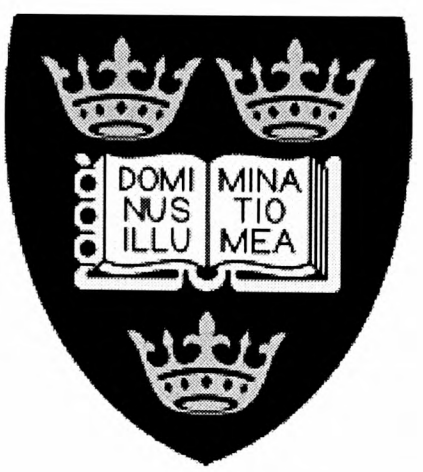
The Molecular Genetics of Haemochromatosis

Jeremy David Shearman

A thesis submitted as partial fulfillment for the examination
of Doctor of Philosophy

Green College and the Faculty of Clinical Medicine
University of Oxford
Trinity 1996.



To Penny,

"You could say I lost my faith in science and progress,

You could say I lost my belief in the holy church,

You could say I lost my sense of direction,

You can say all of this and worse but ..."

(Sting 1993)

Acknowledgements

There are many people to whom I owe great thanks for support over the last three years.

Avoiding comparison with "The Three Tenors", “The Three Profs” (Bell, Powell and Weatherall) gave me the benefit of their time, guile and experience at the times when it was most needed. The trio of Drs Chapman, Jewell and Trowell continue to support my clinical career, for which I am ever appreciative.

Kathryn Robson gave me first class laboratory supervision. Jenny Pointon showed great patience and humility in taking the time to teach me many of the fundamentals of practical laboratory science. Yan Tat Liu introduced me to the Internet. Caro, Alison and members of John Old’s group provided great company.

There are many people with whom I have enjoyed the opportunity to expand on theories of haemochromatosis, mammalian iron metabolism and life. In particular, Alex Kniseley, Michael Chorney and Jeremy Brock stand out as individuals whose broad horizons have influenced my thinking.

William Rosenberg was singularly responsible for giving me the confidence to embrace "new genetics" and to take on a project that has exposed me to the practical application of what I had previously considered to be voodoo to one of the most fascinating problems of clinical medicine.

Most of all, I thank Penny, Aggie and Floyd for both putting up with my ranting and ensuring that I kept things in perspective.

Declaration

The work of this thesis was conducted by the author unless otherwise stated. No part of this thesis has been submitted for another degree at this or any other university.

Nomenclature

Haemochromatosis: the condition, the gene and HLA-H

As stated in the introduction, the term haemochromatosis as used in this thesis refers solely to the recessively inherited, HLA-linked form of primary iron overload. All other forms of iron overload are termed as clearly as possible. The gene harbouring the disease causing mutation had been prospectively entitled HFE (for high iron - Fe). This abbreviation has **not** been used in this thesis to avoid confusion in the discussions that follow the identification of the ancestral haemochromatosis mutation.

In August 1996 Mercator Genetics reported the identification of the ancestral haemochromatosis mutation in a HLA class I-like gene. They called this gene HLA-H (the H presumably referring to Haemochromatosis). Although it is recognised that the name HLA-H had already been given to a pseudogene that maps centromeric to HLA-A, the term HLA-H as used in this thesis refers solely to the gene identified by Mercator.

Sequence Tagged Sites

Many of the sequence tagged sites (STS) used in this study have different names depending on their source and their form (e.g. microsatellites). Wherever possible these are referred to by their Genome Database (GDB) DNA segment number, prefixed by D6S (where 6 refers to the chromosomal location). A comprehensive list of STSs and their aliases is included in Appendix 2.

Diagrams and Illustrations

All illustrations and tables are orientated such that the centromere of chromosome 6 is to the left. Wherever possible the landmarks HLA-A, D6S105 or D6S1260 are included to aid orientation/scale.

Diagrams are included to illustrate biological systems. Where these reflect observations or opinions of others the appropriate reference is made. Where such diagrams are intended to illustrate purely hypothetical processes these will be in pencil line.

Abbreviations

Commonly used abbreviations:

eALAS	erythroid 5-amino laevulinic acid synthetase
CEPH	Centre d'Etude du Polymorphisme Humain
CHLC	Cooperative Human Linkage Center
db	database
EDTA	ethylene diamine tetra-acetic acid
EST	expressed sequence tag
FISH	fluorescence *in-situ* hybridisation
GDB	Genome Database
HFE	haemochromatosis gene (succeeded by HLA-H)
HLA	human leucocyte antigen
OMIM	Online Mendelian Inheritance in Man
MHC	major histocompatibilty complex
NCBI	National Centre for Biotechnology Information
NTBI	Non-transferrin bound iron
PCR	polymerase chain reaction
PAGE	polyacrylamide gel electrophoresis
PFGE	pulsed field gel electrophoresis
RFLP	restriction fragment length polymorphism
SDS	sodium dodecyl sulphate
STS	sequence tagged site
TE	Tris EDTA buffer
WI/MIT	Whitehead Institute/Massachussetts Institute of Technology
WWW	World-Wide Web
YAC	yeast artificial chromosome

Abstract

The Molecular Genetics of Haemochromatosis

Jeremy David Shearman

Thesis presented to the University of Oxford

Faculty of Clinical Medicine

Trinity 1996

Haemochromatosis is the most common single gene disorder to afflict North-West European populations. It is probably the most common genetic disorder of iron metabolism worldwide. As many as 1 in 250 people in the UK are affected and although the phenotype causes only a mild increase in gastrointestinal iron absorption a proportion of affected individuals will accumulate sufficient iron over their life-time to cause cirrhosis and hepatocellular carcinoma. Venesection treatment instituted before cirrhosis has established ensures a normal life expectancy, but clinical presentation is often late in life after irreversible organ injury has occurred. Identification of people at risk in the early, asymptomatic stage by measurements of iron status is unreliable.

The genetic defect responsible for haemochromatosis has been sought in the hope that its identification might facilitate early diagnosis and that studies on the gene product would lead to a greater understanding of the mechanisms of mammalian iron absorption. Genetic linkage to HLA-A3 placed the gene responsible for haemchromatosis in, or close to, the major histocompatibilty complex (MHC) on the short arm of chromosome 6 and a positional cloning strategy has been adopted.

This thesis describes work directed to the identification of the haemochromatosis gene by positional cloning. The region telomeric to the MHC was mapped using yeast artificial chromosomes, from which new microsatellites were isolated. These markers were used in linkage disequilibrium analyses and the mapping of a recombination breakpoint that defined a haemochromatosis gene region. This region was physically mapped in fine detail and positional candidates sought by EST database analysis.

Before a systematic search for genes in the region began a strong positional candidate was reported (Feder *et al* 1996). Analysis of this mutation in patients from the UK confirmed this to be the ancestral haemochromatosis mutation.

Contents

Chapter 2.2

Construction of a yeast artificial chromosome (YAC) contig extending telomeric to the human major histocompatibility complex

Chapter 2.3

The identification of YACs extending telomeric to D6S1260

Chapter 3

Increasing the resolution of the physical map and the recognition of candidate gene sequences

Chapter 3.1

Pages 123-145

Background: The positional candidate principle

Chapter 3.2

Pages 146-155

Targetted database analysis

Chapter 3.3

Pages 156-166

Identification and characterisation of a CpG island within the haemochromatosis gene region

Chapter 3.4

Fine mapping of the haemochromatosis gene region and preparation for genomic sequencing

Chapter 4

Verifying the discovery of the ancestral haemochromatosis mutation

Chapter 4.1

HLA-H "A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis" Nature Genetics 13(4):399-409, August 1996

Chapter 4.2

The physical maps of 6p21.3/22.1

Chapter 1

Introduction: Why map the haemochromatosis gene?

Contents - Chapter 1

Haemochromatosis - primary iron overload

The term haemochromatosis has its origins at the end of the last century when Friedrich von Recklinghausen, the German pathologist, described the post-mortem appearances of patients who had died with the clinical syndrome of "Bronzed Diabetes". This syndrome, as originally described by Troisier, caused clinical presentation late in life with a triad of diabetes mellitus, a slate grey pigmentation of the skin and a pigmented cirrhosis of the liver. In the absence of treatment for hyperglycaemia the condition was rapidly fatal. Von Recklinghausen deduced that the pigment in the liver was iron and, in assuming that this must have been derived from the blood, he called the condition "*hämochromatose*" (i.e. blood colouring). A century after the original description, haemochromatosis is now recognised to represent a subtle but progressive accumulation of dietary iron and furthermore to be the commonest single gene disorder of North West Europeans.

The term haemochromatosis should now be used exclusively to refer to the recessively inherited, HLA-linked, genetic form of primary iron overload. Furthermore, it is reasonable to use the term in referring to individuals homozygous for a haemochromatosis allele regardless of their iron status. The additional terms hereditary, genetic and idiopathic are unnecessary. Other forms of iron overload should be described explicitly stating, where possible, the source of the excess iron (e.g. transfusional iron overload; iron overload secondary to sideroblastic anaemia; Bantu siderosis). Although the gross pathology of haemochromatosis might share some characteristics of other conditions resulting in iron overload, the condition is clearly distinct from these both in terms of pathophysiology and the potential public health considerations.

Haemochromatosis is inherited in an autosomal recessive manner

The genetic basis of haemochromatosis was first alluded to by W.H.Sheldon, a physician from Wolverhampton, who collected data from 345 separate reports from

the world literature in compiling a comprehensive monograph on the condition in 1935 (Sheldon 1935). On the basis of the degree of tissue iron loading in relation to the modest quantities present in the diet Sheldon concluded that haemochromatosis must represent an inherited inborn error of metabolism. In support of this theory he cited 14 separate instances from the literature of familial clustering of the condition although he accepted that demonstration of clear inheritance was restricted to only two families. Familial clustering was also evident in the clinical review of Finch and Finch two decades later, suggesting a dominant mode of inheritance of haemochromatosis in some families but with incomplete penetrance (Finch and Finch 1955).

An excess of liver iron in healthy first degree relatives of patients with haemochromatosis provided further strong support for the familial basis of the condition (Scheuer *et al.* 1962). Others maintained that hepatic iron overload in the presence of liver cirrhosis was not an inherited trait but an environmental feature attributable to excess dietary iron absorption, predominantly facilitated by a common underlying aetiology, alcohol (Macdonald 1965).

It was as late as 1976 that the argument over the genetic basis of haemochromatosis was resolved. Simon and colleagues in Rennes, France, fortuitously studied the frequency of histocompatibility antigens in a group of 51 unrelated individuals with haemochromatosis (Simon *et al.* 1976). They found a statistically significant excess of the Major Histocompatibility Complex (MHC) class I antigens HLA-A3 and HLA-B14.

HLA Antigen	Haemochromatosis (n = 51)	Controls (n = 204)	P value
A3	40 (78.4%)	55 (27%)	$<1x10^{-9}$
B14	13 (25.5%)	7 (3.4%)	$<1x10^{-4}$

Table 1.1 HLA class I antigen frequencies in haemochromatosis (Simon *et al* 1976).

Studies in other populations confirmed the significant association with HLA-A3 (Bomford *et al.* 1977) providing support for the proposed genetic basis of the condition. Although it was thought unlikely that these antigens might play a direct

role in the transmission of the condition (from the relatively high prevalence of A3 in the control population studied), the presence of a gene linked to the MHC on the short arm of chromosome six was postulated.

The inherent polymorphism of HLA antigens provided a very informative tool with which to study the genetic transmission of haemochromatosis in affected individuals and their families. Prior to Simon's seminal observation, both dominant and recessive modes of inheritance had been proposed (Ross *et al.* 1975). An autosomal recessive mode of inheritance was eventually demonstrated and haemochromatosis was shown to be in genetic linkage with the HLA-A locus with a LOD score of 2.24 for a recombination fraction, $\theta = 0.005$ (Simon *et al.* 1977a).

HLA typing in patients with iron overload in the context of alcoholic liver disease failed to show any excess of the A3 antigen above that in the control population distinguishing this common form of acquired liver disease from haemochromatosis (Simon *et al.* 1977b).

Haemochromatosis became established as a distinct disorder of iron metabolism resulting from a genetic defect linked to the MHC on chromosome six.

The clinical problem

Disease prevalence

The true prevalence of haemochromatosis remains something of an enigma. In 1955 Finch and Finch estimated that haemochromatosis constituted only 1 in 20,000 hospital admissions and 1 in 7,000 hospital deaths (Finch and Finch 1955) suggesting that it represented only a limited clinical problem. Since the early 1970's recognition of earlier stages of iron loading has been possible by, amoungst other things, measurement of serum ferritin and a considerably higher prevalence of primary iron overload has been consistently reported. Large studies of healthy blood donors from northern Europe and North America have repeatedly identified excess iron stores at rates of between 0.30% (Worwood and Darke 1993) to 0.45% (Edwards *et al.* 1988). Regular blood donation would lower the diagnostic sensitivity of serum markers of

iron metabolism in such populations and these studies might underestimate the true prevalence of haemochromatosis.

Two large studies have examined the prevalence of haemochromatosis in random cohorts of unselected healthy individuals. Leggett *et al* in 1990 reported a prospective screening of nearly 2,000, predominantly caucasian, employees of two large Australian corporations (Leggett *et al.* 1990). The prevalence of haemochromatosis in that population (after liver biopsy confirmation) was 0.36% implying that as many as 1 in 300 were affected. More recently Smith *et al* screened over 2,000 employees of the Polaroid corporation in Boston, USA (Smith *et al.* 1995). Although 5 new patients were identified amoungst 1919 caucasians in that study (or 1 in 384), 6 individuals with the diagnosis established before the study had been excluded.

Demography

Haemochromatosis is restricted to caucasians of a North-West European extraction. Smith identified no cases of primary iron overload in racial groups other than caucasians (Smith *et al.* 1995) and a recent study examining primary iron overload in Americans of Afro-Carribean descent showed that, when present, this was phenotypically different to haemochromatosis and more akin to dietary iron overload of Sub-Saharan Africa (Barton *et al.* 1995). No figures exist for the prevalence of primary iron overload in Asia.

Haemochromatosis has been described as a "Celtic" disorder with a high prevalence amoung Europeans descended from Celtic peoples (i.e. Irish, Scots and Bretons). Ethnic sub-group analysis of the data generated by Smith identified a prevalence of 1 in 96 for individuals with a pure "British-Irish" ancestry (Smith *et al.* 1995). Within European countries haemochromatosis is often identified in Celtic sub-populations, such as the Bretons in France, although finer ethnic prevalence rates have not been formally defined.

The lowest prevalence rate so far determined within Northern Europe is in Finland. In a population survey of over 22000 Finns aged 26-29 Karlsson found

haemochromatosis at 0.05%, or 1 in 2000 (Karlsson *et al.* 1988). This study might have underestimated the true prevalence in that population due to the young age of their cohort. Even so, the estimate of the prevalence makes haemochromatosis at least 10 times more common than predicted by Finch and Finch (Finch and Finch 1955).

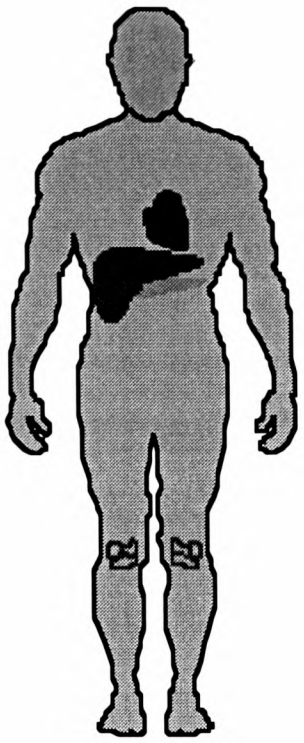
The apparent disparity between prevalence data generated from population studies and hospital admission or death rates for haemochromatosis can only be explained by an appreciation of the haemochromatosis phenotype.

The haemochromatosis phenotype

Haemochromatosis, as identified clinically, is a systemic disorder characterised by iron deposition in parenchymal cells of many organs. Since Powell's demonstration of a persistent and inappropriate increase in gastrointestinal iron absorption, it is now widely accepted that haemochromatosis represents a defect in iron absorption (Powell *et al.* 1970). In the absence of any physiologically relevant iron excretion pathway, persistent positive balance for iron eventually results in progressive iron accumulation. This, in turn, is reflected by a range of clinical presentations covering a spectrum from asymptomatic and physically healthy individuals to patients with established, chronic multiorgan failure.

End organ damage

Although the extreme phenotype of "bronzed diabetes" is rarely seen in clinical practice today, a range of organ pathologies attributable to iron deposition in different organs in haemochromatosis means that affected individuals might present to different medical specialties with different problems.



Pituitary - hypogonadotrophic hypogonadism
Heart - dysrhythmias, dilated cardiomyopathy
Liver - cirrhosis, hepatocellular carcinoma
Pancreas - diabetes
Joints - arthropathy

Non-specific - malaise, lethargy
abdominal pain

Figure 1.1 The range of organ pathologies leading to the clinical presentation of haemochromatosis.

Liver Disease

Iron deposition in the liver is the most important single aspect of haemochromatosis from the point of view of both the clinical presentation and prognosis.

As already outlined, cirrhosis of the liver has been associated with haemochromatosis since its very first description, although the clear distinction of this from other forms of liver injury has only relatively recently been clearly established (Simon *et al.* 1977b). In haemochromatosis homozygotes the liver progressively stores excess absorbed iron as reflected in a progressive increase in hepatic iron concentration (Cartwright *et al.* 1979) which correlates well with total body storage iron (Barry 1974). An hepatic dry weight iron concentration of 400μmol/g or more carries a high risk of the development of cirrhosis (Bassett *et al.* 1986).

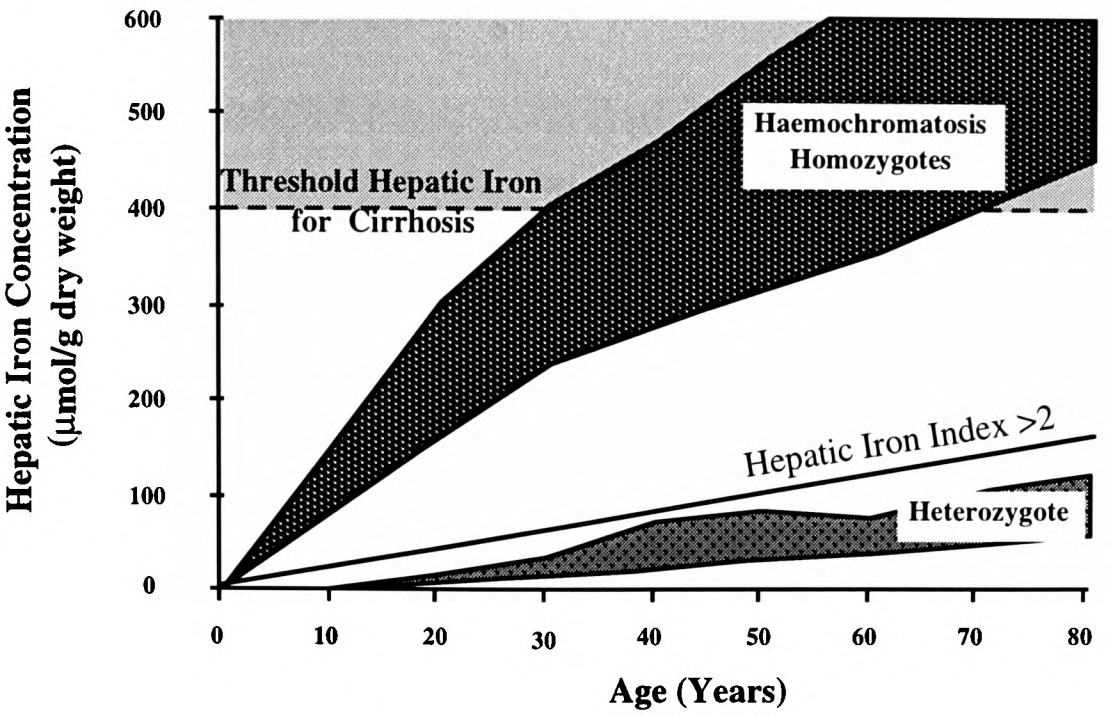


Figure 1.2 Haemochromatosis is characterised by a progressive accumulation of liver iron. Heterozygotes do not accumulate significant iron stores.
(data adapted from Cartwright *et al* 1979 and Bassett *et al* 1986).

Histologically, iron initially accumulates within periportal hepatocytes, reflecting its origin from the gastrointestinal tract. With increasing hepatic iron concentration this progresses to involve zones 2 and 3 of the hepatic lobule (Deugnier *et al.* 1992).

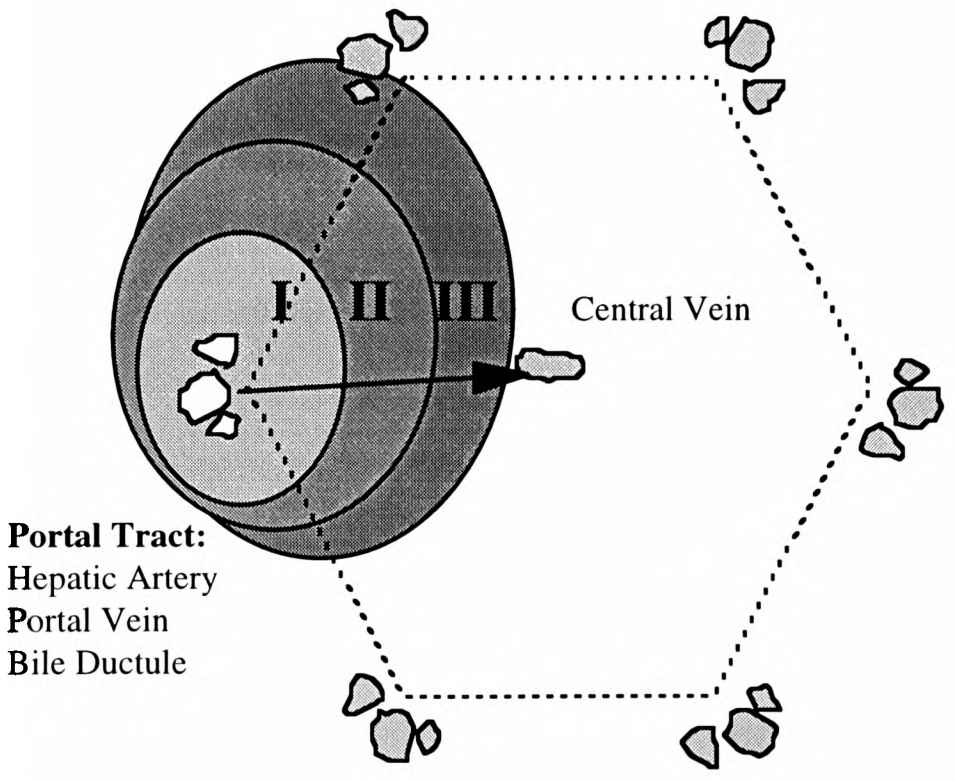


Figure 1.3 The conceptual division of the hepatic lobule into zones reflects the relative exposure of hepatocytes to portal blood derived from the gastrointestinal tract. Zone I is exposed to portal first and shows the earliest accumulation of iron in haemochromatosis.

Compared to secondary iron overload, accumulation of iron in Kupffer cells (the resident reticuloendothelial cell population of the liver) of the hepatic sinusoid is minimal and occurs late in haemochromatosis (Ross *et al.* 1975). This stage of iron loading is accompanied by the progressive deposition of fibrous tissue by transformed hepatic stellate cells (Deugnier *et al.* 1992). Once fibrous septa have divided the hepatic lobule and bridged between portal tracts and central veins the liver is said to be cirrhotic, marking the irreversible stage of liver injury.

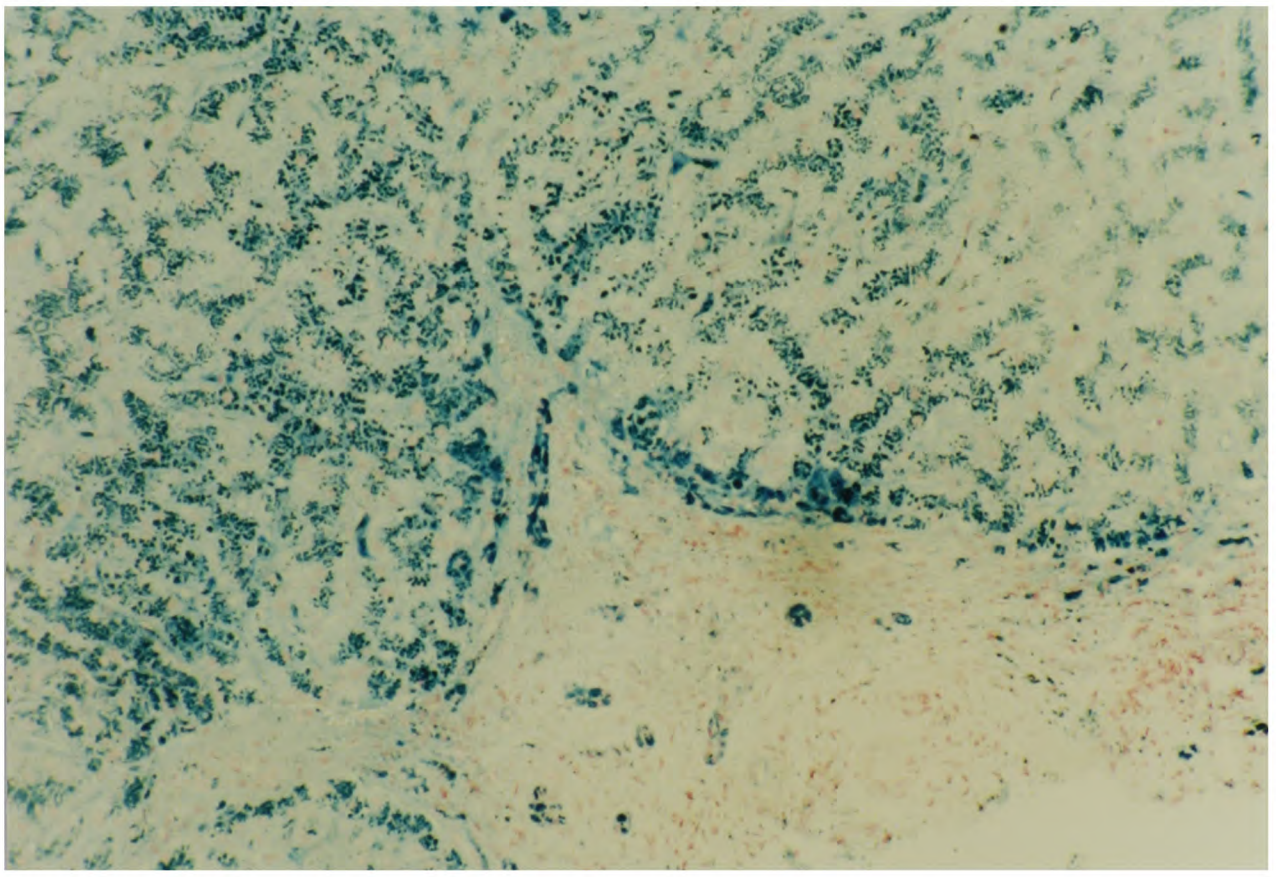


Figure 1.4 A percutaneous liver biopsy from a patient with haemochromatosis. In this Prussian Blue stained section the iron laden hepatocytes stain blue. Each nodule of hepatocytes is encircled by dense fibrous tissue (unstained) indicating that the normal liver architecture has been disrupted and the liver has become cirrhotic (10x magnification).

The molecular mechanisms by which excess hepatocellular iron causes fibrosis in the presence of a modest necroinflammatory reaction are not clearly defined. Kupffer cell iron loading has been demonstrated to induce intracellular adhesion molecule-1 (ICAM-1) on hepatocytes in haemochromatosis, implying subtle inflammatory changes (Stal *et al.* 1995). Many have postulated the production of oxygen free radicals and lipid peroxidation products in causing cellular damage (Evidence reviewed in Bacon and Britton 1990). More direct experimental evidence has supported such hypotheses by the demonstration of a protective effect of an anti-

oxidant (tocopherol, vitamin E) in a rodent model of iron-induced hepatic fibrosis (Pietrangelo *et al.* 1995b).

Hepatic iron concentration can now be measured by non-invasive modalities such as magnetic resonance imaging (MRI - Gandon *et al.* 1994), magnetic resonance spectroscopy (MRS - Dixon *et al.* 1994) and magnetic bio-susceptometry. However, given that other liver pathologies may co-exist with haemochromatosis and that the degree of fibrosis for any given hepatic iron concentration may be influenced by the effect of alcohol(Deugnier *et al.* 1992), liver biopsy remains the standard investigation for both confriming the diagnosis and establishing the degree of liver fibrosis.

Whatever the mechanism, the establishment of cirrhosis marks an irreversible stage of liver injury and is one of the major determinants of prognosis (Niederau *et al.* 1996). Even if excess iron is removed from patients with established cirrhosis they carry a continuing risk of the development of hepatocellular carcinoma (Deugnier *et al.* 1993; Fargion *et al.* 1994) and this now constitutes the commonest cause of death in haemochromatosis (Bomford and Williams 1976; Niederau *et al.* 1996).

The hepatic iron index

The progressive nature of hepatic iron accumulation in haemochromatosis has been used to define an hepatic iron index (HII) as a measure of the severity of iron loading (Bassett *et al.* 1986).

$$\textbf{Hepatic Iron Index} = \frac{\textbf{Hepatic Iron Concentration (}\mu\textbf{mol/g dry weight)}}{\textbf{Age (years)}}$$

An HII of greater than 2 distinguishes homozygous haemochromatosis from heterozygotes (see Figure 1.2, page 9) and patients with other forms of excess liver iron (Bassett *et al.* 1986; Sallie *et al.* 1991; Summers *et al.* 1990).

The heterozygote phenotype

Due to the linkage of the haemochromatosis gene to the MHC, haemochromatosis heterozygotes may be predicted on the basis of sharing a single HLA haplotype with an affected relative (usually sibling). Although haemochromatosis homozygotes can be distinguished from heterozygotes, defined in

this way, by calculation of the hepatic iron index, reliable clinical characterisation of the heterozygote state is not possible. Many reports have indicated partial clinical expression in heterozygotes, but this is usually indicated by modest increases in transport iron (as measured by transferrin saturation) rather than clearly demonstrable increase in storage iron (serum ferritin or hepatic iron concentration Cartwright *et al.* 1979). This creates significant limitations to gene tracking within affected families (see below Targetted screening, page 16).

Treatment

Davis and Arrowsmith were the first to report success in the removal of blood as a rational treatment for iron overload (Davis and Arrowsmith 1950). The basic premise is that with removal of erythrocytes (and hence haemoglobin) excess storage iron is mobilised and transferred to the marrow to make good the loss. The original observation of Davis and Arrowsmith was that patients with haeomchromatosis tolerated repeated bleeding extremely well, not only maintaining reasonable levels of erythrocyte haemoglobin but also feeling symptomatically better for the treatment (Davis and Arrowsmith 1953). Williams went on to demonstrate a marked improvement in 5-year survival in those receiving venesection (Williams *et al.* 1969).

Haemochromatosis Patients	5-year survival
with venesection (n = 40)	89%
without venesection (n = 18)	33%

Table 1.2 Improved prognosis in haemochromatosis following venesection treatment (Williams *et al* 1969)

All the patients included in this limited study had severe disease and the possibility of selection bias in the treatment group is possible. However the benefit of treatment is now established despite the fact that a formal randomised trial has never been reported.

In 1985, Niederau reported the largest study of survival and causes of death in haemochromatosis (Niederau *et al.* 1985). All the 163 patients included in that report had been treated by venesection and followed for a mean period of over 10 years. The

most striking observation was that of the patients identified before the establishment of cirrhosis, life expectancy over the follow-up period was no different to that of age and sex matched controls (Niederau *et al.* 1985). Importantly, early institution of treatment prevents the development of cirrhosis.

A recent update of this series strongly re-affirmed the need for early diagnosis (Niederau *et al.* 1996).

Prospects for early diagnosis

Greater recognition of the clinical condition

Identification of affected individuals by opportuntistic diagnosis in a clinical setting is limited by the non-specific nature of both symptoms and clinical signs. The commonest single symptom at presentation is weakness or lethargy which is present in up to 82% of patients at diagnosis (Niederau *et al.* 1996). Hepatomegaly is the commonest physical sign accounting for a similar proportion of individuals. Despite the clear message of a need for early diagnosis carried by Niederau's earlier publication (Niederau *et al.* 1985), that groups follow-up report 10 years later revealed only a modest increase in the proportion of patients identified before the development of cirrhosis (Niederau *et al.* 1996). In the period from 1982 to 1991 40% of patients in their series were cirrhotic at the time of diagnosis and this was despite the fact that only 70% of their series overall were considered symptomatic (Niederau *et al.* 1996).

Although iron excess in itself causes no consistent abnormality in haematological or simple biochemical indices, the threshold for measurement of liver enzymes in mildly symptomatic patients is falling constantly and patients are often referred for further investigation of abnormal liver tests and 60% of patients had elevated liver enzymes at diagnosis (Niederau *et al.* 1996). However, abnormalities in liver enzymes must be considered a secondary feature of haemochromatosis and are only likely to be present in individuals with significant iron loading. It is worth noting that nearly a third of the patients with established cirrhosis had normal liver enzymes at presentation (Niederau *et al.* 1996).

that nearly a third of the patients with established cirrhosis had normal liver enzymes at presentation (Niederau *et al.* 1996).

Measurement of serum iron indices

Iron status can be measured from peripheral blood (reviewed by Worwood 1994). Transferrin is the specific transport protein for iron. Each transferrin molecule can bind and carry two atoms of ferric (Fe^{3+}) iron. "Transport" iron is measured clinically by serum transferrin saturation, which can be derived from measurements of serum iron and the total iron binding capacity (TIBC).

$$\textbf{Transferrin Saturation} = \frac{\textbf{Serum Iron}}{\textbf{TIBC}}$$

Once transferrin is fully saturated, further iron circulates in less specific forms referred to as "non-transferrin bound iron" (NTBI). This is only present at severe levels of iron overload.

Ferritin is an iron storage protein. It is a large intracellular protein of 440 kiloDaltons which may sequester up to 4500 atoms of iron. The presence of ferritin in the serum in part reflects the release of this storage protein through porous membranes of dying cells. At low levels, ferritin is a sensitive measure of iron stores and a serum ferritin within the normal range excludes significant iron overload. High levels of ferritin are less specific as excess ferritin is secreted by the liver as a feature of the acute phase response.

In normal individuals high levels of stored iron (as reflected by measurements of serum ferritin) present at birth fall in the newborn period, rising after cessation of growth to plateau in mid-life (Yip *et al.* 1984). What regulates the establishment of the final *status quo* is not known.

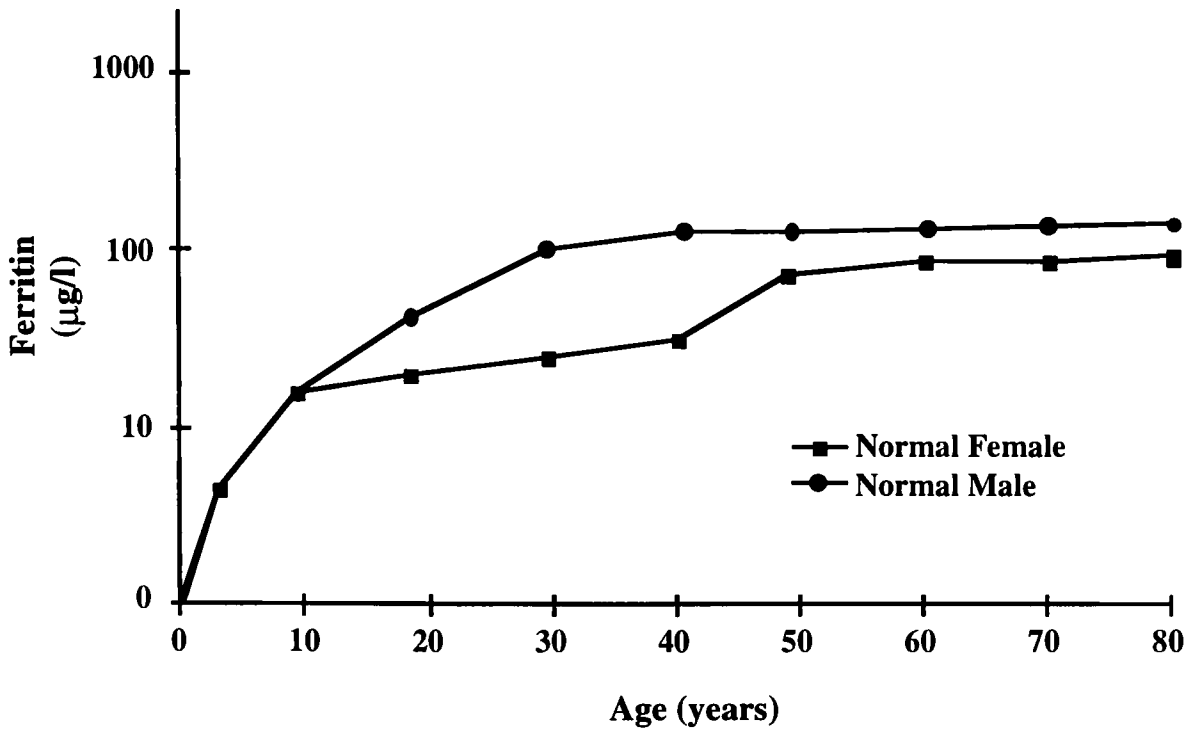


Figure 1.5 Age related changes in storage iron (Data from Yip *et al* 1984).

In haemochromatosis a progressive saturation of serum transferrin is followed by an accumulation of storage iron, as reflected by an increase in serum ferritin and hepatic iron concentration. These sequential changes are in keeping with the proposed underlying defect being a mild increase in gastrointestinal iron absorption. Patients presenting with established end-organ damage from haemochromatosis invariably have an elevation of both transferrin saturation and ferritin. Therefore as confirmatory tests to support a clinical diagnosis prior to liver biopsy these serum iron indices are useful. However it must be stressed that earlier in the natural history of the condition any degree of excess iron may be detected and even normal biochemical indices of iron metabolism are consistent with the diagnosis of haemochromatosis depending on the clinical circumstances (e.g. the age at testing). Furthermore, caution must be exercised when interpreting iron indices in patients with acute illness. In these circumstances serum transferrin saturation falls whilst ferritin may be elevated and hence clinical evaluation of iron status is best made by a combination of both transferrin saturation and serum ferritin.

The rate of iron accumulation in patients with haemochromatosis is variable and the threshold level of hepatic iron concentration at which liver fibrosis is likely may vary depending on other factors such as alcohol. To formally exclude

haemochromatosis in an individual at risk (e.g. family member) requires many years of repeated testing to rule out the progressive accumulation of iron.

Menstruation may delay iron loading in women and symptomatic presentation may be delayed by a decade or two. However, haemochromatosis is an autosomal recessive trait and identification of asymptomatic individuals should not be influenced by sex.

Targeted "screening"

Any first degree relative of a patient with haemochromatosis should be considered at risk of iron loading and should at least have biochemical iron indices measured and interpreted by a clinician with an interest in the condition.

In view of the linkage between haemochromatosis and HLA-A, class I HLA typing can be of use in screening siblings of an affected individual.

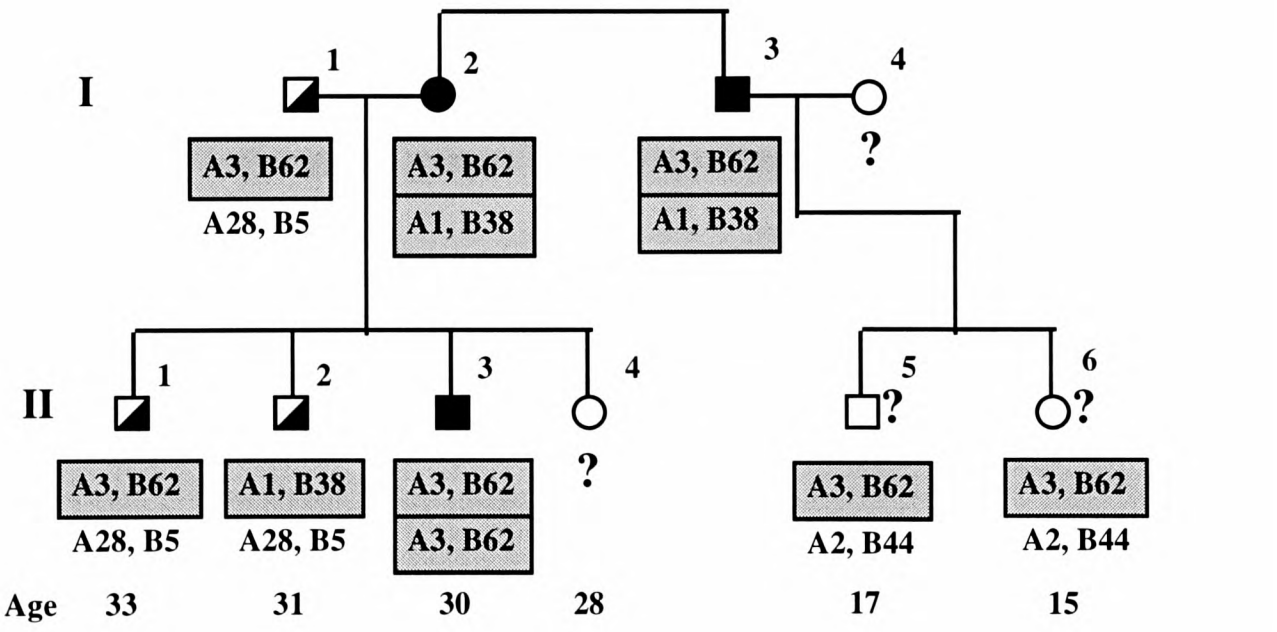


Figure 1.6 The use of HLA class I typing in gene tracking for haemochromatosis. The affected chromosomes are represented by the shaded boxes.

In this example (an actual Oxford haemochromatosis pedigree) individual II-3 has haemochromatosis and has inherited one affected allele from each parent. His younger sister (II-4) has equivocal iron measurements and might not express the homozygote state at this age, being relatively protected from iron loading by menstruation. If on HLA typing she were to prove "haploidentical" to her affected brother (i.e. to carry the same HLA antigens) she should be considered to also carry two haemochromatosis alleles.

On the other side of the extended family, the two younger children (II-5 and II-6) will be at least heterozygotes by virtue of the fact that their father is affected. However, even if their mother was a heterozygote (which cannot be defined clinically), we could not infer the risk to the children as it would be impossible to identify the haemochromatosis allele carrying chromosome.

Although occasionally useful in the risk assessment within well characterised families, the use of HLA typing as a means of early diagnosis has significant limitations.

Population screening

The high prevalence of haemochromatosis in caucasians and the undoubted benefit of early diagnosis make the condition a target for broader screening strategies. Haemochromatosis already meets all the criteria for population screening without the gene having been cloned and strategies for this based on measurements of serum iron indices have recently been published (Adams *et al.* 1995; Witte *et al.* 1996).

The potential benefit of a genetic test for haemochromatosis

The cloning of the gene responsible for cystic fibrosis (CF) was undoubtedly a breakthrough for patients affected by the disorder. This has lead to an understanding of the biology of the gene product (CFTR) and created strategies for somatic cell gene therapy. However, the potential for prenatal molecular diagnosis in CF has been hampered by the enormous range of disease causing mutations (Cystic Fibrosis Consortium 1994), despite the fact that the common Δ508 mutation may account for nearly 70% of affected chromosomes (Kerem *et al.* 1989).

In haemochromatosis, the need for gene therapies is less urgent in view of the proven efficacy of venesection. Identification of the haemochromatosis gene would at least allow the definition of disease causing mutations within individual families which would enhance present attempts at individual risk assessment. Whether this could be extended to population screening would be dependant on the range of disease causing mutations that were defined.

The cellular defect in haemochromatosis

The understanding of disease processes has commonly come from the study of the underlying cellular and molecular pathology. Haemochromatosis exhibits molecular pathological features which distinguish it from other forms of iron overload but before addressing these in particular it is worth examining some general features of iron metabolism in mammalian systems.

Eukaryotes have no physiological mechanism for the excretion of iron. This is true both at the level of the individual cell and of the organism as a whole. Control of cellular iron levels is determined at the point of uptake. This control is largely mediated by a cytoplasmic aconitase that has dual function as an iron response protein (IRP). When intracellular iron supplies are limiting the IRP loses a 4Fe-4S cluster from its enzymatic catalytic site. This infers in a secondary conformational change and a functional switch to an RNA binding protein. In this form the IRP specifically binds iron response elements (IREs) present in the messenger RNA of the proteins controlling iron uptake (transferrin receptor) and storage (ferritin).

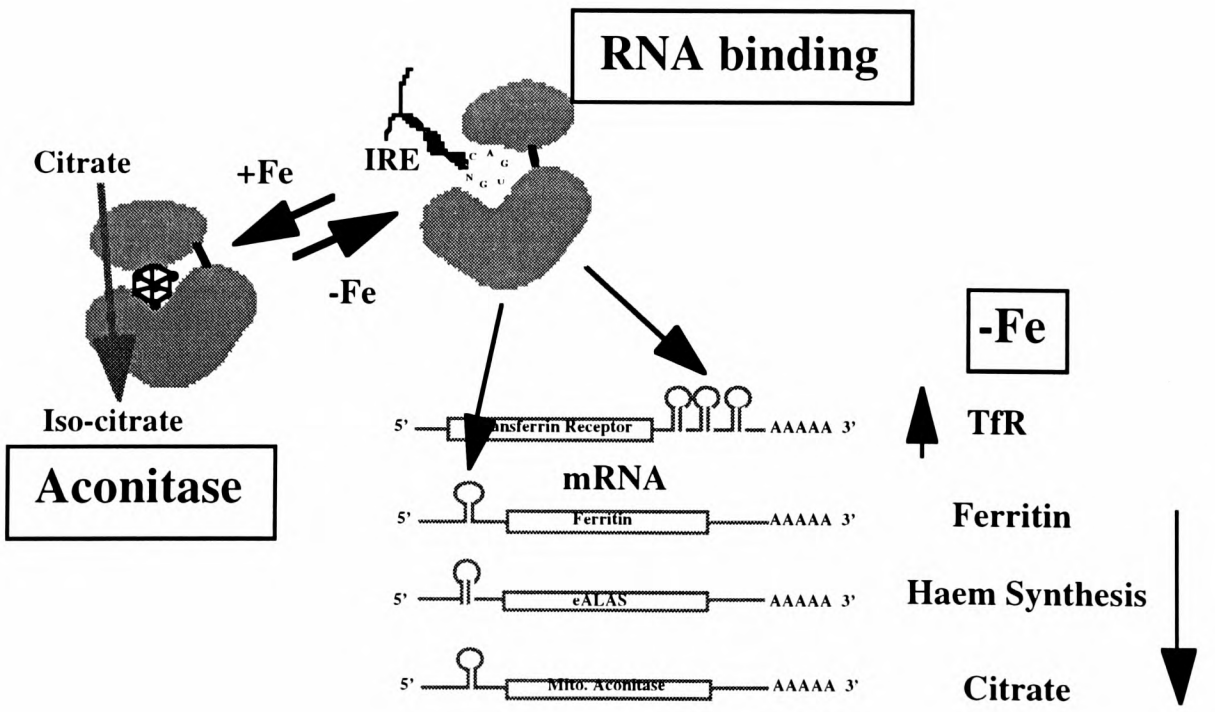


Figure 1.7 The post-transcriptional control of expression of proteins involved in iron metabolism is mediated by the iron response protein (IRP - see text).

The position of the ferritin IRE at the 5' end of the transcript means that binding of the IRP inhibits binding of the ribosomal complex thereby preventing

translation. Conversely multiple copies of the transferrin receptor IRE occupy position in the 3' end of the transcript. In this case binding of the IRP protects the transcript from degradation by ribonucleases prolonging the half-life of the message and resulting in enhanced translation. The net effect of these actions is that when intracellular iron levels are low ferritin levels are coordinately reduced in conjunction with a greater expression of transferrin receptor on the cell surface. (For recent reviews of this subject see Klausner et al 1993 and Kuhn 1994).

The liver

The liver is now considered the innocent bystander in terms of the pathogenesis of haemochromatosis. This partly stems from Powell's demonstration of an increase in absorption of dietary iron from the upper small intestine (Powell *et al.* 1970) and is supported by observations on the response of the hepatocyte to different forms of iron excess. Measuring levels of ferritin and transferrin receptor messenger RNA as a marker of gene expression, Pietrangelo found coordinate and appropriate regulation of these iron metabolism intermediaries in iron loaded hepatocytes of patients with haemochromatosis (Pietrangelo *et al.* 1991).

mRNA	Control	Iron Overload	
		Primary (i.e. haemochromatosis)	Secondary (i.e. transfusion)
Transferrin	+	++	++
Transferrin Receptor	+	-	-
Ferritin	+	++	++

Table 1.3 Coordinate regulation of hepatic iron proteins in haemochromatosis (Pietrangelo *et al* 1991)

Although few haemochromatosis patients meet objective criteria for orthotopic liver transplantation, the instances in which this has been carried out provide clues as to whether an intrinsic hepatic defect is present in haemochromatosis. In his review of the subject in 1992, Powell summarised the data of both normal livers transplanted into patients with haemochromatosis and the inadvertant transplantation of iron-

loaded livers from patients with haemochromatosis into patients with cirrhosis from other causes (Powell 1992). Twenty-two patients with haemochromatosis had received transplants. Hepatic iron concentration post-transplant was only measured in 4 of these (all of which were normal) although serum ferritins fell and remained low in all. Powell predicted that these patients should have demonstrated iron accumulation if the primary pathological defect only lay in the intestine, although he accepted that the mean follow-up period in these patients (2.8 years) was possibly too short. In 3 cases where an iron loaded liver has been transplanted into a normal individual, all demonstrated a reduction in hepatic iron concentration over a 6-12 month period and in one case intestinal iron absorption was reduced to 12% during this period (Powell 1992). Again, the period of follow-up is possibly too short from which to draw conclusions concerning a disease phenotype that usually presents late in life. Other individual case reports of iron loaded livers transplanted into normal recipients have been used to support the haemochromatosis defect lying in the liver (Koskinas *et al.* 1992) and elsewhere (Dabkowski *et al.* 1993). More data is needed.

Enterocyte

The recognition of excessive iron absorption in haemochromatosis has reasonably lead to the belief that the primary cellular defect in this condition lies within the epithelial cells lining the duodenum and upper small intestine (Powell *et al.* 1970). When further considering the possible molecular mechanisms that might be involved in this process, the polarised nature of the enterocyte must be respected.

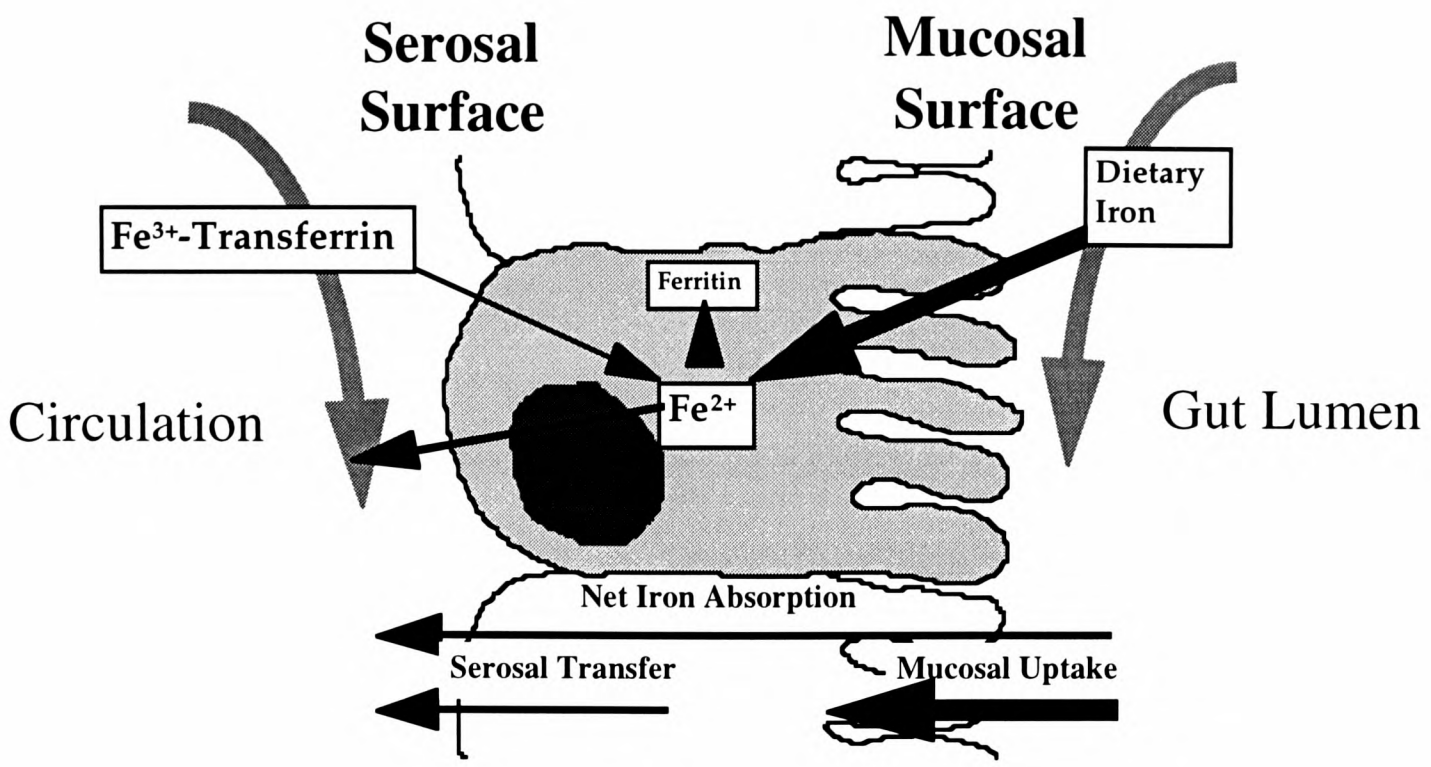


Figure 1.8 The intestinal epithelial cell (enterocyte) is exposed to different molecular forms of iron at its two surfaces.

Dietary iron takes two forms, haem iron and non-haem iron. Although haem iron (as in meat) constitutes the smaller fraction of dietary iron in omnivorous mammals it has been shown to be the more "bio-available" form being readily absorbed (for review of this subject see Skikne and Baynes 1994). Uptake of haem by the enterocyte is mediated by an intestinal haem receptor that has been characterised biochemically but not defined in terms of protein structure or expression. Non-haem iron is present in both animal and vegetable elements of the diet and is present in many molecular forms, the absorption of which may be influenced significantly by other dietary factors (Skikne and Baynes 1994).

In attempting to define the molecular defect in haemochromatosis, many researchers have focused their attention on the luminal (or mucosal) aspect of the enterocyte and have identified factors determining the uptake of dietary non-haem iron. Conrad and colleagues have defined an iron uptake pathway involving intestinal mucins; cell surface integrins and a new intracellular iron transport protein, Mobilferrin (Conrad and Umbreit 1993). Stremmel and co-workers have reported in abstract a new cell surface iron transport protein which appears to be up-regulated in

haemochromatosis (Stremmel *et al.* 1991). The genes encoding these new iron metabolism intermediaries have yet to be cloned and whether these represent the primary cellular defect in haemochromatosis remains to be proven.

Given the variable proportions of the different forms of iron present in the diet at any one time several different pathways for iron uptake by the enterocyte probably exist. Any one of these pathways might be abnormal in haemochromatosis.

Other studies have concentrated on the coordinate regulation of ferritin and the transferrin receptor in the duodenal enterocyte. In haemochromatosis there is a coordinated reduction in cellular ferritin and an increase in transferrin receptor message as detected by northern blot analysis (Pietrangelo *et al.* 1992).

mRNA	Control	Iron deficiency	Secondary iron overload	Haemochromatosis
Transferrin receptor	+	++	-	++
Ferritin	+	-	++	-

Table 1.4 Control of iron proteins in the enterocyte in haemochromatosis (Pietrangelo *et al* 1992)

The difference between this expression pattern in haemochromatosis and that seen in secondary iron overload implies that the enterocyte is behaving as if it (and the organsim as a whole) is iron-deficient. The other possible explanation was that in haemochromatosis there was a tissue specific aberration of ferritin synthesis. Subsequent measurements of the RNA binding activity of the IRP by gel band shift analyses suggested that in haemochromatosis a higher proportion of the IRP was in the RNA binding form reflecting intracellular iron deficiency (Pietrangelo *et al.* 1995a).

