Woman's meat, a man's poison

Peter Little

MANY followers of research into human genetics might be surprised to discover that the gene underlying a single-gene disorder that is arguably the most common in northern Europe has not yet been isolated. Well, perhaps now it has: in this month's issue of *Nature Genetics*, Feder *et al.*[1] report the identification of mutations that occur in a gene in patients with hereditary haemochromatosis (HH). Unlike many candidate genes investigated for their association with hereditary disease, this work has been carried out without recourse to family analysis.

The symptoms of HH are due to a defect in iron metabolism that results in an accumulation of iron in various tissues of the body, leading to failure of liver function, tumours, diabetes and arthritis[2]. The disease can be easily treated — periodic blood-letting is sufficient. Women therefore rarely develop symptoms until after menopause.

Particularly remarkable is the incidence of this recessive disease: it is very common, with about 1 in 10 northern Europeans being carriers. HH is sometimes misdiagnosed as alcoholic cirrhosis; it is also underdiagnosed because most Europeans have a relatively low meat consumption and so do not develop clinically apparent iron loads. To a woman, the selective advantage of HH may be large: stores of body iron could remain high in the face of adverse dietary conditions and this may explain the extraordinarily high HH frequency — one man's genetic disease may be another woman's salvation during pregnancy and childbirth!

Unusually for a common defect, there is evidence that a large proportion of HH genes are derived from a single ancestral source. Generally, such observations are consistent with a genetic founder effect followed by rapid spread of the mutation through the population to retain the characteristic polymorphic patterns of the ancestral HH chromosome. However, familial analysis has shown that there are low levels of recombination over the whole HH region, which also contributes to conservation of the ancestral pattern. Unfortunately, this has also meant that it has been difficult to refine gene location by classical family studies.

Feder *et al.*[1] have tried to overcome this limitation by using linkage disequilibrium (LD) and identity-by-descent to refine the gene location (see box), so neither knowledge of family structures nor identification of recombinants is necessary. They were able to identify a region of 250,000 base pairs thought to contain the HH gene. Sequence analysis showed the region contained, *inter alia*, a gene, which they named *HLA-H* (H for haemochromatosis, presumably), related to the major histocompatibility complex (MHC) class I gene, *HLA-A*. They went on to show that 83 per cent of 187 HH patients carried a tyrosine instead of a cysteine residue at position 282 (a Cys282Tyr mutation) — another mutation at a different site was thought to be a simple polymorphism. (This new *HLA-H* gene is not to be confused with the existing *HLA-H* (OMIM ref. 142925; see also ref. 2), which is an expressed pseudogene within the MHC cluster.)

Feder *et al.* argue that the remaining patients, who did not have any mutations in *HLA-H*, acquired the disease by virtue of mutation in a second, unidentified locus; that is, the genetic basis of HH is heterogeneous and a second HH gene must map to another chromosome region.

How likely is it that an *HLA-A*-related gene causes HH? Curiously, transgenic mice that lack β_2-microglobulin[3] develop iron overload similar to HH, and as β_2-microglobulin complexes with MHC class I proteins, this, rather indirectly, suggests a role for *HLA-H* in iron loading.

The data look compelling, but Feder *et al.* state that "formal proof must await functional studies....". Why this caveat? The principal problem is linkage disequilibrium: if this was complete, then a mutation in the gene, say immediately next to the HH gene, would behave exactly like the HH gene in these analyses because it could not be separated from the HH gene by recombination. We do not yet have a clear idea of the population dynamics of random DNA sequences, and perhaps the LD of the region is purely a chance accident of these dynamics. Experimentally, LD is dependent on differences between

Strategy for tracking down the HH gene

IF a mutation has occurred once in a population and then spread through it, as is the case for many genetic diseases, then the analysis of linkage disequilibrium (LD) can be used to refine the disease gene's location on a particular chromosome. Let us say there are a series of ten polymorphisms along the region of chromosome 6 that contains HH: call these A/a to J/j: in all cases, the upper-case form is found in 90% (P=0.9) and the lower-case form in 10% (P=0.1) of the population (see table below). The actual chromosomal arrangement of polymorphs (AbCDEfghij, for example) is called the haplotype. Further, assume the HH gene is located between E/e and F/f. The original HH mutation must have occurred on a chromosome with a given haplotype — say all lower-case a–j. If we now analyse the polymorphs on the HH-bearing chromosome, we would see a frequency of 1.0 for each lower-case polymorph instead of the expected general-population figure of 0.1: there is a large probability excess over the frequency in the non-HH population. Some recombination would have inevitably occurred in HH chromosomes as they spread through the population: thus, the excess probability of a and j (furthest from HH and therefore most likely to have recombined) might be reduced compared to e and f, which flank the HH gene. In fact, the closer to HH you get, the greater will be the excess. (Good examples of LD mapping and associated mathematics can be found in refs 7 and 8.) Additionally, analysis of individual disease haplotypes can then be used to determine a region which is identical by descent (IBD) even in variant haplotypes and therefore likely to contain the HH gene. P. L.

Ancestral	a	b	c	d	e	HH	f	g	h	i	j
Variant 1	A	B	c	d	e	HH	f	g	h	i	J
Variant 2	A	B	C	D	e	HH	f	g	h	I	J
Variant 3	A	B	C	d	e	HH	f	G	H	I	J
Variant 4	A	b	c	d	e	HH	f	g	H	I	J
$P_{Obs.}$	0.2	0.4	0.6	0.8	1.0		1.0	0.8	0.6	0.4	0.2
$P_{Exp.}$	0.1	0.1	0.1	0.1	0.1		0.1	0.1	0.1	0.1	0.1
IBD	A/a	B/b	C/c	D/d	e only		f only	G/g	H/h	I/i	J/j

Idealized linkage disequilibrium: alleles are indicated in upper or lower case. Frequency (P) of all lower-case alleles in the normal population is 0.1. Five chromosome variants carrying HH are shown; observed and expected P values for lower-case alleles are listed. The peak linkage disequilibrium is over alleles e and f — this is the only region common to all haplotypes and it shows identity by descent. Red, ancestral HH haplotype; blue, non-ancestral.