Mercator are to be congratulated on a such a thorough piece of work. Functional studies confirming that this MHC class I-like gene is indeed the much sought after haemochromatosis gene are now awaited.

### The United Kingdom Haemochromatosis Consortium

KATHRYN J H ROBSON, JEREMY D SHEARMAN, ALISON T MERRYWEATHER-CLARKE, JENNIFER J POINTON, WILLIAM M ROSENBERG, ANN P WALKER, JAMES S DOOLEY, ADRIAN BOMFORD, RUMA RAHA-CHOWDHURY, MARK WORWOOD

Correspondence to: Dr Robson, MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK.

- 1 Simon M, Bourel M, Fauchet R, Genetet B. Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis. *Gut* 1976;17:332-4.
- 2 Feder JN, Gnirke A, Thomas W, *et al*. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996;13:399-408.
- 3 Finch SC, Finch CA. Idiopathic haemochromatosis, an iron storage disease. *Medicine* 1955;34:381-430.
- 4 Leggett BA, Halliday JW, Brown NN, Bryant S, Powell LW. Prevalence of haemochromatosis amongst asymptomatic Australians. *Br J Haematol* 1990;74:525-30.
- 5 Niederau C, Fischer R, Purschel A, *et al*. Long-term survival in patients with hereditary hemochromatosis. *Gastroenterology* 1996;110:1107-19.
- 6 Niederau C, Fischer R, Sonnenberg A, *et al*. Survival and causes of death in cirrhotic and in non cirrhotic patients with primary hemochromatosis. *N Engl J Med* 1985;313:1256-62.
- 7 Simon M, Le Mignon L, Fauchet R, *et al*. A study of 609 HLA haplotypes marking for the hemochromatosis gene: (1) mapping of the gene near the HLA-A locus and characters required to define a heterozygous population and (2) hypothesis concerning the underlying cause of hemochromatosis-HLA association. *Am J Hum Genet* 1987;41:89-105.
- 8 Gasparini P, Borgato L, Piperno A, *et al*. Linkage analysis of 6p21 polymorphic markers and the hereditary hemochromatosis: localization of the gene centromeric to HLA-F. *Hum Mol Genet* 1993;2:571-6.
- 9 Yaouanq J, Perichon M, Chorney M, *et al*. Anonymous marker loci within 400kb of HLA-A generate haplotypes in linkage disequilibrium with the hemochromatosis gene (HFE). *Am J Hum Genet* 1994;54:252-63.
- 10 El Kahoun A, Chauvel B, Mauvieux V, *et al*. Localisation of seven new genes around the HLA-A locus. *Hum Mol Genet* 1993;2:55-60.
- 11 Weber LW, Kwitek AE, May PE, Zoghbi HY. Dinucleotide repeat polymorphism at the D6S105 locus. *Nucleic Acids Res* 1991;19:968.
- 12 Jazwinska EC, Lee SC, Webb SI, Halliday JW, Powell LW. Localisation of the hemochromatosis gene close to D6S105. *Am J Hum Genet* 1993;53:347-52.
- 13 Worwood M, Raha-Chowdhury R, Dorak MT, *et al*. Alleles at D6S265 and D6S105 define a haemochromatosis specific-genotype. *Br J Haematol* 1994;86:863-6.
- 14 Stone C, Pointon JJ, Jazwinska EC, *et al*. Isolation of CA dinucleotide repeats close to D6S105: linkage disequilibrium with haemochromatosis. *Hum Mol Genet* 1994;3:2043-6.
- 15 Burt MJ, Smit DJ, Pyper WR, Powell LW, Jazwinska EC. A 4.5 megabase YAC contig and physical map over the hemochromatosis gene region. *Genomics* 1996;33:153-8.

- 16 Calandro LM, Baer DM, Sensabaugh GF. Characterisation of a recombinant that locates the hereditary hemochromatosis gene telomeric to HLA-F. *Hum Genet* 1995;96:339-42.
- 17 Malfroy L, Coppin H, Calandro LM, *et al.* A new microsatellite marker at the RFP locus on chromosome 6p22 definitely located the haemochromatosis gene at least one megabase telomeric to HLA-A. *Eur J Hum Genet* 1996;4(suppl 1):5.105.
- 18 Raha-Chowdhury R, Bowen DJ, Burnett AK, Worwood M. Allelic associations and homozygosity at loci from HLA-B to D6S299 in genetic haemochromatosis. *J Med Genet* 1995;32:442-52.
- 19 Raha-Chowdhury R, Bowen DJ, Stone C, *et al.* New polymorphic microsatellite markers place the haemochromatosis gene telomeric to D6S105. *Hum Mol Genet* 1995;4:1869-74.
- 20 Hästbacka J, Kaitila I, Sistonen P, de la Chapelle A. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 1992;2:204-11.
- 21 Hästbacka J, de la Chapelle A, Mahtani MM, *et al.* The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. *Cell* 1994;78:1073-87.
- 22 Lehesjoki AE, Koskiniemi M, Norio R, *et al.* Localization of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. *Hum Mol Genet* 1993;2:1229-34.
- 23 Trowsdale J, Ragoussis J, Campbell RD. Map of the human MHC. *Immunol Today* 1991;12: 443-5.
- 24 de Sousa M, Reimão R, Lacerda R, *et al.* Iron overload in  $\beta_2$ -microglobulin-deficient mice. *Immunol Lett* 1994;39:105-11.
- 25 Rothenberg BE, Voland JR.  $\beta_2$  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* 1996;93:1529-4.