As the enterocyte is iron-deficient in haemochromatosis it is important to recognise that this observation counters the suggestion that the molecular abnormality must be on the luminal side of the cell. Any abnormality resulting in increased iron transport across an iron deficient enterocyte is most easily explained by excess iron

release from the serosal surface of the cell rather than inappropriate iron loading from the gut lumen.

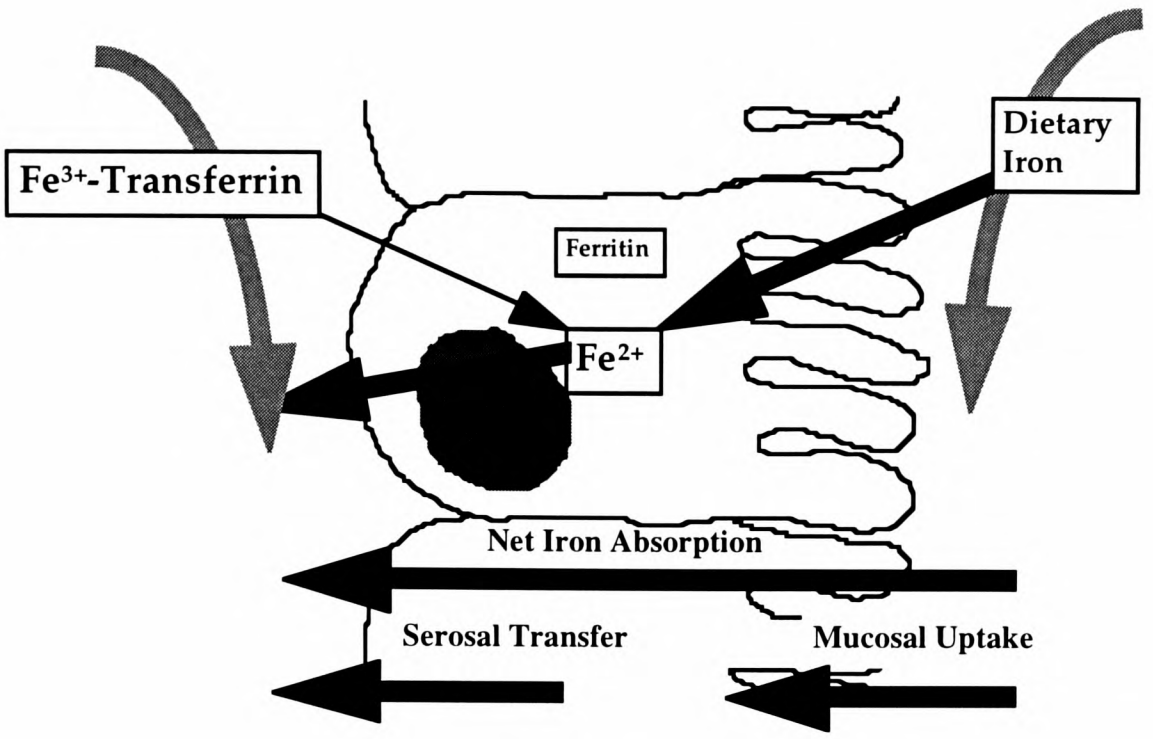


Figure 1.9 Increased net iron absorption in the presence of low intracellular iron suggests that the molecular defect results in excess iron release at the serosa.

Powell's original study of iron absorption had shown that the difference in net iron absorption resulted from an increase in body iron retention rather than mucosal uptake (Powell *et al.* 1970) and an excess release of iron from the serosal surface of the cell into the circulation in haemochromatosis has been suggested by dynamic measurements of radio-iron in patients (McLaren *et al.* 1991). The molecular abnormality causing this is yet to be identifed.

The reticuloendothelial system

Subjectively, many have reported a relative iron sparing of reticuloendothelial cells in haemochromatosis compared to secondary iron overload (McLaren 1989; Valberg 1978). In fact this feature was reported by Sheldon who based some of his pathological hypotheses on reticuloendothelial cell iron metabolism (Sheldon 1935). Some studies have suggested a quantitative difference in reticuloendothelial iron handling in haemochromatosis compared to other conditions of iron overload (Brink *et al.* 1976; Fillet *et al.* 1989), although studies more specifically directed to

identifying differences in reticuloendothelial cells in haemochromatosis have failed to unequivocally prove this suggestion (Adams *et al.* 1991; Bassett *et al.* 1982; Baynes *et al.* 1989; Bjorn-Rasmussen *et al.* 1985; Sizemore and Bassett 1984).

Animal Models

Many important questions regarding the control of intestinal iron absorption remain unanswered and attempts have been made to address these in animal models (Iancu 1993; Raja *et al.* 1994). No entirely satisfactory animal model for haemochromatosis exists.

Hypotransferrinaemic mice become iron loaded but the pathological features of these animals are recognised to differ from haemochromatosis (Simpson *et al.* 1993). By definition, the increased gastrointestinal iron absorption in hypotransferrinaemia is not persisting in the face of saturated circulating transferrin as is the case in haemochromatosis.

β_2-microglobulin deficient mice absorb excess dietary iron and this is deposited in organs in a pattern similar to that seen in haemochromatosis (de Souza *et al.* 1994; Rothenberg and Voland 1996). Without details of transferrin saturation and other haematological indices the mechanism by which these animals accumulate iron cannot be explained by present knowledge of iron metabolism.

Why map the haemochromatosis gene?

Identification of the haemochromatosis gene might facilitate the early diagnosis in individuals at risk of iron overload. Treatment is simple, safe and effective. Although the biology is complex, the gene has been mapped to the short arm of chromosome 6, in, or near to, the MHC. This makes haemochromatosis a candidate for positional cloning.

Positional cloning

Positional cloning (often referred to as "reverse genetics") relies on the identification of a disease gene on the basis of its chromosomal position, with the gene product deduced subsequently. None of the genes coding for the recognised iron proteins map to chromosome 6.

Protein	Gene location
Transferrin	3q21-qter
Transferrin receptor	3q26.2-qter
Ferritin	H subunit Ch 11 L subunit Ch 19
Iron response protein	Ch 9
Erythroid ALAS	Xp21-q21

Table 1.5 None of the recognised iron protein genes maps to the short arm of chromosome 6.

The success of a positional cloning strategy depends on the precise positional localisation of a disease locus. If the positional information is accurate, then the disease gene may be identified even if this ultimately proves to be a novel genetic mechanism in a novel gene (as has recently been demonstrated by the cloning of the gene responsible for Friedrich's ataxia Campuzano *et al.* 1996).

In the case of haemochromatosis the strong linkage with HLA-A implied a position close to the telomeric end of the MHC and groups worldwide adopted positional cloning (Chapter 2.1).

Chapter 2

A haemochromatosis gene region

Contents - Chapter 2

Chapter 2.1

Background: Positional cloning and haemochromatosis

Positional cloning and haemochromatosis

Positional cloning refers to the identification of a disease gene on the basis of its chromosomal location. This is achieved by three inter-related strands of investigation:

Genetic mapping

Physical mapping

Identification of genes

Genetic mapping involves the identification of genetic markers inherited in linkage with the disease phenotype. Physical mapping places the genetic markers into a physical order with respect to chromosomal landmarks (e.g. the centromere or other known genes/markers). The combination of genetic and physical mapping enables the definition of a disease gene candidate region on the basis of a peak of linkage. Such a region is then searched for candidate genes.

In theory, the only essential element for a positional cloning strategy to succeed is a clear and concise definition of the gene position. In practice, many of common disease genes that had been considered exemplary candidates for identification by a positional cloning strategy have ultimately been identified through association with chromosomal rearrangements (e.g. *TSC2* and tuberous sclerosis, The European Chromosome 16 Tuberous Sclerosis Consortium 1993, *PKD1* and adult polycystic kidney disease, The European Polycystic Kidney Disease Consortium 1994), the presence of expanded triplet repeats (Huntington disease, Group 1993) or predicted biological functions (e.g. CFTR and cystic fibrosis Riordan *et al.* 1989) effectively shortening the path to the gene. The recent identification of the genetic mutation resulting in Friedrich's Ataxia (Campuzano *et al.* 1996) suggests that, even in the absence of a clear biological model of a given disease phenotype, accurate positional information can be enough to define the responsible gene.

The key to positional cloning is knowing when the positional data is good enough to support an effective search for candidate genes.

Haemochromatosis and genetic mapping

HLA serotypes

In 1975 Simon and coworkers in Rennes, Brittany, demonstrated a clear association between haemochromatosis and the HLA class I serotypes HLA-A3 and HLA-B14 (first reported in English in Simon *et al.* 1976). This had two fundamental effects. First, it confirmed Sheldon's conviction that haemochromatosis represented an inborn error of metabolism and with further studies the autosomal recessive mode of inheritance was confirmed (Simon *et al.* 1977). Furthermore, confirmation of linkage to HLA-A implied a location for the responsible gene close to, possibly within, the class I region of the human Major Histocompatibility Complex (MHC), on the short arm of chromosome 6.

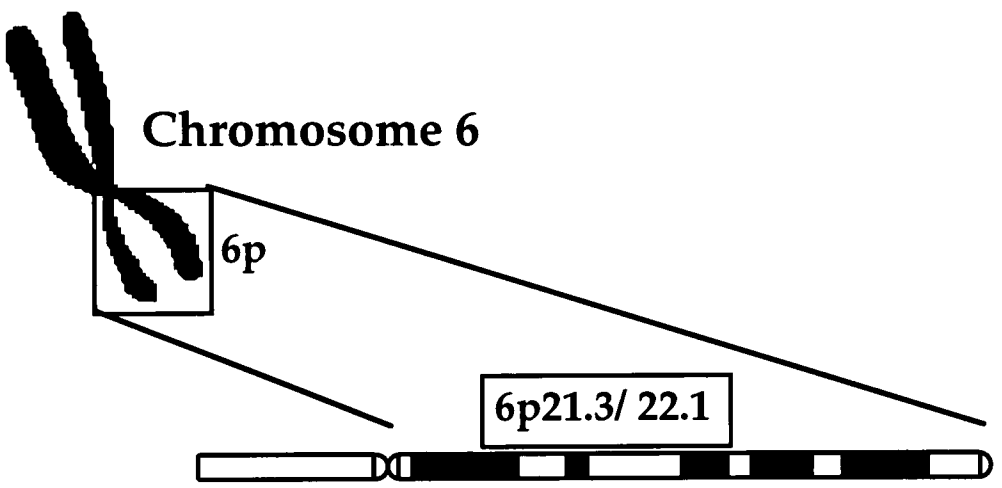


Figure 2.1 Linkage to HLA-A3 placed the haemochromatosis gene on the short arm of chromosome 6.

Genetic markers

HLA serotypes are "surrogate" genetic markers in that they reflect gene products rather than genomic DNA. In the case of haemochromatosis the polymorphism of HLA serotyping provided a very informative tool for population genetics before the application of the emerging technologies of pulsed-field gel electrophoresis (PFGE) and DNA restriction fragment length polymorphism (RFLP).

Attempts to define chromosomal rearrangements by PFGE failed to identify disease associations using four different restriction enzymes and eight DNA probes from the class I region (Lord *et al.* 1990). This suggested that the haemochromatosis

allele was not associated with a rearrangement that would be detectable at this resolution although the choice of probes from within the MHC make the prior assumption that the gene lies within this complex.

Jouanolle found a 14kb EcoRI fragment that appeared more commonly in haemochromatosis than in HLA-A3 controls whilst examining restriction fragment length polymorphisms using two HLA class I probes (Jouanolle *et al.* 1990). Again, this assumes a position for the haemochromatosis gene close to HLA-A. The prospective identification of further anonymous biallelic markers from the HLA-A subregion identified one (i82) which was associated with haemochromatosis, although not as strongly as HLA-A3 (Boretto *et al.* 1992).

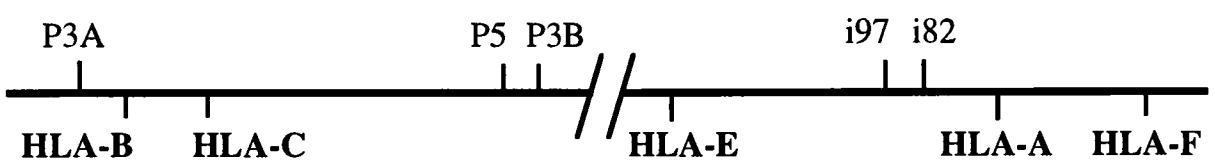


Figure 2.2 Biallelic markers in the MHC class I region (Boretto *et al* 1992)

The conclusion drawn from these data was that the gene either lay within 100kb to the centromeric side of HLA-A or on its telomeric side. These findings were considered sufficient to support a search for candidate genes around HLA-A and the assumptive designation of these loci as "haemochromatosis candidate genes" (ElKahloun *et al.* 1993).

Microsatellites

Two important technological steps lead to the development of a new class of highly polymorphic genetic markers. The first was the identification of regions of DNA sequence repetition or "minisatellites" in which multiple alleles varying in size could be characterised by the use of repeat sequence specific probes (Jeffreys *et al.* 1985). The second was the development of the polymerase chain reaction (PCR), enabling the amplification of specific DNA loci with knowledge of the flanking sequence (Saiki *et al.* 1985). These two strands of technology converged in the report

of PCR amplification of "microsatellites" containing variable lengths of (dC-dA)$_n$·(dG-dT)$_n$ blocks (Litt and Luty 1989; Weber and May 1989). Alleles at these loci could be demonstrated to be inherited in a Mendelian codominant manner and the combination of the potential number of alleles at each locus and the dispersed nature of such loci throughout the human genome lead to the speculation that these might represent a very powerful tool for genetic mapping.

The most common form of microsatellite is the dinucleotide CA repeat which occurs on average every 30kb in genomic DNA. Primers to amplify individual microsatellite loci may be designed to the unique flanking sequence. Size variation in the PCR products derived from human genomic DNA reflect different allelic forms of the microsatellite locus and may be separated by electrophoresis.

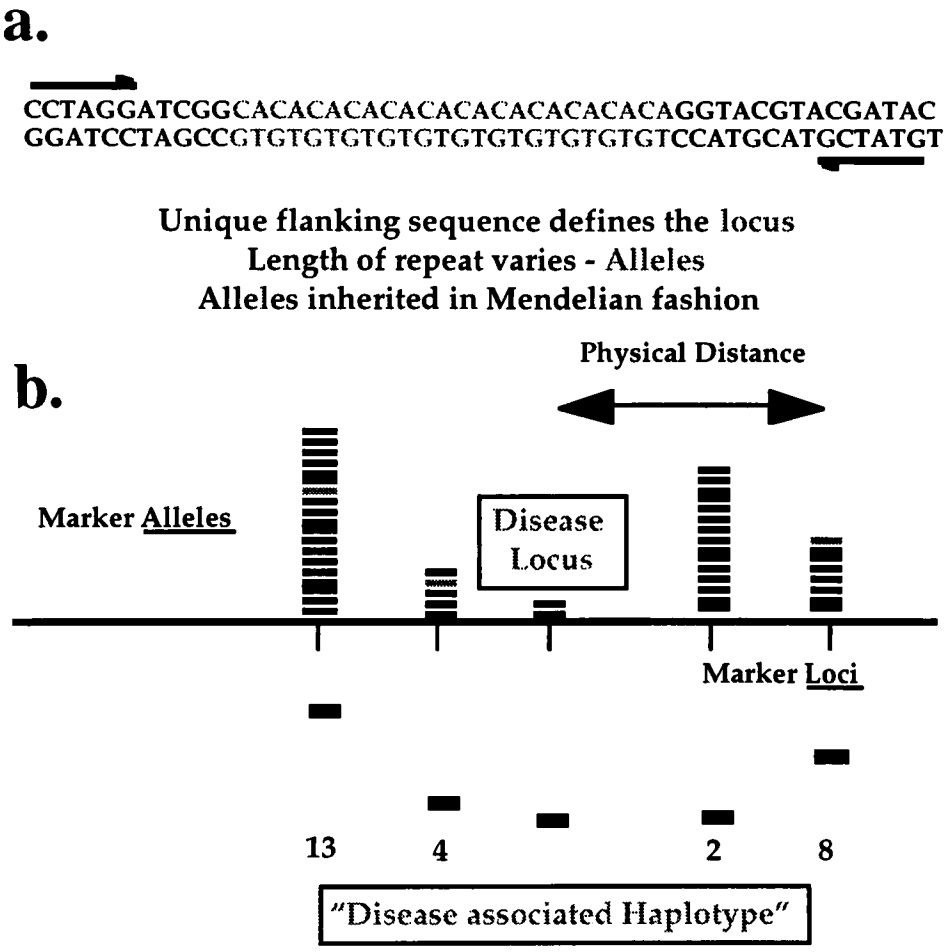


Figure 2.3 Microsatellite loci. Panel a. The principle of designing PCR primers to sequence that flanks a simple sequence repeat (CA). The amplified alleles at any one locus vary in size. Panel b. Identification of microsatellites flanking a disease locus will generate specific alleles associated with the disease.

This approach has greatly advanced mapping of the haemochromatosis gene by contributing polymorphic markers that permitted mapping beyond HLA-A. The first reported use of microsatellite markers in haemochromatosis came in 1993 when

Jazwinska and colleagues in Brisbane reported the analysis of microsatellite markers, known to map to the short arm of chromosome 6, in patients with haemochromatosis (Jazwinska *et al.* 1993). An allele of one of these markers, D6S105 allele 8 (124bp), showed significant linkage to haemochromatosis with a LOD score comparable to that with HLA-A3. Furthermore in an accompanying allele association analysis they demonstrated that allele 8 of D6S105 inferred a greater relative risk of haemochromatosis than HLA-A3.

	HC patients (n = 82)	Controls (n = 82)	p	Allele relative risk
HLA-A3	62%	26%	p<0.001	4.78
D6S105/**8**	93%	21%	p<0.0001	48.43

Table 2.1 Association between the HLA serotype HLA-A3 and allele 8 of the microsatellite D6S105 with haemochromatosis (Jazwinska *et al* 1993)

This suggested that D6S105 might be a closer marker for the disease than HLA-A. D6S105 had been mapped telomeric to the MHC (Weber *et al.* 1991) but the physical distance between the two was not known.

Physical mapping

The order and distance between genetic markers may sometimes be inferred by genetic studies but physical mapping is essential to substantiate both.

Pulsed field mapping

Traditionally physical maps were generated by the arduous process of pulsed field gel electrophoresis of restriction enzyme digested genomic DNA and the identification of overlapping restriction fragments with single copy DNA probes. This process proved of limited success at the telomeric end of the MHC due to the lack of suitable probes and the complexity of restriction fragments generated. Using more limited source DNA from chromosome 6 radiation hybrid cell lines Gruen was able to extend the physical map of the MHC up to 1Mb telomeric to HLA-F, the most telomeric of the MHC class I genes (Gruen *et al.* 1992).

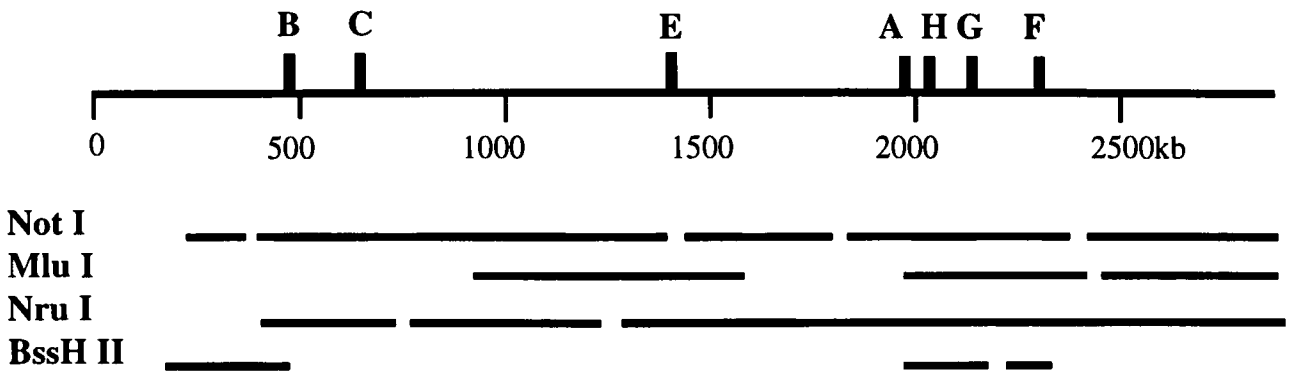


Figure 2.4 The physical map of the HLA class I region (Gruen *et al* 1992)

On the basis of the genetic distances between haemochromatosis and HLA-A and that between each of the class I genes (Koller *et al.* 1989), Gruen considered HLA-F to be the likely telomeric limit to the disease gene candidate region (Gruen *et al.* 1992).

Yeast artificial chromosomes

Physical mapping, and the positional cloning for human disease genes, has been greatly supported by the use of yeast artificial chromosomes (YACs). From the first report of the propagation of human genomic DNA in linear yeast artificial chromosomes the potential for covering large distances in higher genomes increased ten-fold (Burke *et al.* 1987). The great advantage that YACs brought to recombinant DNA technology was the fact that much larger fragments of human DNA could be cloned and stably propagated in lower organisms, maintained as linear chromosomes.

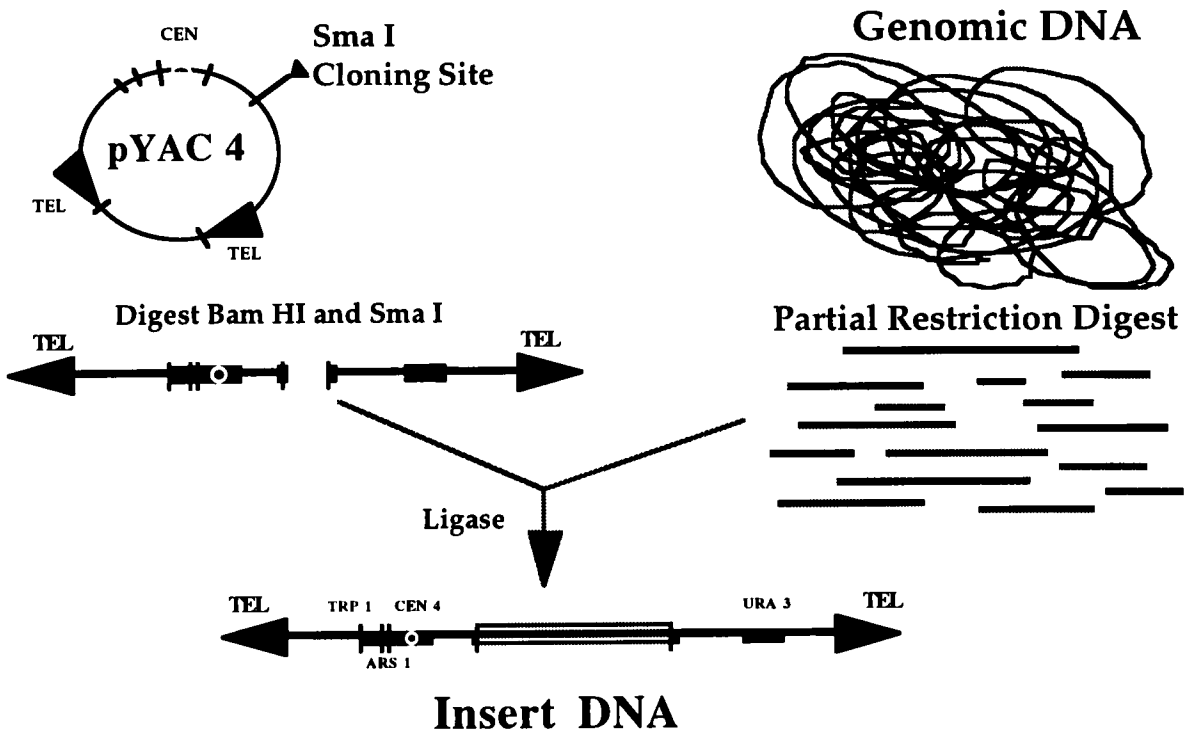


Figure 2.5 The pYAC4 vector.

This led to the construction of human YAC libraries comprising thousands of individual clones, containing variable sized genomic fragments that collectively covered the whole of the human genome (Brownstein *et al.* 1989; Larin *et al.* 1991).

At the time, genomic DNA cloned in YACs conveniently filled a very important gap between the resolution of genetic mapping (to approximately 1Mb) and the largest fragments stably propagated in bacterial cells (i.e. ~40 kb in Cosmids). Hence YACs would allow the rapid generation of physical maps to compliment the increasingly complex genetic maps of microsatellite markers. Soon, contiguous YAC clones covering specific genomic regions (YAC contigs) such as the MHC were used to generate high resolution restriction maps (Abderrahim *et al.* 1994) and more importantly provided cloned DNA in which to look for genes (ElKahloun *et al.* 1993).

The utility of YACs did not come without a significant cost in terms of technical drawbacks. The first problem was that of the low copy number (i.e. one) of the cloned DNA compared to that maintained in independantly replicating plasmids and their derivatives. Clone identification by hybridisation proceedures became significantly less sensitive when applied to pooled DNA from YAC clones. This lead to the use of PCR based screening strategies that were able to identify the target sequences independant of template DNA concentration (Green and Olson 1990).

Using primers to the microsatellite D6S105, Jenny Pointon screened the human YAC libraries and identified a YAC from the ICRF YAC library containing this locus (K43B3). Using dual labelled fluorescence *in-situ* hybridisation (FISH) Caroline Stone subsequently demonstrated that the YAC, and hence D6S105 lay, telomeric to HLA-A at a physical distance of 1-2Mb (Stone 1995). D6S105 lay in an unmapped and uncloned region, telomeric to the MHC. More importantly, it suggested that the haemochromatosis gene lay outside the region that was being studied for candidate genes.

The concept of sequence tagged sites

The need to screen YAC libraries by PCR had very important and far reaching consequences for physical mapping. PCR is based on the amplification of specific

genomic loci through knowledge of relatively small fragments of sequence data (40-60 nucleotides in total). The process is reproducible by the sharing of only that fragment of sequence information contained in the primers. It is both independant of probes or clones and transferable to other experimental settings. Each STS need not be polymorphic (c.f. microsatellites) but serves as a genomic landmark.

Despite initial reservations regarding the cost and technical challenge of the acquisition of sequence data, Olson predicted that the generation of anonymous "sequence tagged sites" (STS) would constitute a very powerful facility for genome mapping (Olson *et al.* 1989). PCR screening of genomic libraries and STS content analysis of to confirm physical overlap between clones has rapidly become the accepted method of physical mapping.

The inevitable link between genetic and physical mapping

Although it may have been inferred that the logical sequence of events in positional cloning is a progression from genetic mapping through physical mapping to the identification of gene sequences, the three are closely inter-related and at times are conducted simultaneously. Physical mapping by YAC contig construction provides both a map onto which to locate genetic markers and a resource of cloned DNA from which to isolate new microsatellites and ultimately genes. By such a complimentary approach a disease gene candidate region can be progressively refined until a rational search for genes can be initiated.

Strategy

For the identification of the haemochromatosis gene by positional cloning to be possible three things were necessary.

Define a telomeric boundary to the haemochromatosis gene region

Clear centromeric and telomeric boundaries for the disease gene region needed to be defined and until this had been achieved the isolation of gene sequences on the

basis of what positional information was already available would be, at best, speculative.

Complete a YAC contig across the region

To establish a YAC contig between HLA-A and D6S105 was the essential first step toward mapping the haemochromatosis gene. The distance between these two points was estimated to be 1-2Mb and chromosome walking using YACs with the generation of new STSs from the ends of YAC clones would allow the interval to be bridged.

More microsatellite markers from the region

Once the region had been cloned in YACs, the isolation of more microsatellites would allow refinement of the likely position of the gene.

Chapter 2.2

Construction of a yeast artificial chromosome (YAC) contig extending telomeric to the human major histocompatibilty complex

Background

Haemochromatosis was more strongly associated with D6S105 than HLA-A suggesting that the gene lay close to the latter (Jazwinska *et al.* 1993). A YAC contig from HLA-A to D6S105 and beyond would greatly facilitate the physical mapping of this region and generate a source of cloned DNA from which new markers could be isolated.

In 1992 a YAC contig was published covering the telomeric end of the MHC including the 300kb between HLA-A and the most telomeric of the Class I genes, HLA-F (Geraghty *et al.* 1992). Sequence from the right arm of the most telomeric of the YAC clones in this contig (A146G11 from the St. Louis YAC library) generated the STS V6 (Jenny Pointon). Library screening with V6 had identified a single non-chimaeric YAC, B188F3, from the St. Louis library.

B188F3 was used as the starting point for a sequential telomeric walk using YACs to cover a physical distance that had been estimated to be between 1 and 2Mb (Stone *et al.* 1994; Volz *et al.* 1994).

Methods

Chromosome walking

The principle of chromosome walking with YACs is illustrated.

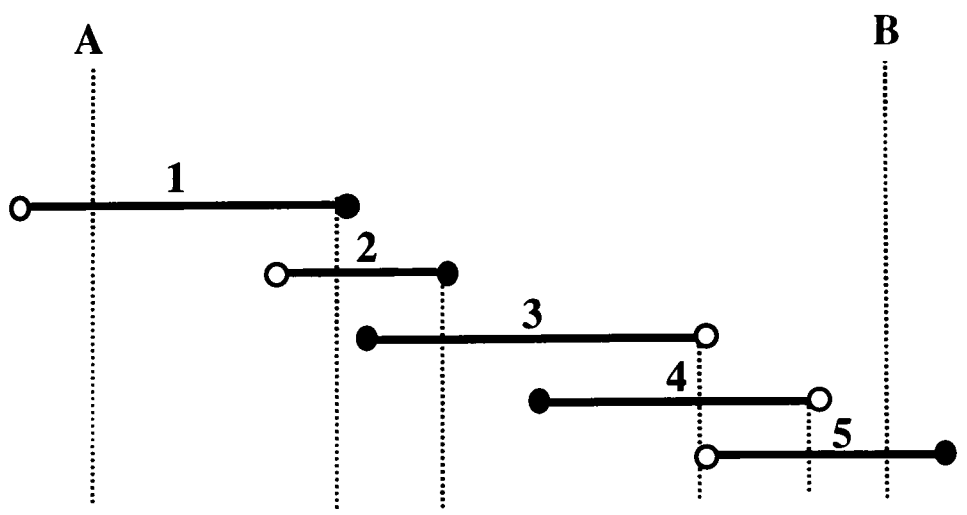


Figure 2.6 Sequential "walking" with yeast artificial chromosomes. Each YAC insert is orientated relative to the right (filled circle) and left (open circle) arms of the vector.

Starting with a YAC clone containing point A, end sequence from the clone is obtained generating an STS identifiable by PCR. The primers to this STS are then used to screen the YAC library to identify another clone containing overlapping insert genomic DNA. This process is repeated until such a time as a clone is identified that contains point B.

Screening YAC libraries

Human YAC libraries comprise thousands of individual yeast clones each containing an artificial chromosome. As described earlier the low copy number of cloned DNA to host means that screening of the YAC libraries is often done by PCR.

Two YAC libraries were held in the Institute of Molecular Medicine. The St. Louis or Washington library was one of the first human YAC libraries reported (Brownstein *et al.* 1989). The ICRF library was constructed to maximise the number of clones with large inserts (Larin *et al.* 1991). Each library has a hierarchical pooling, requiring 2 to 3 rounds of PCR on pooled DNA before identifying the 96 well plate containing the clone of interest.

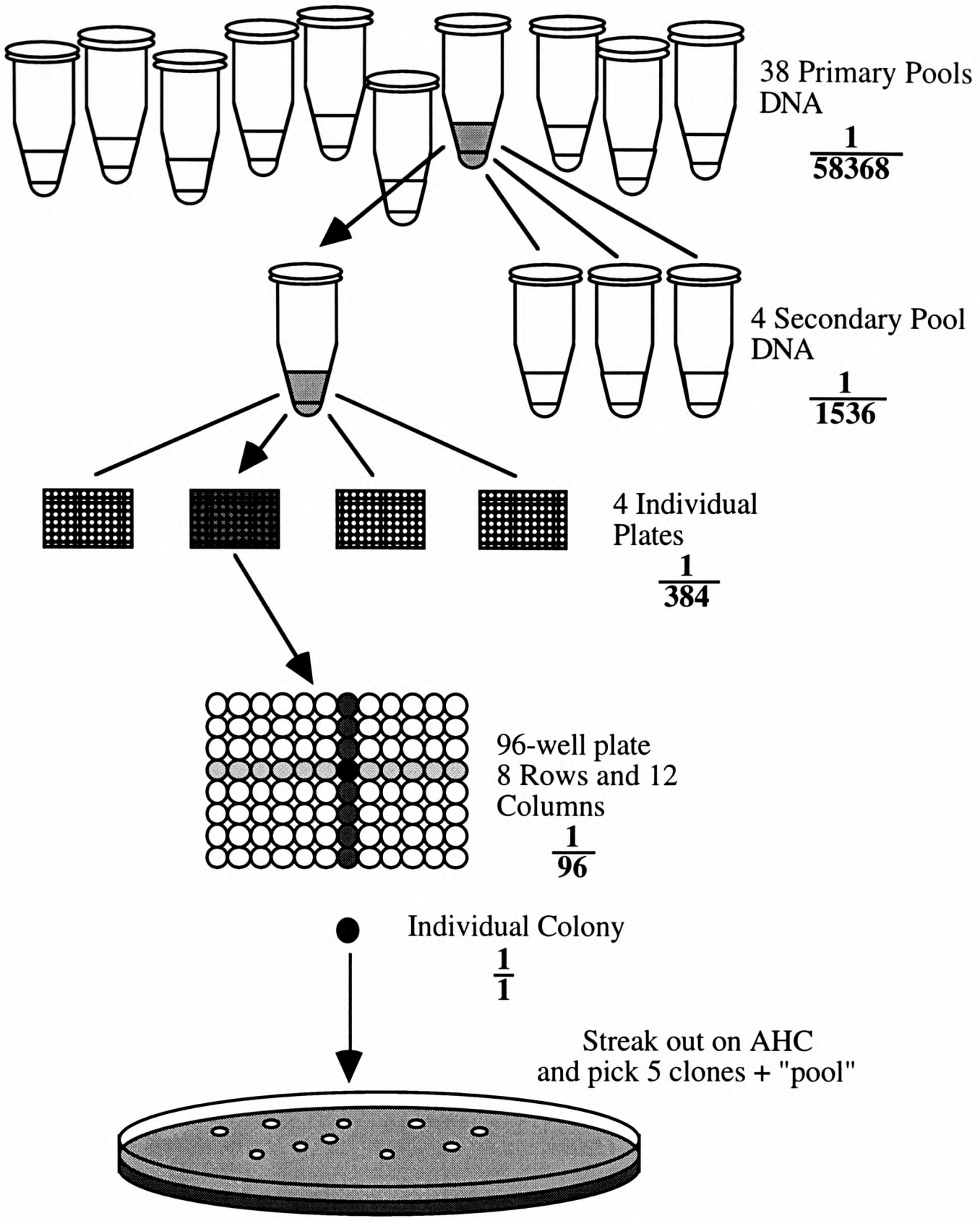


Figure 2.7 Screening the St. Louis YAC library.

At the stage of the single, multiwell plate we have used "rows and columns" PCR to identify the individual clone. This approach involves stamping the grid onto AHC agar. After incubating the plate at 30°C long enough to establish healthy colonies a small sample of each colony is picked and pooled in 100µl of water or T.E. buffer along with the other colonies making up that row or column. PCR, using lysate from boiled yeast, identifies a single row and a single column defining the position of the required clone from that plate.

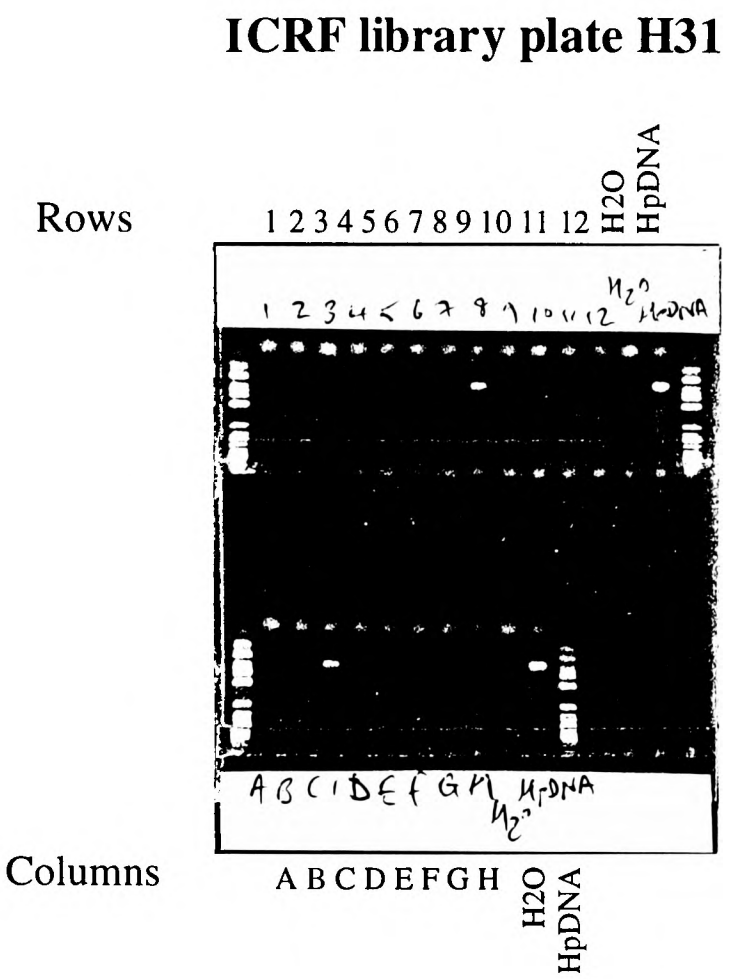


Figure 2.8 "Rows and columns" PCR with the STS V11 on plate H31 from the ICRF library identifying the grid reference C8 as containing the sequence of interest. The size marker is 1kb ladder.

The colony identified as containing the sequence of interest is picked and streaked onto AHC agar. Five individual clones and a pool sample are picked from that plate and the PCR result is ultimately confirmed on prepared DNA from all the samples.

Sizing YAC inserts

The size of the human genomic inserts within the human YAC libraries varies from 100kb to over 1Mb in some libraries. Each insert can be sized by pulsed field gel electrophoresis to separate the host yeast chromosomes. The YAC can often be seen as an additional chromosome band on such a gel after staining with ethidium bromide. However, the integrity of any given YAC clone can only be established by Southern transfer of the DNA to a nylon membrane and subsequent hybridisation of the resulting blot with probes specific to each of the arms of the YAC vector.

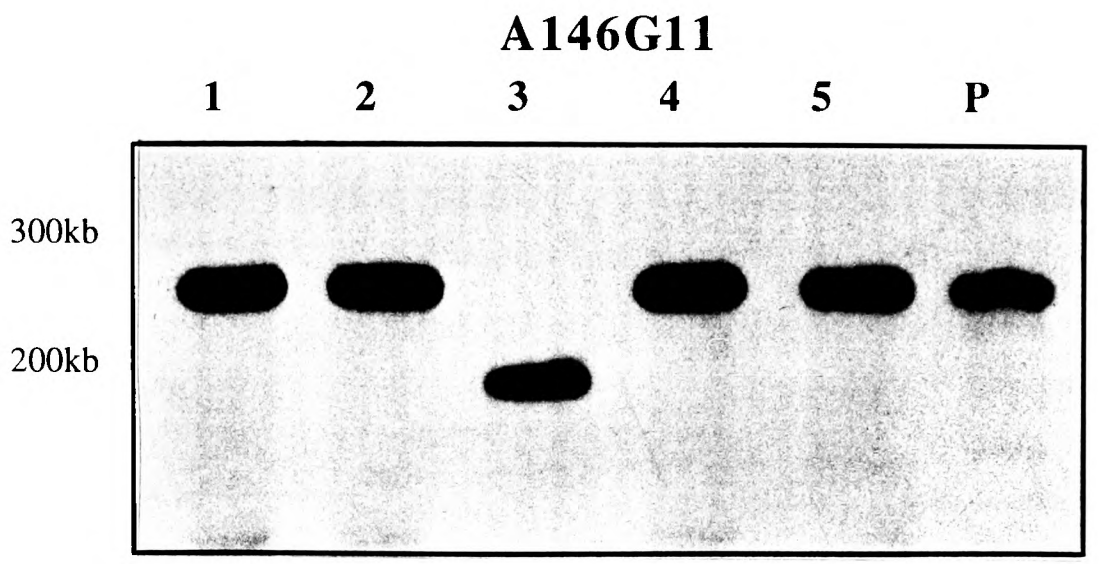


Figure 2.9 A YAC, separated from the host yeast chromosomes by PFGE (1% agarose, 0.5x TBE, 180 volts, switching times 30-120 secs, total run time 36 hrs), transfered to nylon membrane by Southern blotting and probed with a radiolabelled probe to the pYAC4 vector left arm. In this example clone 3 has undergone a rearrangement (probable deletion) resulting in a reduction in the size of the YAC

Such analysis establishes the presence of a single YAC in each clone, confirms YAC stability and allows calculation of the size of the insert by comparison to size standards run in the gel simultaneously.

Isolating YAC insert end sequence

Although it may be possible to rapidly generate insert end probes for hybridisation, the use of PCR to screen the YAC libraries lead us to isolate sequence from the ends of any given YAC insert from which we could design PCR primer pairs. Many methods have been reported for achieving this. We used a vectorette strategy, utilising the known sequence in the YAC vector arms to amplify the ends before either direct sequencing or cloning into a suitable plasmid vector. The vectorette strategy is outlined.

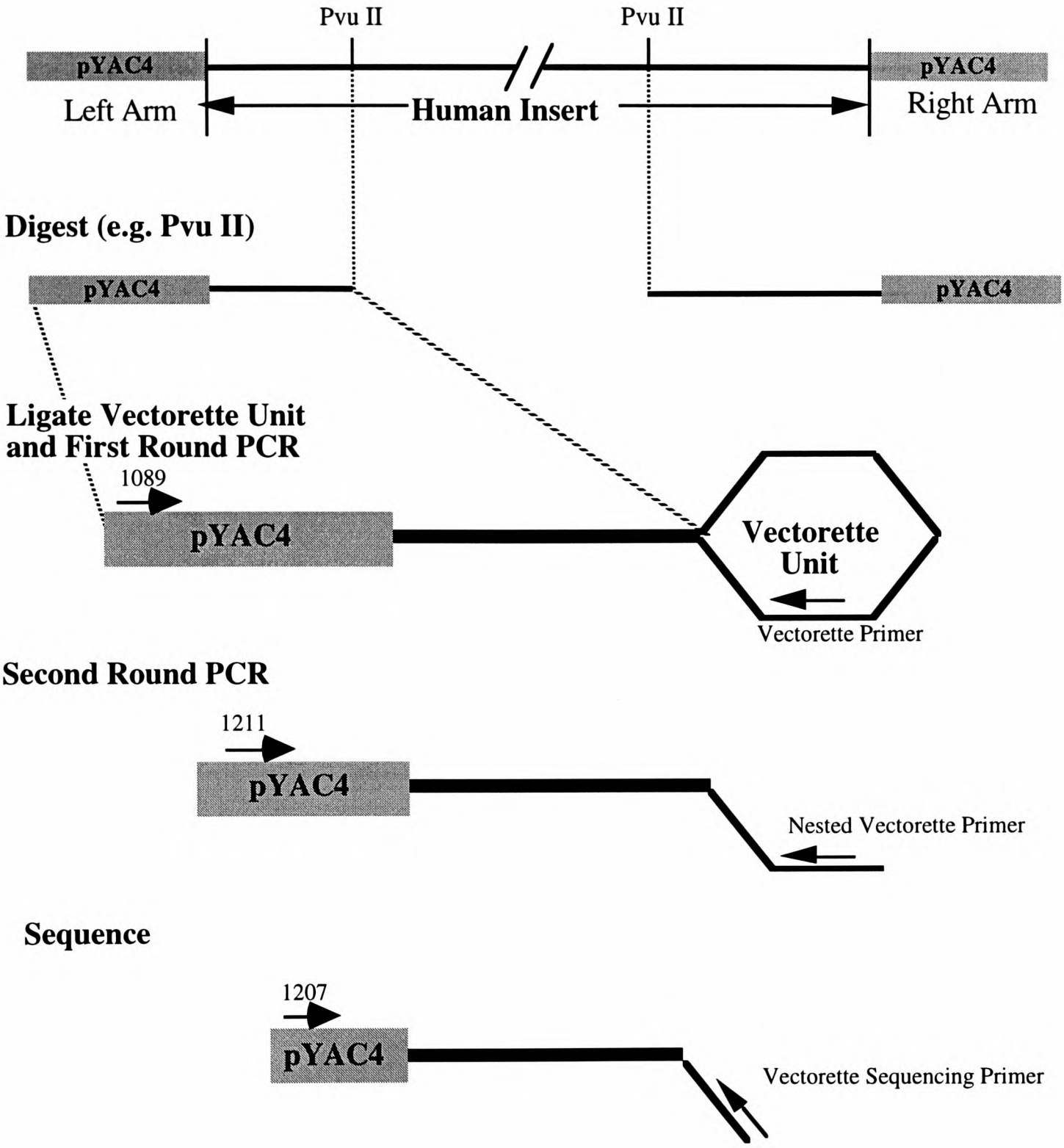


Figure 2.10 The vectorette strategy for isolation of insert end sequence from YACs

For each YAC, total yeast DNA (i.e. including the YAC) is digested with 4 restriction enzymes (*Alu* I, *Pvu* II, *Rsa* I and *Eco* RV). A vectorette unit is ligated to the resulting blunt ends of the digested DNA and two rounds of PCR are performed with nested primers to both the vectorette unit and each of the two YAC vector arms.

The second round PCR products are resolved by gel electrophoresis in 2% agarose and visualised after staining with ethidium bromide. Incorporation of a biotinylated second round PCR primer enabled direct sequencing of clean single band

second round PCR products after separation of the DNA strands by alkali denaturation and capture of single strand DNA on streptavidin coated magnetic beads (Dynabeads, Dynal). Alternatively PCR products were cloned into a suitable plasmid vector (pCR Script SK+, Stratagene) and sequenced. All insert sequence was orientated relative to the *Eco* RI cloning site of the pYAC4 vector and subjected to database comparison using FASTA of the Wisconsin GCG data analysis software identifying repetitive elements such as Alu and LINE repeats. PCR primers were designed to unique non-repetitive DNA.

If the sequence of a given YAC insert immediately adjacent to the vector arms contains one of the many repetitive elements (Alu, LINE, SINE etc.) then the design of PCR primers is precluded due to the likely false positive rate from further library screening. This problem is inherent to all forms of physical mapping in human genomic DNA and provides a further independant argument for a strategy utilising STSs. The generation of sequence data from the end of YAC inserts allows for a intermediary step of sequence analysis which rapidly identifies the common Alu and LINE elements.

Orientation of each YAC clone

PCR primers designed to the two ends of each YAC insert enabled orientation of the clone relative to other YACs in the contig by STS content analysis. On this basis the end of the insert extending out beyond the established contig is considered to extend the chromosome walk.

Clone	STS			
	1	2	3	4
A	+	+	+	
B		+	+	+

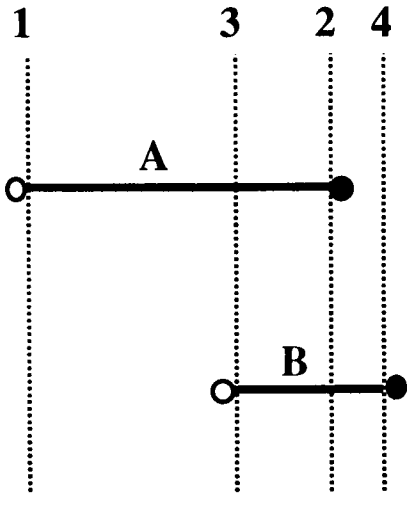


Table 2.2 and Figure 2.11 STS content analysis and its conceptual translation

Chimaeric and co-transformed clones

The major limitation to progressive chromosome walking in this way is the presence of chimaeric YAC clones within the library. Chimaeric clones stem from ligation of two or more genomic fragments to each other prior to incorporation into the YAC vector during creation of the library. If a chimaeric clone is identified by PCR (for example representing its chromosome 6 fragment) isolation of end sequence might result in subsequent steps of the chromosome walk being transferred to another chromosome. Identification of clone chimaerism is achieved by FISH to normal human metaphase chromosome spreads. This confirms the cytogenetic location of the clone and any signal from sites other than that of interest suggests the presence of insert chimaerism.

The second major limitation to sequential walking is the presence of co-transformed clones within the library. Co-transformation results in two or more YACs being present within any single yeast clone. This provides a similiar limit to chromosome walking although the presence of co-transformed clones is usually apparent after PFGE as two separate YAC bands will be demonstrated.

Restriction mapping

Each YAC clone may be restriction mapped. Total yeast DNA is digested with rare cutting restriction endonucleases and the digestion products resolved by PFGE. The DNA is transferred to a nylon membrane which is sequentially hybridised with probes specific to the two YAC arms. The size of fragments identified with each of these probes allows calculation of the distance of each site from the YAC arm.

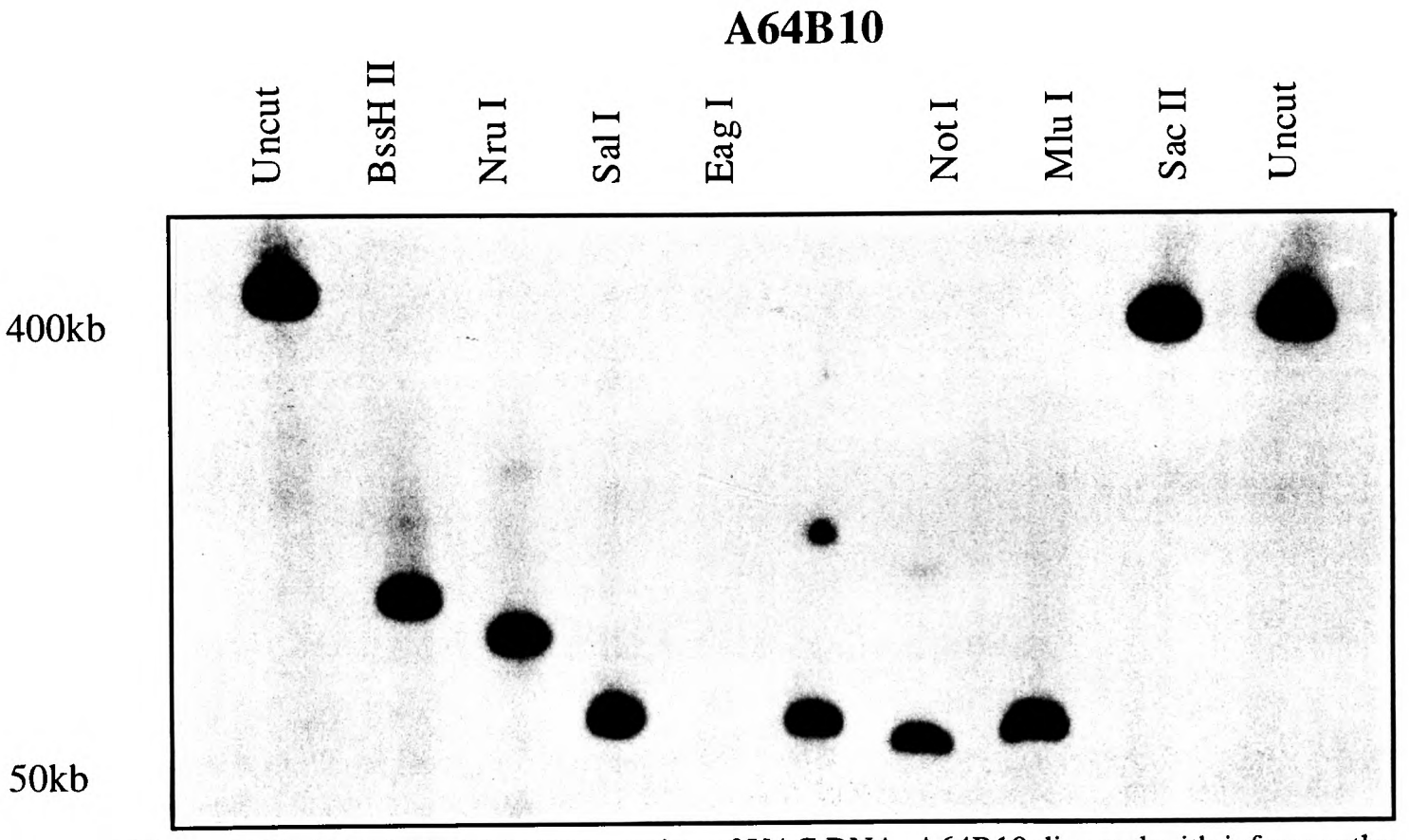


Figure 2.12 Complete restriction mapping of YAC DNA. A64B10 digested with infrequently cutting restriction enzymes. Restriction fragments resolved by PFGE (1% agarose, 0.5x TBE, 180 volts, switching times 10-45 secs, total run time 24 hrs). DNA transferred to nylon membrane and then probed with YAC right arm.

Internal restriction sites are identified by partial digestion with limiting dilutions of the enzyme.

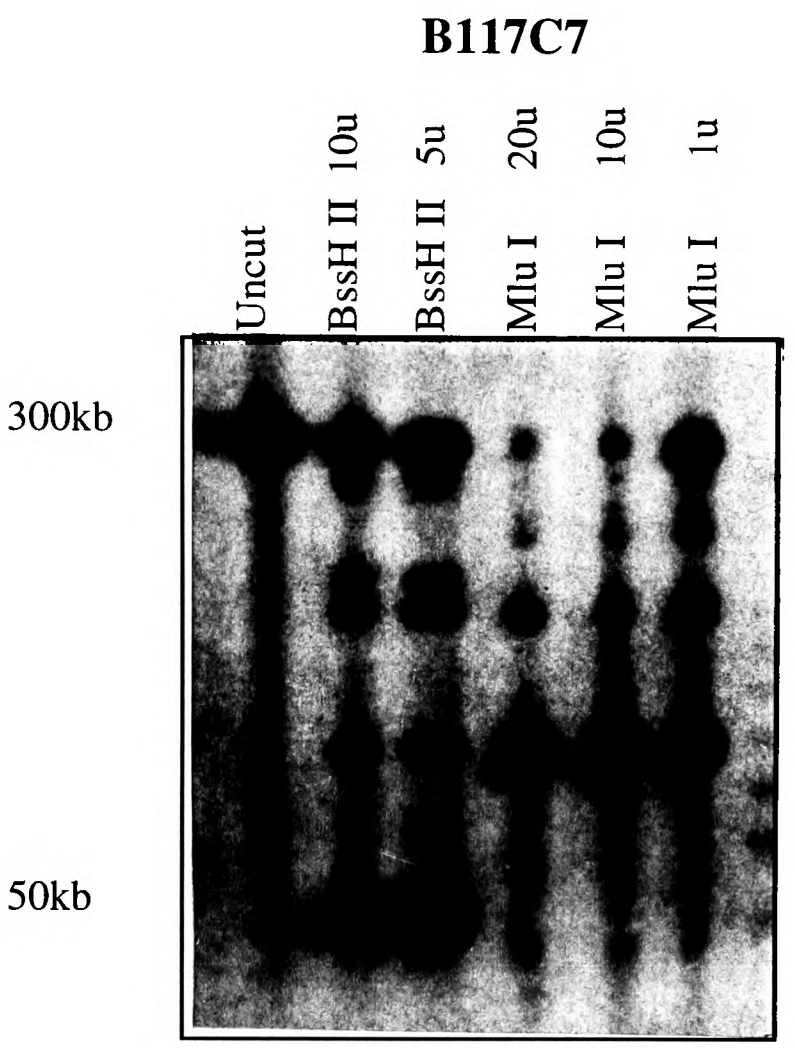


Figure 2.13 Partial restriction mapping of YAC DNA. B117C7 digested with limiting dilutions of the enzymes *BssH* II and *Mlu* I. Restriction fragments resolved by PFGE (1% agarose, 0.5x TBE, 6 volts/cm, switching times 1-45 secs, total run time 24 hrs). DNA fragments transferred to nylon membrane and probed with YAC left arm probe.

Restriction maps of individual YAC clones from the contig can be compiled to generate a consensus restriction map of the region.

Results

YAC contig construction

B188F3 provided the starting point of the contig. It has an human insert of 150kb but does not contain HLA-F by PCR. The two ends of the YAC insert were sequenced using vectorette technology to generate the STSs V10 (left arm) and V11 (right arm). By PCR using A146G11 DNA as a template it was demonstrated that the orientation of B188F3 was such that its right arm was telomeric.

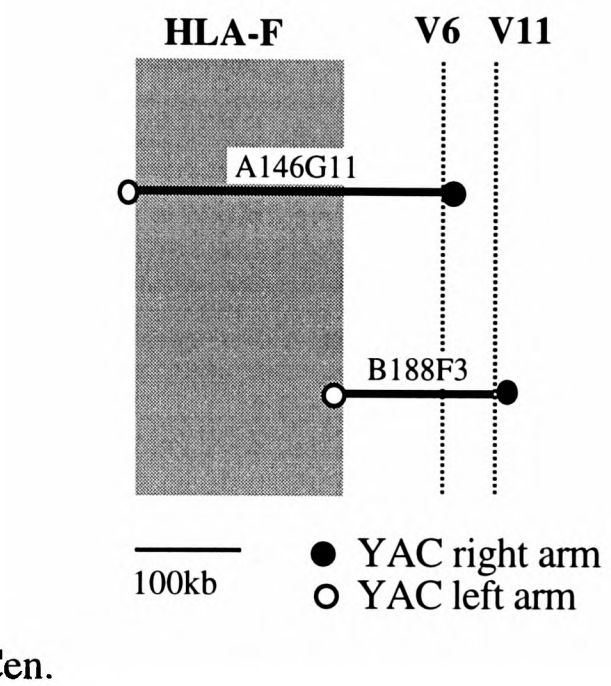


Figure 2.14 B188F3 extends telomeric to the HLA-F

V11 was used to screen the YAC libraries and 4 further YAC clones were identified (A79B7 and A64B10 from the St. Louis library and H31C8 and L46B8 from the ICRF library). Apart from L46B8 all these clones subsequently proved to overlap with the telomeric end of A146G11 and were positive by PCR with V6. L46B8 lacked this STS despite being positive with PCR primers to the two ends of B188F3 suggesting the possibility of an internal deletion and in combination with being unable to clearly define the size of this clone by PFGE it was dropped from further study. A79B7 has an insert size of 370kb and its telomeric left arm generated V15 which on screening the libraries only identified one further small clone B98H7.

Of the other clones H31C8 has an insert of 800kb. This clone was non-chimaeric by FISH (Lyndal Kearney). The two ends of the insert were sequenced. The left arm of H31C8 yielded sequence containing Alu repetitive elements although PCR

with the STS V16 primers identified A146G11, B188F3, A79B7 and A64B10 defining the orientation of the YAC. The right arm vectorette products generated a large *Eco* RV fragment from which the STS V17 was derived. The V17 product was subsequently cloned into pCR Script (SK+) to allow confirmation of its entire 727bp of sequence.

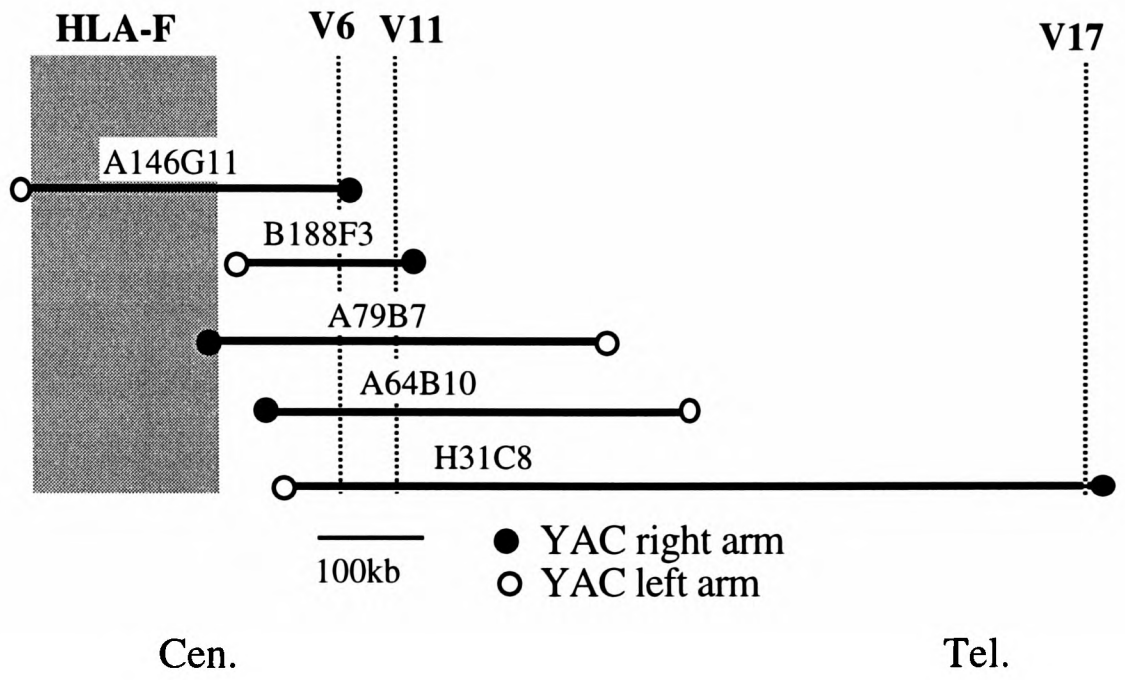


Figure 2.15 Of the 3 YACs identified with V11, H31C8 extended the contig

Library screening with V17 showed great promise with the identification of 6 further YAC clones from the two libraries (F22C7 from the Monaco library and A102H5, A84A5, A216A4, B223A5 and B225A10 from the St. Louis library).

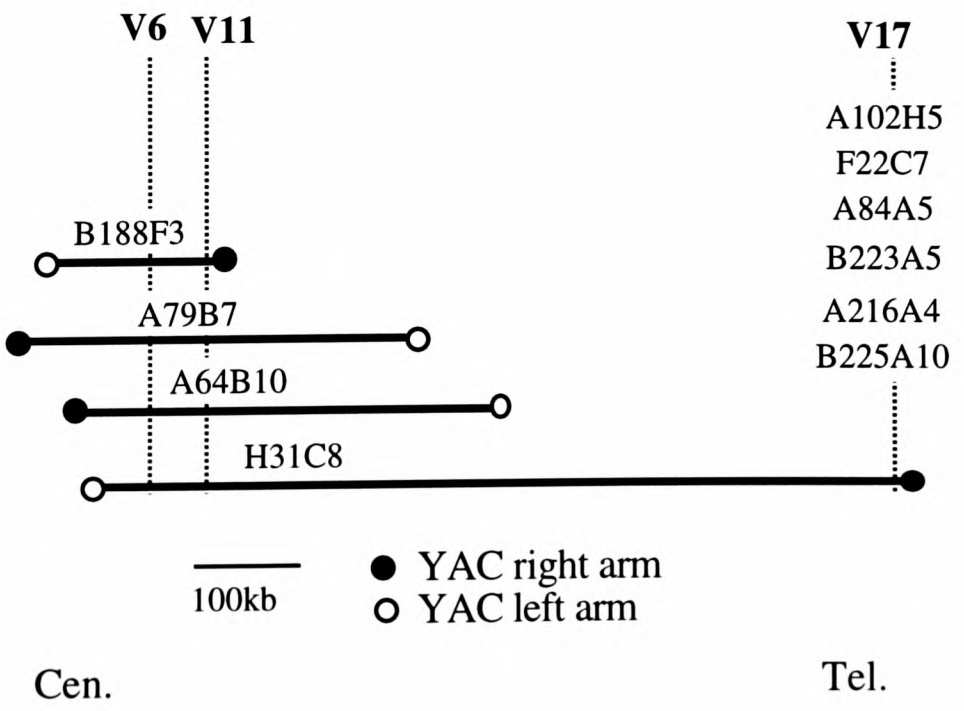


Figure 2.16 6 YACs identified by library screening with V17

At this point sequential chromosome walking became considerably more difficult. Of the 6 YAC clones identified with V17 all proved difficult. A84A5, A216A4, F22C7 and B225A10 were demonstrated to have resulted in co-transformation.

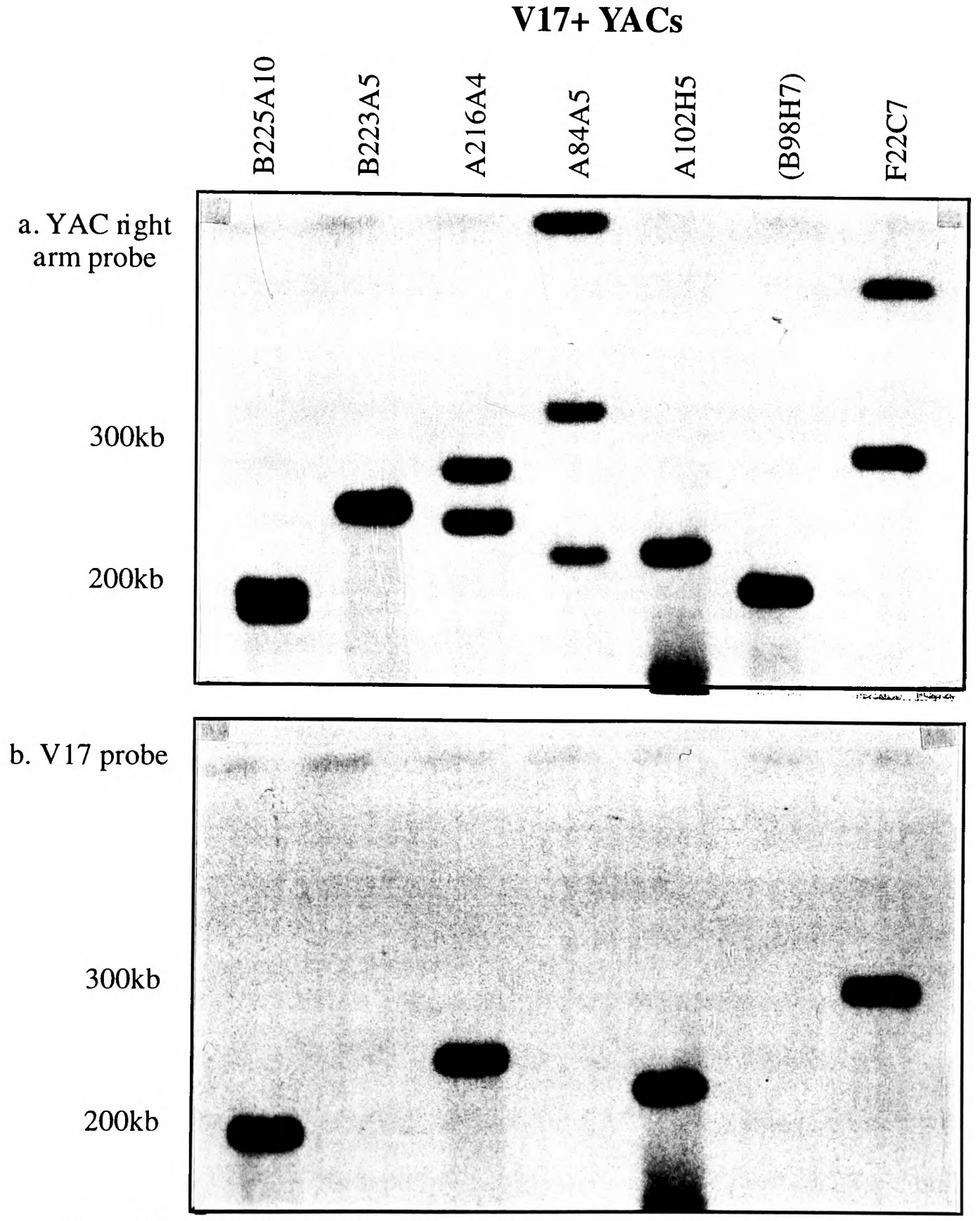


Figure 2.17 YACs identified with V17. Each was prepared as agarose plugs and the separated by PFGE (1% agarose, 0.5x TBE, 180 volts, switching times 10-45 secs, total run time 24 hrs). Panel a. shows the membrane hybridised with a probe to the YAC vector right arm demonstrating that 4 of the V17 clones are co-transformed. Panel b. shows the same blot hybridised with a probe of the V17 sequence. (Although included in this blot the clone B98H7 had been identified with V15 and was not positive for V17 by PCR).

Despite all clones being positive for V17 by PCR, hybridisation of the V17 PCR product to a Southern blot of YAC DNA failed to identify V17 sequence within B223A5 and A84A5 implying that these represented PCR false positives. In view of the pooled screening strategy (see Figure 2.7, page 42), it is unlikely that this could reflect individual clone contamination. Duplication of the primer sequences in such a way that a PCR product of identical size was generated is similarly implausible. However, the absence of V17 sequence by hybridisation was deemed sufficient to drop these two clones from further study.

On the basis of the above observations A102H5 was studied further. The insert of this YAC was sized at 200kb by PFGE. Only a poor signal was obtained by FISH and although this confirmed a signal form 6p another signal from Chromosome 14 could not be excluded (Lyndal Kearney). Sequence from the left arm of A102H5 revealed the presence of Alu repetitive elements which prevented the design of PCR primers. The right arm sequence yielded LINE repeats also preventing progress.

At this time a CEPH megaYAC 960h11 (insert size 1750kb) was identified from the CEPH database as containing the microsatellite D6S258. This marker was shown by PCR to be present on H31C8 but not the more centromeric YACs, suggesting both that 960h11 might extend the contig and that D6S258 might itself be useful for extending the walk. Insert end sequence was obtained for both the ends of 960h11 and by PCR it was shown that the left end of the insert (V26) did not extend beyond H31C8. This STS was used, along with D6S258 to screen the libraries generating 5 further clones. F22A1 and Q65A12 were identified from the ICRF library with D6S258. A222B1, B190C11 and B181H8 were isolated from the St. Louis library with V26. F22A1 was shown to be chimaeric, A222B1 was found to have LINE repetitive elements at both ends of the YAC insert. None of these clones were positive for V17 by PCR, implying that this remained the most telomeric point of the contig.

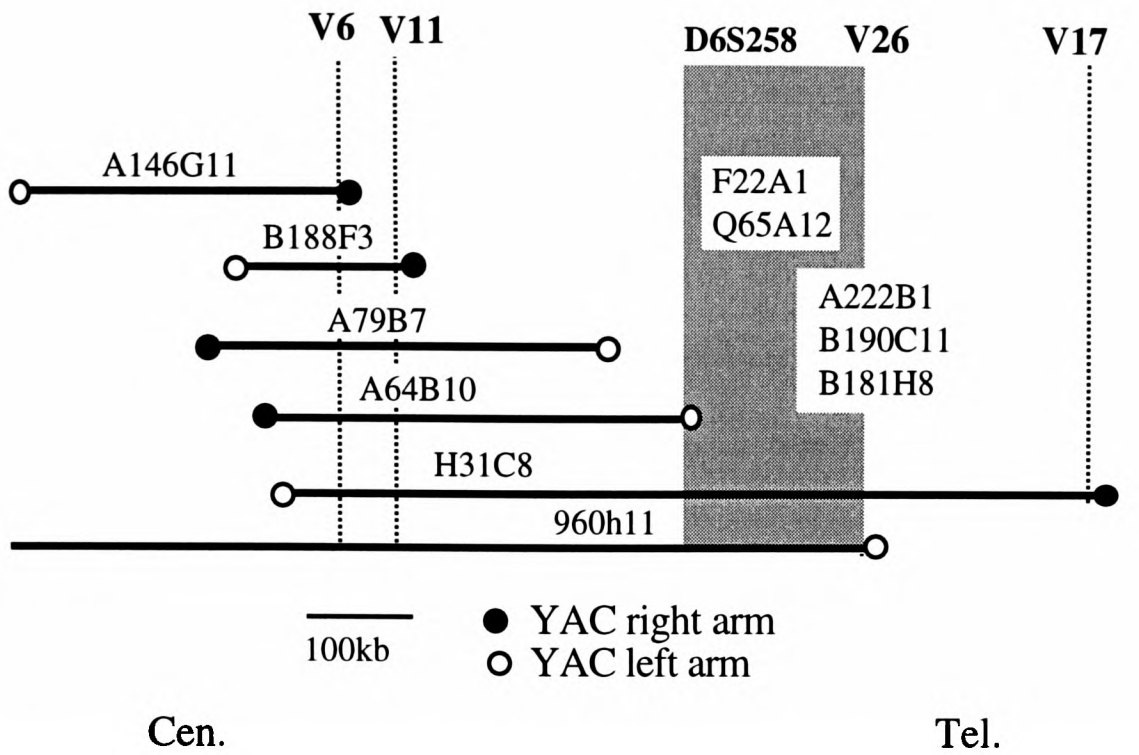


Figure 2.18 D6S258 is placed telomeric to HLA-F but the telomeric end of the CEPH MegaYAC clone 960h11 does not extend the contig.

An attempt was made to use V17 to screen the CEPH megaYAC library. Six primary pools/plates were identified (95, 102, 104, 115, 116, 120). Multiple positive signals were obtained when the secondary pools were examined and on repeating the PCRs at greater stringency the signals were all lost.

YAC manipulation

Use of the ICRF clone F22C7 to extend the contig had been inhibited by the fact that it contained 2 separate YACs, of which the smallest was positive for V17 (300kb see figure 2.17, page 52). In view of the difficulties progressing beyond the telomeric end of H31C8 this clone was re-evaluated.

The YAC of interest was rescued from a sub-clone containing two YACs by exploiting chromosomal segregation at meiosis (Rhona Borts, ref Nemeth *et al.* 1993). The clone containing the 2 YACs was mated with a yeast strain of the opposite mating type (MATα). Diploids resulting from this mating were selected and induced to sporulate. A selection of tetrads from each of several diploids were dissected with a micromanipulator after digestion of the ascus (Rhona Borts). The individual spores were grown on YPD and replicated onto selective media. The diploids that generated 4 ura + tetrads were considered to have probably separated the original two YACs (Nemeth *et al.* 1993). Such segregant yeast clones containing singleton YACs were analysed by PCR with primers to V17 to identify those containing the YAC of

interest. Those clones positive by PCR were analysed by PFGE and a Southern blot of the gel was sequentially hybridised with probes to the two arms of the pYAC4 vector to confirm the presence of a single YAC of the predicted size in each.

A singleton V17 positive YAC of 300kb was isolated from the original clone of F22C7. This singleton clone was confirmed to map to 6p and to be non-chimaeric by FISH (Sharon Horsley/Lyndal Kearney). Sequence generated from the left arm of the insert contained Alu repetitive elements throughout the sequence obtained. The restriction map of F22C7 compared to that from the telomeric end of H31C8 predicted that the right arm of F22C7 was telomeric. The right arm *Eco* RV vectorette fragment (size >1kb) was cloned directly into pCR Script (SK+) and sequenced. V34 PCR primers designed from the F22C7 right arm sequence amplified a product from F22C7 but not H31C8 implying that this STS was telomeric of V17 and extended the contig.

	V26	V17	V34
960h11	+		
H31C8	+	+	
F22C7		+	+

Table 2.3 STS content analysis of YACs at the telomeric end of the contig.

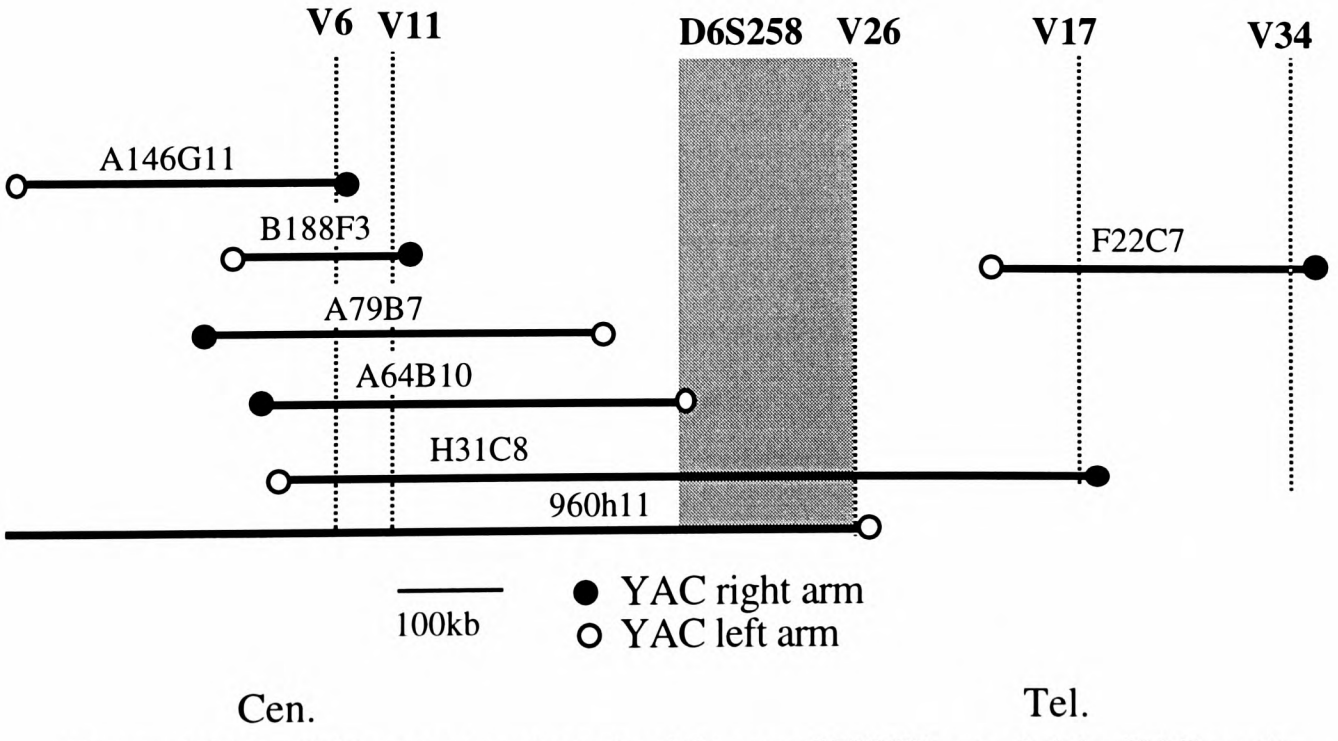


Figure 2.19 By STS content analysis the right arm of F22C7 extended the YAC contig.

Restriction mapping

Each individual YAC clone insert was mapped and a continuous restriction map extending 1.25Mb telomeric to HLA-F was created by compilation of maps from overlapping clones (figure 2.20, page 57).

On the larger clones (e.g. H31C8) clear resolution of some of the small fragments generated by the more frequently cutting enzymes (e.g. *Sal* I, *Sac* II) was impossible and internal sites for these might have been missed. Partial restriction digestion of the megaYAC clone 960h11 was not attempted and the restriction map of this clone was limited to the first site for each restriction enzyme, in from the YACs (telomeric) left arm.

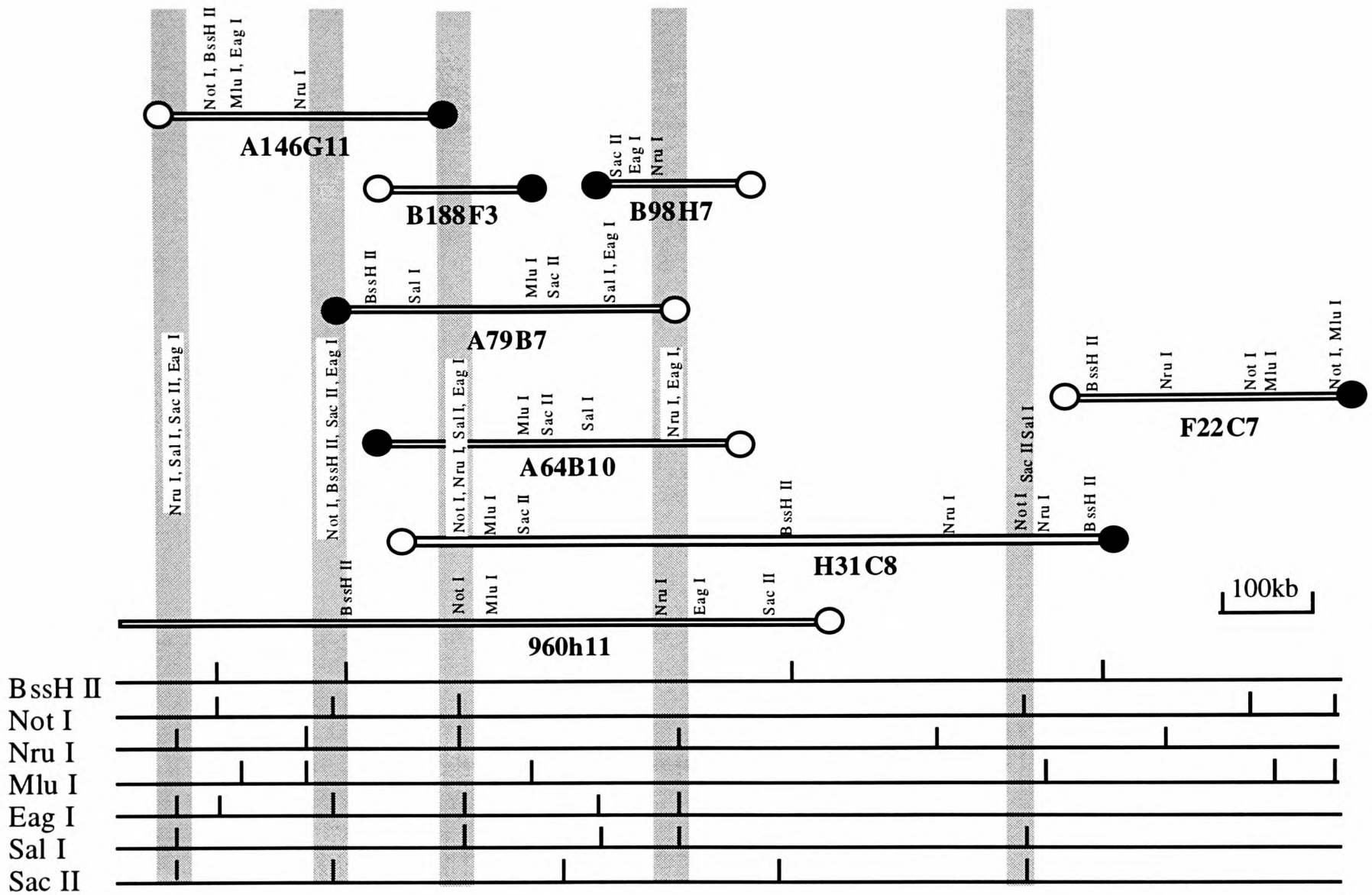


Figure 2.20 Restriction map derived from YAC contig extending telomeric to the MHC. Five rare-cutting restriction site cluster (putative CpG islands) were identified - illustrated by the shaded bars. The total distance covered by the contig (not including the megaYAC 960h11) was 1.35Mb.

Five areas within the map show clustering of rare-cutting restriction sites (indicated in Figure 2.20). These areas might represent CpG islands, associated with house-keeping genes (see Chapter 3.3, page 156).

The composite restriction map covers a physical distance of 1.35Mb. Taking HLA-F to lie within 100kb of the (centromeric) left arm of the YAC A146G11 (Geraghty *et al.* 1992), the most telomeric point of the contig (V34 at the telomeric, right end of F22C7) lies 1.25Mb telomeric to the MHC.

STS content mapping

a. Vectorette STSs

STS content of YAC clones was largely established by the results of the YAC library screening.

	V13	V12	V10	V16	V6	V11	V15	V26	V17	V34
A146G11	+	+	+	+	+					
B188F3			+	+	+	+				
A79B7	+	+	+	+	+	+	+			
A64B10		+	+	+	+	+	+			
B98H7							+			
960h11	+	+	+	+	+	+	+	+		
H31C8					+	+	+	+	+	
F22C7									+	+

Table 2.4 STS content analysis of YAC clones with YAC end STSs

YAC clones A79B7, A64B10 and H31C8 had been identified from the libraries with V11. Each was subsequently shown to be positive with V6 which had been used for the previous round of library screening. The fact that the library screen with V6 had not identified these clones reflects one of the difficulties of screening such large libraries even with a pooled DNA PCR strategy.

b. Microsatellites

Four microsatellite loci mapped into this interval.

	D6S265	D6S131	D6S1683	D6S258
A146G11				
B188F3				
A79B7				
A64B10		+	+	
B98H7		+	+	
960h11	+	+	+	+
H31C8		+	+	+
F22C7				

Table 2.5 Microsatellite STS content analysis of YACs.

D6S265 has been shown to lie within 100kb of HLA-A (Burt *et al.* 1996) and to be in strong linkage disequilibrium with that locus (Worwood *et al.* 1994).

D6S131 and D6S1683 share the same STS content pattern on YACs from this contig and these markers are inseparable by this approach alone. Combining the STS content data with that of the composite restriction map from the YAC contig places these two loci into an interval of less than 100kb within 600kb of the MHC. D6S258 lies telomeric to B98H7 (and hence D6S131 and D6S1683).

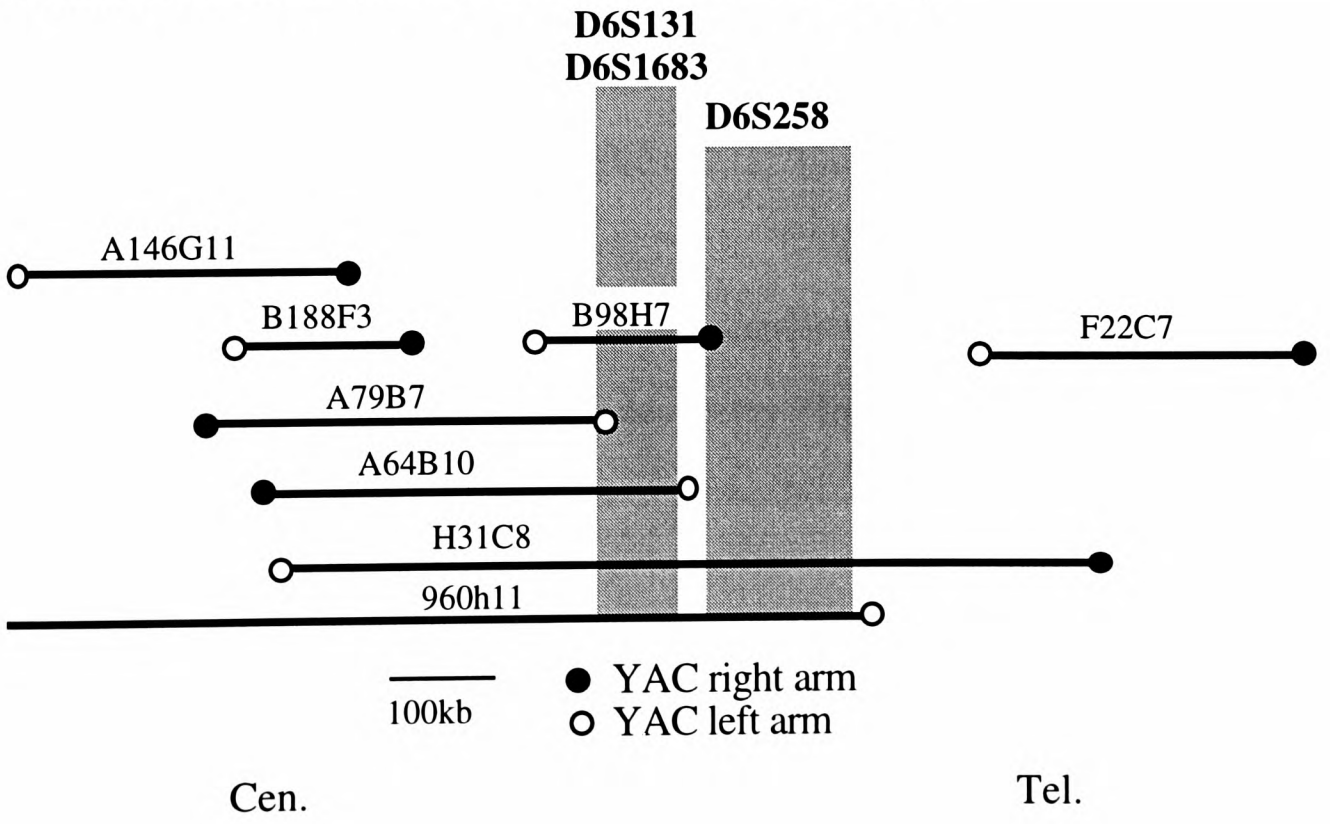


Figure 2.21 Schematic representation of microsatellite STS content on YAC contig

c. Gene Sequences

Similar to microsatellite sequences and STSs, gene sequences provide important genomic landmarks.

	HLA-F	MOG	RFP
A146G11	+	+	
B188F3			
A79B7			
A64B10			
B98H7			
960h11	-	+	
H31C8			+
F22C7			

Table 2.6 STS content analysis of YAC clones with gene sequences

HLA-F had been placed on the Olson YAC clone A146G11 by Geraghty (Geraghty *et al.* 1992). That position was confirmed by PCR, although it is notable that no product was obtained when the CEPH clone 960h11 was used. 960h11 has been reported to extend as far centromeric as HLA-A within the class I region of the MHC (Burt *et al.* 1996). The absence of HLA-F in this clone by PCR suggests an internal deletion.

A myelin-oligodendrocyte glycoprotein (MOG) gene has been placed 100kb telomeric to HLA-F (Roth *et al.* 1995) and this is supported by PCR with primers to the microsatellites described in that report on the YAC A146G11.

In 1995 a Ret-finger protein (transcription factor) was mapped telomeric to HLA-F, MOG and D6S131 (Amadou *et al.* 1995). This locus is telomeric to the end of the CEPH megaYAC 960h11 but is present on H31C8. This STS analysis places RFP at least 700kb telomeric to the MHC.

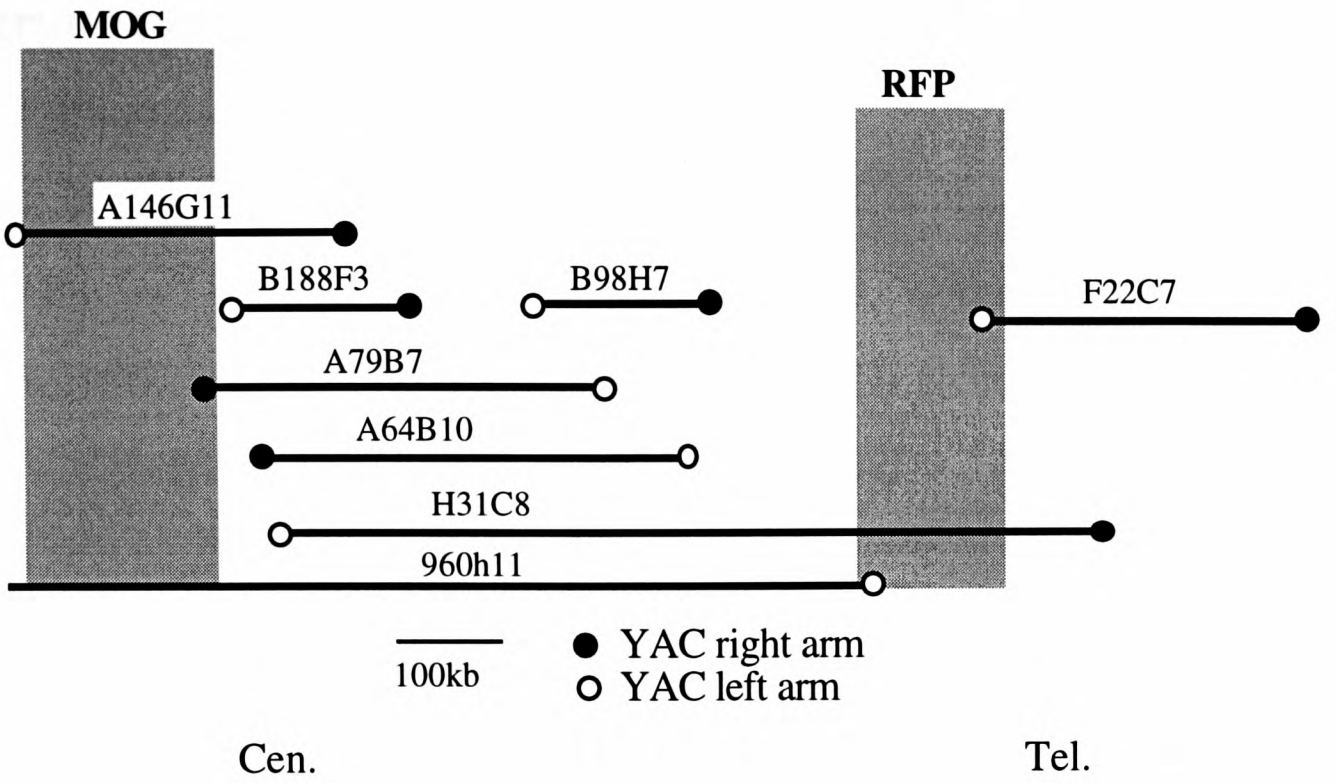


Figure 2.22 By STS content analysis the Ret-Finger Protein maps up to 1Mb telomeric to HLA-F.

Although it has not been proposed that the Ret-finger protein might be implicated in haemochromatosis, the gene locus was to prove of great importance to the positional cloning of the haemochromatosis gene (see page 68).

Subcloning of B188F3 into cosmids

With the intention of isolating CA repeat microsatellites telomeric to HLA-F B188F3 was sub-cloned into cosmids (Supercos - Stratagene). Twenty-one cosmid clones derived from B188F3 were positive with a total human probe. All of these cosmid clones were positive by PCR with the centromeric STSs (V10 and V12) but not for the more telomeric markers (V11 and V6).

Interpretation

The physical distance between HLA-A and D6S105

After 6 rounds of YAC library screening a YAC contig had been extended more than 1.25Mb telomeric to MHC. During this time a contig based on D6S105 had been extended more than 1Mb centromeric to this marker (J.Pointon, C.Stone). The overlap between these two contigs had not been established. The most telomeric STS from F22C7 (V34) did not amplify on the most centromeric of the YACs from the D6S105 contig (C12H9). The most centromeric STS from the D6S105 contig (V29) did not amplify on F22C7. The physical distance between HLA-A and D6S105 was established from these two YAC contigs to be greater than 2.25Mb.

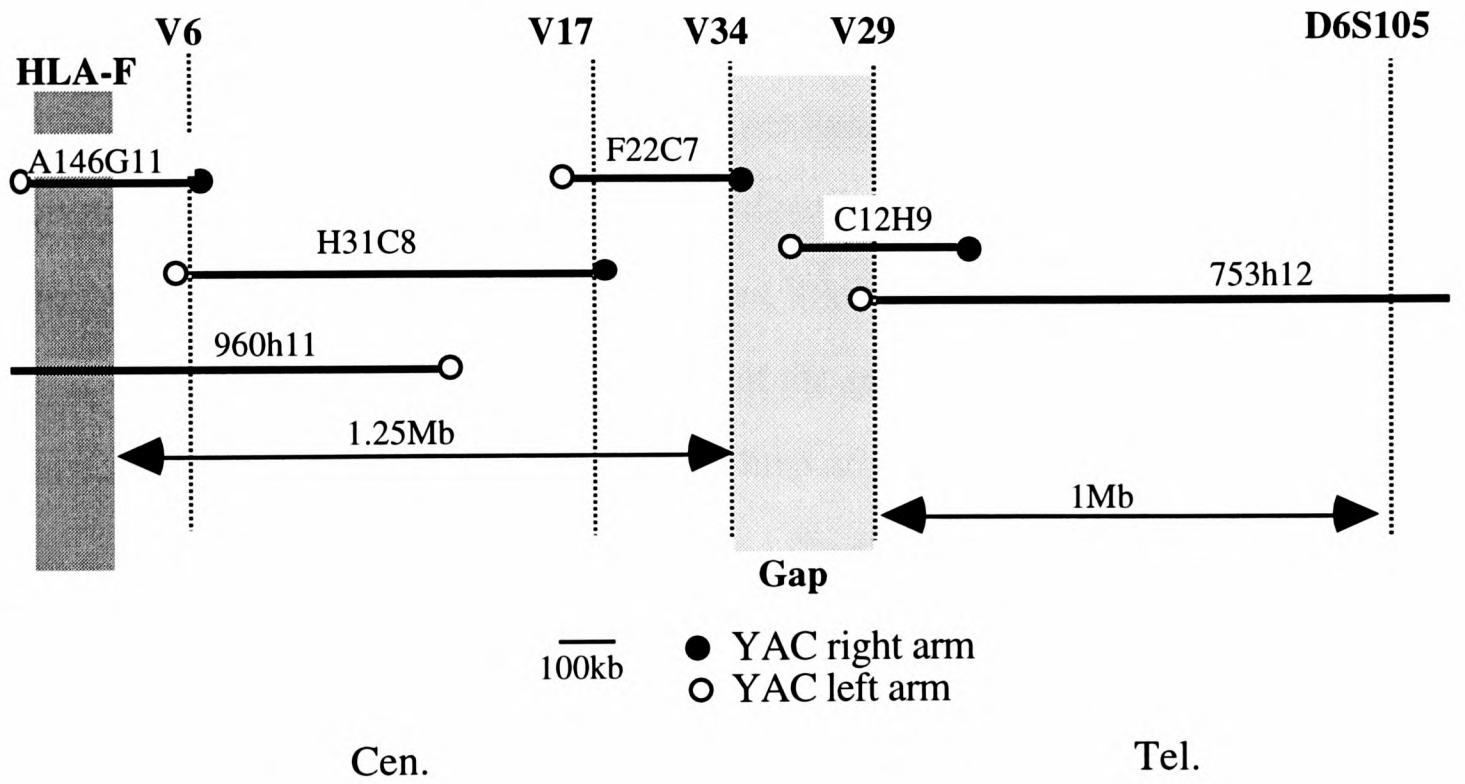


Figure 2.23 The physical distance between HLA-F and D6S105 is greater than 2.25Mb

A gap persists

The primary objective to clone the interval between HLA-A and D6S105 had not yet been achieved. Sequential chromosome walking using two human YAC libraries (St. Louis and ICRF) had so far failed to bridge a physical distance of at least 2.25Mb between the two loci. During this time the initial reports of efforts to use YACs to map the entire human genome were published (Hudson *et al.* 1995). Quite apart from the development of an automated YAC library screening facility, the

principle strategy adopted by Hudson's group at the Whitehead Institute for Genome Research (WIGR) at the Massachussetts Institute of Technology (MIT) was very different. Hudson's principle had been to generate an "STS map" of the genome. Using STS content analysis on CEPH YACs and radiation hybrid panels to order the sequence tags, the YAC clones are themselves ordered into "contigs" with overlap established by the STS content analysis alone. This represents the pinnacle of what had been proposed 8 years earlier when Olson first suggested the universal adoption of sequence tagged sites (Olson *et al.* 1989).

In the release of the STS map of December 1995 (http://www-genome.wi.mit.edu/) Hudson and colleagues had achieved 94% coverage of the human genome with an average marker density of one STS every 200kb. Despite this enormous achievement, overlap between YACs containing known MHC associated STSs and the "contig" around D6S105 had not been established. Therefore the gap that persisted in our contig had not been bridged by the Whitehead Institute team using both a different strategy and a different YAC library.

Assuming complete representation of the genome in any given YAC library (i.e. the CEPH megaYAC library) the STS map of the Whitehead Institute should eventually generate a YAC "contig" covering the entire genome. The fact that a gap persists between the MHC and D6S105 may reflect a relative deficiency in the density of markers from this region and that with more STSs continuity will be established. Alternatively, it is possible that a "hole" is present in the CEPH YAC library (i.e. that a small genomic region remains unrepresented within that library). Certain regions of genomic DNA are more difficult to clone and in theory a region not represented (or under-represented) in the library is possible. Ultimately, more markers and more clones would circumvent this difficulty.

The traditional approach of sequential chromosome walking presents limitations inherent in the necessity to bridge 2 specific loci (as opposed to the saturation mapping of larger regions). Screening two different YAC libraries (ICRF and St. Louis) difficulties were encountered progressing beyond the STS V17. Six

clones identified by library screening with V17 was considered to be a good yield. Each of these clones presented recognised technical limitations of YACs (co-transformed clones, chimaeric clones and insert ends containing repetitive elements). The number of clones that had resulted from co-transformation events seemed particularly high (4 out of 6). Sequence was eventually obtained from the telomeric (right) arm of F22C7 allowing the design of primers defining the STS V34. V34 was used to screen the YAC libraries (Alison Merryweather-Clarke).

A YAC contig bridging HLA-A and D6S105

Using a strategy identical to ours, Burt has recently reported a YAC contig of 4.5Mb covering the interval from HLA-A to D6S105 (Burt *et al.* 1996). That group had been walking telomeric to the MHC and in both directions from D6S105, screening the CEPH mark one, CEPH megaYAC and the ICI libraries. They clearly established two large contigs based on CEPH megaYAC clones (anchored by 960h11 at HLA-F and 753h12 and 950h11 at D6S105). At each of these extremes they established good "depth" of coverage with small CEPH clones. However the two regions are linked by only a single ICI clone (17AH2, the only ICI clone included in the report).

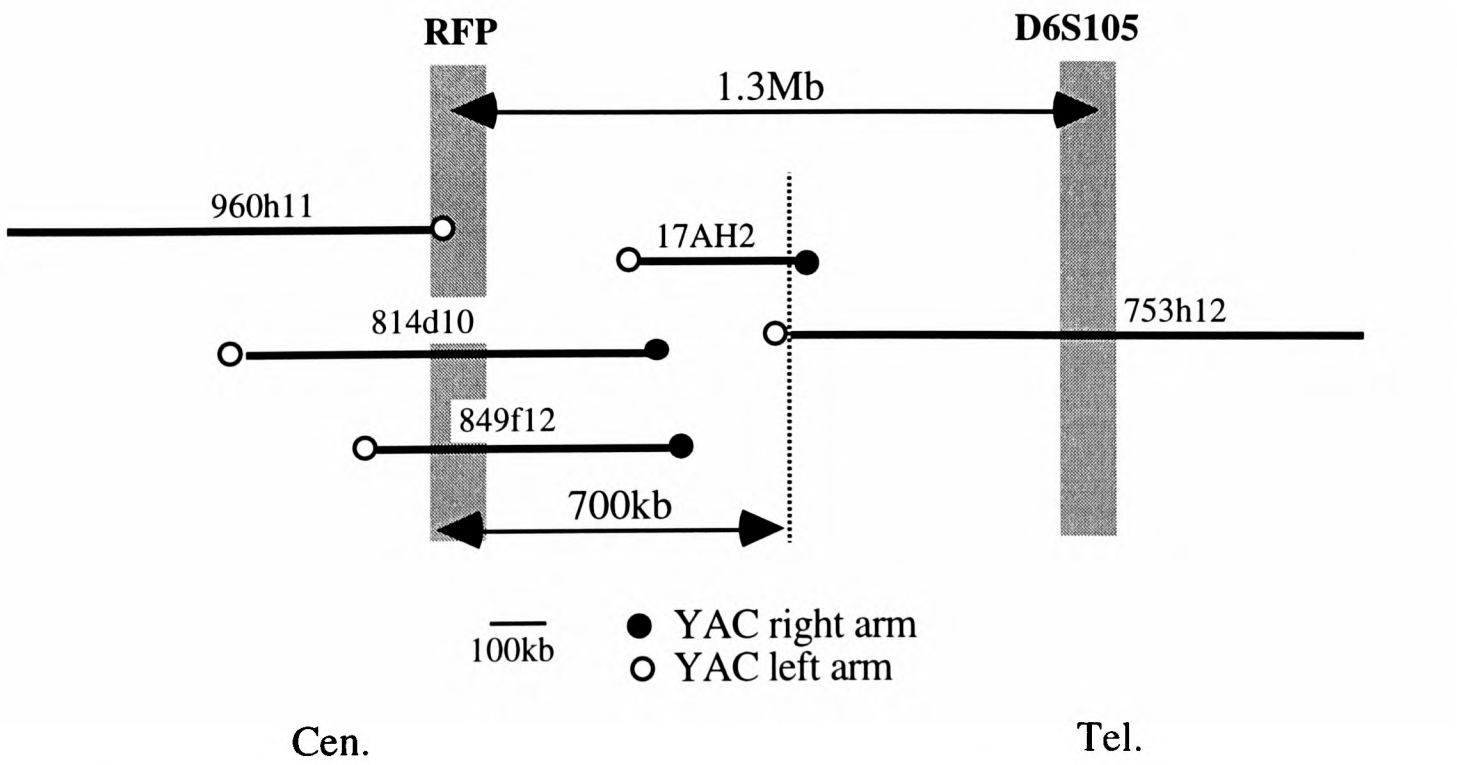


Figure 2.24 A YAC contig spanning the interval from the MHC to D6S105 (Burt *et al* 1996)

The ICI clone 17AH2 was the only clone from the three libraries identified by the most telomeric of their STSs (from the telomeric end of the CEPH clone 814d10). From the accompanying restriction map the distance between RFP and D6S105 was 1.3Mb and that between the ends of the two CEPH YAC clones, 960h11 and 753h12, was 700kb.

The ICI YAC, 17AH2 was found to be positive by PCR with our STS V29 (derived from the centromeric end of the YAC 753h12). However overlap between 17AH2 and F22C7 could not be proven by PCR with V34 primers.

F22C7 might not extend as far as the telomeric end of 814d10 although from the restriction map the distance covered by H31C8 beyond the telomeric end of 960h11 should have ensured that F22C7 (if not H31C8 itself) overlap with 17AH2.

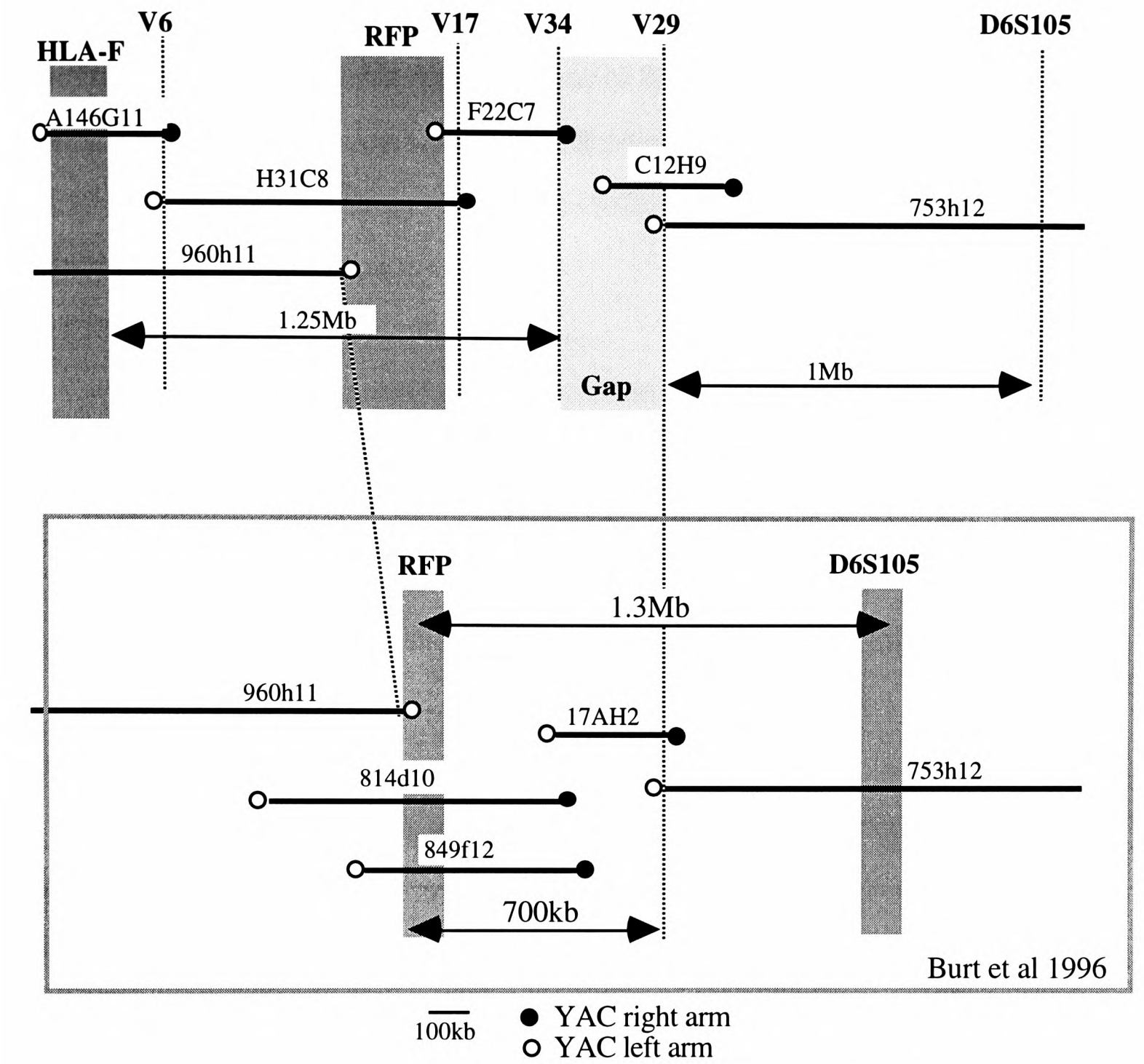


Figure 2.25 Comparison of the YAC contig of this thesis and that of Burt (Burt *et al* 1996)

Without sequence data from the Australian group it is not possible to explore further the relationship between F22C7 and 17AH2.

Restriction mapping and YACs

Comparison of our restriction map with that of Burt raises other important issues regarding the construction of physical maps with YACs. Both F22C7 and H31C8 were isolated from a different library to those used by Burt and the differences in the composite restriction maps may be due to enzyme site polymorphisms between the libraries. However it is noteworthy that, other than 17AH2, Burt's map was generated predominantly from megaYAC clones. To map restriction enzyme sites the crucial step is the consistent resolution and accurate sizing of fragments by pulsed-

field gel electrophoresis. For YACs with large inserts, resolution of smaller fragments can be extremely difficult and even for clones of 800kb (e.g. H31C8) internal restriction sites may be missed unless smaller overlapping clones are identified. Conversely, alignment of large YACs on the basis of restriction sites might duplicate some sites.

Is F22C7 chimaeric?

The inconsistencies between the restriction map and that of Burt could be explained if F22C7 were chimaeric.

Several techniques exist for establishing whether a clone is chimaeric or not. Hybridisation of radiolabelled total YAC DNA to a gridded panel of radiation hybrids representing all of the human chromosomes has been suggested (Zoghbi and Chinault 1994). The particular panel used by that group could not separate the hybrids containing chromosomes 6 and 12.

FISH involves labelling total YAC DNA with a suitable fluorochrome and hybridising this to a spread of normal human metaphase chromosomes. This technique provides several different levels of information. First it confirms the correct chromosomal position of the YAC. Furthermore if the clone is chimaeric the FISH analysis will provide information on which other chromosomal segments are represented. One theoretical drawback to FISH is its resolution. The strength of signal obtained from hybridisation will depend on both the amount of fluorochrome incorporated and on the strength of the hybridisation. Hence there is a limit to the resolution of FISH and a clean signal appearing at the appropriate cytogenetic band can not completely exclude the presence of a small "chimaeric end".

An alternative approach to this problem would be to adopt a principle more in line with STS mapping (such as that of the Whitehead Institute). Even in chromosome walking strategies, YACs are only a means to an end and with the process of screening the YAC libraries by PCR, the most important product at the end of each round of screening is another STS that maps to chromosome 6. To establish whether an STS maps to chromosome 6, or not, is probably easier than to quickly and reliably

evaluate individual YAC clones for chimearism. Using DNA from a monochromosomal radiation hybrid containing human chromosome six in combination with other YACs from the contig would allow the utility of putative PCR primer pairs to be tested very quickly. The chromosome 6 radiation hybrid available through the HGMP Resource Centre (MCP BRA) has a translocation resulting in the loss of much of the short arm (Xqter-Xq13 : 6p21-6qter, Goodfellow *et al.* 1982). V34 did not amplify from this source.

The large MegaYAC clones from the more recent CEPH library allow this concept to be adopted more broadly. Even without formal assessment of chimaerism, if one is confident of the chromosomal position of the initial, anchoring STSs then the more STSs one generates, the more clones that are likely to be identified. A two-way grid evolves (i.e. clones horizontal and STSs vertical) and chimaeric clones and non-chromosome 6 STSs become evident from STS content inconsistencies (see Chapter 4). V34 did amplify product from the CEPH YAC 872f5 which contains other chromosome 6 STSs (D6S1683, data from the Whitehead Institute STS map).

Although this principle is attractive, in practice it applies more to long range contig construction (such as that of the Whitehead Institute) and its use in mapping between two specific points may be limited by the number of STSs in the interval. In such circumstances, the experimental demonstration of individual clone integrity remains vital to the efficient and successful progress of chromosome walking.

Although the primary aim had been to bridge the interval between HLA-F and D6S105 with a YAC contig, the greater goal was the characterisation of centromeric and telomeric boundaries for the haemochromatosis gene.

The centromeric limit to the haemochromatosis gene region lies at RFP

In early 1995 Calandro and collegues in California reported the characterisation of an informative recombination that placed the haemochromatosis gene telomeric to HLA-F (Calandro *et al.* 1995). Previous reports of informative recombinations had placed the likely position of the haemochromatosis gene in the MHC (Edwards *et al.* 1986; Gasparini *et al.* 1993) although each was subsequently

retracted. The report of Calandro defined a recombination breakpoint flanked on either side by at least 4 microsatellite markers.

The family reported by Calandro was subsequently studied with a new marker identified within 4kb of the 5' end of the RFP gene placing the **centromeric** boundary for the haemochromatosis gene region at least 1Mb telomeric to HLA-A (Malfroy *et al.* 1996).

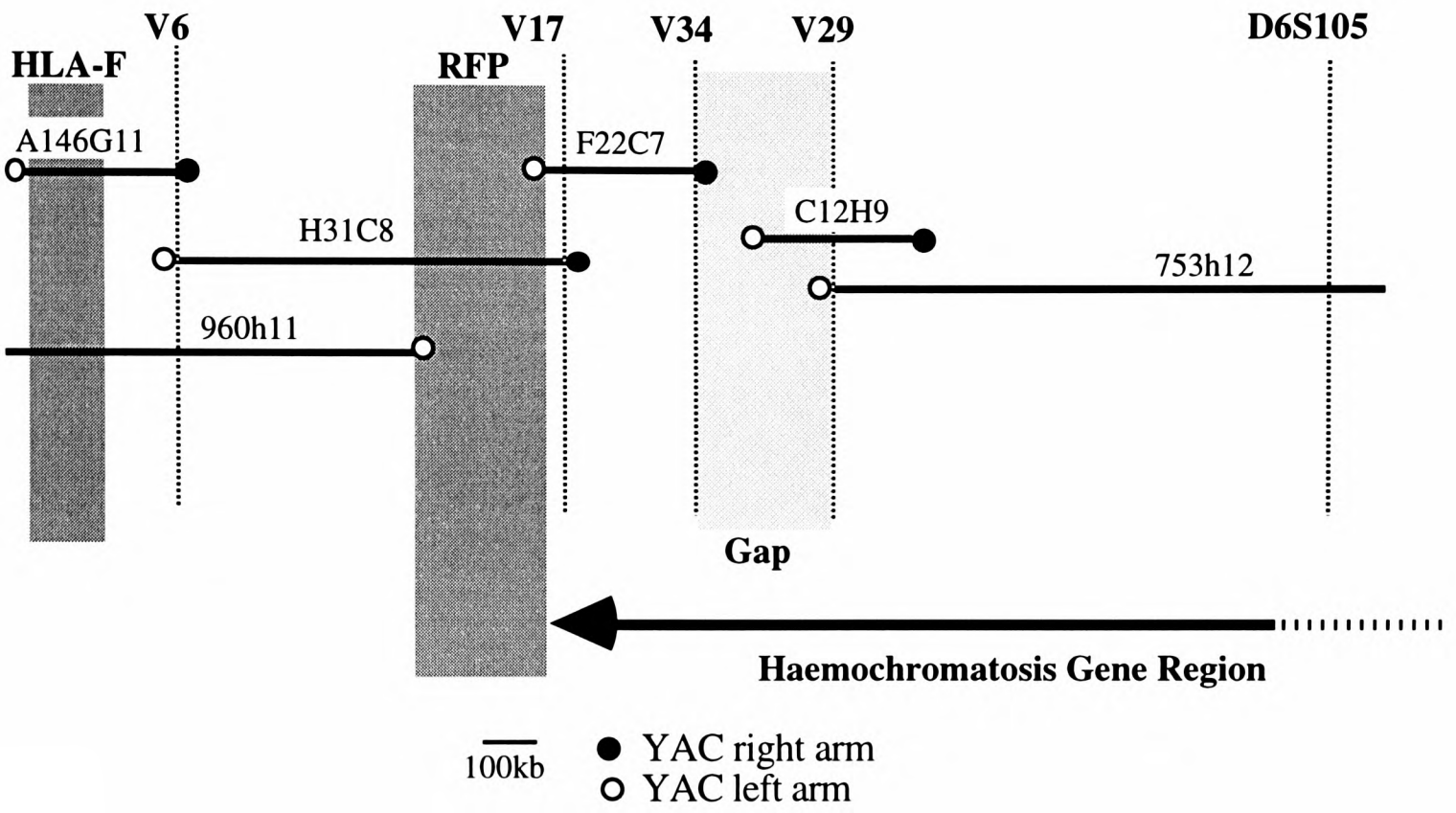


Figure 2.26 RFP marks the **centromeric** limit to the haemochromatosis gene region

The haemochromatosis gene maps telomeric to D6S105

Allele association data generated from new microsatellites isolated in Oxford by Caroline Stone suggested that the haemochromatosis gene lay telomeric to D6S105 (Stone 1995). Using a powerful likelihood method a peak of linkage disequilibrium with the disease was demonstrated at D6S1260 (Raha-Chowdhury *et al.* 1995). This new marker lay 500kb telomeric to D6S105.

In the light of the recombination and allele association data it became imperative to define a telomeric limit for the haemochromatosis gene region.

Chapter 2.3

The identification of YACs extending telomeric to D6S1260.

Background

From the first report of an association between haemochromatosis and HLA-A, a position for the gene in, or close to, the MHC was established (Simon *et al.* 1976). The physical proximity of the haemochromatosis gene to HLA-A was so firmly held that no sooner had the Class I region of the MHC been cloned in YACs then claims were made for haemochromatosis candidate genes from the HLA-A sub-region (ElKahloun *et al.* 1993). The fact that allele 8 of the microsatellite marker D6S105 inferred a greater relative risk for haemochromatosis than HLA-A3 (48 v. 4.8) suggested that the gene lay closer to D6S105 than HLA-A (Jazwinska *et al.* 1993).

To refine the position of the haemochromatosis gene members of the Oxford group isolated more CA repeats telomeric to D6S105 (Pointon 1995; Stone 1995). Markers mapping close to D6S105 were shown to be in linkage disequilibrium with haemochromatosis (Stone *et al.* 1994). In particular, CS-5 (hereafter referred to as D6S1260) was isolated from a YAC extending telomeric to D6S105 (Q68B5 from the ICRF YAC library). Initial calculations of p_{excess} suggested a very strong allelic association between this marker and haemochromatosis. Using a powerful new method of calculating likelihood (Terwilliger 1995), it was demonstrated that D6S1260 was the closest microsatellite yet reported to the haemochromatosis gene (Raha-Chowdhury *et al.* 1995).

As with the HLA-A to D6S105 interval, the region telomeric to D6S105 was at the time almost completely lacking in microsatellite markers. The next telomeric microsatellite to D6S105, D6S461, was 3cM further telomeric.

Figure 2.27 D6S1260 lies telomeric to D6S105 and shows greater allelic association with the haemochromatosis. The next telomeric microsatellite lay 3cM away.

Despite the fact that many groups had already begun to claim the isolation of haemochromatosis candidate genes from the MHC class I region, it was felt that for a positional cloning strategy to be successful a clear and limited gene position had yet to be defined. Although a conservative centromeric limit to the haemochromatosis gene region had been placed at RFP, no telomeric limit had been characterised. It was necessary to identify and map more microsatellites telomeric to D6S1260.

The YAC contig telomeric to D6S105

The YAC K43B3 had been identified by screening the ICRF YAC library with primers to D6S105 (J. Pointon). Primers designed to sequence generated from the left arm of this YAC (V1) had identified a single further clone from the same library (Q68B5). Dual-labelled FISH had demonstrated that this second clone extended telomeric to D6S105 (Stone 1995).

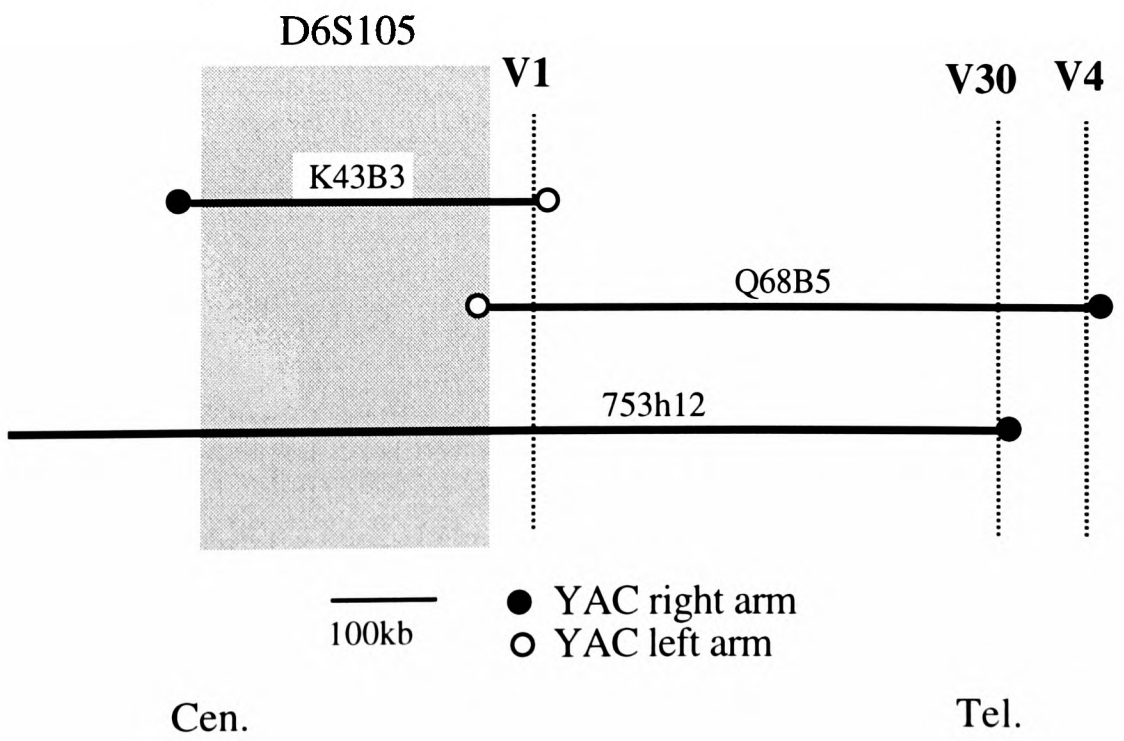


Figure 2.28 The ICRF YAC Q68B5 extends telomeric to D6S105.

Repeated screening of the libraries with primers from the telomeric right arm of Q68B5 (V4) had until this point failed to extend the contig further. The limiting feature had been that the majority of clones identified with this STS had been shown to be chimaeric by FISH. The most recent (fourth) screen of the YAC libraries with V4 identified the clone B117C7 from the St. Louis library and U82E7 from the ICRF library (Jenny Pointon).

Methods, Results and Interpretation

Searching the Whitehead/MIT STS map of the human genome

In view of the need to extend the contig telomeric, the STS based map of the human genome of the Whitehead Institute was searched via its internet site (http://www-genome.wi.mit.edu/) for further YACs to facilitate the telomeric walk.

The Whitehead YAC contig containing D6S105 (Contig WC6.4, release 7 - May 1995) suggested overlap with CEPH YAC clones which extended beyond D6S461.

YAC	Size kb	D6S 306	D6S 105	D6S 464	D6S 1078	D6S 1422	STSG 9945	D6S 1558	WI 3111	D6S 1016	WI 3878	D6S 1281	D6S 1554	D6S 1545
871g6	1190	+	+	+										
753h12	1770	+	+	+	+	+	+							
950h11	1760	+	+	+			+	+	+					
947f6	840								+		+	+		
901a10	1130										+	+	+	+
743f5	1650									+				
905h1	1270							?		+				
727b2	1380						+	+	+					
908d11	1170						+	+	+					
912c11*	540						+	+						
766a3	1320								+					
912c4	1440							+						
771g5*	1360										+	+		+
935d10	1720										+	+	+	
732b9	1110										+	+		
741c9	700										+	+		
792g12	610										+	+		
711g3	610											+		

Table 2.7 STS content of CEPH YAC clones extending beyond D6S105 - Whitehead Institute STS map. The microsatellites are ordered from centromere (left) to telomere (right). The YACs marked with * each have two insert sizes quoted on the database suggesting that these clones might be co-transformed.

The MegaYAC clones 753h12 and 950h11 had already been included in our contig, after having been identified on the CEPH/Généthon database as containing D6S105 (Stone 1995). FISH analysis on 950h11 had suggested this clone was chimaeric and although this clone was positive for V4, it had not been studied further.

From the Whitehead Institute STS map a single YAC clone, 947f6, appeared to link YACs containing D6S105 with a group extending more telomeric.

STS content mapping with new markers and microsatellites

The Whitehead Institute database contained sequence information on other markers provisionally mapped telomeric to D6S105. These markers were analysed on the YACs comprising the contig in addition to the CEPH clones.

YAC	D6S 105	STSG 9945	D6S 1260	D6S 1558	WI 3111	D6S 1016	D6S 1621	D6S 1281	WI 3878
Q68B5		+	+	+					
B117C7				+	+				
U82E7			+	+	+	+			
753h12	+	+	+	+					
947f6					+	-	+	+	+

Table 2.8 STS content on YAC clones with new markers telomeric to D6S105.

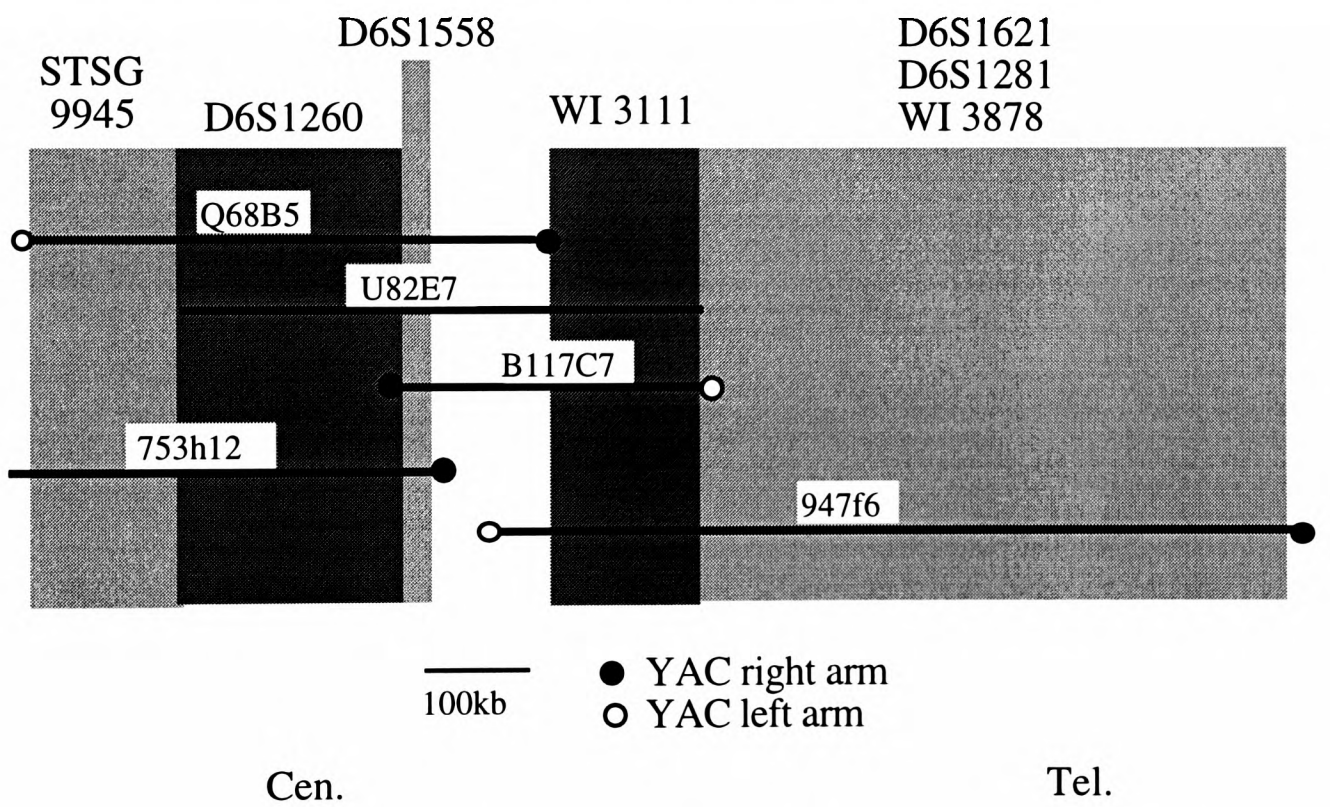


Figure 2.29 Schematic translation of the STS content analysis of new markers telomeric to D6S105.

End sequence STS mapping of new YAC clones

End sequence STSs were obtained from the two arms of B117C7 (V39 - right arm, V40 - left arm), U82E7 (V37 - right arm, V38 - left arm) and 947f6 (V51 - left arm, Jenny Pointon). These were used in STS content analysis to confirm the overlap between YACs. Neither V37 nor V38 were present on any of the other YACs in the contig, implying that U82E7 was chimaeric. Subsequent FISH analysis supported this,

demonstrating signal from both the long and the short arms of chromosome 5 (Sharon Horsley).

YAC	V1	V39	V30	V51	V4	V40
K43B3	+					
Q68B5	+	+	+	+	+	
B117C7		+	+	+	+	+
753h12	+	+	+			
947f6				+	+	+

Table 2.9 STS content analysis with YAC end STSs telomeric to D6S105

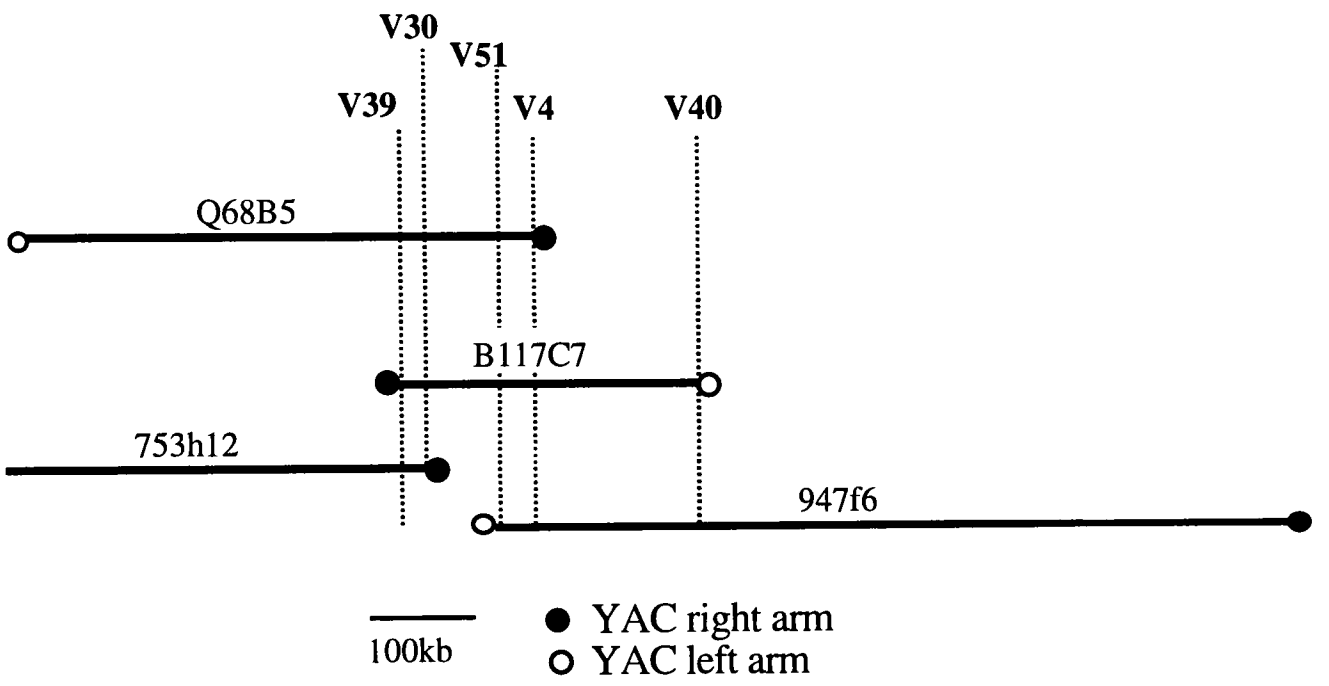


Figure 2.30 Schematic representation of vectorette end sequence STSs content analysis on non-chimaeric YAC clones extending telomeric to D6S105

The STS content analysis with the combination of new markers from the Whitehead Institute database and STSs generated from the ends of the YACs established extension of the YAC contig across the region of interest.

D6S1016

The microsatellite D6S1016 could not be consistently placed on the contig. D6S1016 is a tetranucleotide repeat isolated by the Cooperative Human Linkage Center (CHLC) and mapped by the Whitehead institute between the STSs WI3111 and D6S1621. This locus was not present on 947f6. The proposed position of this microsatellite made it of potential interest with regard to the mapping the haemochromatosis gene.

Amplification of D6S1016 using genomic DNA as template produced multiple alleles from any one individual.

D6S1016

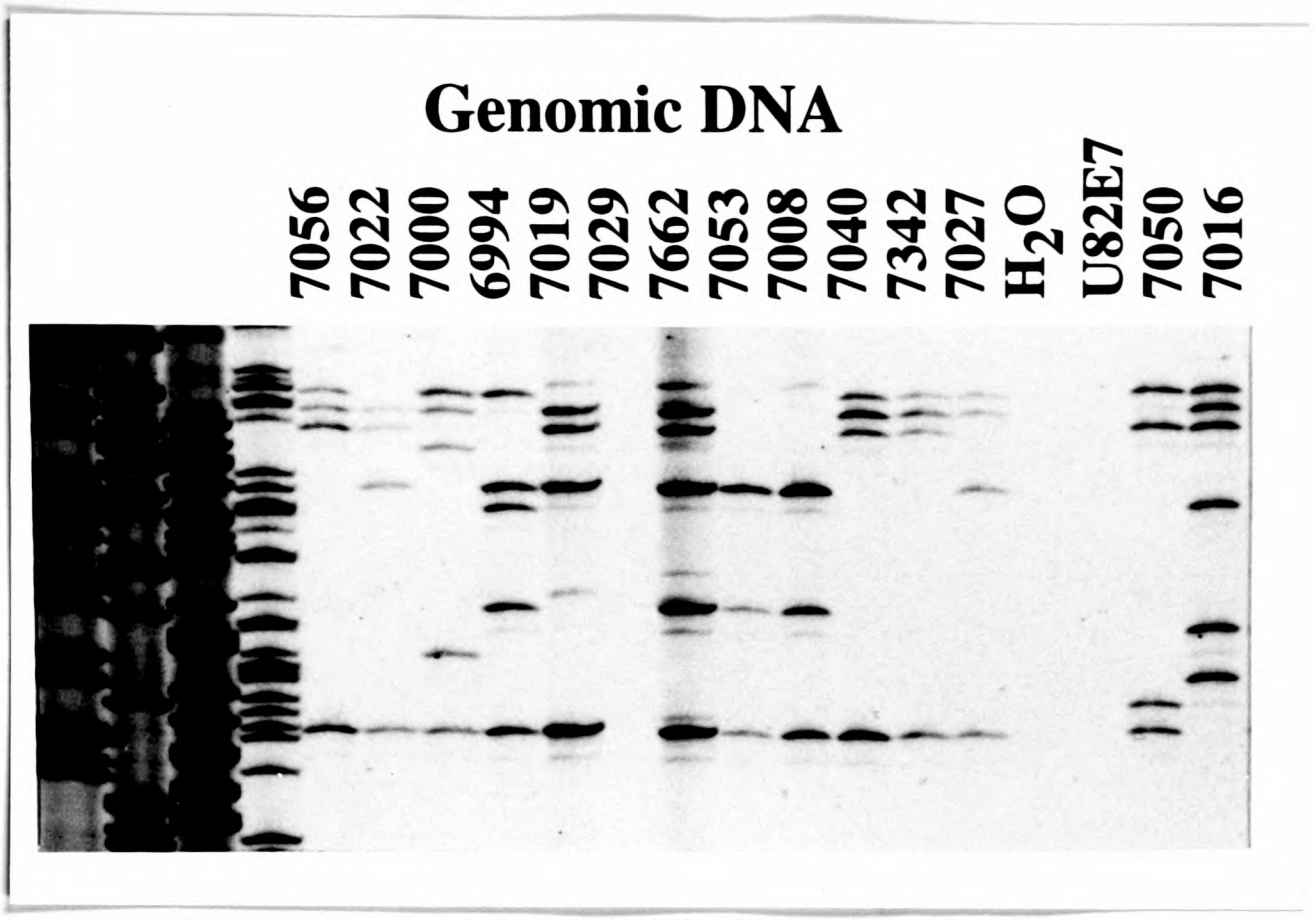


Figure 2.31 Amplification of alleles of D6S1016 on genomic DNA from CEPH reference pedigree samples. The size marker is M13 sequencing ladder.

Despite the complexity of alleles amplified by D6S1016, attempts were made to identify if any of these alleles segregated with other chromosome 6 markers. CEPH, three generation pedigrees were analysed with D6S1016.

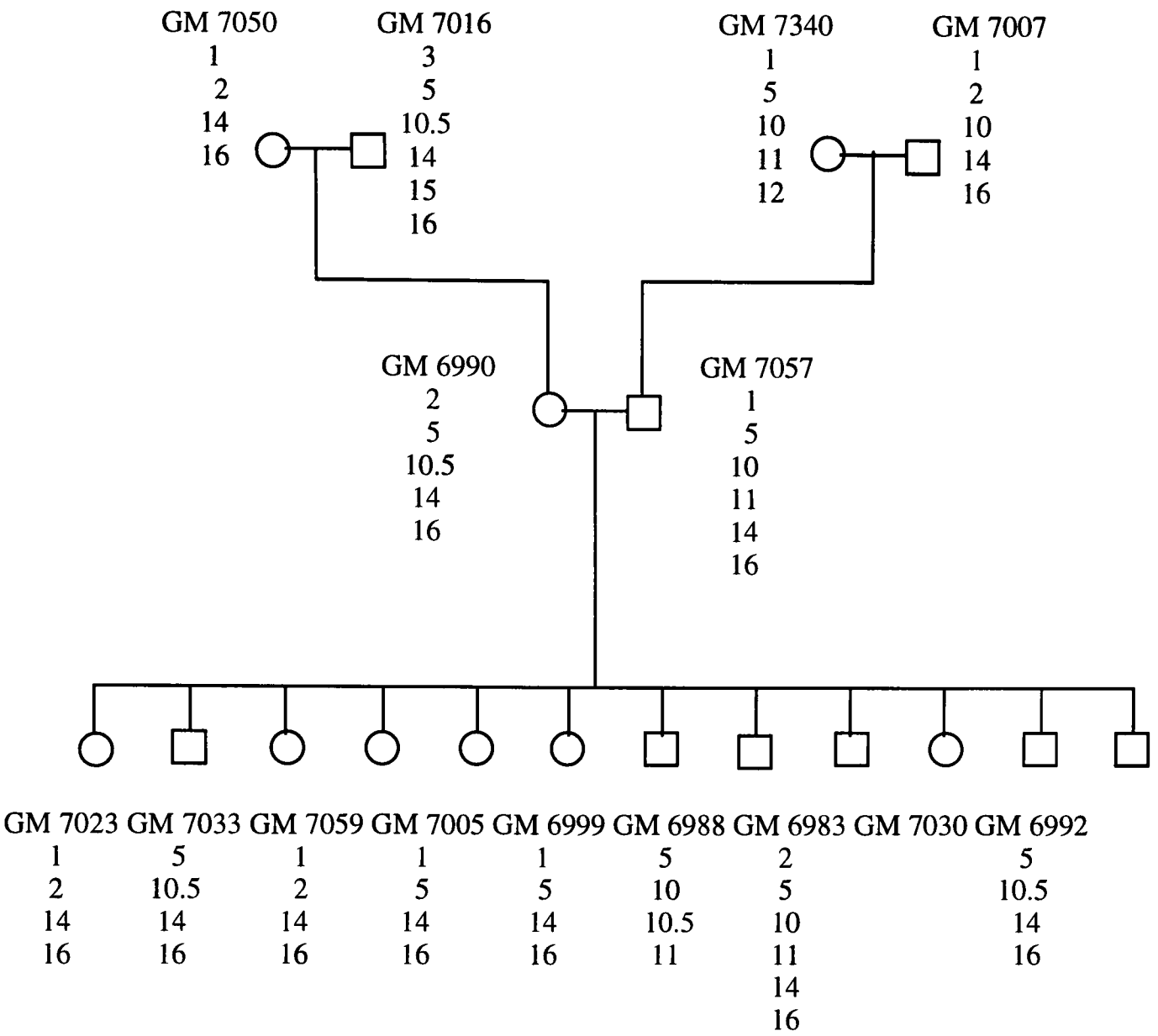


Figure 2.32 Inheritance of alleles at D6S1016 through CEPH reference family 982 (Utah pedigree K-1331). Allele 1 is 231bp. Although the microsatellite is a tetranucleotide repeat the, other alleles are numbered in increasing size by **2**bp each to accomodate the complexity of amplification products (see Figure 2.31).

This analysis confirmed that only some of these alleles segregated with other chromosome 6 markers. In view of this complex pattern of inheritance with a suggestion of concurrent amplification from a locus other than chromosome 6, D6S1016 was not used in allele association studies with haemochromatosis patients.

FISH data

Having established that B117C7 and 947f6 extended the YAC contig, subsequent FISH analysis on these two clones produced an unexpected observation. Prior to the identification of these YACs, Jenny Pointon had commented on the frequency with which YACs identified with V4 had appeared to be chimaeric with chromosome 5 and FISH analysis of both B117C7 and 947f6 demonstrated hybridisation signal from chromosome 5 (Sharon Horsley and Lyndal Kearney).

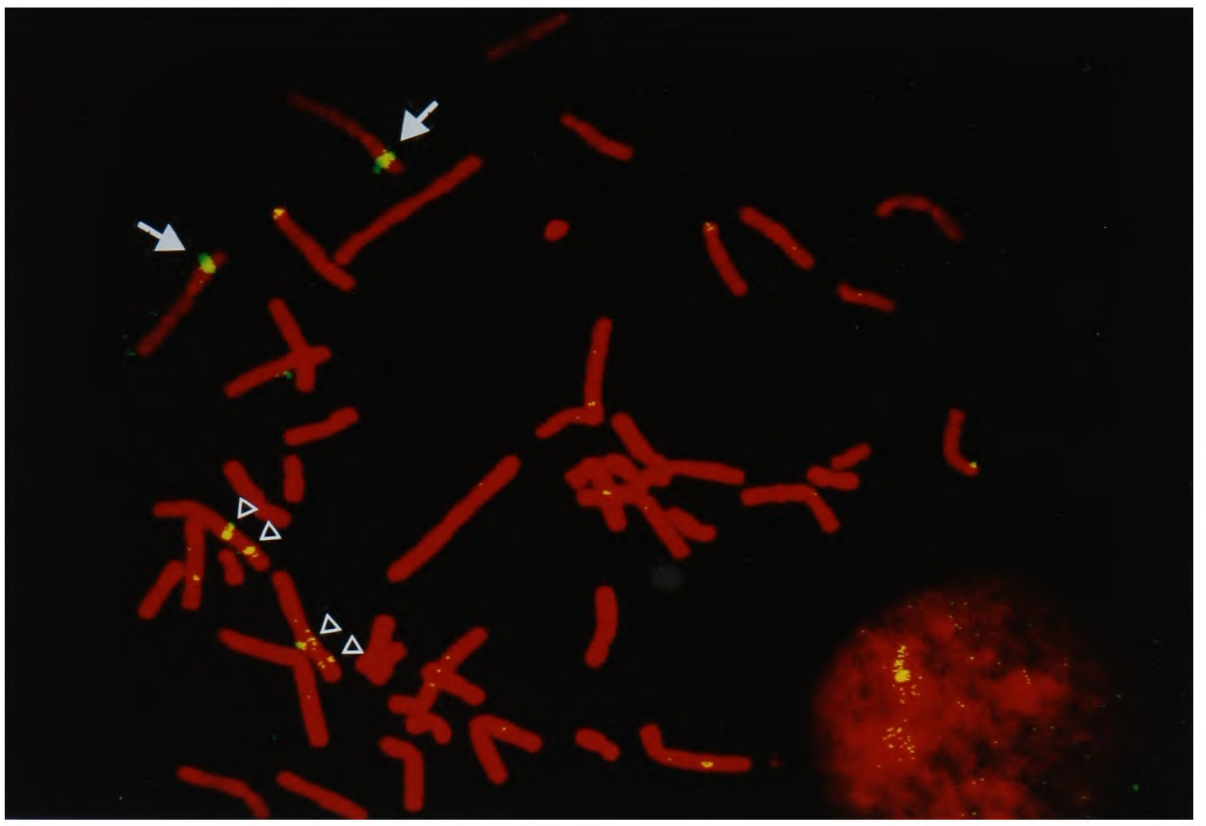


Figure 2.33 FISH of 947f6. This examination demonstrated hybridisation signal from both arms of chromosome 5 (open arrows) in addition to that from 6p21.3/22.1 (solid arrows). (photograph kindly providied by Sharon Horsley and Lyndal Kearney)

In view of this apparent inconsistency in the STS content analysis and FISH data, the results of FISH analysis on clones previously identified with V4 were reviewed.

YACs extending telomeric to V4 commonly showed cross-hybridisation to 5p13-14 and/or 5q12-13. This feature was consistent across clones from three separate YAC libraries (ICRF, St. Louis and CEPH). YACs either side of this region not demonstrating this feature (Q68B5 and 901a10) suggested that this region covered up to 200kb between the markers D6S1558 and D6S1621.

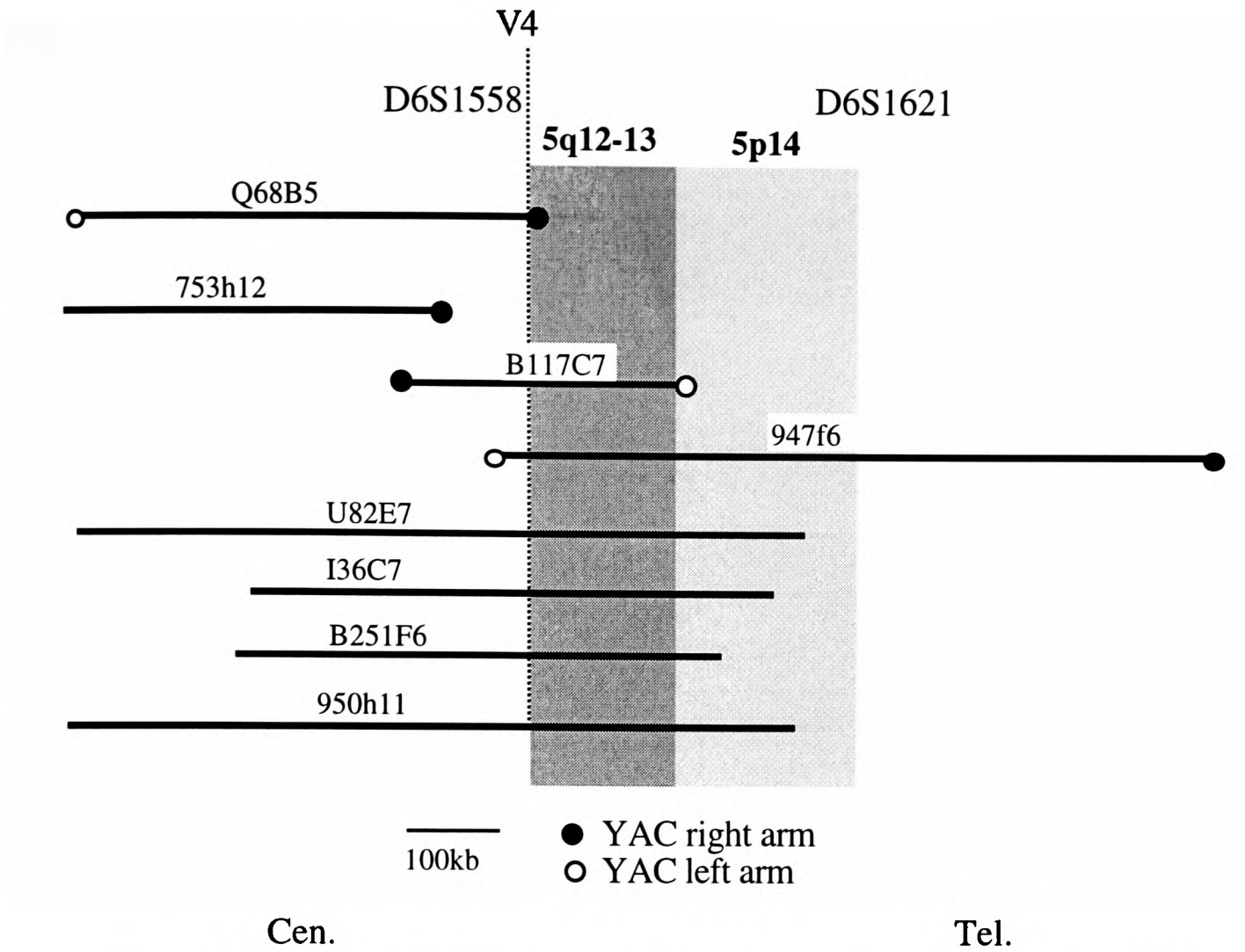


Figure 2.34 Review of FISH data suggested the existance of a region telomeric to V4 that was "duplicated" on chromosome 5

It was into this region that the microsatellite locus D6S1016 had been mapped by the Whitehead Institute and a region of duplication with representation on chromosome 5 was postulated to account for both the FISH results and the behaviour of D6S1016 on genomic DNA.

Duplicated regions of a human β-glucuronidase gene have been reported to map to both arms of chromosome 5 in addition to 6p21.3 (Sargent *et al.* 1994). The PCR primers reported by Sargent to amplify from the β-glucuronidase sequence on 6p21.3 failed to generate product when YACs from the region were used as template (J. Pointon). The source of the "duplicated" signal remains undefined.

Linkage disequilibrium analysis with new markers

Analysis of new microsatellite markers telomeric to D6S105 (D6S1558, D6S1621, D6S1281, D6S1554) suggested a peak of linkage disequilibrium (Mark Worwood). Calculations of λ (the proportion of the maximum possible association between a marker and a disease, Terwilliger 1995) for these new markers showed

D6S1558 to be as strongly associated with haemochromatosis as D6S1260 ($\lambda = 0.71$). This association declines at the next telomeric of the markers, D6S1621 ($\lambda = 0.60$).

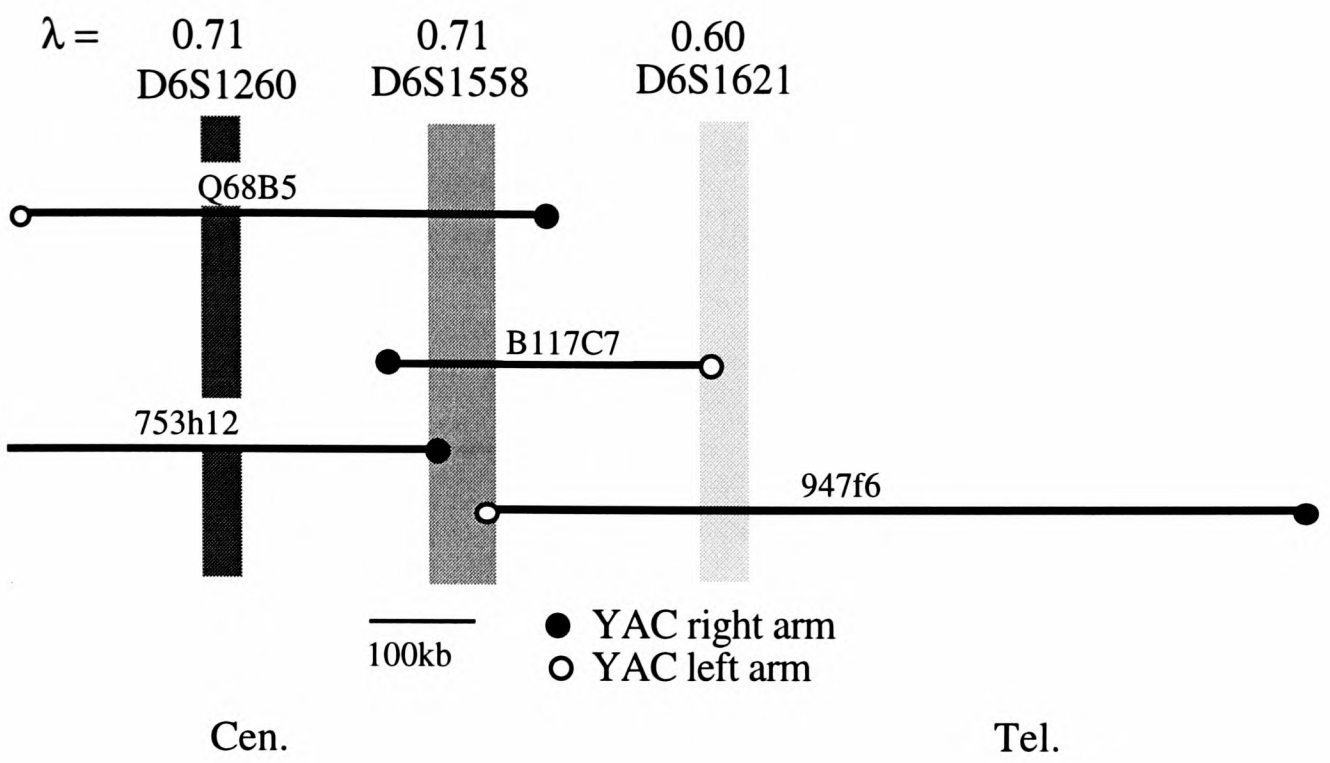


Figure 2.35 Early allele association data suggested that D6S1621 might represent a telomeric limit for the haemochromatosis gene region.

These results suggested a likely location for the haemochromatosis gene centromeric to D6S1621. This prompted the prospective identifcation of further microsatellites from this region to define a peak of association.

Chapter 2.4

The isolation of microsatellite markers telomeric to D6S1260

Background

The data suggested a plateau of disease association between D6S1260 and D6S1558 and more markers were sought from this region to further refine the likely position of the gene. The St. Louis library YAC B117C7 covered the region immediately telomeric to D6S1558 and this clone was selected for both high resolution mapping and the identification of further CA repeat microsatellites.

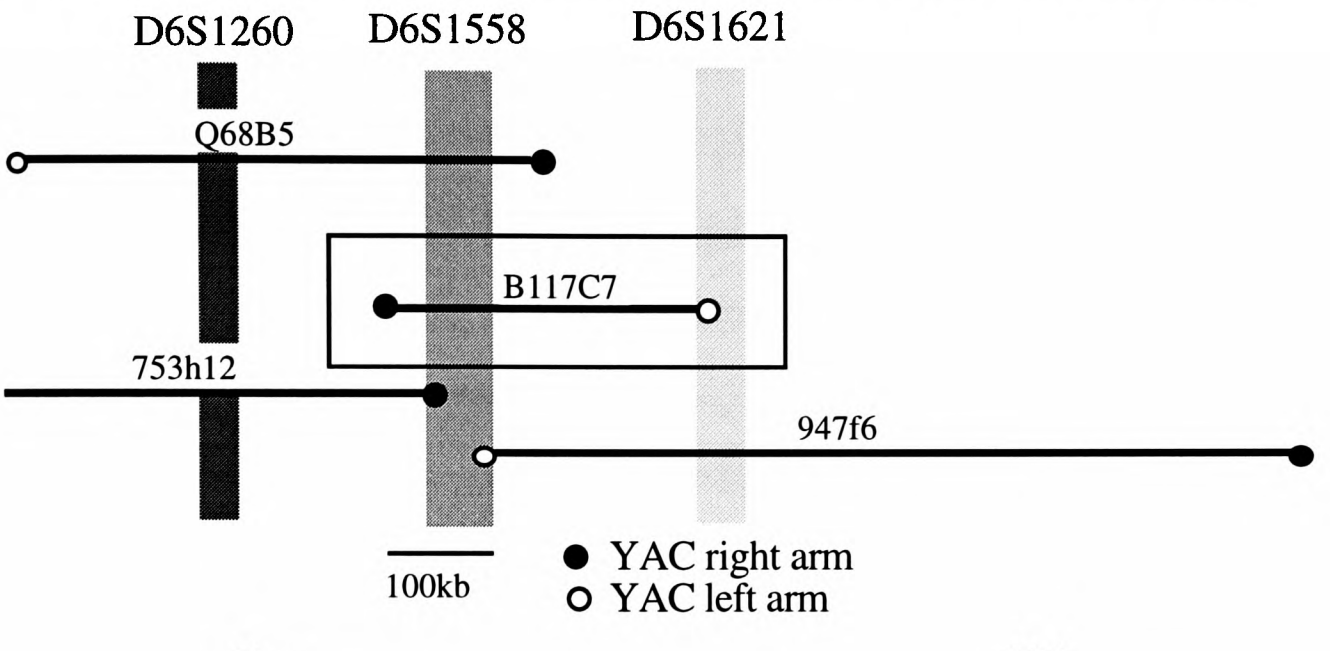


Figure 2.36 B117C7 was selected for mapping and isolation of new CA repeats

CA repeat loci occur on average once every 30kb in human genomic DNA (Hearne *et al.* 1992). As CA repeats occur less frequently in yeast, these microsatellites may be isolated from human DNA cloned in YACs by sequential sub-cloning of the DNA into smaller vectors with selection of CA containing sub-clones by sequential hybridisation to total human and CA probes.

Methods

Restriction mapping B117C7

B117C7 was mapped for rare-cutting restriction enzyme sites as described earlier (Chapter 2.2).

Isolation of CA repeat loci from YACs

YAC DNA was partially digested with a frequently cutting restriction endonuclease and cloned into phage (accomodating fragments average size 15-20kb) or Cosmid vector (average insert size 30-40kb). Sub-clones containing human inserts were selected by hybridisation of filter lifts of phage plaques with a Cot 1 DNA probe. These phage clones were then selected for clones containing CA repeats by hybridisation of either filter lifts or Southern blots of digested phage DNA with a CA probe. Complete digestion with a second restriction enzyme and "shotgun" cloning into plasmid (pUC 18 average insert size 0.2-1kb) followed by another round of selection reduced the cloned DNA to a size which may be sequenced until identification of the CA repeat.

PCR primers are designed to the non-repetitive sequence flanking the CA repeat. These primers enable the physical mapping of the locus by STS content analysis on YACs from the region. Codominant inheritance of the marker is demonstrated by analysis of alleles amplified from three generation CEPH reference pedigree DNA.

Results

Restriction map of B117C7

Restriction mapping of B117C7 revealed a high density of infrequently cutting enzyme sites.

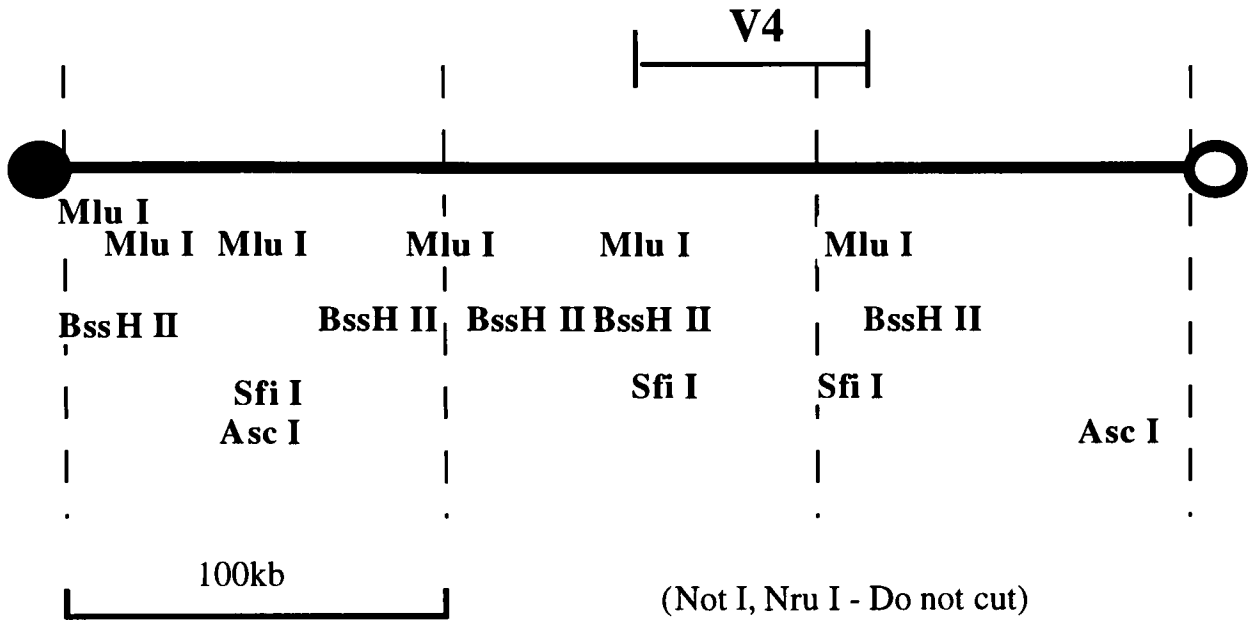


Figure 2.37 Restriction map of B117C7. V4 was mapped by hybridisation of labelled PCR product to a blot of restriction digested DNA.

Such a high density of restriction sites for the infrequently cutting enzymes implies a high GC content of the region, which in turn is associated with coding sequence (i.e. genes).

Sub-cloning YAC B117C7 into λ phage

In view of the fact that the YAC B117C7 was relatively small (300kb) and that we predicted that it might lie in a region that might be of further interest with regard to the position of the haemochromatosis gene it was sub-cloned into λ phage (Lambda DASH II, Stratagene).

Total yeast DNA of B117C7 in an agarose block was partially digested with 20u of *Mbo* I at 37°C for 30 minutes. The reaction was stopped with 0.5M EDTA and the extent of digestion was checked by PFGE. The block was treated with β–Agarase at 37°C overnight.

The digested DNA fragments were treated with 0.1u of calf intestinal alkaline phosphatase in T4 buffer for 30 minutes at 37°C. Again the enzyme was stopped with 0.5M EDTA. These fragments were ligated to *Bam* HI cut Lambda DASH II at 4°C for 24 hours. The ligation product was packaged using Gigapack II XL packaging

extract as per the manufacturers instructions. Packaging was halted with SM buffer and chloroform.

One hundred microlitres of packaged ligation was adsorbed to the bacterial strain XL1 Blue MRA (P2) for 15 minutes at 37°C. All of the ligation was packaged and plated on NZY top agarose producing 10 primary phage plates.

Hybond N membranes were used to lift phage from the surface of the primary plates and, after denaturation and fixation of the DNA, these were probed with radio-labelled C_0t 1 human DNA (Gibco).

Sixty clones were identified with the C_0t 1 hybridisation. The density of plaques on the primary plates was too high to allow individual clones to be picked and in each case the region surrounding the positive signal was picked from the plate with a sterile pipette tip and eluted into SM buffer. The eluted phage were replated in NZY top agarose at 100-fold dilution. A second round of filter lifts and hybridisation with C_0t 1 allowed 36 individual positive clones to be picked.

After identifcation of the human containing clones, the primary and secondary filters were stripped and re-probed with a CA probe (Pharmacia). Twenty-seven of the clones showed positive hybridisation with CA.

Identification of CA repeats

All of the 36 λ phage clones were plated to near confluence in top agarose. Phage were eluted and DNA prepared by polyethylene glycol precipitation (Qiagen <lambda> miniprep). DNA was digested overnight with *Eco* RI and *Hin*d III. Fragments were resolved by electrophoresis in 1% agarose and transferred to Hybond N nylon membranes (Amersham) by Southern blotting. These blots were hybridised with a $(CA)_n$ probe.

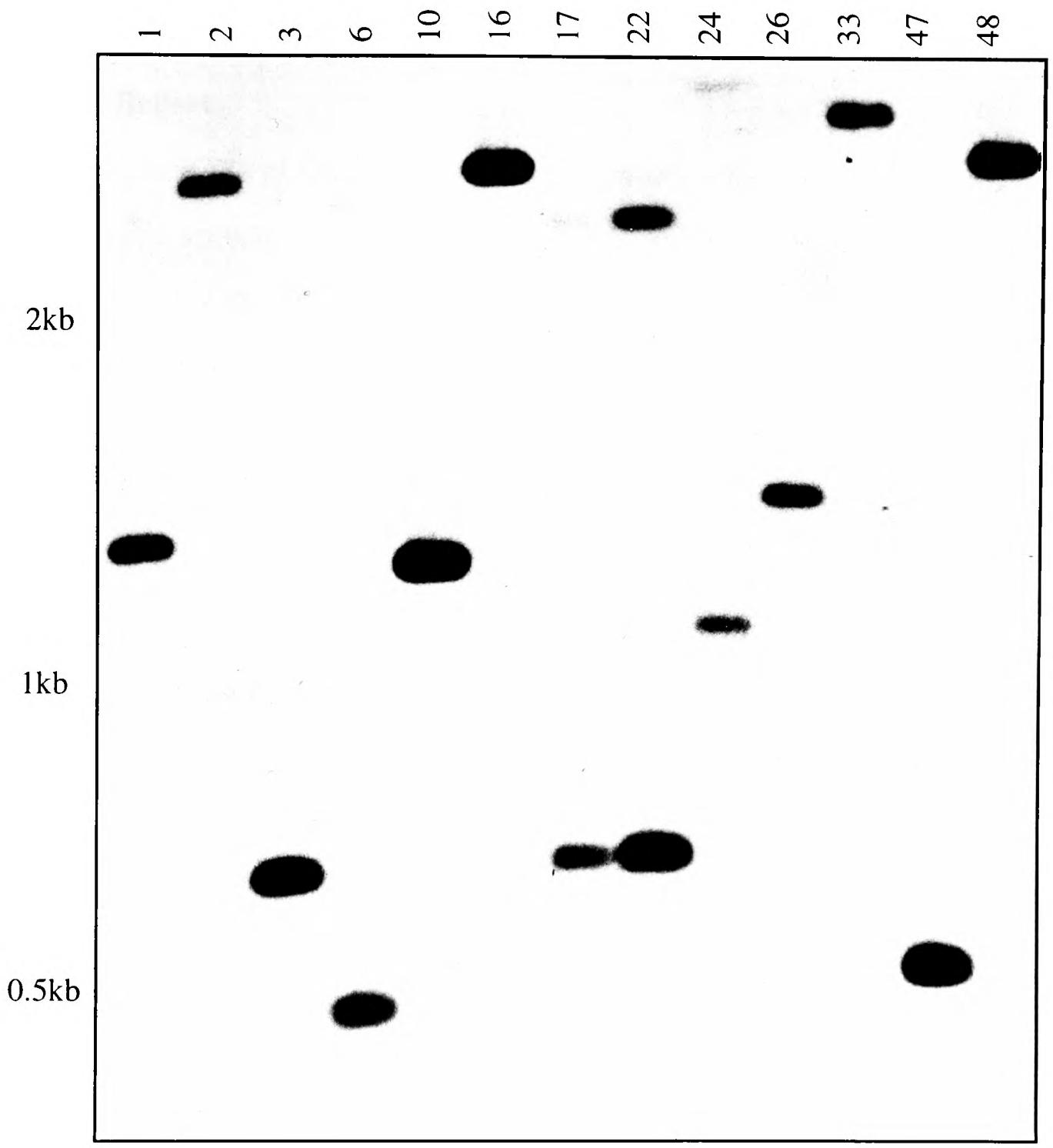


Figure 2.38 *Eco* RI/*Hin*d III digested phage DNA probed with CA. Washed in 2xSSC at 65°C.

This process identified at least 12 $(CA)_n$ repeat bands. Most phage clones contained a single $(CA)_n$ repeat.

CA containing phage clones (initially clones 1, 3, 10 and 16) were digested to completion with *Sau3A*I and cloned into pUC18. DH5α competent cells were transformed with pUC18 and plated onto Hybond membranes placed on top of agar (Luria broth and ampicillin). After incubation overnight at 37°C, colonies were lifted from these filters. The bacterial colonies on the copy filters were lysed, the DNA denatured and fixed to the membrane. These copy filters were hybridised with a

radiolabelled CA probe. CA positive pUC18 clones were sequenced with M13 forward and reverse primers until the CA repeat was identified.

Yeast CA Repeats

Sequence of a pUC clone derived from Phage clone 16 (C7p16.1) revealed a long CA repeat sequence. However, FASTA analysis of the sequence scored 100% identity with a *Saccharomyces cerevisiae* sequence (SCF9765.Emfun).

Although less abundant, CA repeat sequences are known to occur in yeast and this observation implied that by using total yeast DNA for the creation of the phage sub-library some clones with yeast rather than human inserts had inadvertantly been picked. The library was reviewed to clarify what proportion of phage clones this applied to.

Having undertaken two rounds of selection with hybridisation to human C_0t 1 DNA, there was no doubt that C7p16 was unequivocally positive on both hybridisations.

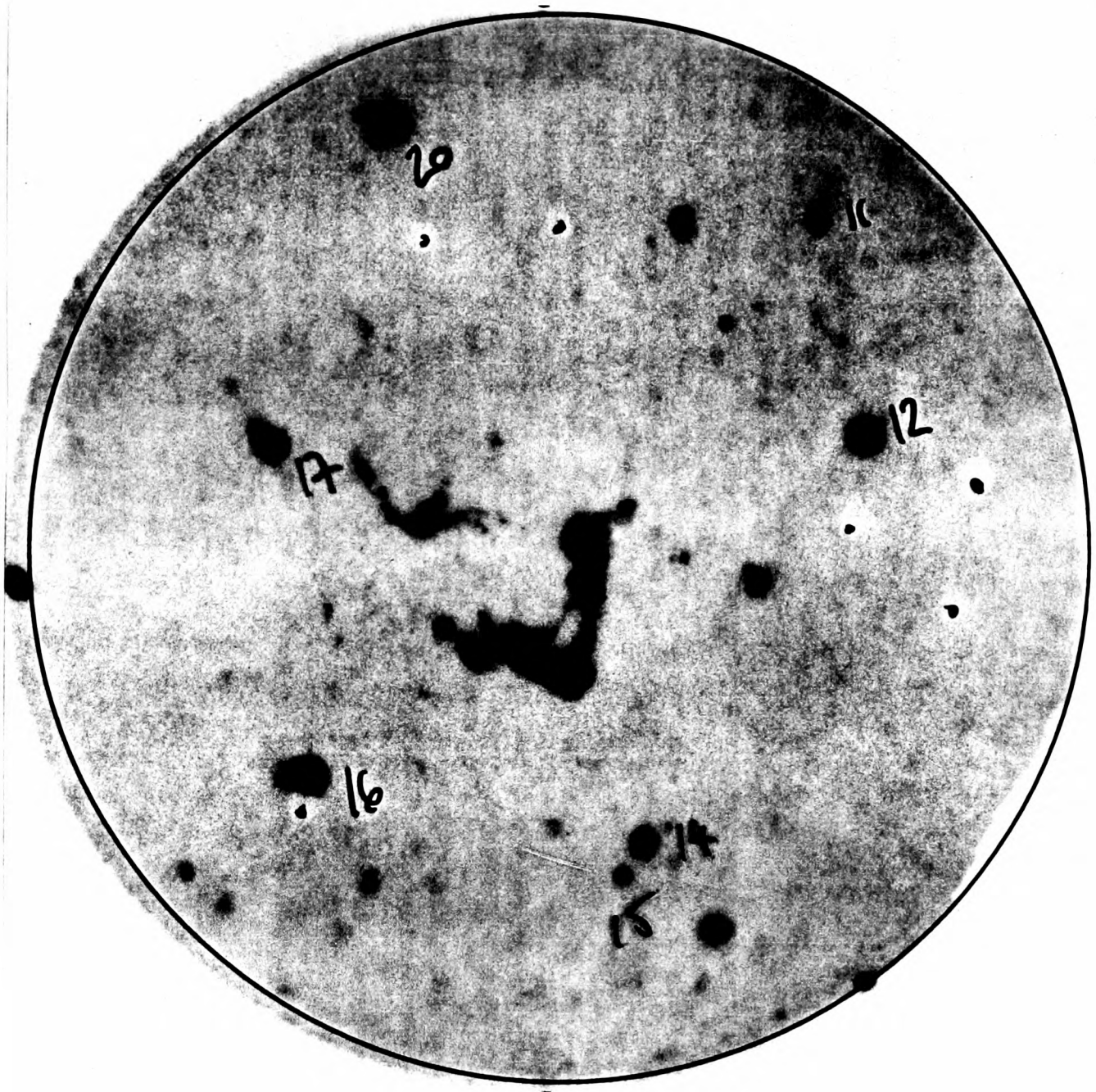


Figure 2.39 Phage clone C7p16. First round hybridisation with C_0t 1 probe. Filter washed to 2xSSC, 0.1% SDS at 65^0C.

This signal represented cross-hybridisation of the C_0t 1 probe to yeast CA sequence.

The Southern blots of the *Eco* RI and *Hin*d III digested phage DNA were stripped and re-probed with total yeast DNA from AB1380 (the YAC library host strain). Yeast has fewer repetitive elements other than CA and the result of this hybridisation was that only CA bands were identified.

Ultimately the "fingerprint" generated for each phage clone from hybridisation of these blots with human C_0t 1 DNA provided sufficient information on which to identify the clones that contained yeast. Yeast containing clones identified from sequence analysis (C7p16 and C7p2) showed only a single band when probed with

C_0t 1 DNA and this band always corresponded to the band identified by the CA probe.

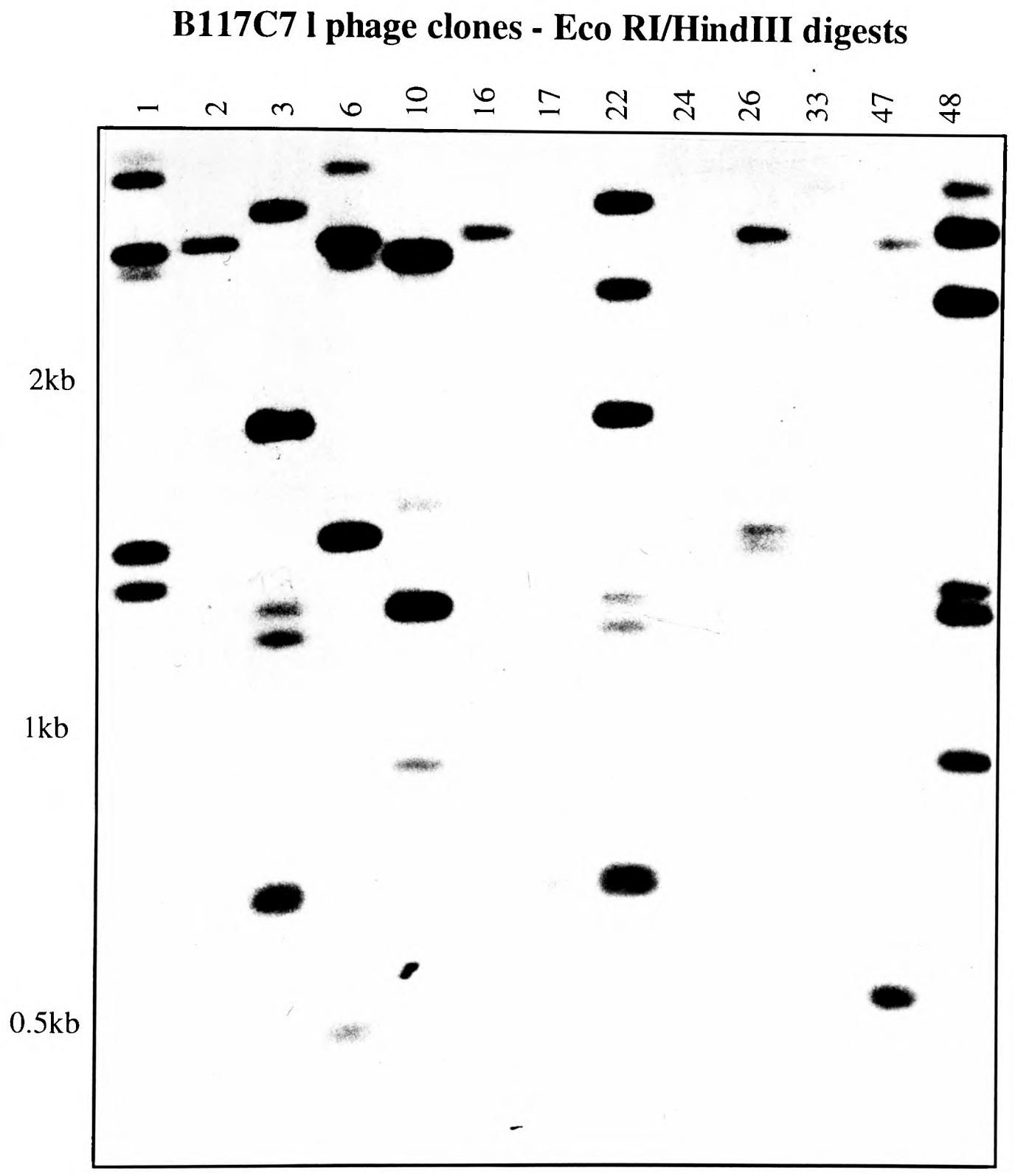


Figure 2.40 *Eco* RI/*Hin*d III digested phage DNA probed with C_0t 1 DNA. Clones 2 and 16 contained yeast (as determined by sequence analysis). In each case only a single band appears with the Cot 1 probe and this is the same band identified by the CA probe (figure 2.38, page 87). Filter washed to 2xSSC, 0.1% SDS at 65^0C.

On this basis twelve of the phage clones were concluded to have yeast inserts. None of these clones was sequenced further.

New Microsatellite Loci

Three human CA repeats were identified in the initial screen.

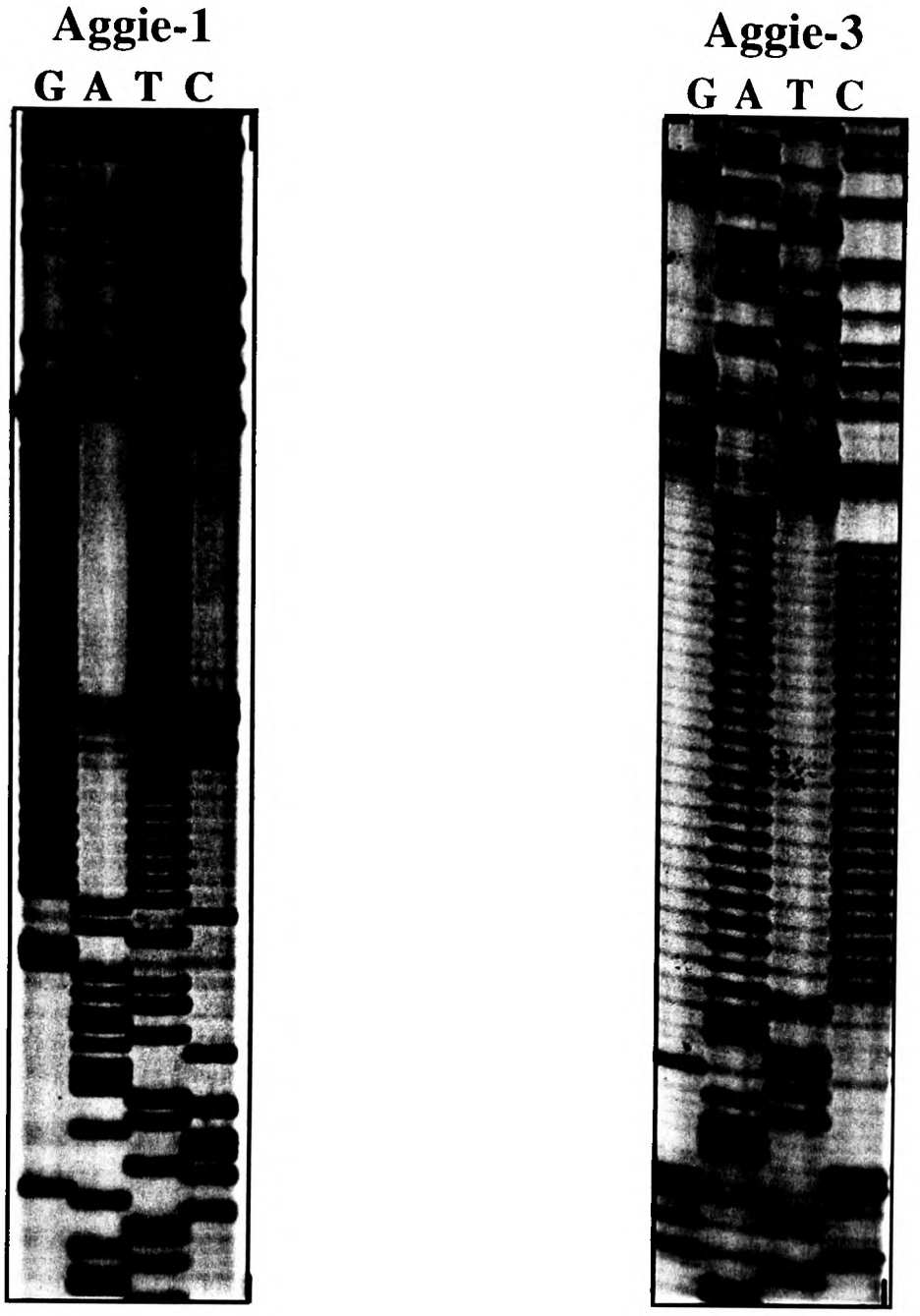


Figure 2.41 Human CA repeat loci isolated from B117C7

Aggie-1.

Phage clone 10 (C7p10) yielded a long CA repeat interupted by an apparent compression ($(CA)_9GCA\ CGG(CA)_{24}$). By STS content analysis on YAC clones this repeat was found to lie telomeric to D6S1558.

YAC	D6S1260	D6S1558	Aggie-1	D6S1621
Q68B5	+	+		
B117C7		+	+	
753h12	+			
947f6			+	+

Table 2.10 STS content analysis of Aggie-1 on YACs

PCR on YAC, phage or cosmid template DNA with primers to this locus produced a single band. When human genomic DNA was used no distinct bands could be identified.

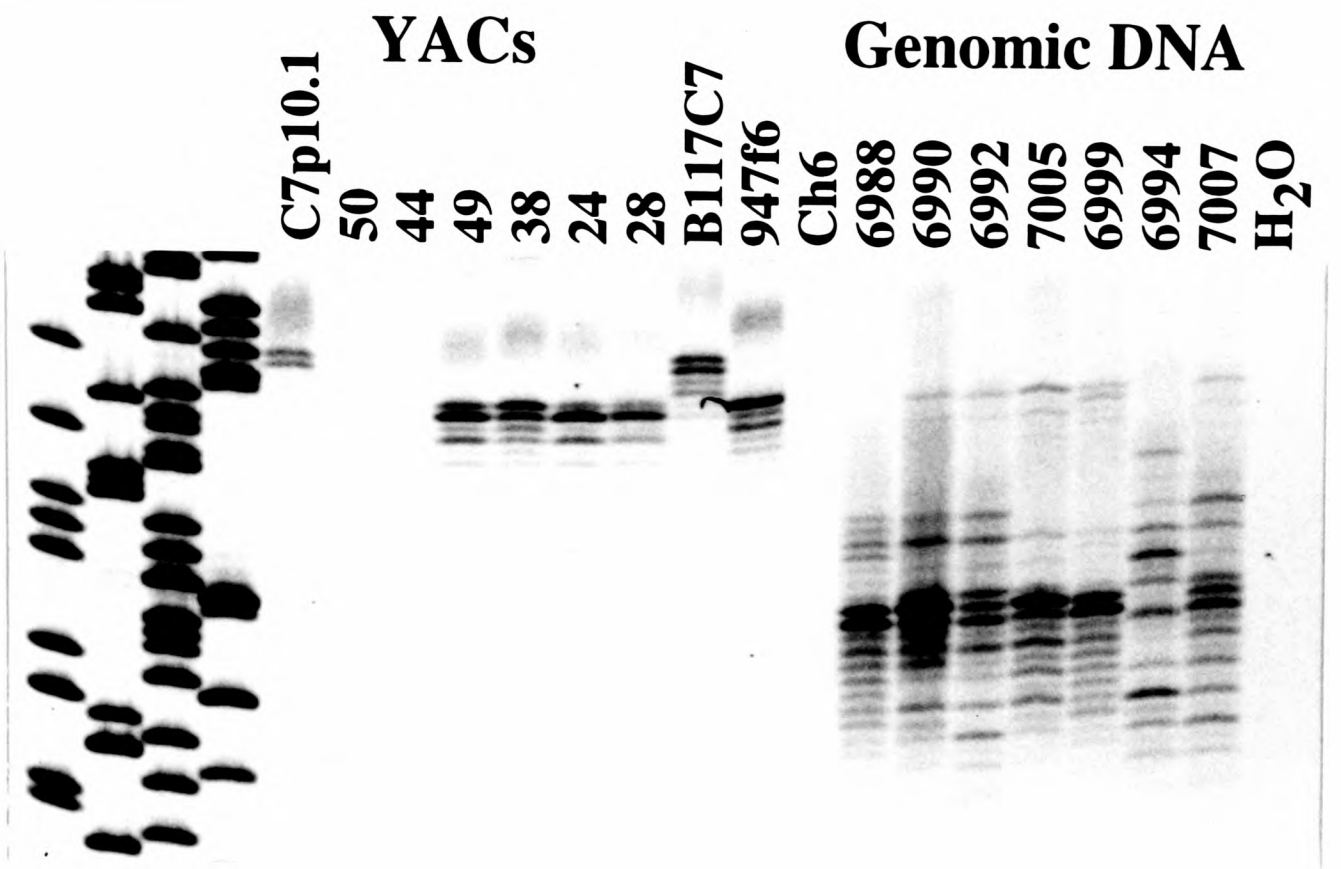


Figure 2.42 Amplification of Aggie-1 on genomic DNA. Despite all adjustments to reaction conditions no distinct alleles could be defined.

The predicted physical position of Aggie-1 was similar to that of D6S1016 (see Chapter 2.3, page 76). However, whereas D6S1016 generated multiple distinct products from genomic DNA, Aggie-1 produced a smear. This suggested that the sequence from which the primers had been designed contained a repetitive element (Alu, LINE etc). Repeated FASTA analysis of the flanking sequence failed to identify any recognised human repetitive element. Despite many measures (re-designing the primers, increasing the stringency of the reaction or the use of Amplitaq Gold - a heat-stable DNA polymerase marketed for its potential to increase the specificity of PCR priming) no clean product could be obtained from genomic DNA. Ultimately allele scoring for Aggie-1 was not possible and this microsatellite could not be used in patient analyses.

Aggie-2.

Phage clone 1 (C7p1) produced a complex repeat - $(TA)_9(CA)_9CG(CA)_5$. This locus lay centromeric to D6S1558.

YAC	D6S1260	Aggie-2	D6S1558	Aggie-1	D6S1621
Q68B5	+	+	+		
B117C7		+	+	+	
753h12	+	+			
947f6				+	+

Table 2.11 STS content analysis of Aggie-2 on YACs

FASTA analysis demonstrated Alu repetitive sequence flanking **both** sides of this sequence precluding the use of this locus as an informative marker.

Aggie-3.

Aggie-3 was isolated from phage clone 3 (C7p3.4). This uninterupted $(CA)_{24}$ repeat also mapped between the markers D6S1260 and D6S1558.

YAC	D6S1260	Aggie-3	Aggie-2	D6S1558	Aggie-1	D6S1621
Q68B5	+	+	+	+		
B117C7		+	+	+	+	
753h12	+	+	+			
947f6					+	+

Table 2.12 STS content analysis of Aggie-3 on YACs

Primers designed to this locus amplified alleles in human genomic DNA which showed co-dominant inheritance in CEPH reference pedigrees.

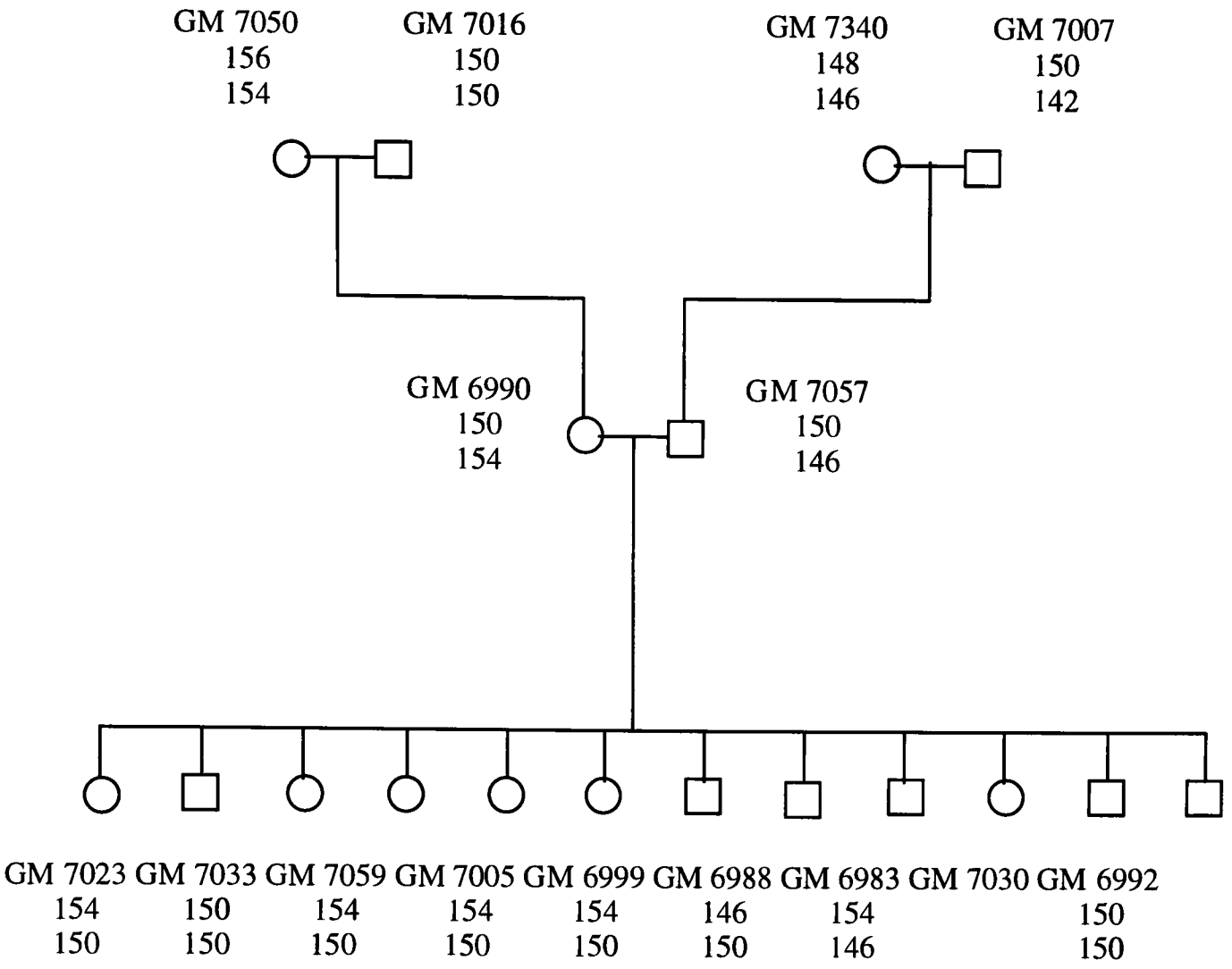


Figure 2.43 Inheritance of Aggie-3 through CEPH reference family 982 (Utah pedigree K-1331). Alleles are identified by size in base-pairs.

A total of 10 alleles were seen at this locus ranging in size from 138bp to 156bp.

All CA repeat loci present on B117C7 accounted for

Although the initial blot of *Eco* RI/*Hin*d III digested phage hybridised with the CA probe suggested that there might be up to 12 CA repeat loci on B117C7 (see Figure 2.38, page 87), subsequent analysis was only able to confirm the presence of 5. Some of the bands on the original blot proved to be of yeast origin (see discussion above). Of the human CA repeats all were ultimately identified as being one of either Aggie-1, Aggie-2, Aggie-3, AM-2 or D6S1558 (see Chapter 3.4, page 171).

The combination of mapping new Généthon markers and the isolation of a new CA repeat (Aggie-3) established an even distribution of polymorphic markers immediately telomeric to D6S1260. Of these Aggie-3 mapped between the two microsatellites that had hitherto demonstrated the peak of linkage disequilibrium with haemochromatosis.

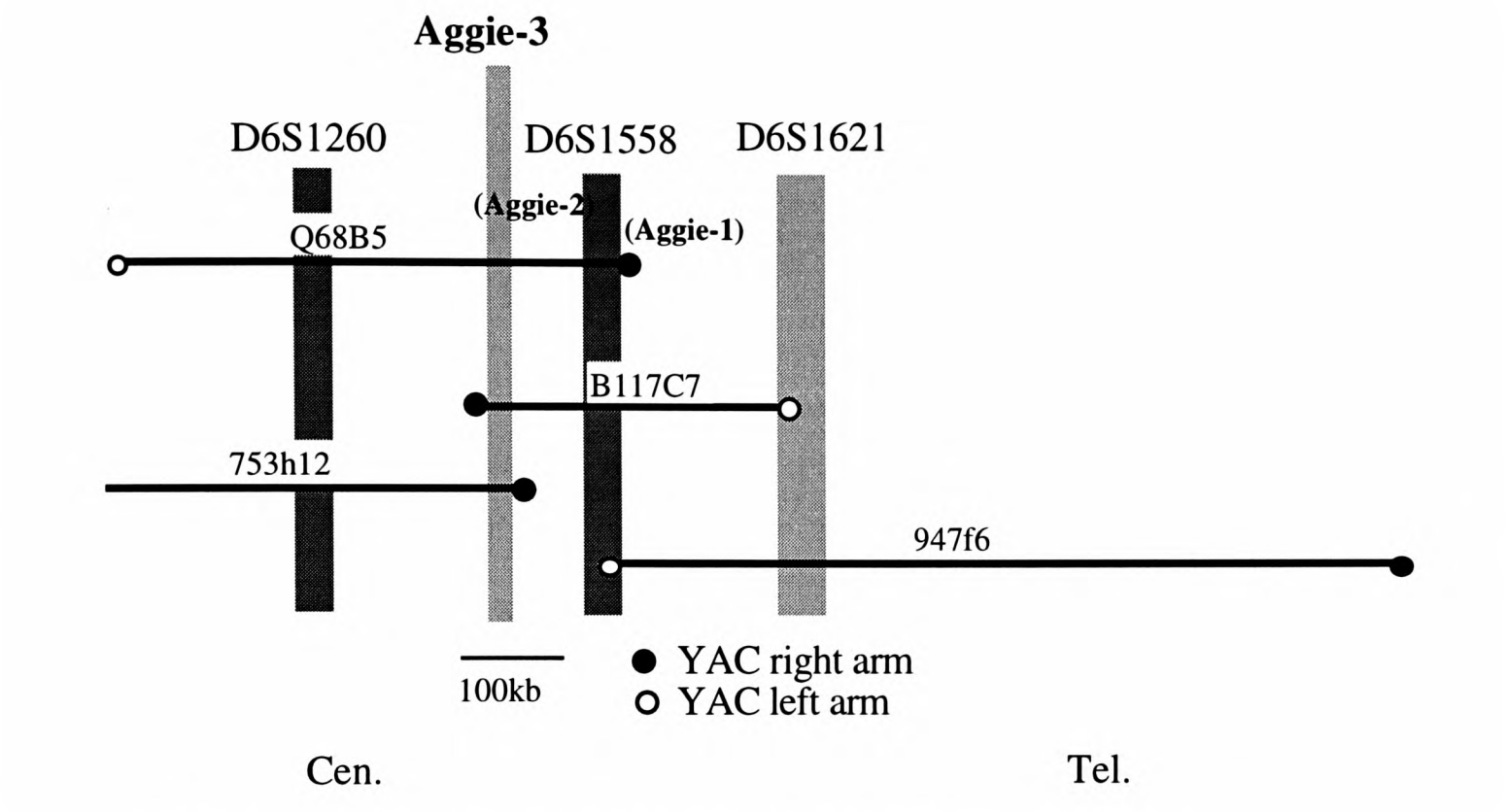


Figure 2.44 New microsatellites telomeric to D6S1260.

Chapter 2.5

Linkage disequilibrium mapping with new markers in the haemochromatosis gene region

Background

Association

Traditional genetic linkage analyses observe the frequency with which a marker allele and a disease phenotype co-segregate within families affected by a given disorder. Such an approach has the resolution to map a disease gene to within 1cM (i.e. identifying marker loci that are separated from the disease locus by as little as a single recombination in 100 meioses). In the case of haemochromatosis this degree of linkage has been shown at two loci that are separated by a physical distance of greater than 2.5Mb of DNA (HLA-A and D6S105). Finer mapping of disease loci is possible using association studies that examine the coexistance of a marker allele and the disease in large populations of unrelated individuals. In principle this increases the number of meioses available for examination but this approach does require certain assumptions be made about the population under study. The most fundamental of these is that there is a common disease allele that was introduced into the population on a single chromosome. The alleles at the loci flanking the disease gene then make up the "ancestral haplotype" of markers associated with the condition. During successive generations recombination causes this haplotype to disintegrate and ultimately only alleles at the marker loci closest to the gene remain.

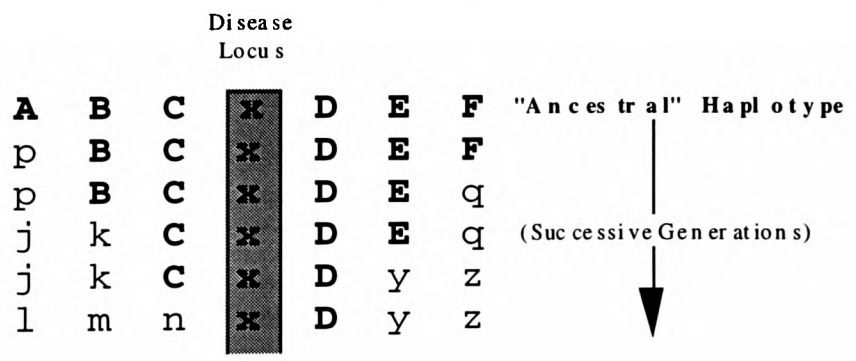


Figure 2.45 The principle of association. An original disease causing mutation (x) is associated with an "ancestral" haplotype of alleles at flanking marker loci (A, B, C, D, E and F). With successive generations recombination will gradually separate these alleles leaving those that are close to the disease locus (C and D).

Where marker alleles remain associated with a locus more often than would be predicted by the physical distance between them they are said to be in linkage disequilibrium. When this effect is very marked it implies that the physical region into

which the loci map is less subject to the (non-random) process of chromosomal recombination. In such circumstances genetic distances (as calculated from the observed number of recombinations) may underestimate the physical distance separating the loci.

Although association studies are often reserved for confirmation of genetic linkage and more recently the fine mapping of disease loci, Simon's original study of HLA class I types and haemochromatosis was an association study (Simon *et al.* 1976). In demonstrating a significant excess of the HLA-A3 antigen in unrelated patients over that seen in controls, he predicted genetic linkage and placed the responsible gene on the short arm of chromosome 6.

The principle of allele association of genetic markers was adopted by Jazwinska in a study of HLA-A and the microsatellite D6S105 in unrelated individuals with haemochromatosis (Jazwinska *et al.* 1993). In the Australian population allele 8 of D6S105 (124bp) inferred a greater allele relative risk for haemochromatosis than the HLA-A3 serotype suggesting that the microsatellite was a closer marker to the gene than the class I locus. Subsequently more sophisticated methods have been developed to quantify the extent of linkage disequilibrium between marker loci and a disease.

Quantitative assessment of linkage disequilibrium

The present day population of Finland was founded by a small number of settlers approximately 2000 years ago. This was sufficiently recent for an ancestral haplotype to be recognisable for many single gene disorders, but sufficiently long ago for the process of recombination to cause all but the most marked regions linkage disequilibrium to have dissipated. On this basis, Hästbacka and colleagues have applied an adaptation of calculations originally derived from observations on mutation rates in bacterial populations to quantify the extent of linkage disequilibrium between a marker and a disease locus. Their parameter (p_{excess}) attempts to infer the rate of rare events (in this case recombination) based on the frequency in the present generation (Hastbacka *et al.* 1992).

$$p_{excess} = (p_{affected} - p_{normal}) / (1 - p_{normal})$$

Where $p_{affected}$ and p_{normal} denote the frequency of the disease associated allele on disease-bearing and normal chromosomes respectively. The statistic p_{excess} should be directly related to proximity to the disease locus (Hastbacka *et al* 1994).

When applied to the Finnish population, calculation of p_{excess} suggested a location for the gene responsible for diastrophic dysplasia within 0.06cM (or about 60kb) from a polymorphism at the colony stimulating factor 1 receptor (CSF1R) locus on the long arm of chromosome 5 (Hastbacka *et al.* 1992). A novel sulphate transporter gene was subsequently cloned and found to lie 70kb from CSF1R (Hastbacka *et al.* 1994).

More statistically sophisticated and essentially conservative methods have subsequently been reported that allow the application of linkage disequilibrium mapping to less well defined human populations (Kaplan *et al.* 1995). Terwilliger reported a likelihood-based approach to testing for linkage disequilibrium that accommodates data from multiple, multi-allelic markers such as microsatellites without having to reduce this to a bi-allelic system (Terwilliger 1995). The derived parameter λ reflects the proportion of the maximum possible association between a marker and disease.

Application of the Terwilliger method to allele association data from markers extending telomeric to HLA-A showed a peak disease association at D6S1260 (Raha-Chowdhury *et al.* 1995) telomeric to D6S105. At the time of that study few microsatellite markers had been mapped to this region. Physical mapping and the isolation of new microsatellites now promised a clearer definition of the likely position of the gene.

Methods

Knowing the physical order of new markers from the region telomeric to D6S1260 (Aggie-3, D6S1558, D6S1621, D6S1281), allele association analyses were carried out in a population of UK haemochromatosis patients to refine the location of the gene.

Patient and control DNA

Microsatellite analyses were carried out at the Department of Haematology at the University Hospital of Wales in Cardiff (Ruma Raha-Chowdhury and Mark Worwood). The patient group for this study consisted of 42 unrelated individuals with haemochromatosis as defined by an hepatic iron index of greater than 2 or more than 5g of mobilisable iron stores by quantitative phlebotomy. Control samples had been obtained from a group of healthy blood donors from South Wales (Worwood and Darke 1993).

Linkage disequilibrium

Lambda (λ) was calculated as a measure of the proportion of the maximum association between a marker and the disease locus using the Unix program software of Terwilliger (Terwilliger 1995).

Results

Linkage disequilibrium - haemochromatosis and D6S1558 and D6S1621

Lambda values calculated for the markers D6S1558 and D6S1621 were higher than that at D6S105 consistent with a likely gene location telomeric to D6S105.

Locus and Allele	Patients/Controls	λ	p
D6S265/**1** (HLA-A3)	42/47	0.30	$3x10^{-5}$
D6S105/**8**	42/47	0.43	$6x10^{-6}$
D6S1260/**4**	42/47	0.71	$2x10^{-7}$
D6S1558/**3**	42/46	0.71	$7x10^{-4}$
D6S1621/**5**	42/47	0.60	$1x10^{-15}$

Table 2.13 Linkage disequilibrium between microsatellites and haemochromatosis.

The two markers D6S1260 and D6S1558 shared a value of λ of 0.71. Beyond this (D6S1621, D6S1281, D6S1554) the calculated value of λ decayed.

From the physical map D6S1260 and D6S1558 are separated by a physical distance of just over 200kb. Moreover, a value of λ of 0.71 was recognised as "biologically" very significant. In Terwilliger's original report of his method he had taken raw data from the mapping of the cystic fibrosis locus on chromosome 7. Calculating λ for each of the 23 markers surrounding the disease locus he demonstrated a peak value of 0.69 which corresponds well to the obeserved frequency of the common Δ508 mutation (68% in that data set, Kerem *et al.* 1989). Variability has been reported in the haemochromatosis phenotype and, although many cite environmental factors as accounting for this, the prospect of a common mutation in the gene associated with 70% of the affected chromosomes was considered very high.

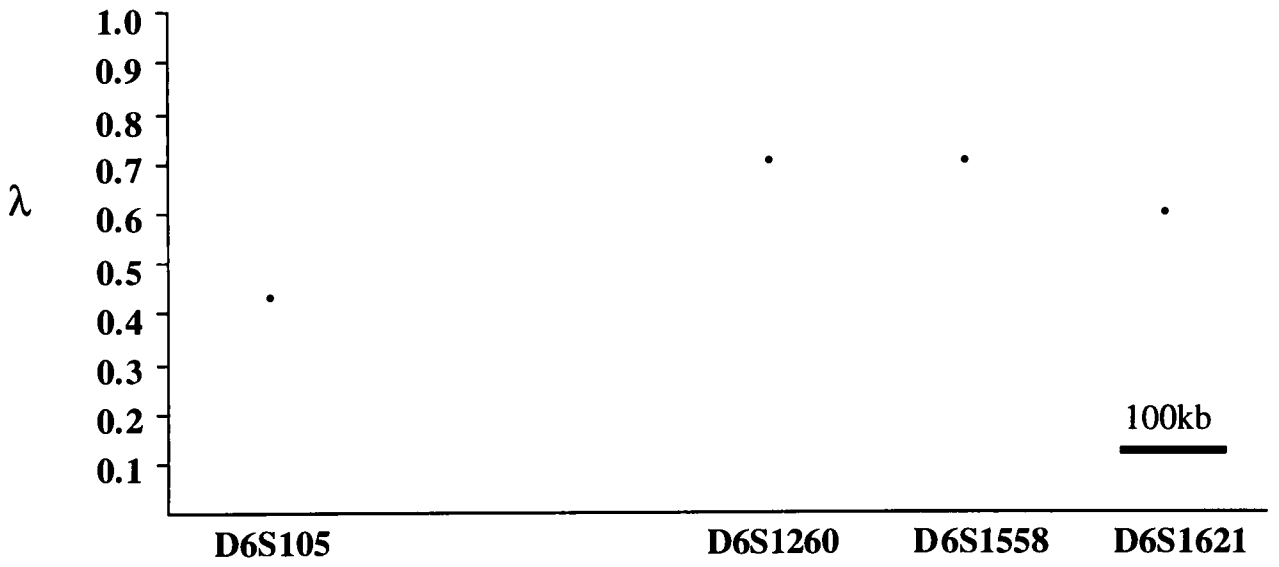


Figure 2.46 Values for λ plotted against physical distance telomeric of D6S105

The isolation of a polymorphic CA repeat physically mapping between D6S1260 and D6S1558 was predicted to either confirm a plateau of association or to further define a peak within this region.

Linkage disequilibrium - Haemochromatosis and Aggie-3

In total 10 alleles of Aggie-3 were seen in control subjects. Of these allele 9 (154bp) is strongly associated with haemochromatosis.

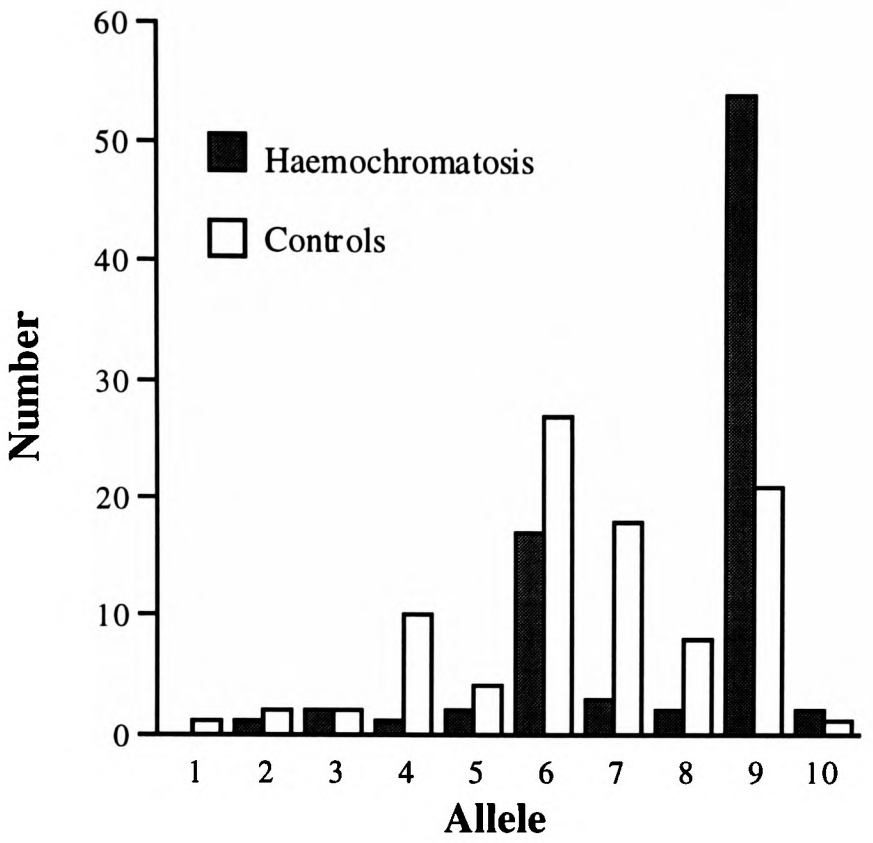


Figure 2.47 Distribution of alleles of Aggie-3 in haemochromatosis and controls

The calculation of λ for Aggie-3 generated a score lower than that at the flanking loci D6S1260 and D6S1558.

Locus and Allele	Patients/Controls	λ	p
D6S265/**1** (HLA-A3)	42/47	0.30	$3x10^{-5}$
D6S105/**8**	42/47	0.43	$6x10^{-6}$
D6S1260/**4**	42/47	0.71	$2x10^{-7}$
Aggie-3/9	**42/47**	**0.53**	**$2x10^{-8}$**
D6S1558/**3**	42/46	0.71	$7x10^{-4}$
D6S1621/**5**	42/47	0.60	$1x10^{-15}$

Table 2.14 Linkage disequilibrium between Aggie-3 and haemochromatosis

Preliminary observations on haemochromatosis haplotypes suggested that alleles 6 and 9 at Aggie-3 defined two distinct disease associated haplotypes in the UK which each carried the common alleles at the flanking loci (D6S1260 allele 4 and D6S1558 allele 3).

	D6S265	D6S105	D6S1260	**Aggie-3**	D6S1558	D6S1621
Patient O.T.	1	8	4	**9**	3	5
	5	6	4	**6**	3	5

Figure 2.48 Alleles at Aggie-3 appeared to split the ancestral haemochromatosis haplotype

This particular haplotypic configuration of alleles at D6S105 and Aggie-3 occured in a total of 7 of 37 patients in which the haplotype was studied. Allowing for the possibility of slippage of mis-scoring (and any one allele varying by a single repeat unit) this number would increase to 12 out of 37 or 32%.

Interpretation

Possible causes of a low allele association

Unlike the flanking markers (D6S1260 and D6S1558), the disease associated allele of Aggie-3 is not the common allele seen in the control population. This raises the possibility that Aggie-3 might define two separate disease associated haplotypes (possibly carrying distinct mutations), that both share the common alleles at D6S1260 and D6S1558. This could account for why the value for λ for this marker is lower.

Microsatellites are by nature unstable. The alternative explanation for these observations is that many generations ago Aggie-3 had undergone mutation on the disease haplotype. The commonest mechanism of mutation at CA repeat loci is allele slippage with variation in a single repeat unit. The two disease associated alleles at Aggie-3 vary by 6 base pairs. Even if the difference in the two alleles was confirmed to represent 3 repeat units by sequencing it is likely that this represents the more likely explanation.

Is disequilibrium analysis and the assumption of a "founder effect" reasonable for studying haemochromatosis in the UK?

Within isolated populations the terms "founder effect" and "ancestral haplotype" imply that a disease causing mutation had been introduced to the "gene pool" of that population on a single affected chromosome carrying a specific haplotype of alleles at marker loci flanking the gene. As the present Finnish population was founded a suitable period ago, they have become the ideal European population for testing hypotheses concerning the application of linkage disequilibrium

analysis to human populations. In addition to the landmark studies of Hastbacka and colleagues in fine mapping the diastrophic dysplasia gene (Hastbacka *et al.* 1992; Hastbacka *et al.* 1994), other genes responsible for rare disease phenotypes that are especially prevalent in Finland have been mapped using this approach (congenital chloride diarrhoea, Hogland *et al.* 1995, progressive myoclonic epilepsy Lehesjoki *et al.* 1993). The statistic used by Hästabacka was an adpatation of the Luria-Delbruck method for estimating mutation rates in bacteria to accommodate the fact that, since its founding, the Finnish population would not have remained stable but grown rapidly. Other likelihood methods have been devised that apply to similar non-equilibrium populations (Kaplan *et al.* 1995). In the worked examples, this method appeared to be useful for diastrophic dysplasia and cystic fibrosis but worked less well for Friedrich ataxia and Huntington disease (Kaplan *et al.* 1995). Hence the utility of this approach would appear to be potentiated by either a recently founded and homogenous population (i.e. the Finns and DTD) or the very high prevalence of a common mutation with a specific haplotype (i.e. Δ508 and CF).

Haemochromatosis is present in Finland, but at a prevalence 10 to 100 times less than in other European populations (Karlsson *et al.* 1988). Haemochromatosis has not been reported to be excessively prevalent in any of the other isolated populations (Amish, Hutterites etc). Australian researchers have argued that their population is ideal for studying linkage disequilibrium in the haemochromatosis gene region (Jazwinska *et al.* 1995). The Australian population stems from the recent settlement of a limited number of European (predominantly British) migrants and haemochromatosis is as prevalent as in the UK (Leggett *et al.* 1990). An "ancestral haplotype" extending over 2.5Mb (D6S248 to D6S105) has been identified in Australian haemochromatosis patients (Jazwinska *et al.* 1995). This ancestral haplotype was present on a third of affected chromosomes, was exclusively associated with haemochromatosis and inferred a haplotype relative risk of over 900. A parallel study of disease phenotype in the same population made the somewhat tautological conclusion that the ancestral haplotype "may be" associated with a "common"

mutation in the gene (Crawford *et al.* 1995). Informative microsatellites mapping telomeric to D6S105 were not included in that study and the authors conclusion that the HC gene region extended from D6S248 to ("and including") D6S105 could be viewed as conservative in the light of the data from the UK with D6S1260 (Raha-Chowdhury *et al.* 1995). More importantly, the MHC has been demonstrated to be under influence of recombination suppression (Martin *et al.* 1995) and the founding of the Australian population is almost certainly too recent for linkage disequilibrium to have had time to dissipate. The profound association of such an extended haplotype might reflect a "secondary founder" effect where the haplotype in that population only in part reflects the original ancestral haplotype and limits the potential for mapping the disease locus.

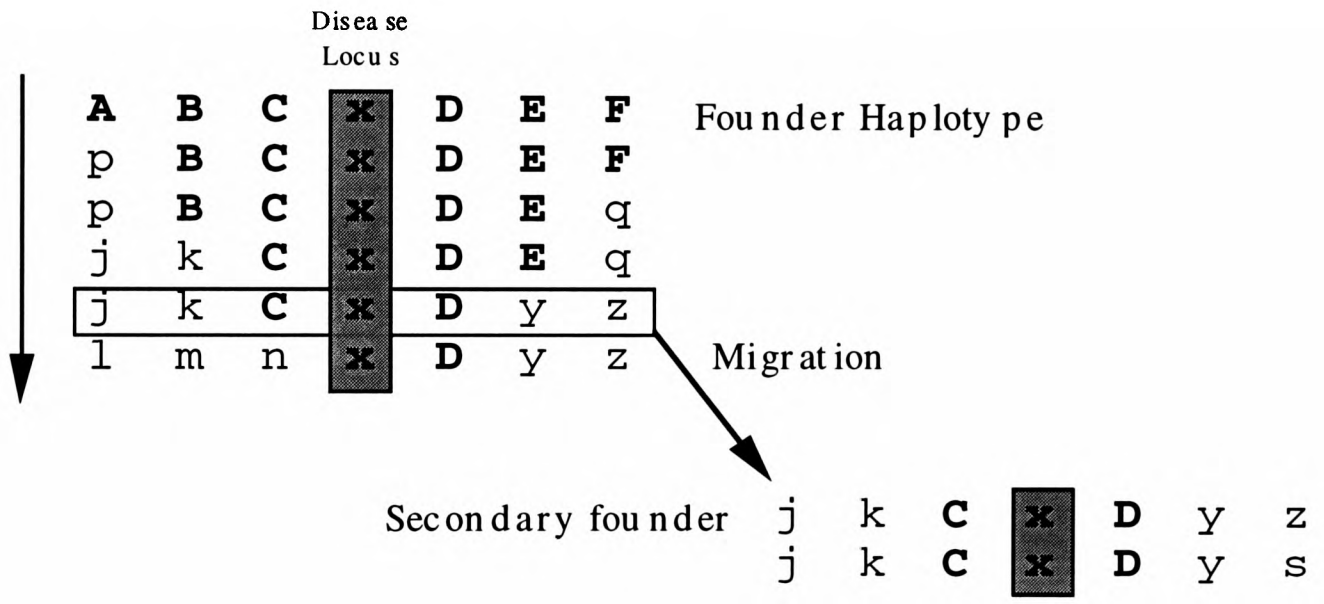


Figure 2.49 Migration and the foundation of a new population creates a "secondary founder" haplotype associated with a disease allele.

Analysis of HLA class I haplotypes in European haemochromatosis patients has suggested that the haemochromatosis mutation was a rare if not unique event (Simon *et al.* 1987). This implies that the linkage disequilibrium model is appropriate for studying haemochromatosis in European populations. Haplotype analyses using microsatellite markers from Europe have confirmed the presence of an ancestral haplotype extending from HLA-A at least as far as D6S105. In Italy this haplotype is present in 30% of patients with haemochromatosis (Camaschella *et al.* 1996) whereas

in Brittany the prevalence of the "ancestral haplotype" reaches 50% (Gandon *et al.* 1996).

Using microsatellites mapping further telomeric to the MHC, analysis in UK patients suggest that the British "ancestral haplotype" extends at least as far as D6S1260 (Raha-Chowdhury *et al.* 1995). The most conserved part of this haplotype lies at the telomeric end. Such haplotype data supports both the existence of a predominant mutation and a location for the gene telomeric to D6S105.

Does the low allele association at Aggie-3 alter the likely position of the haemochromatosis gene?

The new microsatellite, Aggie-3, was isolated from between D6S1260 and D6S1558. Allele association for this marker (as measured by λ) was lower than that at each of the flanking loci and two sepaprate alleles (9 and 6) appeared on ancestral haemochromatosis haplotypes. This most likely represents an historical mutation at Aggie-3 rather than the existance of two separate common haemochromatosis haplotypes. Mechanisms such as this can explain markers that demonstrate lower association but do not detract from the very high association at D6S1260 and D6S1558.

Chapter 2.6

Mapping a recombination that defines a telomeric boundary for the haemochromatosis gene region

Background

A position for the haemochromatosis gene telomeric to the MHC was increasingly likely. However, a clear haemochromatosis gene region had still not been defined. Linkage disequilibrium analyses such as that of Terwilliger are inherently conservative and the unequivocal definition of bondaries for the gene region on the basis of confidence intervals is impractical. The demonstration of a recombination separating the gene locus from the MHC (Calandro *et al.* 1995), in the light of the accumulating association data with more telomeric microsatellites, suggested that HLA-F and subsequently RFP (Malfroy *et al.* 1996) lay as a centromeric boundary for the gene region.

The historical focus on markers from the MHC and the apparent suppression of recombination extending telomeric to the class I region (Martin *et al.* 1995) meant that informative recombinations defining a distal (telomeric) boundary to the gene region had not been reported.

During the course of haplotype studies of UK haemochromatosis families Ruma Raha-Chowdhury and Mark Worwood identified an informative recombinant chromosome within a family under the care of the Liver Unit at King's College Hospital.

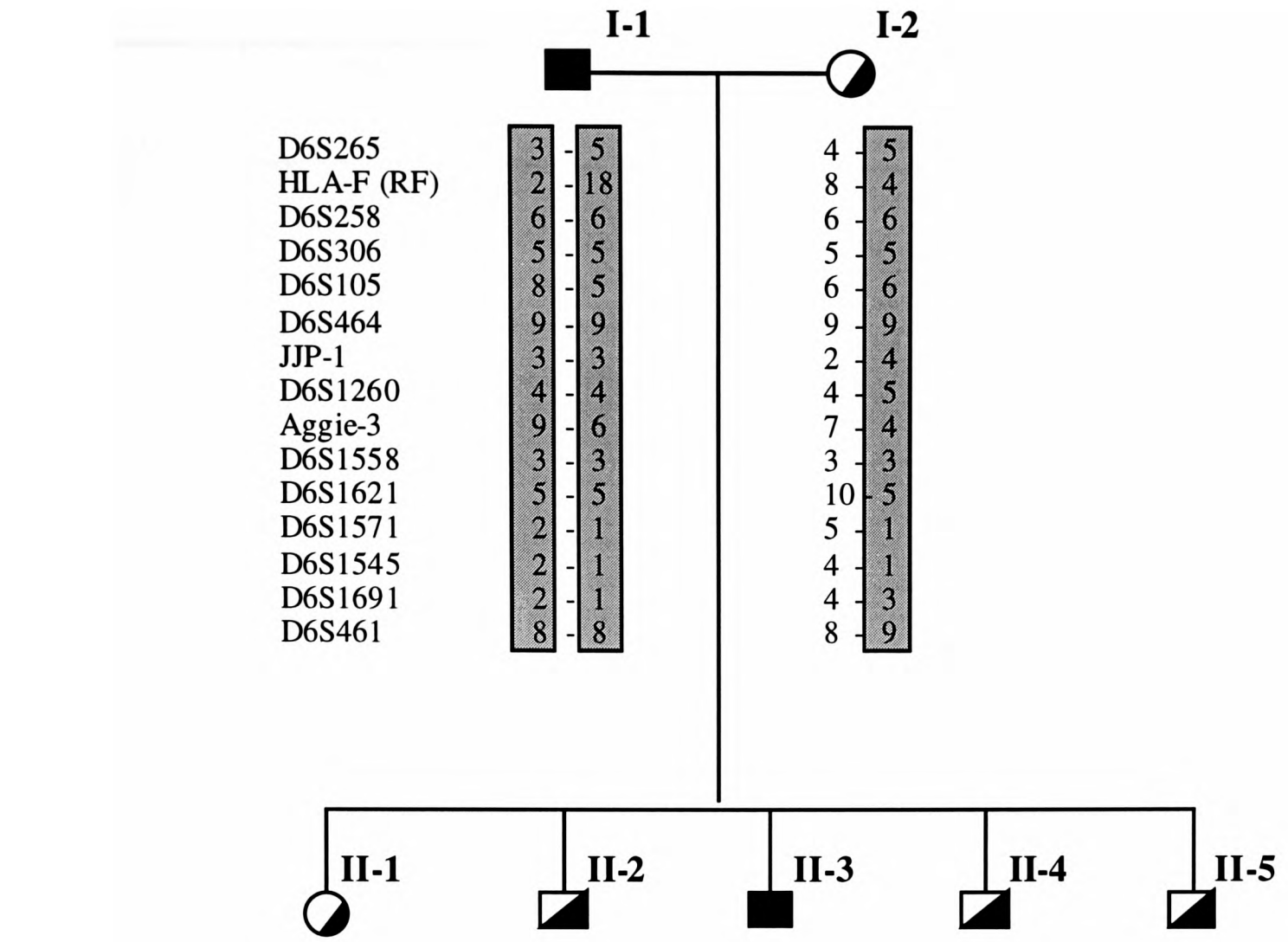


Figure 2.50 Extended haplotypes of markers flanking the haemochromatosis locus in a family with an informative recombination

Patient	Age	Ferritin (mg/l)	Transferrin Saturation (%)	Liver Biopsy
I-1	61	605	96	Grade IV Siderosis
I-2	54	106	30	
II-1	29	41	34	
II-2	27	57	44	
II-3	22	662	67	Grade III Siderosis
II-4	16	42	24	
II-5	14	51	36	

Table 2.15 Iron indices for each of the members of the pedigree outlined in Figure 2.50. I-1 and II-3 both have haemochromatosis on the basis of serum markers of iron status and liver biopsy. No other member of the family has evidence of significant iron loading.

Haemochromatosis is present in two generations of this family. II-3 was diagnosed as having haemochromatosis following a liver biopsy demonstrating grade III siderosis and an hepatic iron index of over 10. His siblings have all been followed clinically for over five years and none has shown evidence of iron loading with current serum iron indices within the normal range.

Subject II-2 (the proband's elder brother) has inherited a recombinant maternal chromosome carrying the distal half of the haemochromatosis haplotype. The fact that he is unaffected implies that the haemochromatosis gene locus lies centromeric to the recombinational breakpoint.

D6S1558 is uninformative in this pedigree as all the family members are homozygous for the allele 3. On the basis of the haplotype data the recombination has occurred between Aggie-3 and D6S1621.

Efforts were directed to refining the position of the recombination breakpoint. From the original haplotype analysis it was predicted that this would be consistent with the allele association data (i.e. a lower value for λ at D6S1621) and it was hoped that the physical mapping of the breakpoint might support the definition of a telomeric limit to the haemochromatosis gene region within which a rational gene search could commence.

Results

Other than D6S1558 (which was uninformative in this pedigree), 3 further microsatellite loci mapped to this interval.

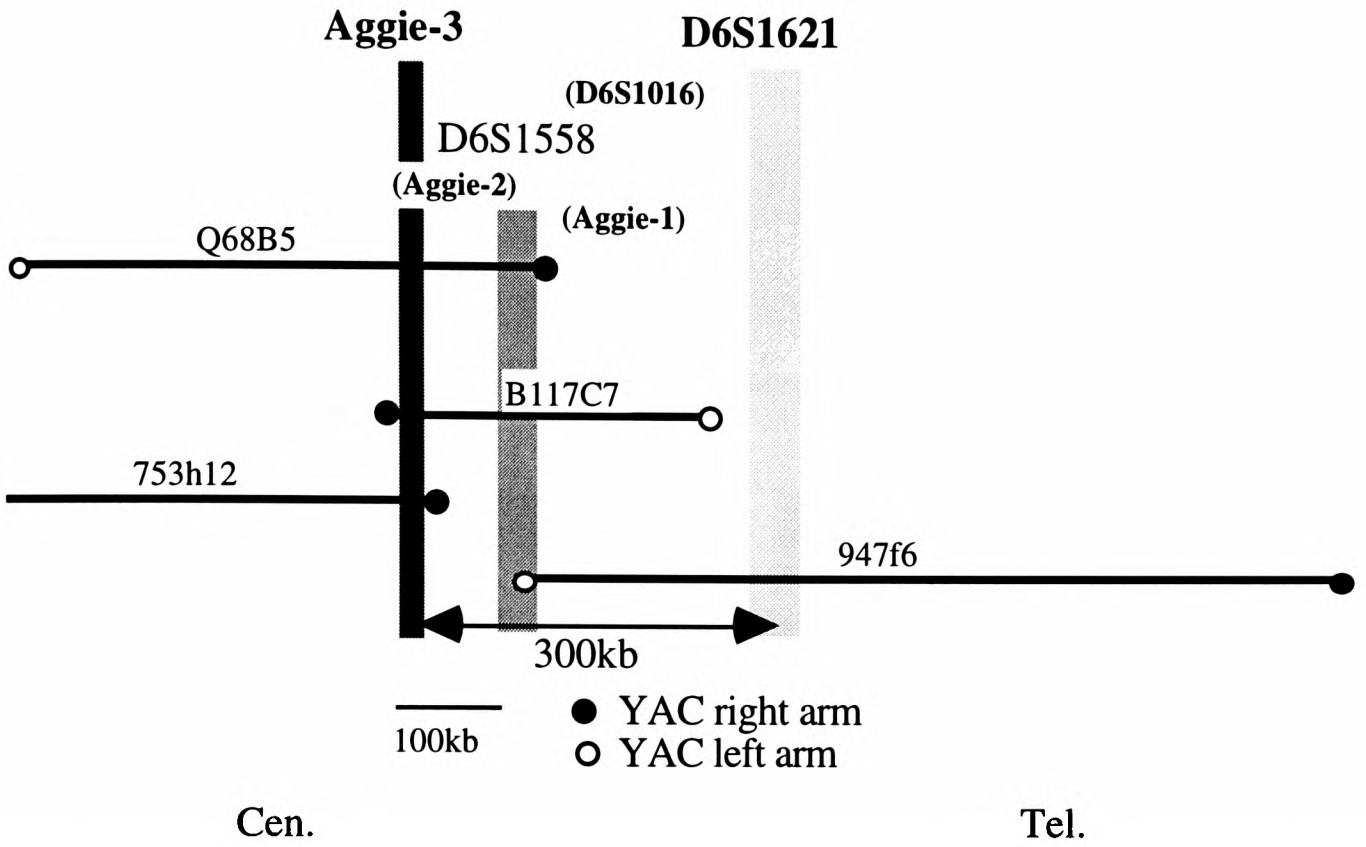


Figure 2.51 Microsatellites mapping into the region of the proposed recombination

Aggie-1 and Aggie-2

Aggie-2 is flanked by Alu repetitive elements precluding its use. Aggie-1 mapped between D6S1558 and D6S1621. As discussed in Chapter 2.4 (page 91) this marker was suspected to be flanked by repetitive sequence. Attempts to type members of this potentially informative pedigree using a recently marketed form of heat activated DNA polymerase (Amplitaq Gold, Perkin Elmer) failed to generate interpretable data.

D6S1016

Accepting the difficulty to interpreting the data (see Chapter 2.3, page 76), individuals from the pedigree were typed for D6S1016.

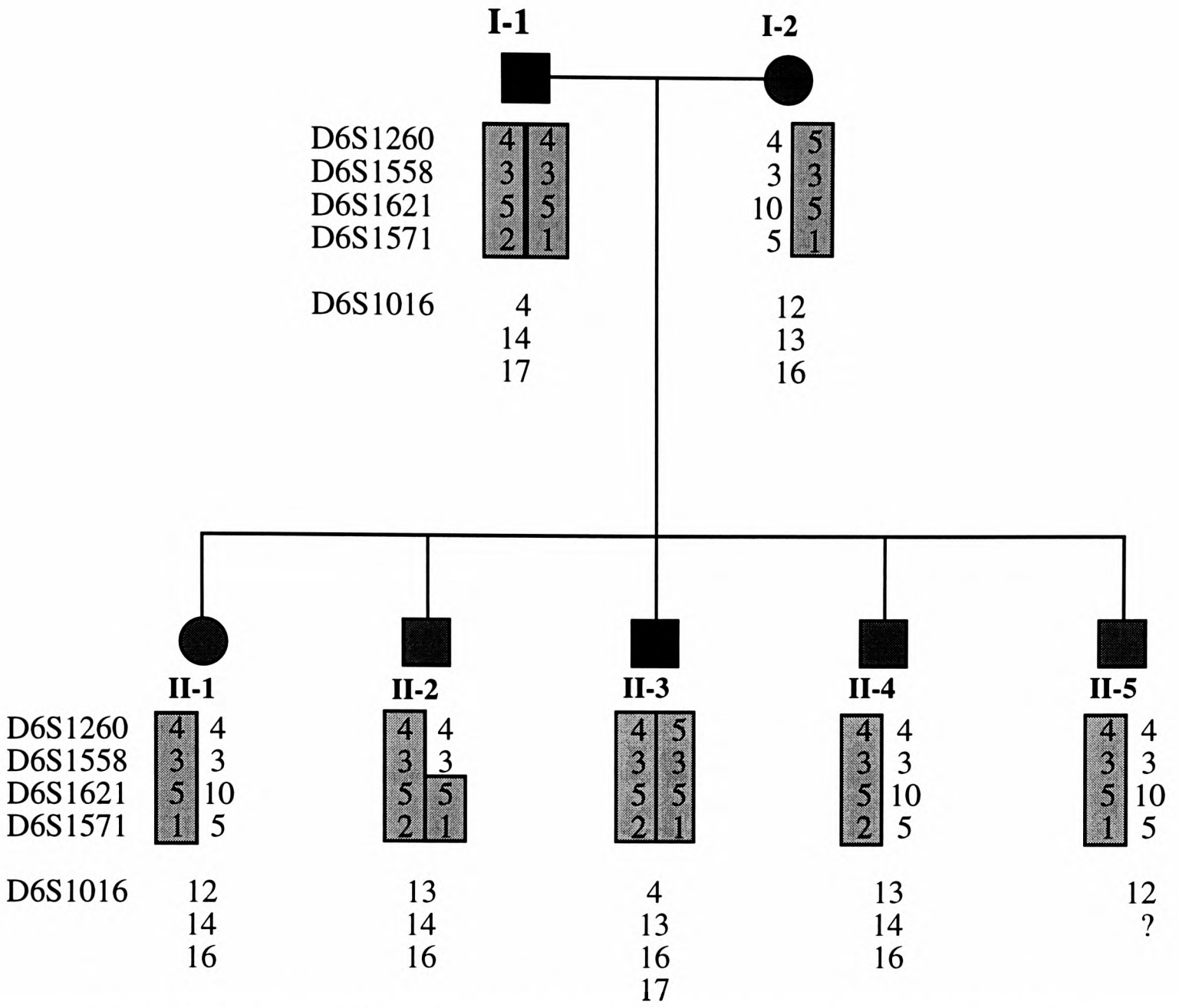


Figure 2.52 Alleles of D6S1016 in family with informative recombination. Allele 1 is 231bp. The other alleles are number increasing size by 2bp each. See Figure 2.31, page 77.

Allele 16 does not segregate with the other chromosome 6 markers as it is shared by I-2, II-1 and II-3. The same argument can be made for allele 13 being shared by I-2, II-2 and II-3. If allele 12 is present on the unaffected chromosome of I-2 then, as this is lost in the individual bearing the recombinant chromosome, this would suggest that the disease locus lies centromeric to D6S1016.

As discussed in Chapter 2.3 this microsatellite could not be placed on our physical map, although it was hypothesised that it might lie in the region duplicated on chromosome 5.

The proposed recombination breakpoint

The haplotype and physical mapping data suggest that D6S1621 marks the telomeric boundary for the haemochromatosis gene region. From the evidence of the YAC contig this defines a boundary that lies within 1Mb of D6S105.

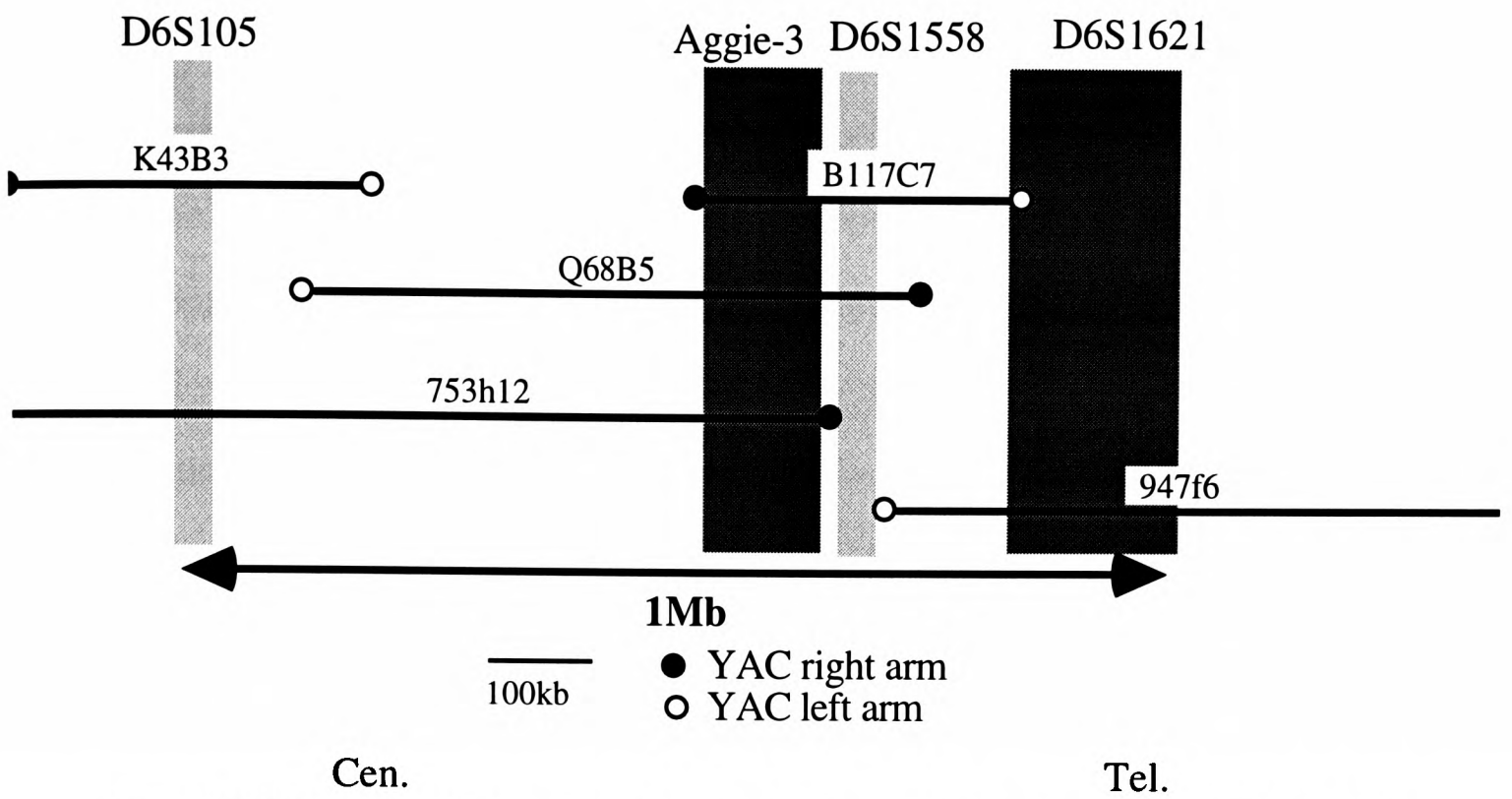


Figure 2.53 The telomeric boundary for the haemochromatosis gene region as defined by this informative recombination lies at D6S1621, within 1Mb of D6S105.

Finer definition of this boundary was precluded by homozygosity at D6S1558 and by failure to obtain interpretable data form the other repeat loci (Aggie-1, Aggie-2 and D6S1016).

Interpretation

The relative value of data from individual recombinations

The first report of a recombination defining the telomeric limit of the haemochromatosis gene region was based on data from HLA serotypes implying a location for the gene between HLA-A and HLA-B (i.e. within the Class I region of the MHC - Edwards *et al.* 1986). Re-examination of this pedigree by restriction fragment length polymorphism (RFLP) with molecular probes from the class I region failed to support the suggestion of a crossover (Radisky *et al.* 1994).

Using microsatellite markers that mapped telomeric to the MHC, Gasparini reported the identification of a complex pedigree suggesting HLA-F as the telomeric limit to the disease region (Gasparini *et al.* 1993). The pedigree included reconstructed parental chromosomes, a "non-expressing homozygote" and defined the breakpoints of the proposed double recombination on the basis of alleles at the two serologically determined loci (HLA-B and HLA-A). It was later recognised that the

typing data was erroneous and the claim was subsequently retracted at the First International Workshop on the Molecular Genetics of Haemochromatosis (P. Gasparini, oral presentation, Gargnano, Italy, September 1994).

In the family reported here the recombination is supported by data from at least four separate polymorphic microsatellites either side of the proposed breakpoint. There is little doubt that the individual concerned bears a recombinant chromosome. The reservation to drawing conclusions regarding the physical location of the haemochromatosis gene is from interpretation of the phenotypic expression of the condition. Haemochromatosis has a mild phenotype. The heterozygote state cannot be reliably defined by any investigation (biochemical or other) and exclusion of the condition can only be based on repeated measurements of iron status over many years. In the pedigree presented the individual carrying the recombinant chromosome is older than his affected sib, who has severe iron loading as reflected by the HII of over 10. He has now been followed for over five years with no increase in either of his serum iron indices.

For the haemochromatosis gene to lie distal (telomeric) to the breakpoint, the individual bearing the recombinant chromosome would have to express haemochromatosis at a later date. If that were the case, then his mother and his other 3 sibs would also be obligate homozygotes and would be expected to demonstrate some degree of iron loading. None of these indivdiuals presently shows evidence of iron accumulation after 5 years of observation. The data supports a position for the haemochromatosis gene centromeric to D6S1621.

Chapter 2.7

Overview:

Is the Map Sufficient to Support a Rational Search for Gene Sequences?

Overview

During the course of the mapping the following points had been established:

1. A position for the haemochromatosis gene physically close to HLA-A was refuted.

The long held conviction that haemochromatosis must lie physically close to HLA-A was abandoned by all groups worldwide. This derived from the extension of the physical map/YAC contig telomeric to the MHC with the identification of informative microsatellites and the subsequent recognition of linkage disequilibrium extending over a large physical region.

2. A position for the gene centromeric of D6S105 was unlikely.

That the haemochromatosis gene might lie centromeric to D6S105 had been borne out of a naïve interpretation of the linkage analysis data and the lack of information from markers mapping more telomeric.

3. On the basis of recombinations, conservative boundaries for the gene region lie at RFP (cen) and D6S1621 (tel).

Despite the chequered history of recombinants in this disease, two important recombinations were reported during this time. Both of which appeared to be well characterised clinically and each was supported by data from sufficient polymorphic marker loci either side of the proposed recombinational breakpoint. The first of these separated the gene from the MHC placing a centromeric limit for the disease region at the RFP locus, 1Mb telomeric to HLA-F (Calandro *et al.* 1995; Malfroy *et al.* 1996). The second established the first telomeric limit for the gene region at the microsatellite locus D6S1621, 1Mb telomeric to D6S105.

4. A plateau of linkage disequilibrium as defined in UK patients extended from D6S1260 to D6S1558, a physical distance of approximately 200kb.

A "critical region" of linkage disequilibrium was defined on the basis of the calculation of λ for polymorphic microsatellite markers extending over the entire region covering more than 5Mb from HLA-A.

5. Greater refinement of this "critical region" was unlikely

Identification of genes within even limited regions of genomic DNA can be an arduous task consuming resources, manpower and time. The mapping data had to be as good as possible before the transition to this important phase of positional cloning.

More YACs?

Sequential chromosome walking telomeric to D6S105 had been limited by the number of clones demonstrating cross-hybridisation by FISH. The alternative approach of STS content analysis had identified continuity between Q68B5 (ICRF library), B117C7 (StLouis) and 947f6 (CEPH). These three clones contained all the markers showing peak allele association with haemochromatosis (D6S1260 to D6S1281). The only STS reported from the region that was not present on one of these three YACs was D6S1016, which was demonstrated to amplify alleles from sites other than 6p21.3.

More microsatellites?

From D6S105 to D6S1281 we had mapped 15 microsatellite loci. The one microsatellite that had been reported to map to this region that we were not able to place on our contig was D6S1016.

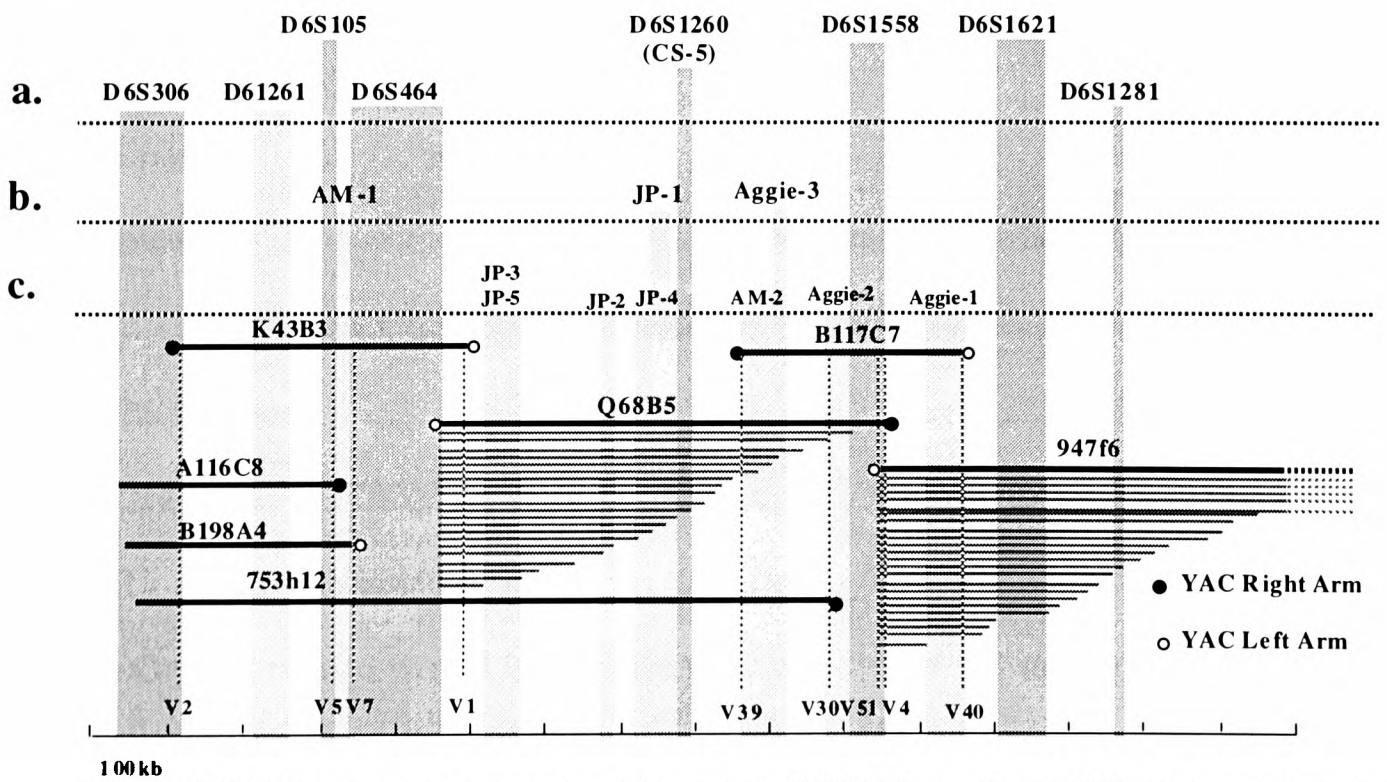


Figure 2.54 CA repeat loci mapping telomeric to D6S105. a. microsatellites reported on the public access databases, b. informative microsatellites identified during the project (JP = Jenny Pointon; CS = Caroline Stone; AM = Alison Merryweather-Clarke), c. other CA loci that were ultimately uninformative.

More patients?

Association analyses require analysis of genetic markers in large groups of affected individuals compared to control groups without the condition. Although haemochromatosis has been shown to be mediated by a single gene, many of the features of phenotypic classification reflect similar difficulties to those encountered in mapping polygenic conditions. Haemochromatosis has a mild phenotype. The heterozygote state cannot be identified and homozygotes present late in life and may remain undiagnosed.

Controls

The high prevalence of the haemochromatosis gene in North-West Europeans and the inability to identify carriers by any biochemical test means that any control group will be "contaminated" by heterozygotes and undiagnosed homozygotes (i.e. those who will develop iron loading later in life). This is theoretically a greater problem in using blood donors as the control group. These individuals will tend to be young and, by having blood taken on a regular basis, undiagnosed haemochromatosis might be further obscured. Contamination of any control group will lessen the degree of association between a marker and the disease phenotype by presenting a false prevalence of the disease associated allele in the control group. At the very least this will reduce the statistical significance of any observations.

Patients

Screening of asymptomatic individuals with markers of iron status has consistently established a disease prevalence of 0.3% to 0.45% in caucasian populations (Edwards *et al.* 1988; Leggett *et al.* 1990; Worwood and Darke 1993). The John Radcliffe Hospital serves a population of 1 million. Taking a conservative estimate for the homozygote frequency in the UK as 0.003 (Worwood and Darke 1993) the Oxford region would have somewhere near to 4000 individuals with haemochromatosis. According to hospital records less than 1% of that number currently attends either the gastroenterology or haematology departments for supervised venesection treatment. Taken together with the range of liver iron

concentrations in affected individuals (Bassett *et al.* 1986; Cartwright *et al.* 1979), this suggests that the common haemochromatosis phenotype is very mild and may pass clinically undetected. Strict diagnostic criteria based on heavy iron stores at the time of diagnosis (i.e. HII greater than 2, more than 5g mobilizable iron by quantitative phlebotomy) is essential for studies of this nature. Patients treated by phlebotomy have no persistent abnormality (either in terms of clinical signs or biochemical/haematological tests) to allow a retrospective confirmation of the diagnosis once iron stores have been reduced.

Of the two variables considered (patients and controls) it is correct to be stringent over the definition of the patient group, accepting the limitations to which one can realistically influence the purity of the control group.

6. The concerted search for gene sequences was to begin in the D6S1260 to D6S1558 interval

The extent of linkage disequilibrium demonstrated at the microsatellite markers D6S1260 and D6S1558 was very high, predicting a common mutation present on over 70% of haemochromatosis chromosomes. The physical distance spanned by these two markers is just over 200kb. This interval was chosen for the start of efforts to identify candidate genes.

Chapter 3

Increasing the resolution of the physical map and the recognition of candidate gene sequences.

Contents - Chapter 3

Chapter 3.4

Fine mapping of the haemochromatosis gene region and preparation for genomic sequencing

Chapter 3.1

Background:

The positional candidate principle.

A needle in a haystack

Chapter 2 outlined the physical mapping of the haemochromatosis gene region. The most conservative boundaries (as defined by well-substantiated recombinations) lie at the Ret-Finger protein locus (centromeric - (Malfroy *et al.* 1996)) and D6S1621 (telomeric - unpublished). Data from YAC contigs across this region suggested that this covered an interval of approximately 2.5Mb. Within this we defined a critical region of high linkage disequilibrium between D6S1260 and D6S1558.

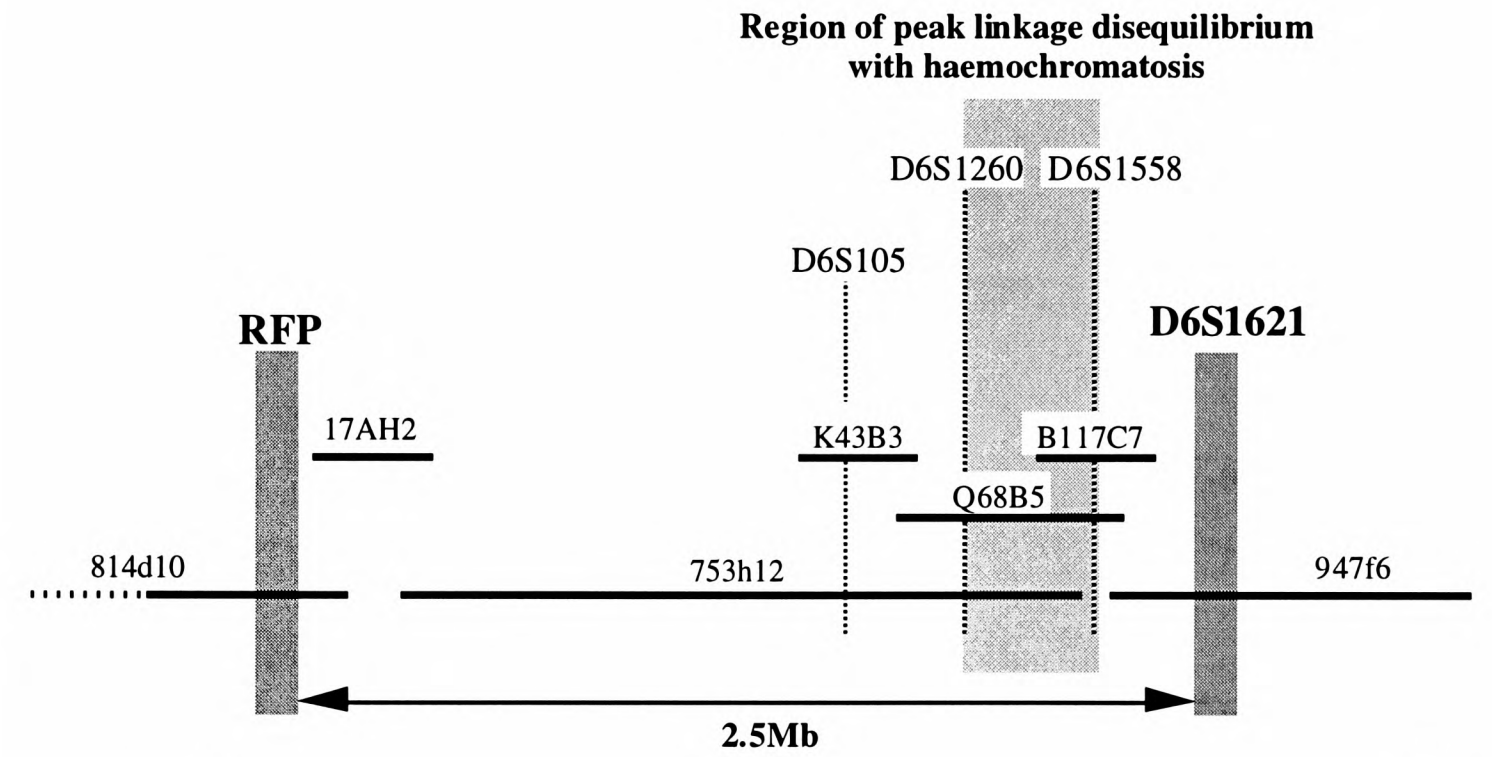


Figure 3.1 The haemochromatosis gene region

This region of 200kb of DNA was chosen for the initial search for gene sequences. Early observations from restriction mapping had suggested that this region was GC rich implying a high density of genes and constitutively expressed genes such as histones and transfer RNAs had already been identified around D6S105 (Stone 1995).

This chapter will address strategies for the identification of a disease specific candidate gene from within a gene dense region of genomic DNA.

Methods for the identification of genes within cloned DNA

cDNA Selection

The traditional approach to identifying human disease genes is the isolation of disease specific messenger RNA (mRNA) from affected tissues and synthesis of its complimentary DNA (cDNA). These stable sequence fragments can then used for screening genomic libraries and the identification of the defective gene.

Such an approach may be used in reverse. Cloned genomic DNA in the form of YACs or cosmids can be used to screen libraries of cDNA generated from affected tissues to identify expressed genes from a disease candidate region.

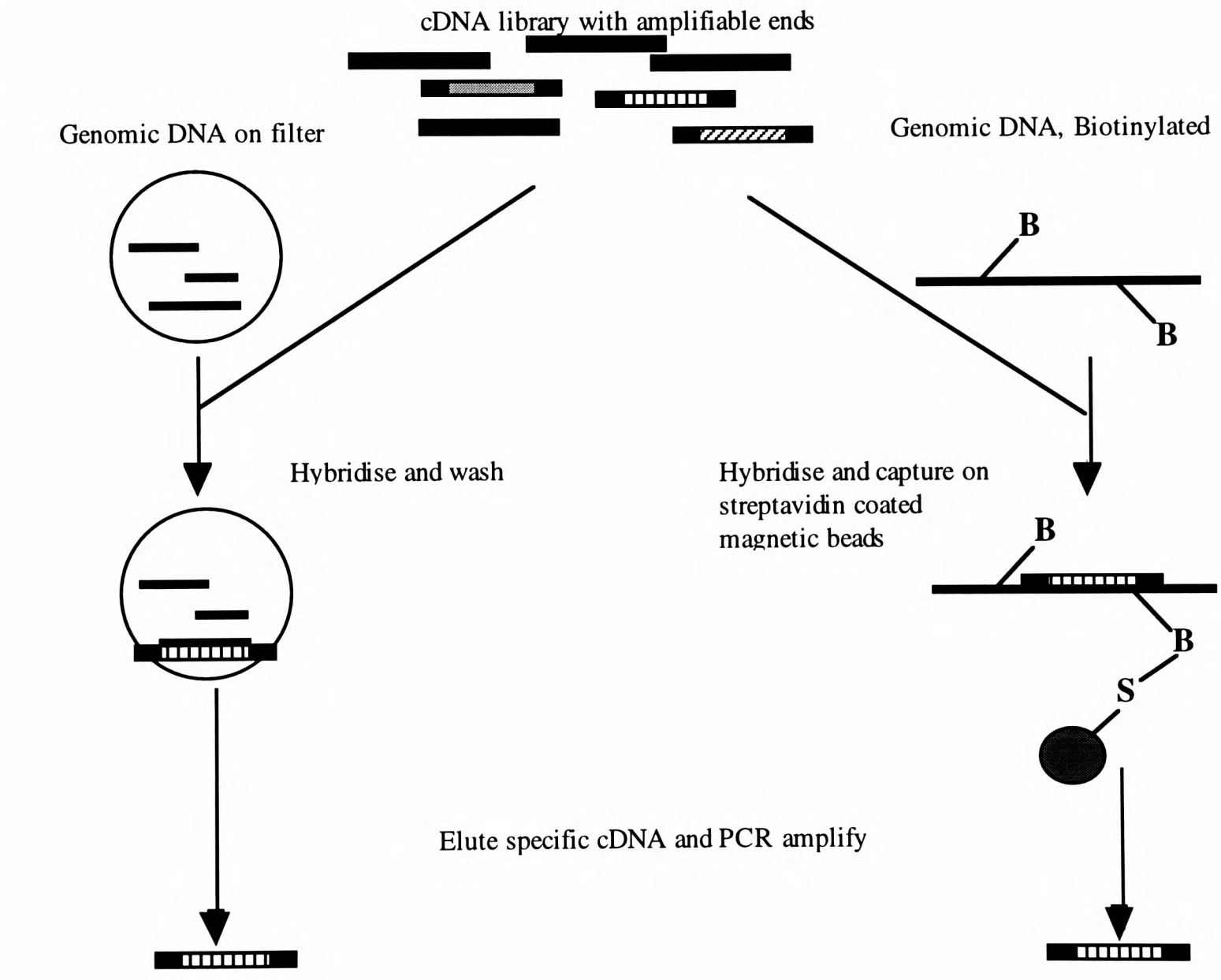


Figure 3.2 cDNA selection involves the identification of cDNAs with sequence complimentary to that in the target genomic clone (YAC or cosmid).

This approach is "expression dependant" in that the chance of identifying a rare, tissue specific, transcript is dependant on the tissue of origin of the cDNA

library. The best library to screen in order to increase the prospect of identifying a specific disease gene requires at least a prediction of the tissue expression of the gene product. The stringency of the selection process will ultimately determine success.

Searching for haemochromatosis candidate genes, Goldwurm used direct cDNA selection (Goldwurm *et al.* 1996). Using CEPH megaYAC clones containing D6S105 for the selection he identified 314 cDNA clones. Only 14 of these were subsequently confirmed to map to the region, illustrating the potential inefficiency of this approach. Of the 14 transcripts, half were constitutively expressed genes or repetitive elements. Seven novel expressed sequences were identified. The abstract reporting this work did not specify which megaYAC clones were used and the total physical distance covered by these experiments nor which cDNA libraries were screened. Most megaYAC clones contain inserts of over 1Mb and the identification of 7 expressed sequences might represent meagre return for the effort invested.

The same approach, using megaYAC clones containing D6S105, has been adopted by other groups (Jouanolle *et al.* 1996) although none has yet reported a single candidate gene.

Exon trapping

Where cDNA selection is influenced by tissue-specific patterns of gene expression, exon trapping is independant of this. Exon trapping (or exon amplification) identifies coding sequence from cloned genomic DNA by exploiting conserved DNA sequence motifs necessary for exon splicing. Cloned genomic DNA (YAC or bacterial clone) is sub-cloned into a permissive exon trapping vector that contains splice donor and acceptor sequences. Transient transfection into a mammalian cell line results in a splicing event when the inserted DNA itself contains splice sites. The "spliced" or "trapped" exon is identified by a novel sized product generated by reverse transcription PCR. Such trapped exon cDNA can be cloned, sequenced and used for the identification of full length cDNA clones by conventional library screening.

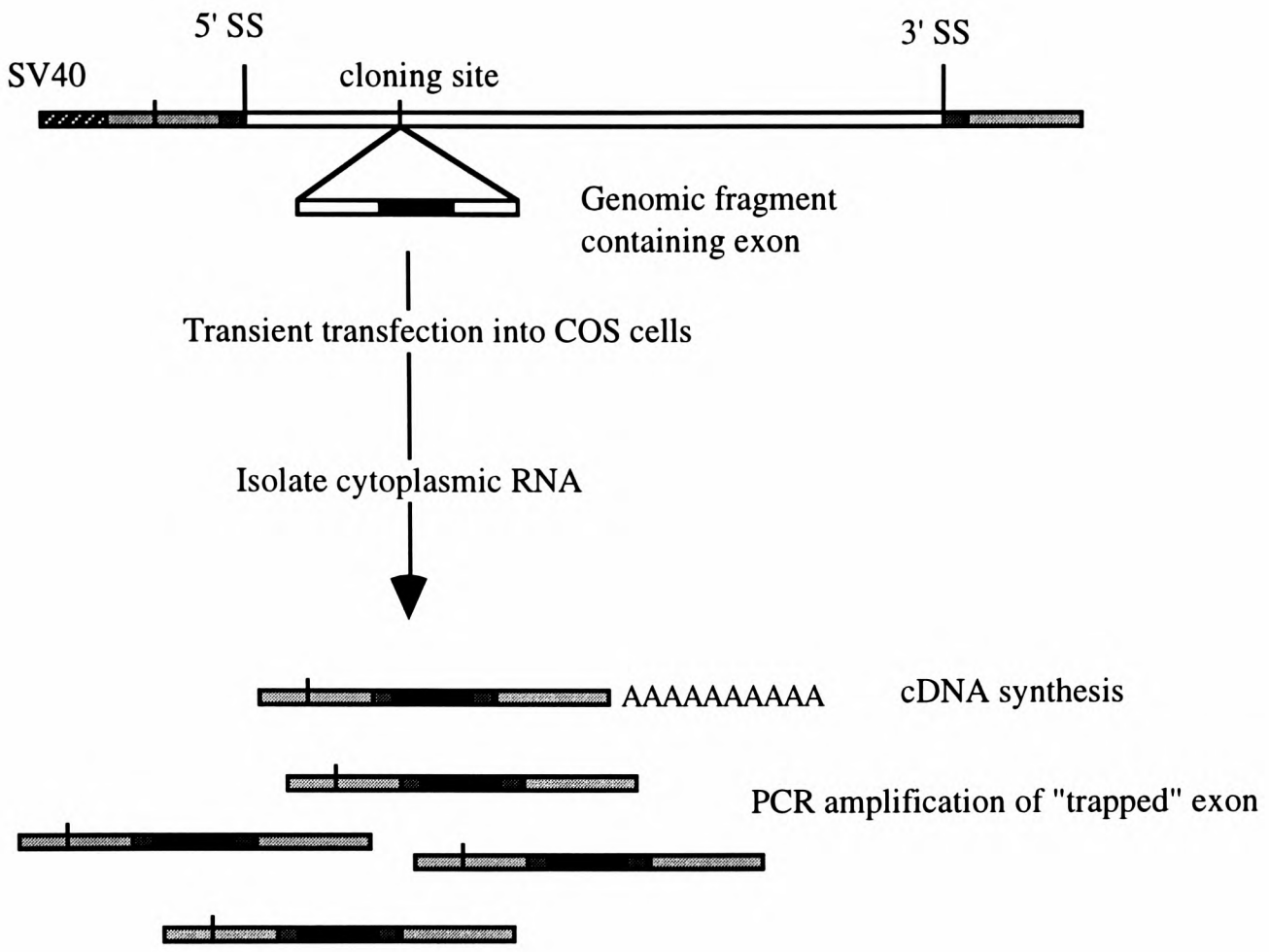


Figure 3.3 Exon trapping (adapted from (Buckler *et al.* 1991))

Some groups have reported the use of exon trapping around D6S105 (Jazwinska *et al.* 1996), but to date no haemochromatosis candidate gene has been identified by this strategy.

CpG island cloning

CpG islands are short stretches of genomic DNA rich in the CG dinucleotide that remain hypomethylated in mammalian cells. It has been hypothesised that they might convey important secondary chromatin changes that facilitate gene transcription (Tazi and Bird 1990). Such islands have been found associated with 99% of contitutively expressed or "house-keeping" genes and 40% of tissue specific genes (Larsen *et al.* 1992). On this basis CpG islands have been proposed as useful markers of genes within cloned DNA.

Caroline Stone used a CpG island cloning strategy in the search for haemochromatosis candidate sequences around D6S105 (Stone 1995). YAC DNA was digested with the rare cutting restriction endonucleases associated with CpG

islands, cloned into plasmid vector and sequenced to identify the associated gene. Using this approach only histone sequences were identified in that region.

A more recent strategy to exploit the association between CpG islands and genes has been the generation of a specific CpG island library (Cross *et al.* 1994). This was achieved by the selection of hypomethylated genomic fragments on a methylated DNA binding column and the subsequent cloning of hypomethylated fragments into a plasmid vector (pGEM-5Zf(-)). This is intended to provide an efficient resource for the identification of human genes and it has been suggested that sequencing of all the estimated 45,000 CpG islands in human genomic DNA was a realistic goal. A proportion of tissue specific genes are not associated with CpG islands which must be bourne in mind when applying this approach in the search for a specific disease gene.

Genomic sequencing

An alternative to the methods described above is the direct sequencing of a disease candidate region. A decade ago this may have seemed an impractical approach but three recent developments have combined to make this increasingly attractive (Chen *et al.* 1994). First, the resolution of physical maps of disease gene candidate regions has been greatly improved by the combined technologies of PCR, microsatellite analysis and cloning of human genomic in DNA in a range of size specific vectors (YACs, P1 clones, BACs, Cosmids, λ phage and M13 phage). Secondly, facilities for automated sequencing have become commonplace (Hunkapiller *et al.* 1991). The value of this approach has been proven in the sequencing of the entire genomes of *Haemophilus influenzae* (Fleischmann *et al.* 1995) and more significantly the eukaryote yeast *Saccharomyces cerevisiae* (see Walsh and Barrell 1996 and Dujon 1996). Finally, sequence data analysis has become considerably more successful due to a combination of more sophisticated raw sequence analysis programs (BLASTX, GRAIL, Altschul *et al.* 1994) and the exponential growth of the sequence databases against which sequences might be compared.

Sequencing the whole of the human genome is not only considered technically feasible but considered by some to now be the primary goal of the Human Genome Project (Olson 1995).

The positional candidate approach

One of the fastest growing areas of sequence data is that of the "expressed sequence tag" or EST. Parallel to the concept of sequence tagged sites, ESTs are short fragments of sequence data, identified by PCR, that are obtained from the systematic sequencing of cDNA clones. ESTs therefore represent sequence tagged landmarks for transcribed genes. In the simplest application of such data, sequence identity between a random fragment of genomic sequence and an EST data suggests the identification of a gene. Mapping of EST loci (by PCR on either YACs or radiation hybrid panels) is expected to eventually generate a "transcript map" of the human genome (c.f. the STS map of the Whitehead Institute, Boguski and Schuler 1995).

Recognising the power of EST information some researchers have adopted the term "positional candidate" approach, to reflect the intersection of positional information with the increasing bank of EST data (Ballabio 1993; Collins 1995). As the mapping of ESTs increases it will become possible, after defining a disease candidate region, to then identify the ESTs (and hence expressed genes) from that region, selecting those which provide most plausible models for the condition and limit the sequence mutation analysis necessary for confirmation of causality. As of June 1996, 82% of human disease genes identified by positional cloning were represented by exact matches with one or more ESTs in dbEST (NCBI data). Ultimately this approach might surpass all other forms of disease gene identification.

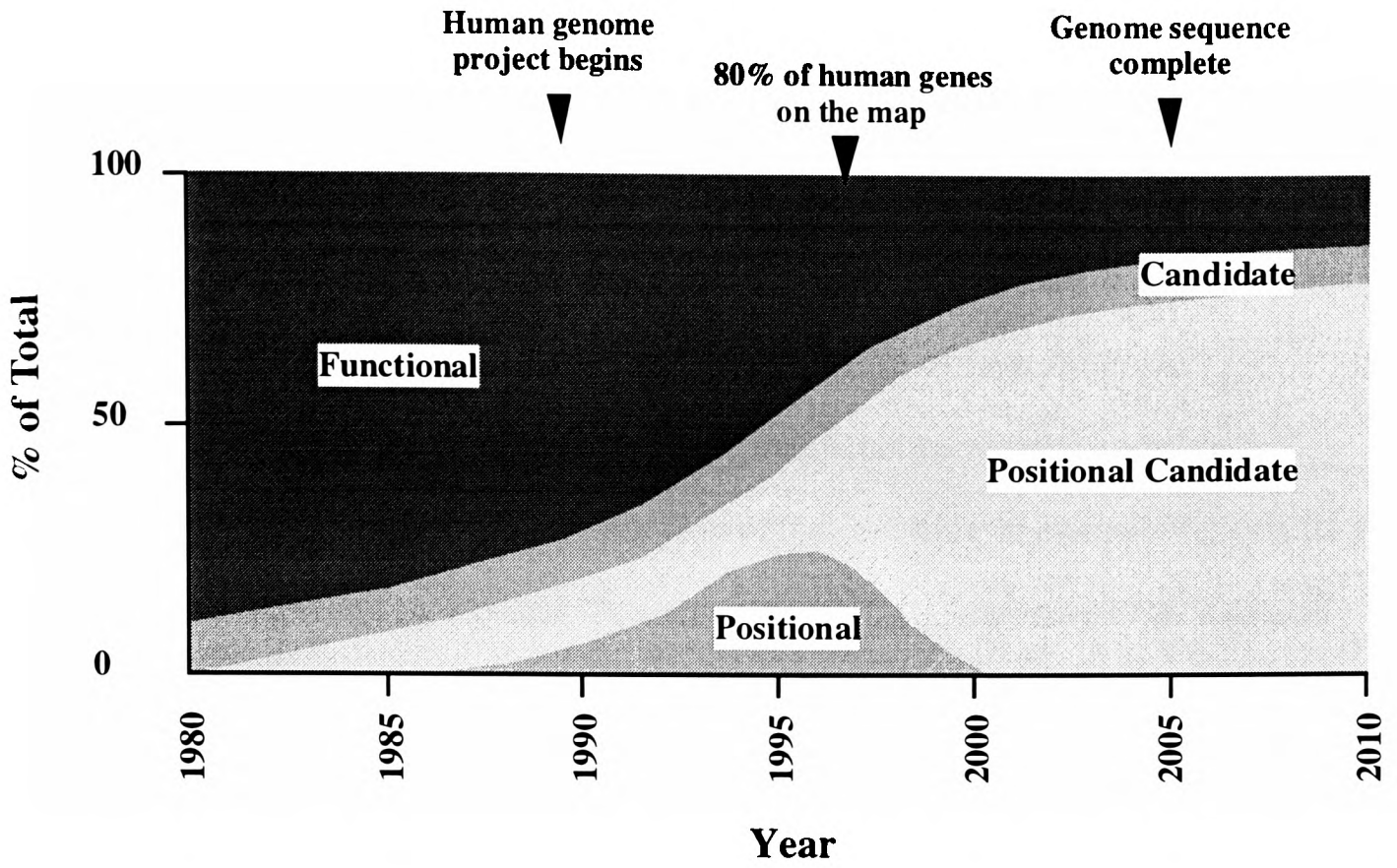


Figure 3.4 Trends in methods for cloning human disease genes (Collins 1995)

For this to be realised would require near completion of genetic, physical (clone and STS) and transcript maps of the genome. Before such a time the improvement in the quality and size of sequence databases has important bearing on any approach to the identification of new genes. All the cloning strategies outlined above (cDNA selection, exon trapping, CpG island cloning, genomic sequencing) have sequence data as their end product. Interpretation of this data or the rational identification of "positional candidates" requires some insight into the phenotype of the disease.

"The needle" - Can predictions be made about the nature of the haemochromatosis gene product?

The work outlined in chapter 2 of this thesis defined the likely genomic location of the haemochromatosis gene ("the haystack"). To critically evaluate any gene (novel sequences or members of multigene families) from this region as reasonable haemochromatosis candidates it is important to closely examine the disease phenotype in the context of the current understanding of the molecular regulation of mammalian iron metabolism.

Haemochromatosis homozygotes absorb excess dietary iron

There is no universally accepted, physiologically relevant, route for iron excretion in eukaryotes. Even mild excess in dietary iron absorption over that needed to make good any losses will inevitably lead to iron overload. This implies that the mechanisms that normally exert tight regulatory control of iron stores act predominantly at the point iron uptake. These mechanisms remain undefined.

The essential haemochromatosis phenotype resulting in excess absorption of dietary iron was defined by Powell (Powell *et al.* 1970). Measuring both mucosal uptake and retention of radio-labelled iron from a test meal, Powell demonstrated that iron-loaded haemochromatosis patients showed uptake of iron into the gut mucosa similar to normal control subjects but that a greater proportion of this was retained at day 14. After completion of venesection treatment, both mucosal uptake and retention were greatly increased in the haemochromatosis patients with values similar to those obtained from patients with iron-deficiency anaemia. These figures were used to calculate a mucosal transport index for iron, reflecting the proportional fraction of the absorbed iron that was ultimately transferred to the plasma.

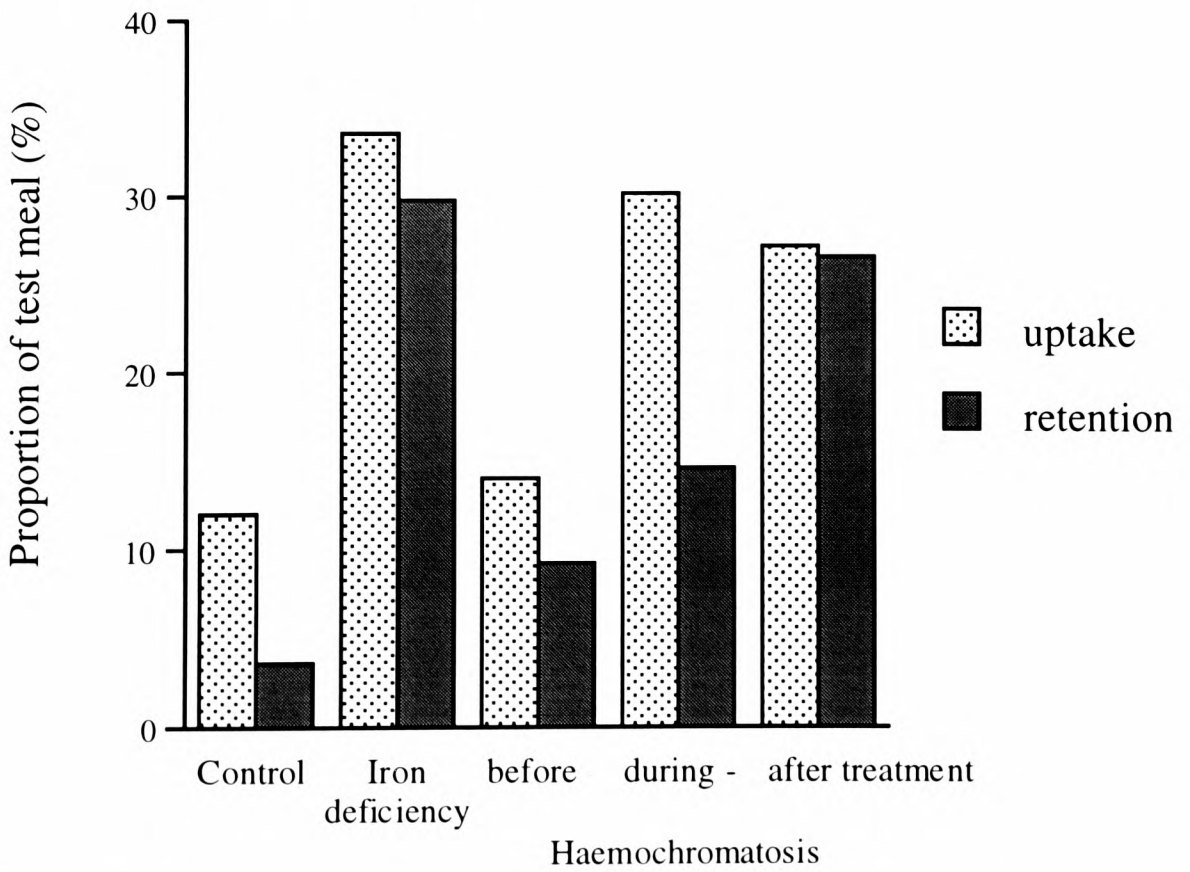


Figure 3.5 Mucosal uptake and retention of dietary iron (data from Powell *et al* 1970)

Walters also demonstrated that iron absorption was inversely proportional to iron stores as measured by serum ferritin and that in haemochromatosis this relationship persisted although for any level of serum ferritin iron absorption was higher than that seen in control subjects (Walters *et al.* 1975).

These findings suggest two important principles. First that the control of iron absorption might act at the level of **release** of iron from the mucosal cell into the circulation, explaining the observed differences in mucosal transport index. Secondly, some degree of control persists in haemochromatosis given that iron retention increases after venesection treatment.

Using both oral and intravenous tracer doses of radio-iron, McLaren studied mucosal iron kinetics in relation to serum ferritin measurements in haemochromatosis (McLaren *et al.* 1991). His results demonstrated an increase in the rate constant for transfer of mucosal iron to the plasma over that seen in normal control subjects. No difference was demonstrated in either the uptake rate constant or the rate constant for incorporation of iron into the mucosal storage pool.

The coordinate regulation of "Iron Proteins"

Elemental iron, although vital for many biological processes in mammals, is extremely toxic through its propensity to generate oxygen free radicals. The transport, uptake and storage of iron by mammalian cells is conducted in a very closely controlled manner mediated by iron-specific proteins. Iron transport is chaperoned by transferrin, a glycosylated protein of 79.5kDa, synthesised and secreted by the liver. Each transferrin molecule binds up to 2 atoms of ferric (Fe^{3+}) iron. Transferrin delivers iron to cells via receptor mediated endocytosis after interaction with a specific cell surface transferrin receptor. Ferritin is the specific iron storage protein. It is a large multimeric complex composed of L and H subunits, the relative proportions of which define ferritin isoforms characteristic of different tissues. Each ferritin molecule has the capacity to store up to 4500 atoms of iron. Cellular iron uptake and storage is regulated by a specific post-transcriptional mechanism mediated by a

bifunctional intracellular protein, the iron response protein or IRP. In the presence of low levels of intracellular iron this cytoplasmic aconitase enzyme initially loses iron and then an iron-sulphur cluster from its catalytic site. This infers a conformational change in the protein effecting a functional switch from aconitase activity to an RNA binding role. This RNA binding form of the protein recognises and binds to specific stem loop structures adopted by a specific sequences motif (the iron response element or IRE) in the non-coding regions of the messenger RNA of ferritin and transferrin receptor.

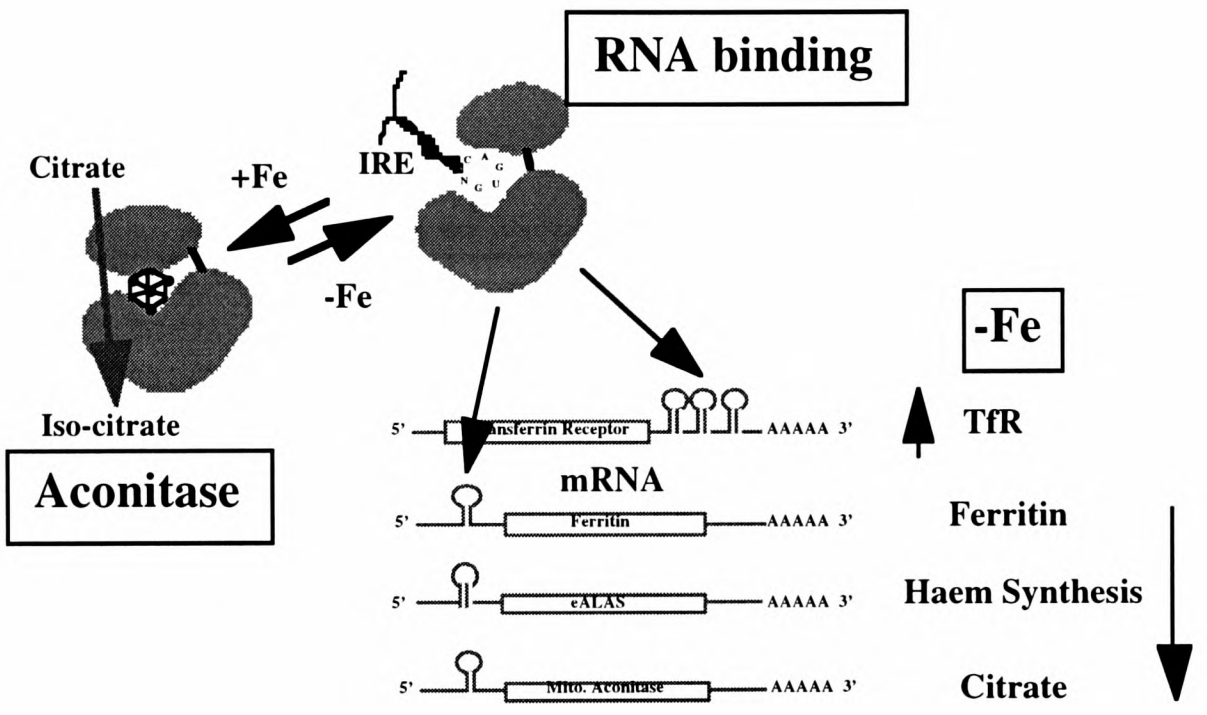


Figure 3.6 The iron response protein (IRP) regulates the coordinate expresion of proteins involved in both iron uptake (transferrin receptor) and storage (ferritin).

Binding of IRP to an IRE in the 5' non-coding region of the ferritin transcript prevents binding of the ribosomal complex to this message effectively reducing ferritin translation. Conversely binding of IRP to multiple copies of the IRE structure in the 3' non-coding region of the transferrin receptor message protects this transcript from degradation by intracellular ribonucleases resulting in enhanced transferrin receptor expression. These combined effects result in reduction in ferritin production and an enhanced expression of transferrin receptor in response to intracellular iron deficiency.

As none of these iron proteins maps to the haemochromatosis gene region on 6p, they have been excluded as harbouring the specific disease causing mutation. However, analysis of the regulation of these proteins in haemochromatosis has lead to important observations concerning the cellular phenotype of the condition.

The molecular regulation of iron absorption

The recognition of excessive iron absorption in haemochromatosis has reasonably lead to the belief that the primary cellular defect in this condition lies within the epithelial cells lining the duodenum and upper small intestine(Powell *et al.* 1970). When further considering the possible molecular mechanisms that might be involved in this process, the polarised nature of the enterocyte must be respected.

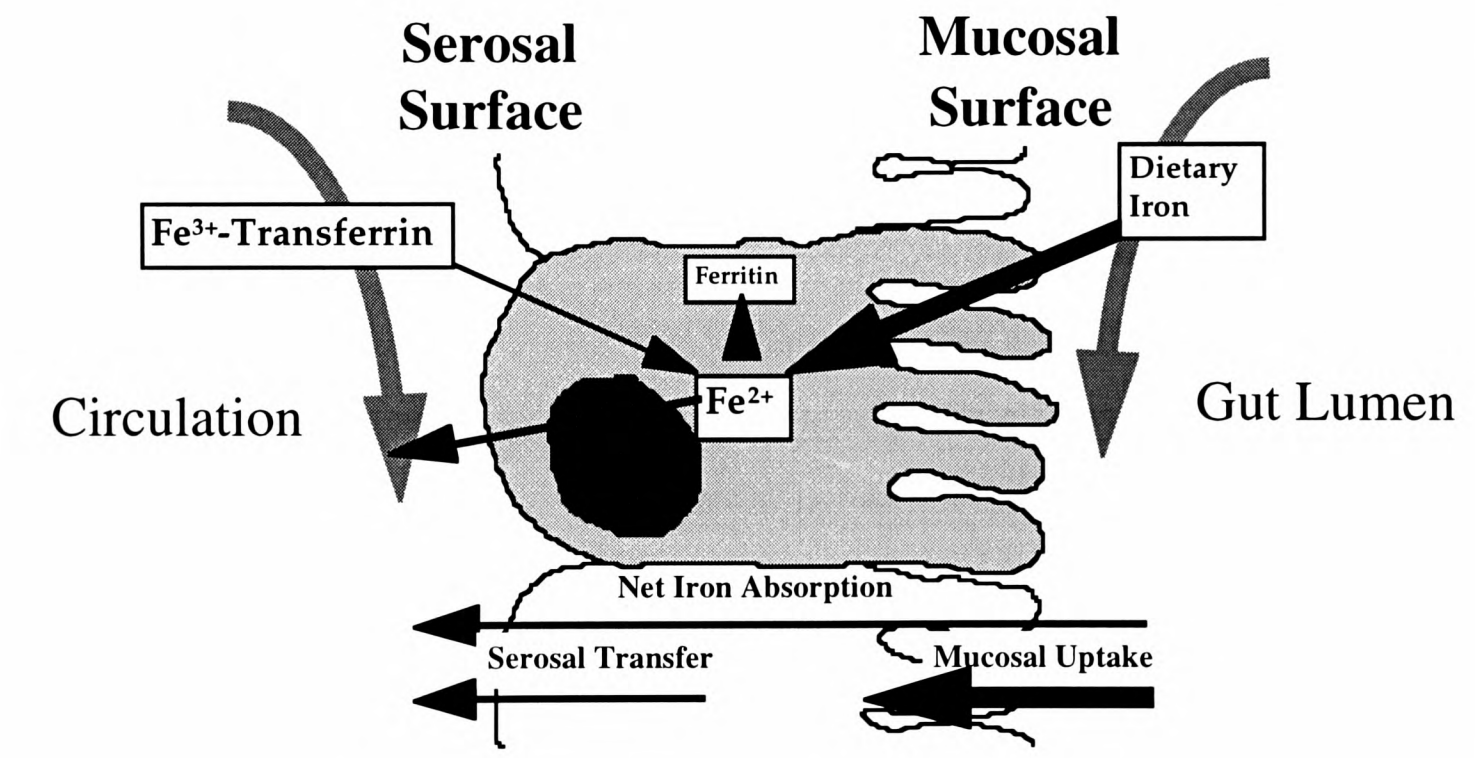


Figure 3.7 The intestinal epithelial cell (enterocyte) is exposed to different molecular forms of iron at its two surfaces.

Dietary iron takes two forms, haem iron and non-haem iron. Although haem iron (as in meat) constitutes the smaller fraction of dietary iron in omnivorous mammals it has been shown to be the more "bio-available" form being readily absorbed (for review of this subject see Skikne and Baynes 1994). Uptake of haem by the enterocyte is mediated by an intestinal haem receptor that has been characterised biochemically but not defined in terms of protein structure or expression. Non-haem

iron is present in both animal and vegetable elements of the diet and is present in many molecular forms, the absorption of which may be influenced significantly by other dietary factors (Skikne and Baynes 1994).

Despite the implications of the observations made by Powell and McLaren, many researchers have focused their attention on the luminal (or mucosal) aspect of the enterocyte in their attempts to define the molecular defect in haemochromatosis. Most of these have concentrated on the characterisation of factors determining the uptake of dietary non-haem iron. Conrad and colleagues have defined an iron uptake pathway involving intestinal mucins; cell surface integrins and a new intracellular iron transport protein, Mobilferrin (see Conrad and Umbreit 1993). Stremmel and co-workers have reported in abstract a new cell surface iron transport protein which appears to be up-regulated in haemochromatosis (Stremmel *et al.* 1991; Teichmann *et al.* 1991). The genes encoding these new iron metabolism intermediaries have yet to be cloned and whether these represent the primary cellular defect in haemochromatosis remains to be proven.

Other studies have concentrated on examining the coordinate regulation of ferritin and the transferrin receptor in the duodenal enterocyte. A lack of ferritin within intestinal epithelial cells of patients with haemochromatosis demonstrated by electron microscopy predated Powell's physiological observations (Crosby 1963). This characteristic feature of the condition has been supported by immunohistochemical staining with monoclonal antibodies to the ferritin subunits (Francanzani *et al.* 1989). Whittaker demonstrated that the mucosal ferritin concentration in haemochromatosis was inversely related to iron absorption and although he was unable to determine whether this was the cause or the result of the absorptive abnormality (Whittaker *et al.* 1989). Pietrangelo addressed this issue using semi-quantitative analysis of ferritin gene expression (Pietrangelo *et al.* 1992). In haemochromatosis he found a coordinated reduction in cellular ferritin and a parallel increase in transferrin receptor message .

mRNA	Control	Iron deficiency	Secondary iron overload	Haemochromatosis
Transferrin receptor	+	++	-	++
Ferritin	+	-	++	-

Table 3.1 Tranferrin receptor expression is upregulated in haemochromatosis (Pietrangelo *et al* 1992)

The difference between the expression pattern in haemochromatosis and that seen in secondary iron overload implies that the enterocyte is behaving as if it (and the organism as a whole) is iron-deficient. The other possible explanation is that in haemochromatosis there is a tissue specific aberration of ferritin synthesis. Subsequent measurements of the RNA binding activity of the IRP by gel band shift analyses suggested that in haemochromatosis a higher proportion of the IRP was in the RNA binding form reflecting intracellular iron deficiency (Pietrangelo *et al.* 1995). Another report using similar techniques appears to contradict these findings stating that RNA binding activity of IRP in haemochromatosis intestine was normal (Flanagan *et al.* 1995). Conceptually, this issue is vital to a clear understanding of haemochromatosis and it is worth considering the study design of the two reports in more detail. In particular, the selection of the study and control groups needs to be put in context. Both studies used gel retardation analysis with radiolabelled RNA probes transcribed from the same plasmid containing the iron-response element of human ferritin H chain (pSPT-fer Mullner *et al.* 1989). Both studied IRP activity in cell lysates prepared from homogenised intestinal biopsies and the methodologies are broadly comparable. The main difference between the two studies is the choice of control group. Pietrangelo studied 19 haemochromatosis patients in addition to 6 healthy controls, 8 patients with iron deficiency anaemia and a group of 10 patients with secondary iron overload (thalassaemia major, porphyria cutanea tarda and alcoholic liver disease). Interpreting the results of the band shift analysis he noted that patients with haemochromatosis showed higher activity of the IRP (i.e. RNA binding activity) compared to patients with secondary iron overload (Pietrangelo *et al.* 1995). Flanagan studied 14 iron loaded patients with haemochromatosis and found the

activity of IRP to be no different to that measured in 16 normal (i.e. not iron loaded) controls (Flanagan *et al.* 1995). To describe this activity as "normal" is not in itself incorrect and indeed the authors describe it "appropriate to the intracellular iron concentration". However in the context of iron overload this observation simply implies that the IRP response is intact but "inappropriately normal" in the face of excess iron stores. In this light both studies support the basic premise that the enterocyte is iron-deficient in haemochromatosis.

Accepting the enterocyte to be iron-deficient in haemochromatosis, counters the suggestion that the molecular abnormality might lie on the luminal side of the cell. Any abnormality resulting in increased iron transport across an iron deficient enterocyte is most easily explained by excess iron release from the serosal surface of the cell rather than inappropriate iron loading from the gut lumen.

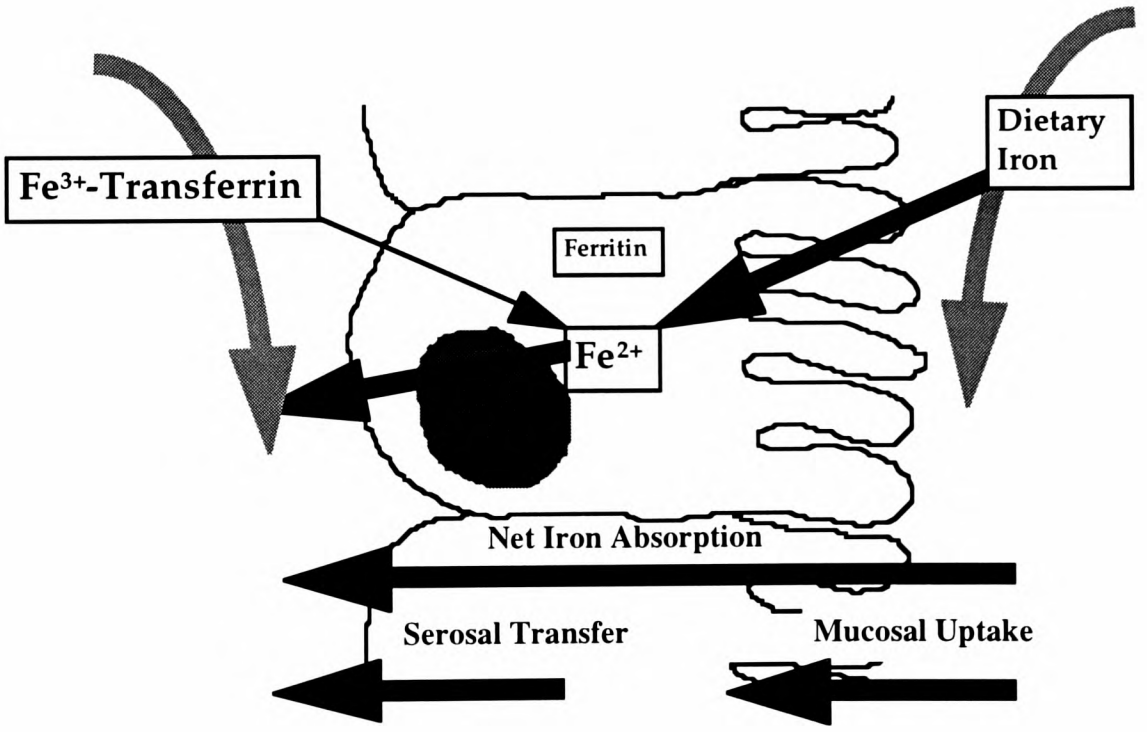


Figure 3.8 Increased net iron absorption in the presence of low intracellular iron suggests that the molecular defect results in excess iron release at the serosa.

This principle is consistent with the observations of Powell (Powell *et al.* 1970) and McLaren (McLaren *et al.* 1991). The molecular abnormality causing this is yet to be identifed.

Mechanisms of iron release

Iron release in eukaryotic cells is highly complex. As there is no universally accepted physiological excretion pathway for iron, most models consider that iron balance is controlled at the point of uptake (hence the relevance of the coordinate regulation of ferritin and transferrin receptor by the IRP). However in multicelluar organisms iron release by a limited number of cell types is central to the distribution of iron.

Once taken up by the enterocyte, iron (whether it be derived from haem or non-haem sources) initially occupies a hypothetical "labile" or "chelatable" pool within the cell cytoplasm. Whether this represents free ionic iron or iron bound to low molecular ligands such as citrate has not been defined. From this point iron is either incorporated into cellular ferritin stores or exported from the baso-lateral surface of the cell into the circulation. Factors determining the proportional division of iron between these two fates and the molecular intermediaries in iron export are not known.

Reticuloendothelial cells are relatively iron deficient in haemochromatosis

The enterocyte releases absorbed dietary iron into the portal circulation. Quantitatively, more iron is transferred through the reticuloendothelial cells of the marrow and spleen, releasing iron from effete erythrocytes and making this iron available to developing erythroblasts in the marrow (Noyes *et al.* 1960). There is an intriguing but unproven possibility of a defect in iron handling within cells of the reticuloendothelial system in haemochromatosis.

Subjectively, many have reported a relative iron sparing of reticuloendothelial cells in haemochromatosis compared to secondary iron overload (Deugnier *et al.* 1992; McLaren 1989; Valberg 1978). In fact this feature was reported by Sheldon who based some of his pathological hypotheses on reticuloendothelial cell iron metabolism (Sheldon 1935). Few studies have been able to demonstrate a quantitative difference in reticuloendothelial iron handling in haemochromatosis compared to other conditions of iron overload.

Bantu siderosis is a distinct form of iron overload, more accurately termed dietary iron overload of sub-saharan africa (Gordeuk *et al.* 1992). Until recently this condition was believed to be purely an environmental iron overload as many of the affected Bantu tribesmen imbibe an alcoholic brew made in iron drums. To evaluate the utility of marrow iron as a marker of body iron stores, Brink measured the iron concentration in the liver, spleen and marrow in a series of post-mortem cases of Bantu siderosis (Brink *et al.* 1976). The marrow iron concentration correlated well with hepatic iron concentration suggesting this to be a useful marker of total iron stores in these patients. The mean hepatic/marrow iron ratio was 1.9 in this group. Brink extended his observations by measuring the marrow iron concentration from posterior iliac crest trephine in a group of 8 caucasians with haemochromatosis who were having liver biopsies to confirm the diagnosis. In these individuals the hepatic/marrow iron ratio ranged from 7.3 to 102.4 indicating that the reticuloendothelial organ (marrow) was iron deficient.

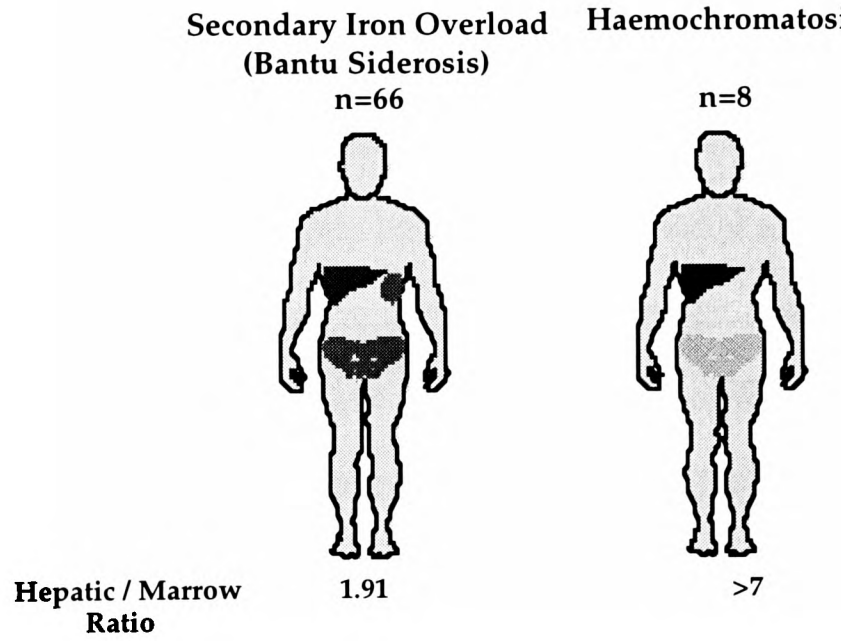


Figure 3.9 Hepatic/marrow ratio in haemochromatosis exceeds that seen in Bantu siderosis. (Brink *et al* 1979)

In a more dynamic experimental setting Fillet studied ferrokinetics (i.e. the distribution of iron) in patients with differing iron status including patients with haemochromatosis (Fillet *et al.* 1989). Measuring the distribution of radioactivity in patients after transfusion of heat-damaged erythrocytes, labelled with radioiron (^{59}Fe), Fillet defined a biphasic model of reticuloendothelial iron release.

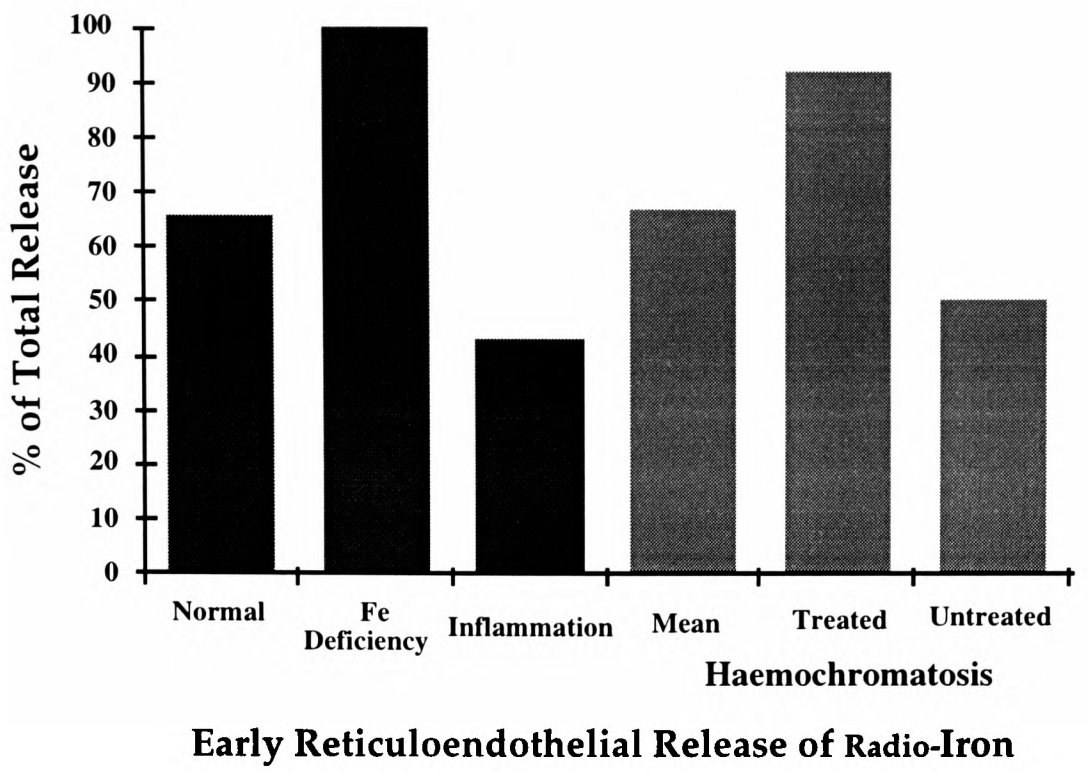


Figure 3.10 Cartoon of reticuloendothelial cell iron release profiles in iron deficiency and iron overload (Fillet *et al* 1989)

In normal volunteer subjects 64% of radioiron was released early ($t_{1/2}$ - 33min). The remaining iron was released in a late phase over several days. In iron deficiency the early phase release accounted for 100% of the label. In patients with chronic inflammation (and accompanying normochromic normocytic anaemia) early release was surpressed. In haemochromatosis the mean profile was "normal". However the haemochromatosis group contained both treated and untreated individuals and if the two groups are analysed separately the treated haemochromatosis patients clearly have an iron release profile akin to patients with iron deficiency with an almost complete release of iron in the early phase.

Of the studies specifically directed to defining a difference in reticuloendothelial cell iron handling in haemochromatosis most have focused on limited aspects of iron metabolism (transferrin receptor expression and ferritin synthesis) and confined their observations to the relatively undifferentiated peripheral blood monocytes (Adams *et al.* 1991; Bassett *et al.* 1982; Baynes *et al.* 1989; Bjorn-Rasmussen *et al.* 1985; Sizemore and Bassett 1984). Cairo has recently reported studies of IRP activity in reticuloendothelial cells from haemochromatosis patients (Cairo *et al.* 1996). In peripheral blood monocytes he demonstrated that IRP activity

was significantly increased compared to patients with secondary iron overload with similiar iron burden. Furthermore, in mature macrophages derived from peripheral blood monocytes, IRP remained up-regulated. This was correctable by increasing the ambient iron concentration, implying that the control mechanism was intact. Overall, the results support an inability of the macrophages to retain iron.

Animal models of increased iron absorption

Hypotransferrinaemic mice become iron loaded as a result of increased gastrointestinal iron absorption, although these animals do not show sparing of reticuloendothelial cells (Simpson *et al.* 1993). Hereditary hypotransferrinaemia has been reported in humans (Goya *et al.* 1972; Hayashi *et al.* 1993). Phenotypically hypotransferrinaemia differs markedly from haemochromatosis presenting with fatal iron loading in childhood. Furthermore, in hypotransferrinaemia excess iron aborption occurs in the face of low (or even absent) circulating transferrin. This suggests that transferrin saturation might contribute to the control of intestinal iron absorption, possibly providing the afferent arm of a feedback loop. In haemochromatosis iron absorption is increased despite normal levels of transferrin and a high transferrin saturation.

Mice deficient in β_2-microglobulin have also been shown to become iron loaded through increased gastrointestinal iron absorption (de Souza *et al.* 1994). Although no data has been published on transport iron in these animals, the organ distribution of iron in these animals is similar to that in haemochromatosis. Rejecting the possibility that MHC class I gene products might influence iron absorption, de Souza interpreted her findings as suggesting that specific T lymphocyte subsets, that are also subsequently deficient in these animals, might provide the mechanistic link.

Rothenberg predicted that the haemochromatosis gene product might be homologous to an MHC class I product on the basis of the iron loading of the β_2-microglobulin knockouts and the recognition of a globin-like promotor sequence (β–GAP) in association with non-classical class I genes in the mouse (Rothenberg and Voland 1996). However, the assumption that β-globin is under iron-regulated

transcriptional control is not proven and there is no evidence that the motif identified by Rothenberg mediates any form of an "iron response". Furthermore, the proposed hypothesis of an abherent class I like gene product being expressed on the luminal border of the enterocyte conferring increased iron uptake counters the observation of enterocyte iron deficiency in haemochromatosis (see above). The mechanism by which these mice become iron loaded has not been clearly defined.

Cellular iron uptake in *Saccharomyces cerevisiae*

Over recent years many of the molecular mechanisms of eukaryotic iron metabolism have been elucidated from genetic studies of the baker's yeast *Saccharomyces cerevisiae*. This has culminated in the characterisation of a membrane bound multi-molecular complex responsible for iron uptake (see Askwith *et al.* 1996).

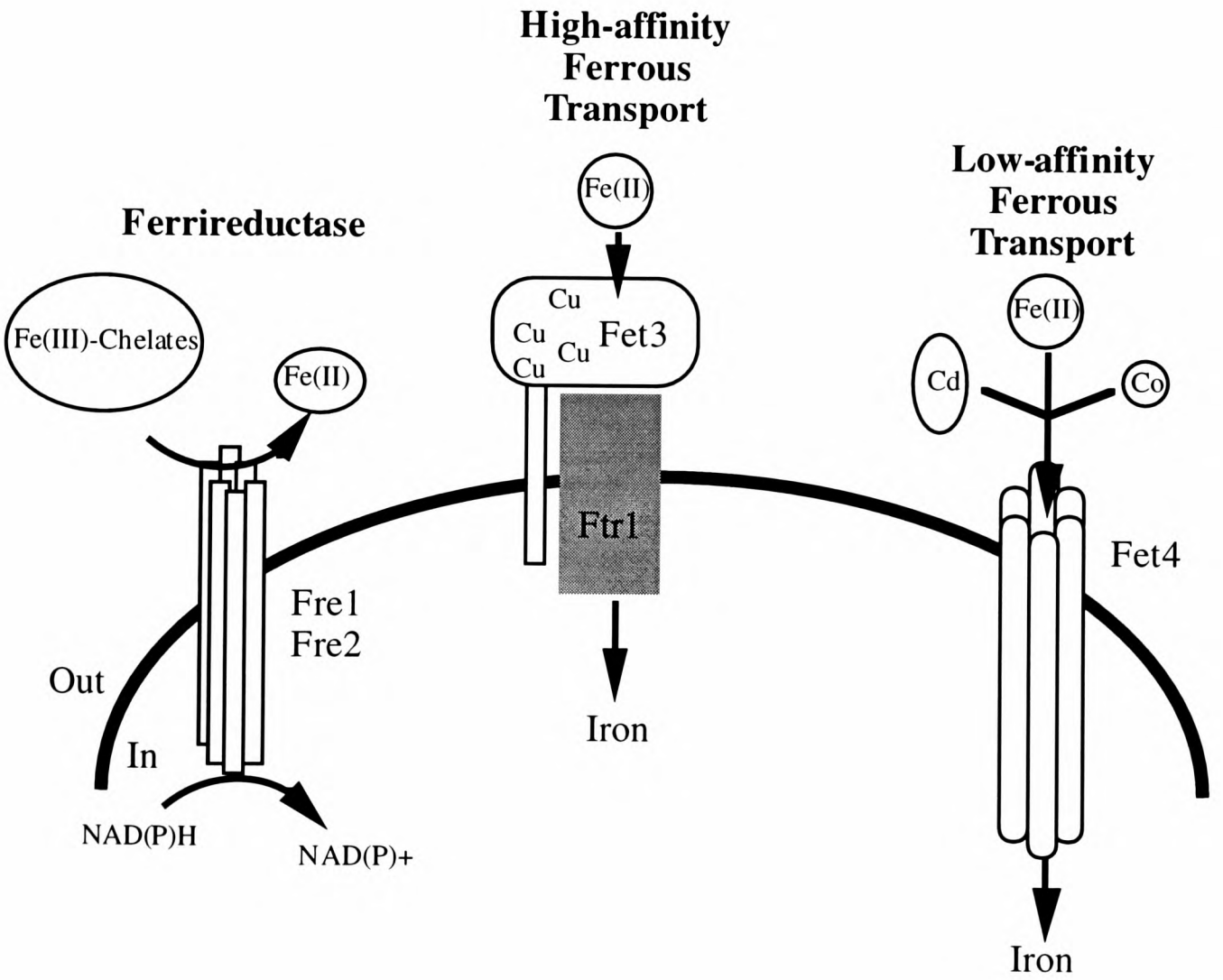


Figure 3.11 Pathways of iron acquisition in *Saccharomyces cerevisiae* (Askwith *et al* 1996)

The experimental route leading to the discovery of this complex has been a genetic one in which the genes have been defined initially, with the protein structure and function inferred by the translation of the sequence and subsequently confirmed

experimentally. Ferric iron reduction (Fe^{3+} to Fe^{2+}) is catalysed by one (or both) of two cell surface ferric reductases (fre1p and fre2p) and then transported through a transmembrane "permease" (ftr1p, Stearman *et al.* 1996). Thus transmembrane iron transport in eukaryotes depends on at least two critical steps - reduction and the presence of an iron specific pore.

Ferric reduction has been demonstrated to play a critical role in transferrin-independant mammalian iron uptake *in vitro* (Jordan and Kaplan 1994; Pountney *et al.* 1994; Raja *et al.* 1992) and an abnormality of ferric reduction has been demonstrated in mucosal fragments from patients with haemochromatosis (Raja *et al.* 1996). A role for ferric reduction in iron release in mammalian cells has not been investigated. The human homologues of either the yeast ferric reductases (fre1p, fre2p) or the iron pore (ftr1p) have not been identified. Either would be attractive functional candidates for haemochromatosis.

Iron absorption - the copper connection

One feature of mammalian iron metabolism that has already benefitted from genetic studies in yeast is the connection between iron and copper metabolism. It has been recognised that pigs fed on a copper deficient diet develop the microcytic hypochromic anaemia characteristic of iron deficiency. If these animals are treated with iron (either orally or by intramuscular injection) they fail to utilise this until circulating caeruloplasmin levels are restored by correction of the copper deficiency (Lee *et al.* 1968).

Certain *Saccharomyces* disruption mutants require supplementation of the media with copper in order to utilise iron. The gene responsible for this phenotype, *FET3*, was found to code for a copper dependant ferroxidase (fet3p Askwith *et al.* 1994). The functional homologue of fet3p in humans is caeruloplasmin. This serum protein in man had been noted to possess ferroxidase activity (Osaki *et al.* 1971) but low levels of the protein in patients with Wilson disease had implied a function in the transport of copper.

Individuals with an hereditary deficiency in caeruloplasmin due to inherited defects within the gene have recently been identified (Harris *et al.* 1996; Harris *et al.* 1995; Logan *et al.* 1994; Takahashi *et al.* 1996; Yoshida *et al.* 1995). These patients have normal copper metabolism (reflecting normal Wilson and Menkes gene products) but present clinically in mid-life with diabetes, movement disorders and dementia due to the deposition of iron in pancreas and brain. The pathology reflects the loss of the specific role of caeruloplasmin catalysing the oxidation of ferrous iron (Fe^{2+} to Fe^{3+}) for its efficient incorporation into transferrin. This is a very specific defect in iron transport. Despite toxic tissue iron levels, patients with acaeruloplasminaemia have a mild "iron deficient" anaemia with low mean corpuscular volume (MCV) and low mean cell haemoglobin concentration (MCHC).

Copper and caeruloplasmin levels have never been specifically studied in haemochromatosis. Similar to hypotransferrinaemia, this iron-loading phenotype differs significantly from haemochromatosis in that excess iron absorption occurs in the presence of a low circulating transferrin saturation.

The search for candidate gene sequences: an open mind

Function

The haemochromatosis phenotype may be defined as a mild, excess absorption of dietary iron over that needed to make good daily iron loses through desquamation of skin and loss of intestinal mucosal cells. The defect in the putative haemochromatosis gene product disturbs the tight regulation controlling iron absorption. Iron deficiency in the duodenal enterocytes and possibly reticuloendothelial cells suggest that an excess release of iron is a characteristic feature. Mechanisms controlling iron release by mammalian cells are not known.

A eukaryote transferrin independant iron transport complex has been defined in yeast in which ferric reduction is an essential step prior to transfer of iron through a transmembrane pore. Proteins fulfilling these functional roles would be plausible candidates for the haemochromatosis gene product.

To be considered as candidates for haemochromatosis, genes coding for proteins with roles different to those described above would require the definition of entirely novel pathways of iron transport.

Expression

The demonstration of a functional disturbance in iron handling in either duodenal enterocytes and/or reticuloendothelial does not confine the likely distribution of the haemochromatosis gene product to these cell types alone. Placental trophoblast and breast epithelium also transfer iron (Baynes *et al.* 1991; Jankovic *et al.* 1991).

The defect might represent an aberration in a ubiquitous pathway of ion transport.

If the haemochromatosis gene encoded an obvious iron metabolism intermediary then its identification would have been completed long before now. Having overcome the difficulties imposed on the positional cloning approach by linkage disequilibrium and the historical reverence placed on linkage to HLA-A we have defined a genomic region of 200kb in which the gene maps. With a conservative estimation of one gene every 20kb we would expect at least 10 genes to be contained within this region. Each will require rational evaluation in light of present interpretation of human iron absorption until such a time as the genetic mutations causing haemochromatosis are unequivocally defined.

Chapter 3.2

Targetted database analysis

Background

Such has been the rate of progress in biological information technology that Mark Boguski of the National Center for Biotechnology Information (NCBI) reviewing the state of bioinformatics for Current Opinions in Genetics and Development in 1994 proclaimed that his review would be "dated before the ink was dry" (Boguski 1994). Even in such a short period of time, the utility of the nucleotide and protein sequence databases have benefitted both from more efficient "user friendly" search tools and more importantly from a continued exponential expansion of the data itself.

The power that PCR brings to raw sequence data has already been briefly discussed in the context of sequence tagged sites (STSs) and their utility in genome mapping (Chapter 2.2). An associated but distinct development over recent years has been the establishment of EST sequence databases such as dbEST at the National Center for Biotechnology Information, NCBI (Boguski *et al.* 1993). Although the majority of ESTs are incomplete sequence fragments, the fact that they represent expressed genes and are identifiable by PCR makes the EST databases an extremely useful tool for gene identification.

Methods

Having defined both a chromosomal location and a theoretical biological model for haemochromatosis, the databases were searched for candidate ESTs. All sequence fragments generated from the ends of genomic clones from the region were searched by FASTA and BLASTX analysis against EST data as well as genomic sequence database (GenEMBL). This will be discussed in more detail in a later chapter (chapter 3.4).

Using Netscape Navigator™ Version 2.0 (Netscape Communications Corporation), the public EST databases held at NCBI (http://www.ncbi.nlm.nih.gov/) were speculatively searched in two ways. First, ESTs that have been mapped to

human chromosome 6 were analysed for potential haemochromatosis candidates. Secondly, human homologues were sought in the EST database for each of the components of the *Saccharomyces* iron transport complex.

Results

1. Allignment of short genomic sequence fragments

See Chapter 3.4, page 173.

2. Searching by chromosomal location - UniGene at NCBI

Transcript maps of defined physical regions are traditionally constructed by mapping of cDNAs on YAC contigs (for examples see Totaro *et al.* 1996 and Gruen *et al.* 1996). In principle this could be achieved for the whole of the human genome (Hochgeschwender 1992). To this end many groups have taken to the random mapping of ESTs on radiation hybrid panels or YACs by PCR. UniGene, the Unique Human Gene Sequence Collection at NCBI, is accessible through the internet (http://www.ncbi.nlm.nih.gov/Schuler/UniGene/index.html). It provides a database of mapped ESTs, grouped by chromosome. Given that most ESTs are only fragments of the expressed transcript and that some genes might be represented more than others, sequence redundancy is a potential problem of this type of database (i.e. multiple sequence depositions representing the same gene). UniGene addresses this problem by grouping together EST sequences with significant sequence similarity such that their final database contains "loci" that represent the pooled 3' transcription products of distinct genes (Boguski and Schuler 1995).

At the time of writing (July 1996) the chromosome 6 site of UniGene contained a total of 629 EST loci. Of these 457 are "anonymous" ESTs (i.e. deposited sequence without homology to other known sequences). Twenty-six separate loci have homology to HLA genes. Of the other genes reported to map telomeric to the MHC (prolactin, MOG, RFP, olfactory receptor, phospholipase D, butyrophilin, Col11A2) all were represented by ESTs.

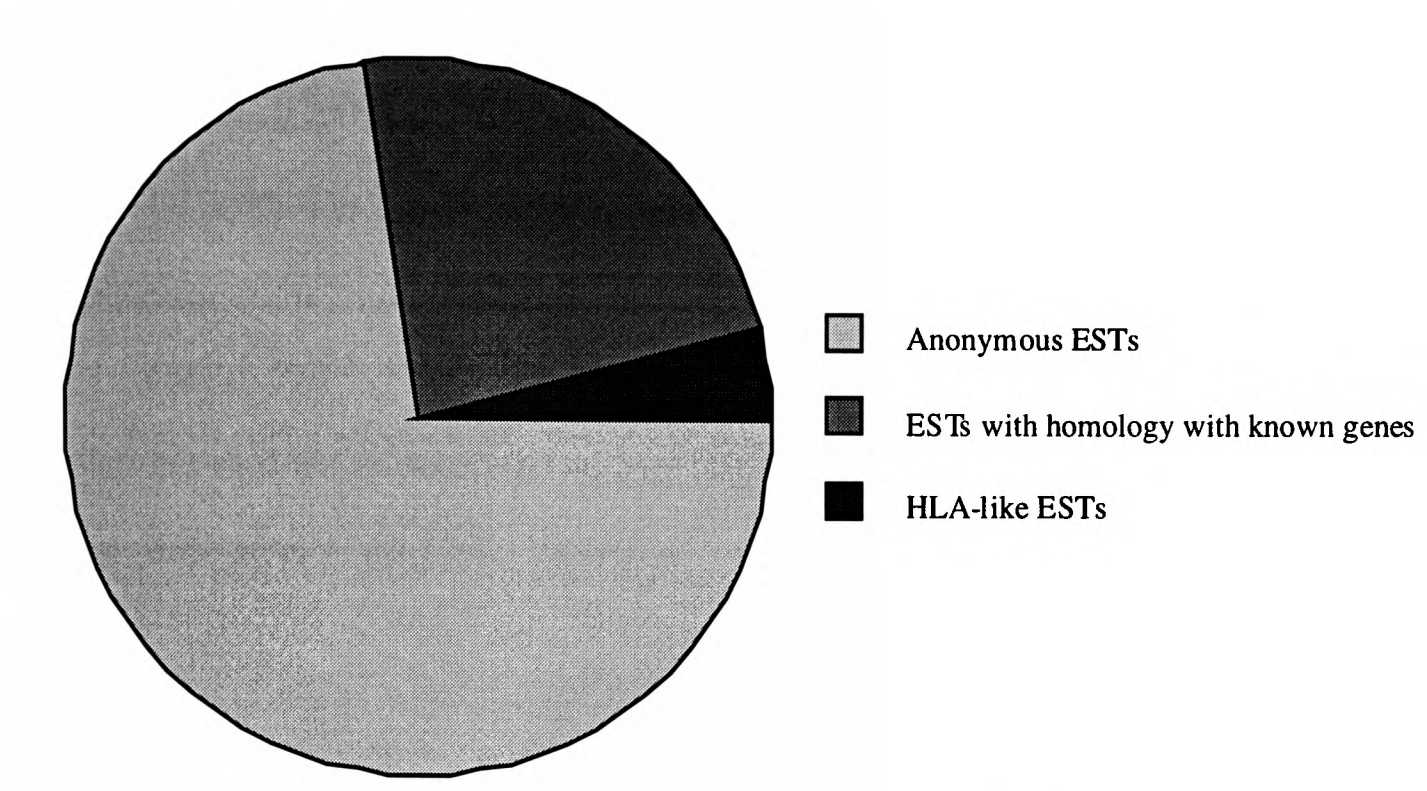


Figure 3.12 Expressed sequence tags mapped to chromosome 6 (UniGene at NCBI).

Of the 172 EST loci showing homology to known genes, none had an obvious role in ionic reduction, binding or transport.

3. Searching by cross-species homology - XREFdb

Sequence conservation between species tends to be more marked for genes involved in fundamental biological pathways. Although there are likely to be important differences in the molecular pathway determining the controlled iron uptake by a unicellular organism and the intermediaries responsible for iron distribution in mammals, the iron transport mechanism in the eukaryote yeast *Saccharomyces cerevisiae* provides an opportunity to identify haemochromatosis candidates by comparative genomics (Tugendreich *et al.* 1994).

XREFdb (http://www.ncbi.nlm.nih.gov/XREFdb/) is a cross-referencing database established by NCBI to expedite the process of homology searches between genes characterised in model organisms and those in man. Entering a search query in the form of the recognised gene name or accession number, XREFdb performs a BLAST analysis translating the query sequence into all 6 reading frames and searching for peptide sequence allignment in both dbEST and a non-redundant protein database (Bassett *et al.* 1995). Recognising the continued growth and refinement of a

database such as dbEST, XREFdb automatically researches with the query every month and issues an updated report by e-mail.

XREFdb was used to search for human ESTs with homologies to the components of the multi-molecular iron uptake complex of *Saccharomyces cerevisiae* (*FRE1*, *FRE2*, *FET3*, *FET4* and *FTR1* - Askwith *et al.* 1996), in addition to other yeast genes defined as determining iron dependant growth in yeast (*AFT1* (Yamaguchi-Iwai *et al.* 1996), *GEF1* (Greene *et al.* 1993), *MAC1* (Jungmann *et al.* 1993)).

Gene Name	Organism	SwissProt Accession	Protein Matches			**EST Matches**			Probes Mapped
			New	Old	Total	New	Old	Total	
FRE1	Sce	P32791	0	10	10	0	1	1	0
FRE2	Sce	P36033	0	6	6	0	1	1	0
FET3	Sce	P38993	3	17	20	0	0	0	0
FET4	Sce	P40988	0	2	2	0	0	0	0
GEF1	Sce	P37020	20	0	20	7	0	7	1
AFT1	Sce	P22149	0	6	6	0	0	0	0
MAC1	Sce	P35192	10	0	10	0	0	0	0

Table 3.2 Results of homology searching. No record was found for the recently identified saccharomyces transmembrane "permease" FTR1 (Stearman *et al* 1996)

The human homologue of the yeast gene GEF1 maps to chromosome 3q

The yeast gene *GEF1* identified 7 human ESTs. *GEF1* had been identified in *Saccharomyces* from studies of petite (respiration defective) mutants (Greene *et al.* 1993). Petite mutants are unable to grow on non-fermentable carbon sources, due to intrinsic mitochondrial dysfuction. Oxygen consumption (both whole cell and mitochondrial) is reduced in *gef1*$^-$ mutants and the petite phenotype was suppressible by supplementation of the growth medium with iron despite no demonstrable defect in iron uptake by these cells (Greene *et al.* 1993). The translated amino acid sequence of the *GEF1* gene suggests that its product spans a cell membrane. The demonstration of normal iron uptake by the *gef1*$^-$ mutants suggests that this might act at an intracellular

location although this is at the level of the mitochondion or intracellular vacuoles has not yet been determined.

Of the ESTs demonstrating homology with *GEF1*, one has been mapped to human chromosome 3q (GenBank Accession number F07799). Taken together with the data on the behaviour of *gef1*$^{-}$ mutants (i.e. the apparent disruption in cellular respiration), this would not appear to be a functional or positional candidate for haemochromatosis.

FRE1 and FRE2 show strong homology with a single Human EST

The *Saccharomyces* cell surface ferric reductase genes *FRE1* and *FRE2* both identified the same EST (N63208). N63208 contains a sequence fragment of 338bp. The TBLASTN alignment with *FRE1* was modest (P value $2.5e^{-2}$) but the alignment of the same part of the EST sequence (nucleotides 181-333) with *FRE2* is striking (P value $7.3e^{-60}$).

FASTA analysis of the N63208 sequence against the EST database identified sparse alignments with three other ESTs (H10464 59.8% identity over 92bp - bases 49-139, R83735 and H13746 58.3% over 96bp - bases 132-225 and L40416 67.4% over 43bp - bases 15-57). FASTA against the nucleotide database (GenEMBL) demonstrated strong identity with coding sequence from Yeast (Z28220 *S. cerevisiae* chromosome XI ORF YKL220c 97.9% identity over 338bp) with no significant alignments with mammalian sequence.

The strong nucleotide identity with yeast genomic DNA raised doubts as to whether this clone contained a human cDNA or represents a contamination of the original cDNA library with yeast. The N63208 sequence had been deposited in the XREF database in March of 1996 by R.K. Wilson of the WashU-Merck EST project having sequenced the clone from the Morton Fetal Cochlea cDNA library. Contamination of the library with yeast DNA is not a risk that would have been inherent in the experimental proceedure. In review of the TBLASTN alignments of yeast protein queries against human EST sequences in 1994, Tugendreich identified

37 yeast proteins (or 2.13% / out of a total of 1733) represented by human ESTs with TBLASTN P values less than 10^{-47} (Tugendreich *et al.* 1994). Furthermore, as of January 1996, 25% of human disease genes identified by positional cloning matched yeast genes in GenBank with a P value of less than $P = 2.0x10^{-46}$ (Bassett *et al.* 1996). The cystic fibrosis cDNA sequence (M28668) matches the yeast gene *YCF1* with a BLASTX P value of $1.3x10^{-167}$. The Wilson disease cDNA (U11700) and *CCC2* match with a P value of $5.9x10^{-161}$ (Bassett *et al.* 1996).

Despite the remarkable nucleotide alignment with the yeast gene *FRE2*, N63208 was considered as being human.

Attempted mapping of N63208

The partial cDNA pBluescript SK- clone N63208 was available as part of the IMAGE Consortium library (Lennon *et al.* 1996) and it was obtained for further study. The insert sequence was confirmed by sequencing the clone using primers to the T3 and T7 promotor sequence of the vector.

Primers designed to the insert sequence of N63208 were used in an attempt to map the sequence by PCR. Amplification of a product of the predicted size from all YACs studied was interpreted as cross-amplication of the yeast gene of the YAC host strain given the extent of the nucleotide match.

PCR on human genomic DNA failed to generate product precluding the mapping of this locus using this technique. Furthermore, a probe derived from the PCR product from yeast failed to demonstrate specific hybridisation to Southern blots of human genomic DNA and human multiple tissue northern blots.

That the sequence contained in N63208 is human remains to be proven.

Interpretation

Could a defective ferric reductase explain the haemochromatosis phenotype?

In the yeast model defined in *Saccharomyces cerevisiae* the ferric reductases (fre1p and fre2p) are bound on the exterior of the cell surface.

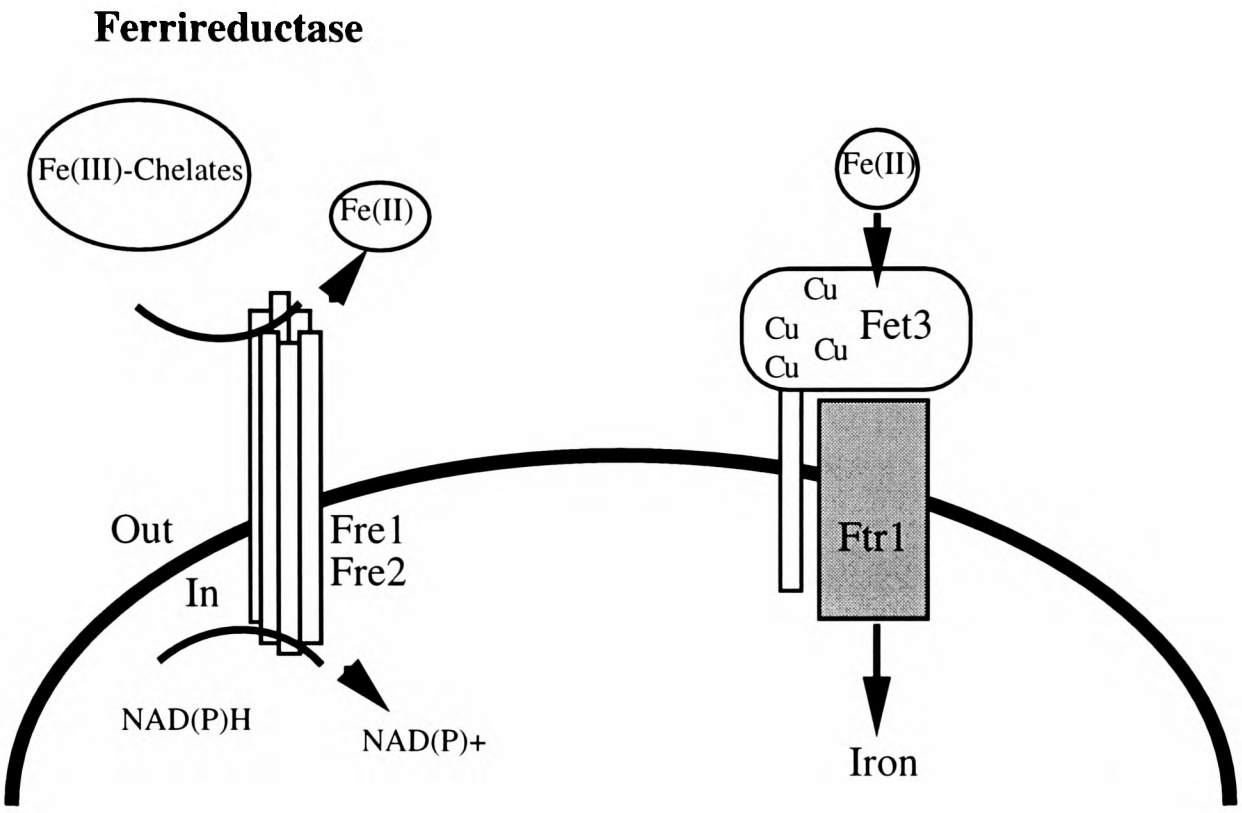


Figure 3.13 The Saccharomyces cell surface iron uptake system (reproduced with permission from Askwith *et al* 1996)

Reduction of ferric iron (Fe^{3+} to Fe^{2+}) appears to be the first critical step in iron uptake by either the high affinity (*FET3*) or low affinity (*FET4*) pathways. The first ferric reductase, *FRE1*, was identified in yeast by complimentation studies on mutants unable to grow on iron limited media (Dancis *et al.* 1990). The gene identified in this strain was shown to be transcriptionally regulated by iron and confirmation that this represented the gene of interest was established by the demonstration of a specific Δfre1 disruption strain which shared the phenotype of the original mutant (Dancis *et al.* 1992). Persisting surface ferrireductase activity in the Δfre1 disruption strain suggested the existance of a second ferric reductase and this was confirmed by the identification of sequence similar to *FRE1* in an open reading frame on saccharomyces chromosome XI (*FRE2* Georgatsou and Alexandraki 1994). In combination *FRE1* and *FRE2* account for 90-98% of cell surface ferric reductase

activity in yeast of which *FRE2* may account for up to 40% , appearing to be more important after the early exponential phase of growth (Georgatsou and Alexandraki 1994).

No mammalian ferric reductase gene has been cloned although ferric iron reduction has been inferred in several *in vitro* models of mammalian iron absorption. Raja studied rates of ferric reduction (as measured by increases in ferrozine chelatable iron) from incubation of mouse duodenal fragments (Raja *et al.* 1992). He demonstrated that this reducing capability was induced in iron deficient mice but the variation in rates of reduction did not parallel the increase in iron absorption (Raja *et al.* 1993).

Duodenal ferric reductase activity has been measured in endoscopic fragments from man (Pountney *et al.* 1994). Reduction and uptake measured in this way is increased in iron loaded patients with haemochromatosis and remains elevated after iron stores are corrected by phlebotomy treatment (Raja *et al.* 1996). Demonstration of the down regulation of this process in patients with secondary iron overload or iron loaded mice would further enhance the claims that ferric reduction might confer a degree of control over intestinal iron absorption in mammals.

The quantitative contribution of surface ferric reduction and iron uptake in mammalian systems must be kept in perspective. For the majority of mammalian cells iron is delivered and taken up by a specific carrier mediated iron transport and uptake system (i.e. transferrin and transferrin receptor). This is not the case for the lining epithelium of the gut, hence Raja's initial hypothesis. Two points must be re-emphasised with regard to inferring a defect in luminal ferric reduction as possibly being important in the increased gastrointestinal iron absorption in haemochromatosis. First, ferric iron chelates are not the only form in which dietary iron is presented to the gut and pathways in which iron remains more heavily complexed (e.g. to haem) must also be considered. The second, and more fundamental issue, is that concerning the apparent iron deficiency of the duodenal enterocyte in haemochromatosis. As

discussed earlier, this observation implies that the control point for iron absorption (at least that which might be aberrant in HC) lies at the serosal side of the cell.

Although Raja's experiments examined ferric reductase activity from the exterior of the cell it is possible that reduction is a common feature of different forms of membrane iron transfer (Nunez *et al.* 1990; Watkins *et al.* 1992) and that such a step is necessary in the process of iron release from mammalian cells.

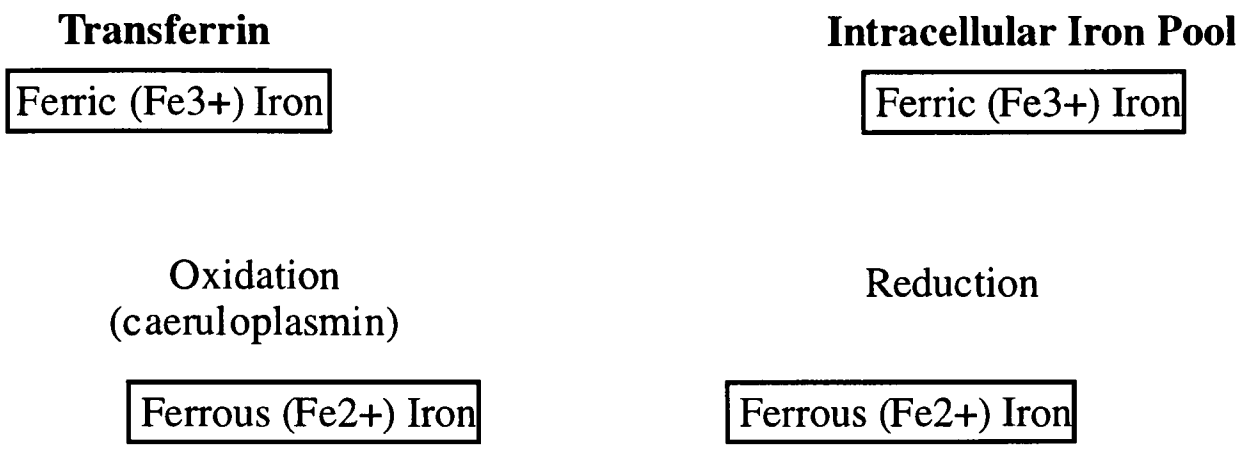


Figure 3.14 Could ferric reductase activity facilitate iron release from the enterocyte?

Whether or not it represents the primary defect causing haemochromatosis, ferric reductase activity is likely to play at least a part in the control of intestinal iron absorption.

Chapter 3.3

Identification and characterisation of a CpG island within the haemochromatosis gene region

Background

CpG Islands and genes

The term CpG island is now the accepted term for what was originally referred to as an HTF (*Hpa* II tiny fragment) island. The original term reflected the clustering of sites for the methylation sensitive restriction enzyme *Hpa* II (recognising the sequence CCGG) by which these regions had first been recognised. Each island may span up to 2kb of DNA and it has been estimated that there are up to 45000 CpG islands in man accounting for about 2% of the genome. Interest in CpG islands has followed from the recognition that these regions that are identifiable in genomic DNA, lie at the 5' end of many genes. In particular CpG islands are associated with all constutively expressed or "house-keeping" genes and a high proportion of tissue specific genes (Larsen *et al.* 1992). Recognition of CpG islands can provide an efficient route to characterisation of genes.

Identification of CpG islands in cloned DNA

CpG islands are characterised by an excess of the CpG dinucleotide over bulk genomic DNA and that they remain unmethylated in genomic DNA. The CpG "excess" of these regions is a reflection of the under-representation of this dinucleotide in "non-island DNA" due to the insidious deamination of 5-methylcytosine to thymine. The exact role of methylation in the preservation of the CpG dinucleotide content of CpG islands is not understood.

Human genomic DNA cloned within yeast becomes unmethylated and identification of CpG islands (as defined above) in YACs is not possible. However, as CpG islands represent regions that are very GC rich they show clustering of sites for other methylation sensitive rare-cutting restriction endonucleases that have a high GC content in their recognition sequence (e.g. *Not* I, *Sac* II). This feature of CpG islands has previously been exploited to identify CpG islands in cloned DNA and in YACs *Not* I sites have often been considered to represent the location of "putative" CpG islands. Sequencing through and beyond the CpG island identifies the associated gene.

Aggie-3 lies close to a CpG island

The CA repeat microsatellite Aggie-3 was isolated from phage clone C7p3 (Chapter 2.4, page 93). This clone had been shown to cut with several rare cutting restriction enzymes (*Mlu* I, *Asc* I, *BssH* II, *Sac* II, *Eag* I). Direct cycle sequencing of phage ends was being undertaken to facilitate the fine mapping of the region (see Chapter 3.4) and sequence from the T3 end of C7p3 was identical to a sequence deposited in GenBank from the CpG island library (Z63994 H. sapiens CpG DNA, clone 94b5, reverse read - 100% identity over 212bp). From these two independant pieces of information (restriction site clustering and inferred hypomethylation in genomic DNA) a CpG island was defined.

Recognition of a CpG island within the haemochromatosis gene region naturally lead to further studies to establish the nature of the associated gene.

Methods

Sequencing the CpG island

The initial CpG island sequence (217bp) had been obtained from cycle sequencing (ThermoSequenase, Amersham) the λ phage clone C7p3 using a T3 sequence primer.

Mapping of C7p3

C7p3 DNA was digested with *Eco* RI and *Hin*d III (complete and partial digestion). Half of each digest was "end-labelled" with ^{35}S αdATP using Klenow fragment of DNA polymerase. Both the labelled and the unlabelled DNA fragments were separated by electrophoresis in two halves of the same 1% agarose gel. After electrophoresis, the half of the gel containing the unlabelled DNA was blotted onto a nylon membrane (Hybond N, Amersham). The radiolabelled half of the gel was fixed and dried onto paper under vacuum. This dried gel was exposed directly to x-ray film.

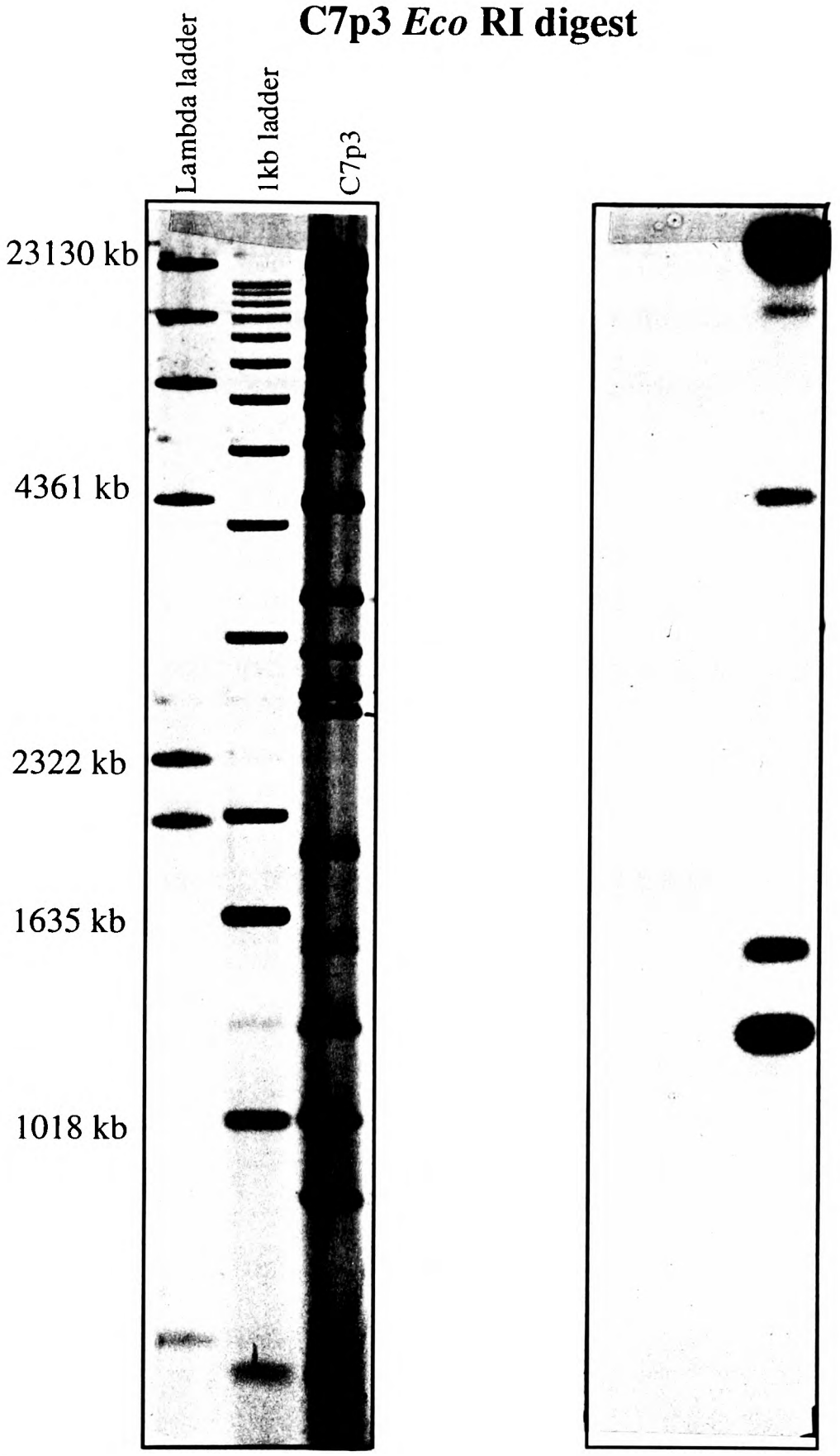


Figure 3.15 End-labelling of restriction fragments of C7p3. The left hand panel shows an autoradiograph of the partial restriction digest products of C7p3 digested with *Eco* RI, end-labelled with ^{35}S and resolved in 1% agarose. The right hand panel shows the other half of the gel (unlabelled) transferred to Hybond N membrane by Southern blotting and probed with the C7p3T3 PCR product.

This procedure was undertaken to generate an accurate fingerprint of the clone. Hybridisation of C_0t 1, or human placental, DNA to a Southern blot of digested genomic DNA produces a hybridsation signal proportional to the repetitive elements within the fragment (see Chapter 2.4, page 90). Fragments containing non-repetitive (i.e. single copy) DNA will be under-represented in such hybridisations. By "end-

labelling" the fragments, the derived signal as recorded in the autoradiograph is proportional to the amount of DNA only, regardless of repetitive elements.

The Southern blot (i.e. the other half of the original gel) was probed sequentially with PCR products derived from sequence form both the T3 and T7 end of C7p3 and from sequence flanking the Aggie-3 CA repeat (not containing the repeat). From this a restriction map of the clone was established.

Subcloning C7p3 *Eco* RI digested DNA into pUC

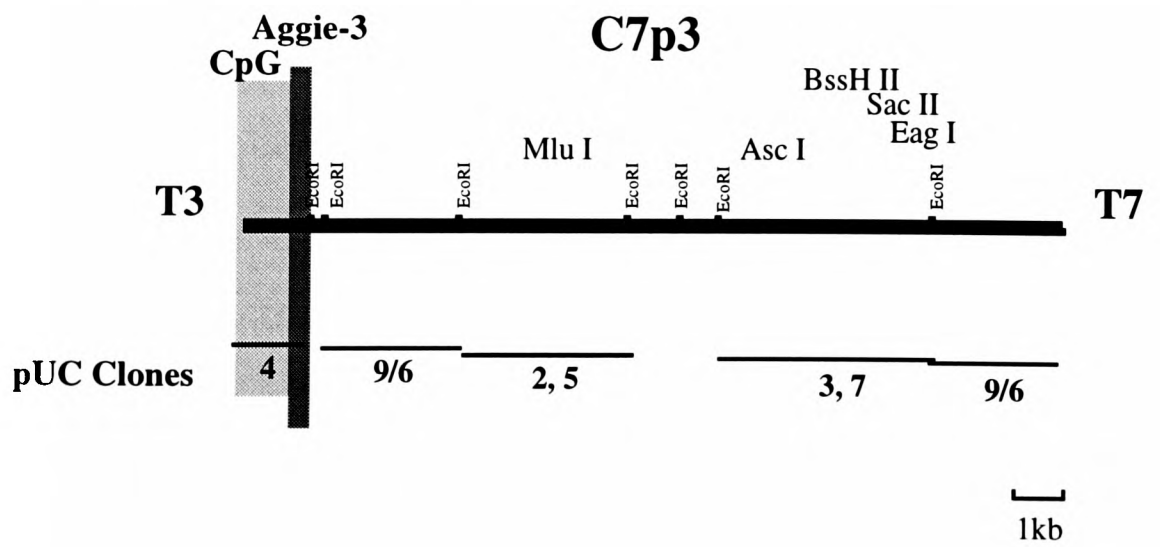
To facilitate sequencing of the CpG island, C7p3 *Eco* RI fragments were cloned into pUC13. The pUC clone containing the T3 end of C7p3 (C7p3plas4) was identified by PCR using primers designed to the sequence previously obtained (C7p3T3.seq).

Sequence extending further into the phage clone was obtained with sequential primers designed to sequence previously obtained (C7p3T3.seq) and by sequencing the pUC clone with M13 forward and reverse primers.

Results

The map of C7p3

An *Eco* RI map of C7p3 was generated



Sequence

pUC clone 4 insert was sequenced totalling 1270bp of sequence including the Aggie-3 repeat.

C7p3T3F
(T3) GATCTT ATCACACTGA GTAATGCATG CAGTGCTTGT AATCGAGTCT GGCTAAGGAG
PenR1.3
AAGCCACATA CACTTTGTTA GTACTGAGTG AGGAATGTGG CGCCTAAAAA ATTAGGGATG
TGCAGGAAAC TTGTGTTAAC AGAAAGTGCT TCCTGGCTCT TGGCGCGAAA ATGGATACTT
CCGTTGTGTC GAGTGCTGTG ACTTCCTGTT AGAACTTGTT GAAAGCCTAT TGTGTCACGT
PenF1
GTACTTTCCA CCATGTAATG GCGTTCTAAC GTGAGTTTAT ATATTAGTAA ATTTGGTTCC
C7p3T3R
CGATGGCGTT CCAGTCAGTA TGGCTTCGCT CGAGAGTACT TATTTAACCT CTTCTAGGCT
GTACCCAGCT CTGCCTTGCT TTGCTTCCCC CTCACCCCCT GATTCCATCT CTGCAGCCCA
CGGGACATCG GGAAGATGCC AGACCACCAG GCAGGATAGT CTGCAAGGGT AAGGGGAGGA
PenF1.5
GCACAGCTTT TCTCAGGAGT GAAGAAACGG AACATGAAAT CATCACGGAG ATTTTTGGAA
TAGTTTACTA GTTTTTAAGG CATGTTCAGC GTCAATTGGT CCAATCAGTA ATTTTTCCTT
TTGACAAGAA AGCTTATAGA AAAGGAAAAT GAAAAGCAAG TAGATTCAGA GGGGAGACAA
CAGACAGACT AGAGACAGTC CGCCCCGAAG CAAGAAGTTG GAGTAGAGGA GAAGGGAACT
TCCCCAGCAG CACAGGGTTT GAGCTAAGTG TTAGGTGGTG GGTTAGAATT AGGTGGTGGG
TTTATTGTGT GATTAGAGGC TAGAATGACT ATGCCGGTGA TGAGCAGTGT GAAACTTTTT
TGGAAGAACC AAGTGAAAGA AAGAGTCCAC GAATTTTAAG CTAGGTTGTT ACCTGGCTAA
Aggie-3 GT
ATTTGTGATT TAGAAATGTA GGCCAGTTTC ACAGGGGAAT TTT**GTGTGTG TGTGTGTGTG**
TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG ATTTAACATA TTTGGCCACG CATTGTTCTA
PenF2
GTCTTTGGGG ATATTGCCCG CCCTTAGGCA TTTGGAGATT AGCAGGGTCA TATTGAAAAA
Aggie-3 CA
CTTCATACAT GCAAAGCATA TGTTACACTG TTGATCCAAG TCTTATTTAT CCTACCAGTA
ATTTTGGGGT TACATTAATA TTCTAAGAAA TGTATCAACA AATAATTATA TTTATTGTTT
ACCTATATAG GAGGCACTGA TCTATATCTA TACCTATATA GGAGGGCAGA TCACTTGAAG
TCAGGAATTC (T7)

Figure 3.16 Sequence obtained from λ phage clone C7p3. The primer pair C7p3T3 were used for STS content mapping (Chapter 3.4). The Aggie-3 primer pair was used to amplify the microsatellite. The primers PenF1, PenF1.5 and PenF2 were used for sequential sequencing into the clone. PenR1.3 was eventually used to sequence toward the T3 end of the overlapping clone C7p22. Reverse primers were designed to sequence on the opposite strand (not shown).

Analysis of this sequence by FASTA against the GenEMBL and EST databases identified no significant alignments. Analysis of CG dinucleotide content

showed a progressive fall in CG content suggesting that the sequencing was moving away from the CpG island.

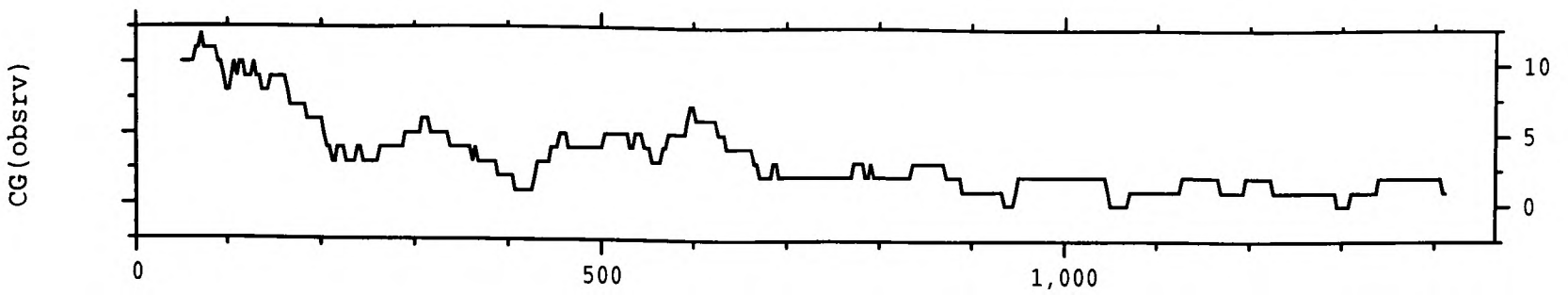


Figure 3.17 Statplot of observed CG dinucleotide content of the sequence from C7p3plas4.

Identification of an overlapping phage clone

An *Eco* RI/*Hin*d III fingerprint of the B117C7 phage library clones had suggested that C7p3 shared bands with C7p22 (see figure 2.40, page 90). The CA containing band in the two clones appeared to be of different size, although Aggie-3 (identified from C7p3) was present on C7p22 by PCR. Direct cycle sequencing of the phage ends of the C7p22 demonstrated that it shared the same T7 end with C7p3 (identical sequence although this sequence did contain Alu repeat elements).

Restriction mapping of C7p22

Phage clone C7p22 was mapped and found to extend out beyond the T3 end of C7p3.

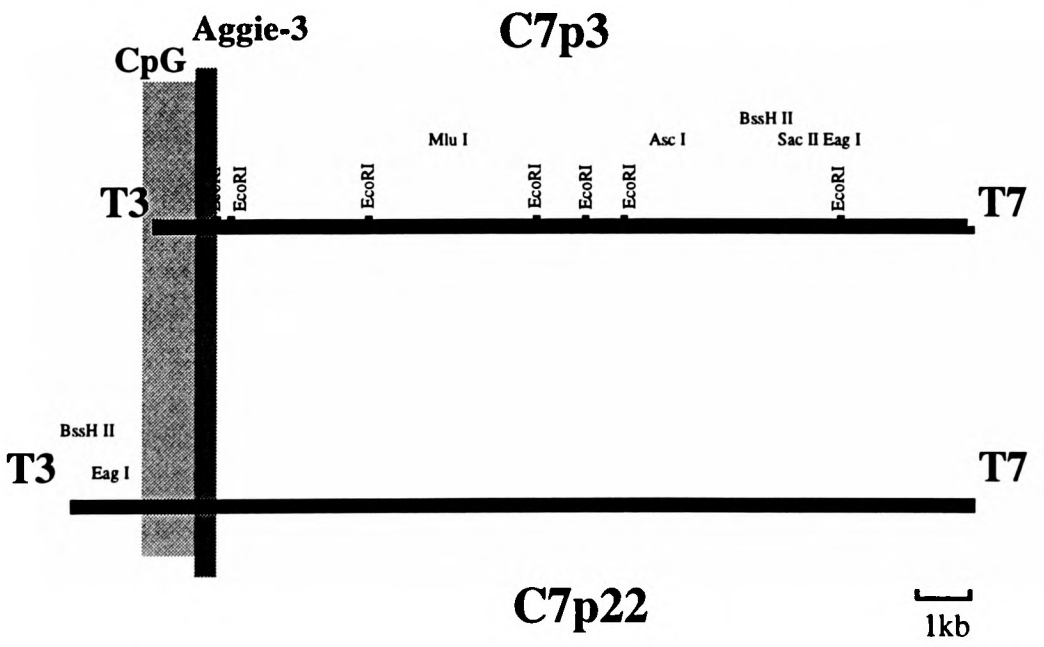


Figure 3.18 Phage clone C7p22 overlaps the T3 end of C7p3

With the design of a reverse sequencing primer from C7p3 (PenR1.3), sequencing was extended toward the T3 end of C7p22. This sequence (C7p22PenR1.3.seq, 339bp in total) not only confirmed a progressive rise in GC content but identified a histone H2B sequence in both the genomic (GenEMBL) and expressed sequence (TAGS) databases.

Title	Species	Accession	Identity	Over	Range	Comments
Histone H2B.1	H.sapiens	M60751	84.4%	307bp	32-339	Ch 1
Histone H2b	H.sapiens	X00088	84.3%	281bp	58-339	
Histone H2B.1	H.sapiens	M60756 J05322	90.6%	203bp	136-339	Ch 1
Histone H2B.2 and H2A.1	H.sapiens	X57138	85.7%	258bp	81-339	
CpG Island	H.sapiens	Z57636	93.8%	194bp	116-310	
CpG Island	H.sapiens	Z58190	86.4%	250bp	89-339	
Histone H2B.1 and H2A	H.sapiens	X57985	90.6%	203bp	136-339	Ch 1

Table 3.3 Sequence allignments with sequence obtained from the T3 end of C7p22 (C7p22PenR1.3.seq). Range refers to the part of the C7p22PenR1.3.seq sequence demonstrating allignment.

The orientation of this histone gene sequence was such that the sequence was extending into the gene from the 3' end.

Interpretation

Histones

Histones comprise the major structural elements of chromatin. Histones are broadly divided into 5 classes. The H1 histones are the largest class both in terms of sequence variants and protein size (23 kiloDaltons). The 4 other classes are often referred to as "core histones" due to the fact that they aggregate to form the central "kernal" of the nucleosome. H3 and H4 histones are highly conserved across eukaryotic species whereas the H2A and H2B histones show species-specific sequence variation.

Histone gene clusters

Multiple copies of genes for each class of histone are present in any one species. Furthermore the genes tend to cluster. In man, histone genes have been mapped to chromosomes 7 (Chandler *et al.* 1979), 1, 6 and 12 (Carozzi *et al.* 1984). All the H1 sequences have been mapped to chromosome 6 in the region 6p21.1-6p22.2 (Albig *et al.* 1993).

Sequencing putative CpG islands around D6S105 Caroline Stone identified histone sequences from four classes (H1, H2B, H3 and H4) suggesting the presence of a histone cluster (Stone 1995). Another cluster of histone genes has been identified several megabases telomeric to D6S105 (Volz and Ziegler 1996).

Another core histone sequence had been identified within the haemochromatosis gene region. Evidence that this sequence was expressed was that the sequence was associated with a CpG island and that it was represented in the EST database. No other histone sequences were identified in that region by PCR.

Relationship between histone genes and CpG islands

The presence of CpG islands close to 99% of house-keeping genes suggests that this association is true for histones. Larsen described CpG islands associated with all classes of histone genes (ubiquitous and tissue specific) in addition to a histone H3.3 pseudogene (Larsen *et al.* 1992).

The sequence data from C7p3 suggested that the CpG island that had been identified might lie predominantly over the 3' end of the histone gene. This raised the possibility that another neighbouring gene 3' to the histone might have been overlooked. The sequence 3' end of the histone was rigorously examined for another coding sequence or other sequence motifs suggestive of a neighbouring gene (TATA and CAAT boxes indicative of promoters and consensus splice site sequences reflecting intron-exon boundaries). No exons were identified with the kilobase of sequence that had been obtained.

Could a histone gene mutation account for the haemochromatosis phenotype?

Intuitively a histone gene does not appear to be a very attractive candidate for the haemochromatosis gene. Most of these genes are constituitively expressed and their products play an intrinsic role in nuclear DNA packaging. Furthermore the apparent redundancy in terms of expressed genes within a given histone class might be expected to protect against mutation at a single locus being transmitted to tissue specific cellular dysfunction. UniGene at NCBI (see Chapter 3.2, page 148) has a total of 61 histone loci including 11 for histone H2B expressed sequences. Two of these expressed histone H2B loci have been mapped to chromosome 6.

For histone dysfunction to account for a very mild and extremely specific defect in iron absorption is unlikely. This histone gene was not considered further as a candidate gene.

Does the recognition of a histone gene have bearing on the search for haemochromatosis candidates?

Although this histone gene might be rejected as a haemochromatosis candidate, its identification has a significant bearing on the strategy adopted in the search for other gene sequences in this region. Histones are constituitively expressed and represent an extended multigene family with significant sequence homologies. Techniques to identify genes on the basis of expression (e.g. cDNA selection) will identify histones in all libraries. Furthermore, two other constitutively expressed multigene families are also represented within this region. A serine transfer RNA has been identified by sequence analysis close to D6S1260 (Pointon 1995). Database analysis using the UniGene site at NCBI revealed a zinc-finger transcription factor gene (zinc finger protein HF.12, EST accession 60204) contiguous with the sequence of STSG9945.

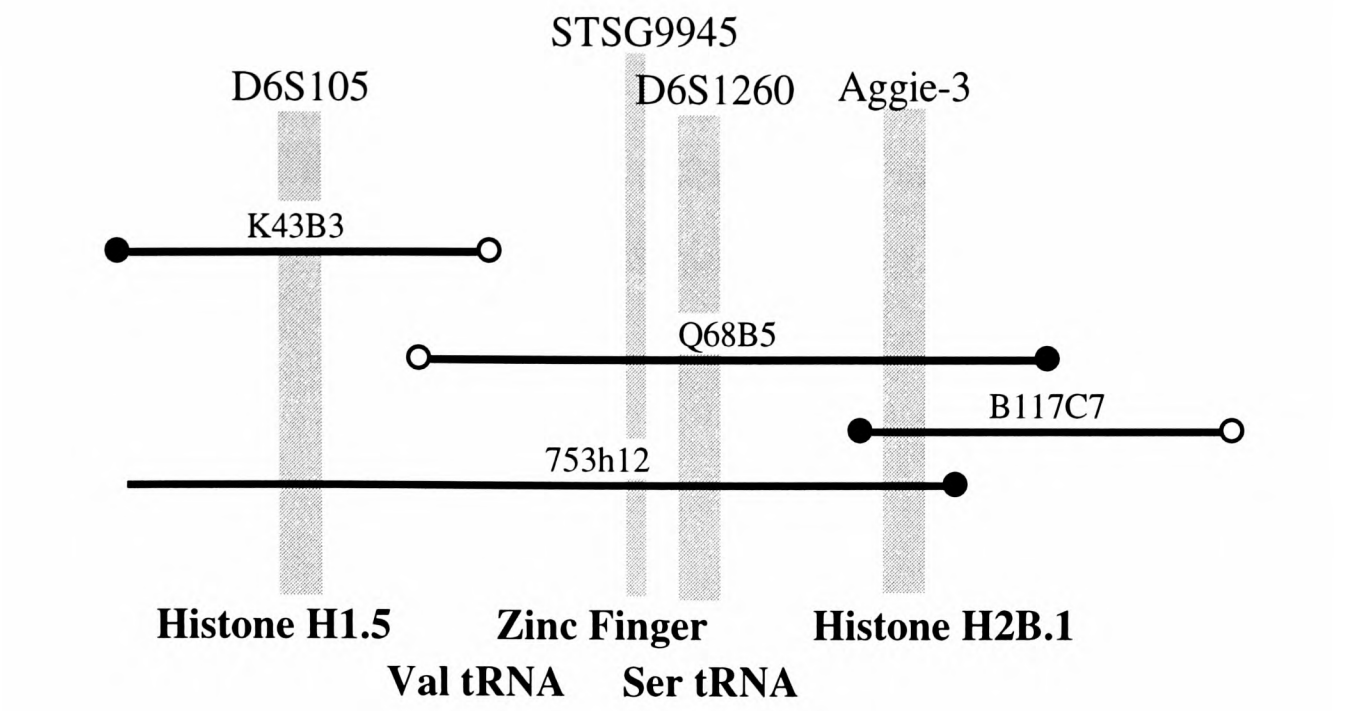


Figure 3.19 Members of the histones, tRNAs and zinc finger multigene families are all represented in the region telomeric of D6S105.

Two important features are characteristic of these multi-gene families and worth considering in more depth. First, histone and tRNA genes lack introns. This means that they would not be identified by exon trapping. This approach is attractive as a gene identification technique in this region, although this would assume that the haemochromatosis gene does contain introns. The more obvious feature is that members of conserved multigene families can be relatively easily identified by sequence comparison to the genomic and EST databases.

Having defined a gene region of 200kb and having identified the presence of members of constituitively expressed multi-gene families, the favoured approach to identifying novel genes was a combination of genomic sequencing and exon trapping.

Chapter 3.4

Fine mapping of the haemochromatosis gene region and preparations for genomic sequencing.

Background

In the twenty years since the chance association with HLA-A3 was reported, haemochromatosis has remained a major challenge of positional cloning. The effect of linkage disequilibrium had at last been overcome by the combined approaches of microsatellite analysis and statistical likelihood methods. A haemochromatosis gene region of 2.5Mb (bounded by RFP and D6S1621) had been defined in a previously unmapped region telomeric to the MHC. This region had been recognised to be potentially gene-rich and members of constitutively expressed multi-gene families had already been identified. A search for haemochromatosis candidate genes was planned using a combination of exon trapping and genomic sequencing.

A YAC contig spanned the entire gene region and within this we had identified a critical region based on a plateau of linkage disequilibrium at the microsatellites D6S1260 and D6S1558. Selecting the next STSs outside these two points we planned to convert this into a "bacterial" clone (i.e. cosmid, PAC, BAC) contig as a prelude to sequencing.

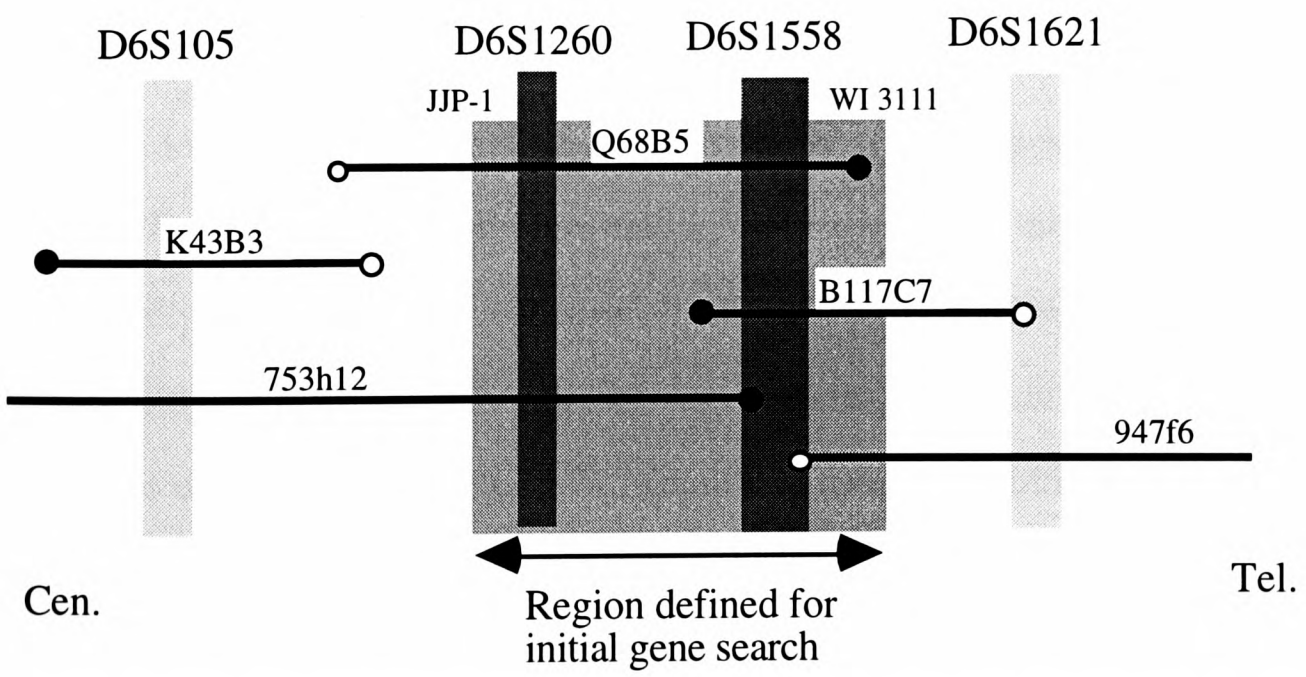


Figure 3.20 The haemochromatosis "critical region" lay between D6S1260 and D6S1558.

Methods

During the search for microsatellite loci in this region the YACs K43B3, Q68B5 and B117C7 had already been subcloned. These cosmid and λ phage clones were used to generate short end sequence fragments for mapping and for screening cosmid, BAC and PAC libraries, to identify further clones from the region.

Phage end sequencing

Each of the human containing λ phage clones obtained from B117C7 was sequenced directly with T3 and T7 sequence primers (ThermoSequenase, Amersham). Primer pairs were designed to each sequence fragment. These phage insert STSs were used in STS content mapping on other phage clones from the library and YACs from the region.

Each sequence fragment was also analysed by FASTA and by BLASTX to identify repetitive elements and for significant sequence allignments with known genes.

Refining the physical position of microsatellite loci

PCR based STS content analysis on intact YAC clones from the region had provided a crude order of these new microsatellite loci (Chapter 2.4). Fine mapping of CA repeat loci by hybridisation to digests of YAC DNA has two drawbacks. First, microsatellite loci, by nature, contain repetitive sequence and hybridisation must by competed with C_0t 1 DNA, potentially limiting the resolution of the hybridisation signal. Secondly, the density of restriction sites within large YAC clones means that the unequivocal identification of small internal restriction fragments may be impossible (as discussed in Chapter 2.2). Identifying these difficulties we adopted an alternative strategy to fine mapping of STSs within this region using YAC fragmentation.

Using the fragmentation vector pBCL 8.1 two large YACs telomeric to D6S105, Q68B5 and 947f6, were fragmented (Jenny Pointon). The fragmentation process anchored the derivative clones by the left arm, which in both cases lay centromeric.

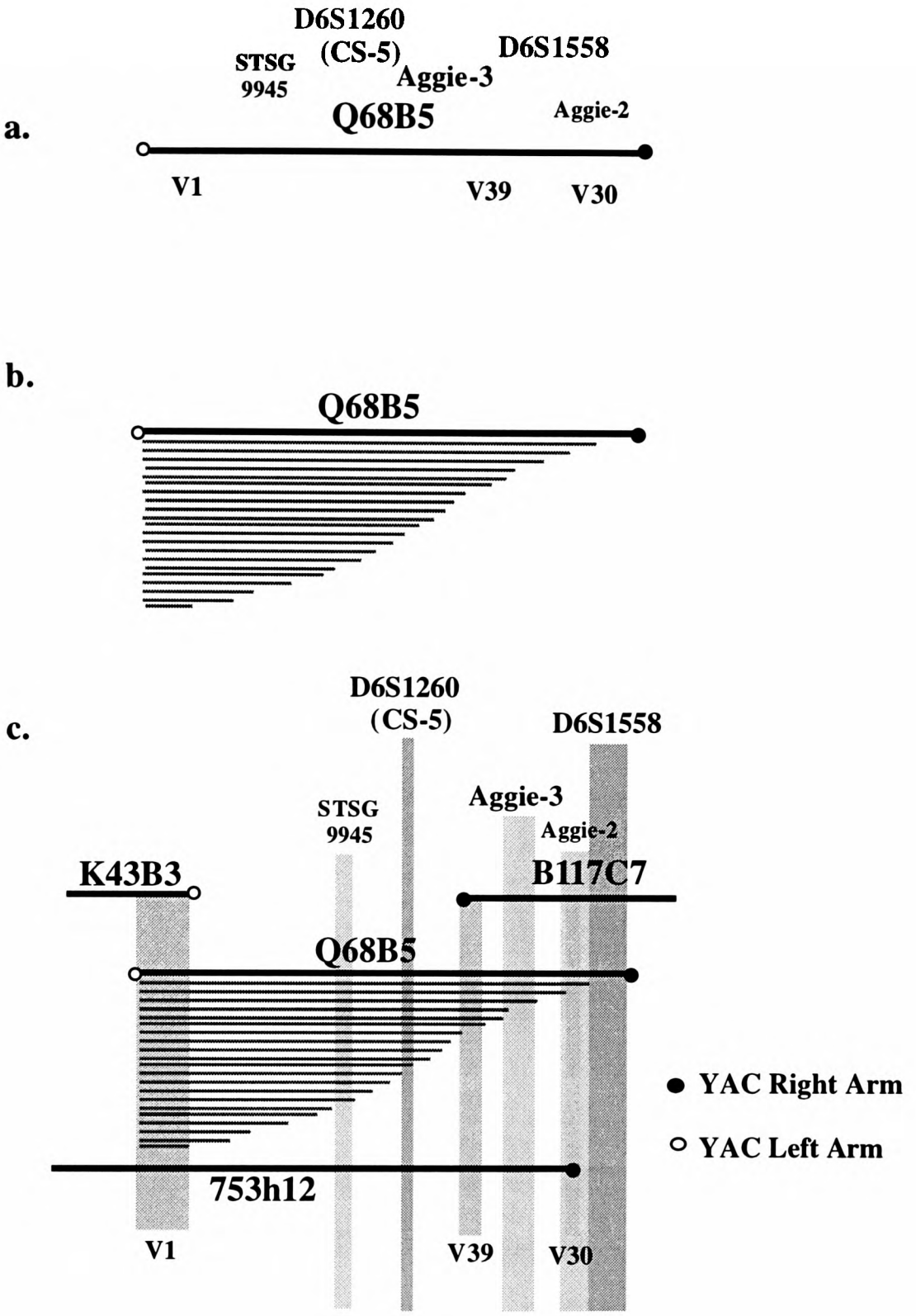


Figure 3.21 The principle of YAC fragmentation and its use in STS content mapping as illustrated by Q68B5. Three microsatellites and 4 STSs mapped to this clone by PCR. Although an approximate order was established from STS content on overlapping YAC clones, fine mapping of these loci was established by PCR on smaller clones generated from the fragmentation of YAC.

The fragmented clones were sized by PFGE and DNA from each sub-clone was used in high resolution STS content mapping by PCR.

Screening the cosmid and PAC libraries

The ICRF chromosome 6 cosmid library and the human PAC library (HGMP resource centre) were screened by PCR and hybridisation using the STSs generated from across the interval (Alison Merryweather-Clarke).

Results

B117C7

A composite sub-clone and restriction map of B117C7 was constructed.

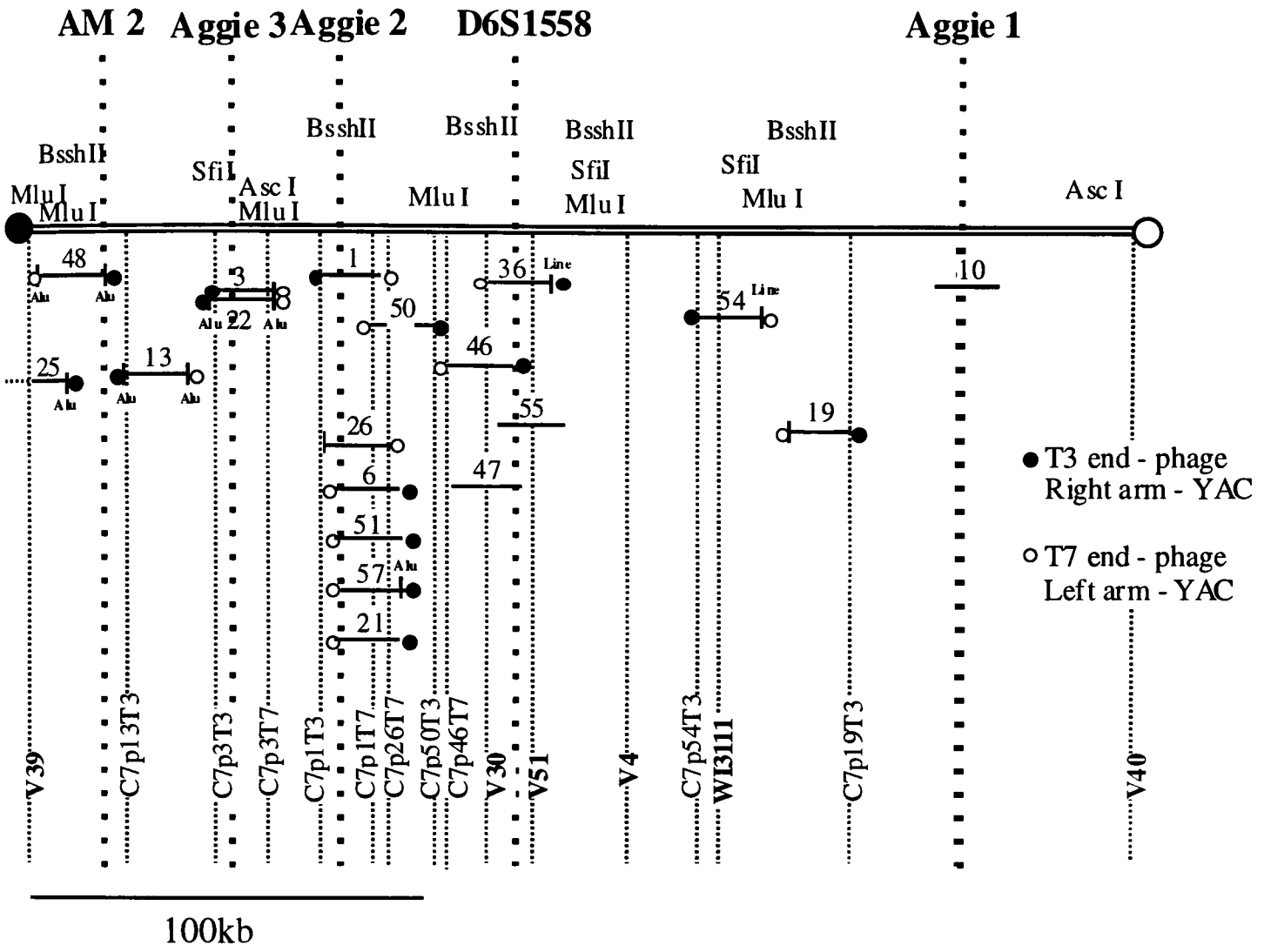


Figure 3.22 Composite restriction, phage clone and STS content map of B117C7.

An STS map of the D6S1260 to D6S1558 interval

An STS content map from D6S1260 to D6S1558 was constructed using STSs generated from sequence fragments obtained from YACs, cosmids and phage clones from the region. This facilitated the rapid assembly of a bacterial clone contig between the two points (Figure 3.23, page 172).

Figure 3.** The STS content map of the haemochromatosis critical region

The region between D6S1260 and D6S1558 was covered by 42 STSs (including 3 microsatellites, AM2, Aggie-3 and Aggie-2) and a contig of cosmids and a single phage clone was rapidly assembled between the two points (Alison Merryweather-Clarke).

BLAST analysis on sequence fragments across the region

BLAST analysis on the sequence fragments from which the STSs had been designed identified no ESTs with significant homology.

Interpretation

The need to convert the YAC contig into a bacterial clone contig is partly a reflection of the low copy number problem inherent in YACs and the difficulties in obtaining "clean" human DNA free of host cell contamination. Others have expressed reservations to simple sub-cloning of YAC DNA on the basis that sequence variations either in the original genomic DNA or acquired in the successive rounds of reproduction might be carried over and confuse interpretation of the final raw sequence data.

We constructed a high resolution STS map across the D6S1260 to D6S1558 interval. This facilitated the rapid assembly of a bacterial clone contig and furthermore would provide a "sequence tag frame" within which fragments of sequence from direct genomic sequencing could be orientated. Preparations were made to start exon trapping and efforts were made to have the interval sequenced as soon as possible.

Chapter 4

Verifying the discovery of the ancestral haemochromatosis mutation.

Contents - Chapter 4

Chapter 4.5

Pages 208-219

Haemochromatosis and heterogeneity

Chapter 4.1

HLA-H

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

Nature Genetics 13(4):399-409, August 1996

Background

Whilst preparations were being made for our haemochromatosis candidate interval to be sequenced and other groups worldwide were beginning to reproduce the data placing the gene telomeric to D6S1260, the August 1996 issue of Nature Genetics carried a report of a haemochromatosis candidate gene from a commercial biotechnology company, Mercator Genetics Inc. (Feder *et al.* 1996). The methodology adopted by Mercator was essentially the same as ours but their results differ significantly. They report a very strong positional candidate.

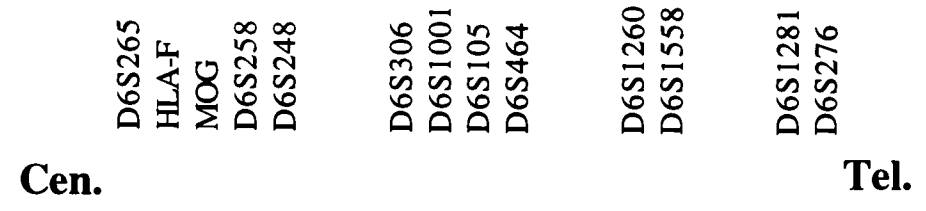
Physical map

Citing the UK results with D6S1260, Mercator Genetics extended their search for the haemochromatosis gene telomeric to D6S105. They had initiated a YAC contig by screening the CEPH YAC library with the markers D6S265; HLA-F; D6S258; D6S306; D6S105; D6S464 and D6S276 and like ourselves this contig was supplemented by the identification of CEPH YAC clones containing these markers from the CEPH/Généthon and Whitehead/ MIT databases. The contig was validated by the generation of STSs generated from the YAC ends by vectorette sequencing and holes in the contig filled be re-screening the CEPH library. With greater resources than ourselves the telomeric end of the contig was complimented by a walk centromeric to the marker D6S276, which lies centromeric to D6S461.

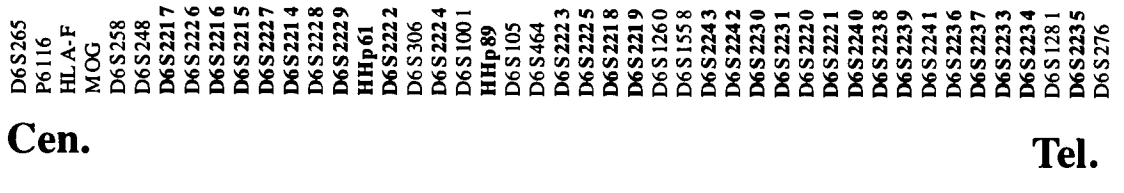
Their final YAC contig extended from HLA-F to beyond D6S276 covering 8Mb and comprised 50 YAC clones and 87 STSs. From this contig 14 clones that spanned the region were selected for RecA-assisted restriction endonuclease (RARE) cleavage mapping. This provided a framework for the physical overlap between clones from which physical distances could be derived. Parallel screening of large insert bacterial clone libraries created a contig covering the central 3Mb extending telomeric from D6S105.

Linkage Disequilibrium Data

As the YAC contig was evolving, Mercator used the physical data to order the 13 genetic markers known to map to the region.



This was supplemented by the isolation of a further 29 polymorphic microsatellites.



These markers (and three simple sequence polymorphisms) were studied in 101 haemochromatsosis homozygotes and 64 controls. Using p_{excess} and a method based on the observed deviation from Hardy-Weinberg equilibrium (F, proportionate increase in homogotes over expected), they defined a peak of linkage disequilibrium with haemochromatosis between two new markers, D6S2240 and D6S2233.

The premise that the haemochromatosis gene might lie in this region was substantiated by meticulous haplotype analyses on 40 haemochromatosis chromosomes separated into somatic cell hybrids (determining marker phase). A minimal region of 250kb between the markers D6S2238 and D6S2241 was defined as being identical by descent (IBD). This region lay within that defined by the observed excess of homozygosity and was considered as the candidate region likely to contain the haemochromatosis gene.

Genes in the region

Mercator Genetics' approach to the identification of genes within the region was as comprehensive and thorough as their mapping. First they used cDNA selection and exon trapping. This identified 3 genes - the 52kD ribonucleoprotein Ro/SSA, a sodium-phosphate transporter and a gene showing sequence homology to HLA-A2.

None of these immediately appeared to be a likely functional candidate gene for haemochromatosis. To ensure against having missed any gene sequences in the region the company sequenced the entire 250kb. In doing this they identified 12 histone genes in addition to the three genes described above.

Mutations

Having not identified a strong functional candidate for haemochromatosis, Mercator Genetics looked for sequence polymorphisms in all 15 of the genes from this region in haemochromatosis patients and controls. Eighteen sequence polymorphisms were identified, of which 15 were silent changes. The remaining 3 sequence polymorphisms were present in two genes. Two changes were present in a histone H1 gene and analysis of further control chromosomes revealed a high prevalence of these changes on control chromosomes arguing against these representing the ancestral haemochromatosis mutation.

Only one nucleotide change was consistent with the ancestral haemochromatosis mutation and that occurred within the MHC class-I like gene.

The HLA-H gene

The HLA class I-like gene was termed HLA-H. The use of the term HLA-H was predictable (the H presumably to reflect Haemochromatosis) but unfortunate as this name had already been given to an expressed pseudogene from the class I region (OMIM ref. 142925). International convention had been that the haemochromatosis gene be called HFE.

The HLA-H 2.7kb cDNA (GenBank accession number U60319) contains an open reading frame of 1029 bases encoding a predicted polypeptide of 343 amino acids. Database analysis identified significant sequence similarity with classical (HLA-A2) and non-classical class I genes (HLA-G) and the human Fc receptor.

Mercator's search for sequence variation identified a G to A transition at nucleotide 845 resulting in a cysteine to tyrosine substitution at position 282 of the predicted polypeptide (Cys282Tyr). By analogy with other class I gene products this cysteine is predicted to be involved in the formation of an intrachain disulphide bond

formation crucial to the secondary structure of the α3 domain of the protein. By inference, the mutation is predicted to disrupt the proteins interaction with, and stabilisation by, β_2-microglobulin.

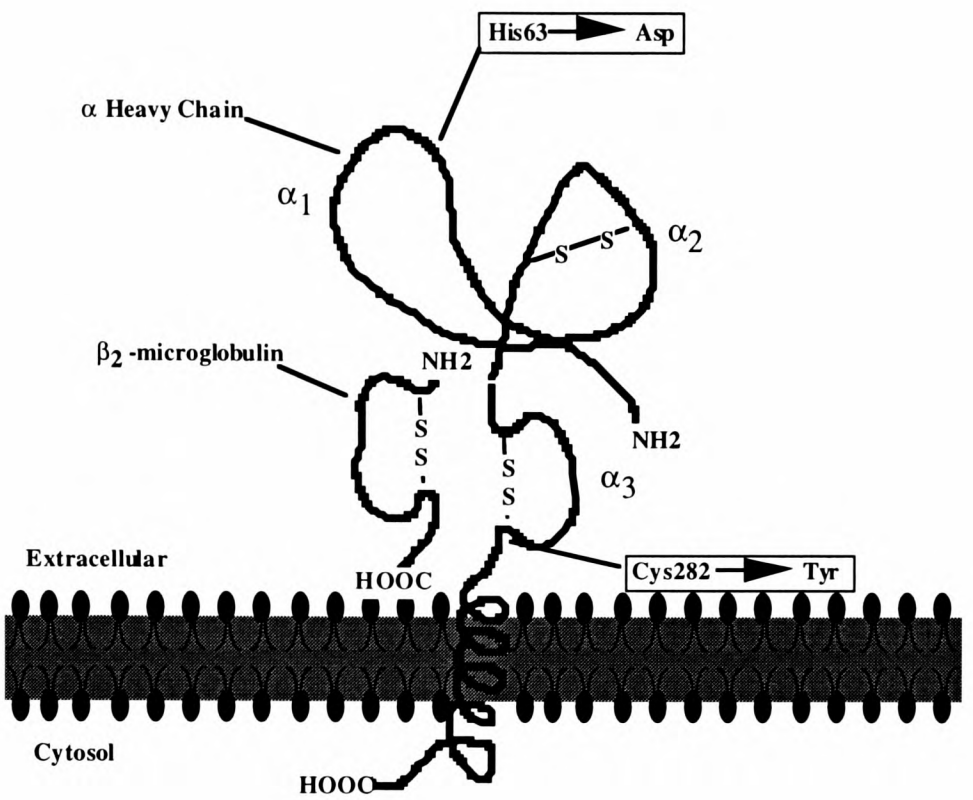


Figure 4.1 HLA-H (from Feder *et al* 1996). The cysteine to tyrosine substitution at position 282 of the polypeptide is predicted to disrupt the secondary structure of the α3 domain preventing the molecule from combining with β_2-microglobulin.

The evidence that the Cys282Tyr mutation represented the ancestral haemochromatosis mutation was that it was found on 85% of haemochromatosis chromosomes compared with only 3.2% of controls. In total 83% of 178 patients studied were homozygous for this mutation.

Of 9 patients found to be heterozygous for the Cys282Tyr mutation 8 were subsequently found to carry a second missense mutation, a C to G change in exon 2 that results in a histidine to aspartic acid substitution at position 63 in the putative "peptide" binding groove (His63Asp).

Although the HLA-H gene product has no obvious role in iron metabolism the prevalence of the common mutation in haemochromatosis chromosomes was interpreted as compelling evidence that this was the gene defect responsible for the condition.

Chapter 4.2

The physical maps of 6p21.3/22.1

Background

The candidate gene (HLA-H) reported by Mercator maps 4Mb telomeric to HLA-F and 1.75Mb telomeric to D6S105 (Feder *et al* 1996). On the basis of physical distance alone this would place HLA-H outside the gene region that we had defined and beyond the recombinational breakpoint at D6S1621.

Comparison of the physical map of Mercator with ours is vital to evaluate their claim based on the positional data alone (i.e. in the absence of a testable biological model of the disease).

Results

Although the maps are not easily comparable, as they contain different YAC clones and STSs, there is a very significant difference between the physical maps generated by Mercator Genetics and ourselves. The peak of linkage disequilibrium defined by their values for F and p_{excess}, and supported by their haplotype analysis, clearly lies between the two CEPH YACs 950h11 and 901a10, both of which we had been studying.

950h11 and 901a10

We had studied 950h11 following its identification on the CEPH database as containing D6S105 (Stone *et al* 1995). FISH analysis had shown a hybridisation signal from chromosome 5 (Lyndal Kearney) and an end sequence STS from its telomeric right arm (V42) did not amplify from 947f6. On this basis we had felt that 950h11 was probably chimaeric.

901a10 had been identified on the Whitehead Institute STS database. It did not appear to be chimaeric by FISH and its left (centromeric) arm was mapped within 1Mb of D6S105 by PCR on the fragmented clones from 947f6.

Mercator placed the left (centromeric) arm end of 901a10 2.5Mb from D6S105. This in turn raised doubts about the integrity of one of the clones anchoring our map - 947f6.

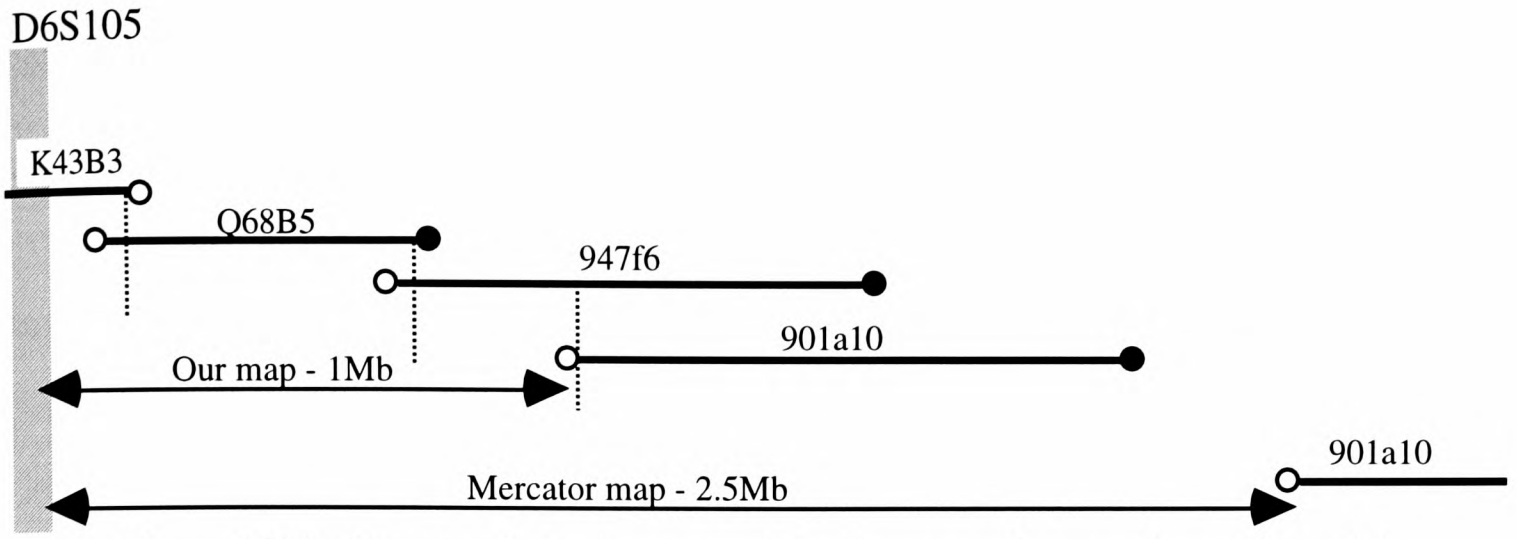


Figure 4.2 The Mercator physical map revealed a discrepancy between the positions of the CEPH YAC clone 901a10.

947f6

The CEPH clone 947f6 had been identified from the Whitehead Institute database. The most recent release of the STS map of the human genome (May 10th 1996) made it the only clone bridging contigs centred on D6S105 and D6S276.

YAC	Size kb	D6S 306	D6S 105	D6S 464	D6S 1078	D6S 1422	STSG 9945	D6S 1558	WI 3111	D6S 1016	WI 3878	D6S 1281	D6S 1554	D6S 1545
871g6	1190	+	+	+										
753h12	1770	+	+	+	+	+	+							
950h11	1760	+	+	+			+	+	+					
947f6	840								+		+	+		
901a10	1130										+	+	+	+
743f5	1650									+				
905h1	1270							?		+				
727b2	1380						+	+	+					
908d11	1170						+	+	+					
912c11*	540						+	+						
766a3	1320								+					
912c4	1440							+						
771g5*	1360										+	+		+
935d10	1720										+	+	+	
732b9	1110										+	+		
741c9	700										+	+		
792g12	610										+	+		
711g3	610											+		

Table 4.1 CEPH YAC clones extending telomeric to D6S105 (Whitehead Institue STS-based map of the human genome). The clones marked * have two sizes quoted on the database suggesting the possibility that these have resulted from co-transformation.

This single clone, 947f6, has an insert size of 840kb. FISH analysis had shown this clone to cross-hybridise with chromosome 5, but this was after the recognition that this feature was shared by many of the YACs from this immediate area (Chapter

2.3, page 78). STSs generated from the YACs two ends mapped back to clones within the contig (V51 onto Q68B5 and B117C7; and V52 onto 901a10 Jenny Pointon).

The only STS not to amplify on 947f6 was D6S1016. As discussed previously we had significant reservations about this marker and suspected that this might be duplicated on another chromosome (Chapter 2.3, page 77). 947f6 was fragmented (Jenny Pointon) and used for high resolution STS content mapping. On this basis D6S1281 was placed within 1.1Mb of D6S105. With D6S1281 showing a significantly lower disease association than D6S1621 this was considered as supportive evidence to the recombinational data that the disease gene lay centromeric to D6S1261, within 1Mb of D6S105 (Chapter 2.6).

The physical map generated by Mercator Genetics is quite different. The position of 950h11 on the Mercator contig is consistent with our experience but 901a10 (and hence D6S1281) lies at least 1Mb further telomeric than would be suggested from our map. 947f6 was not included in their contig. The only possible explanation for this observation is that 947f6 lacks 1Mb of sequence from its middle (i.e. an internal deletion).

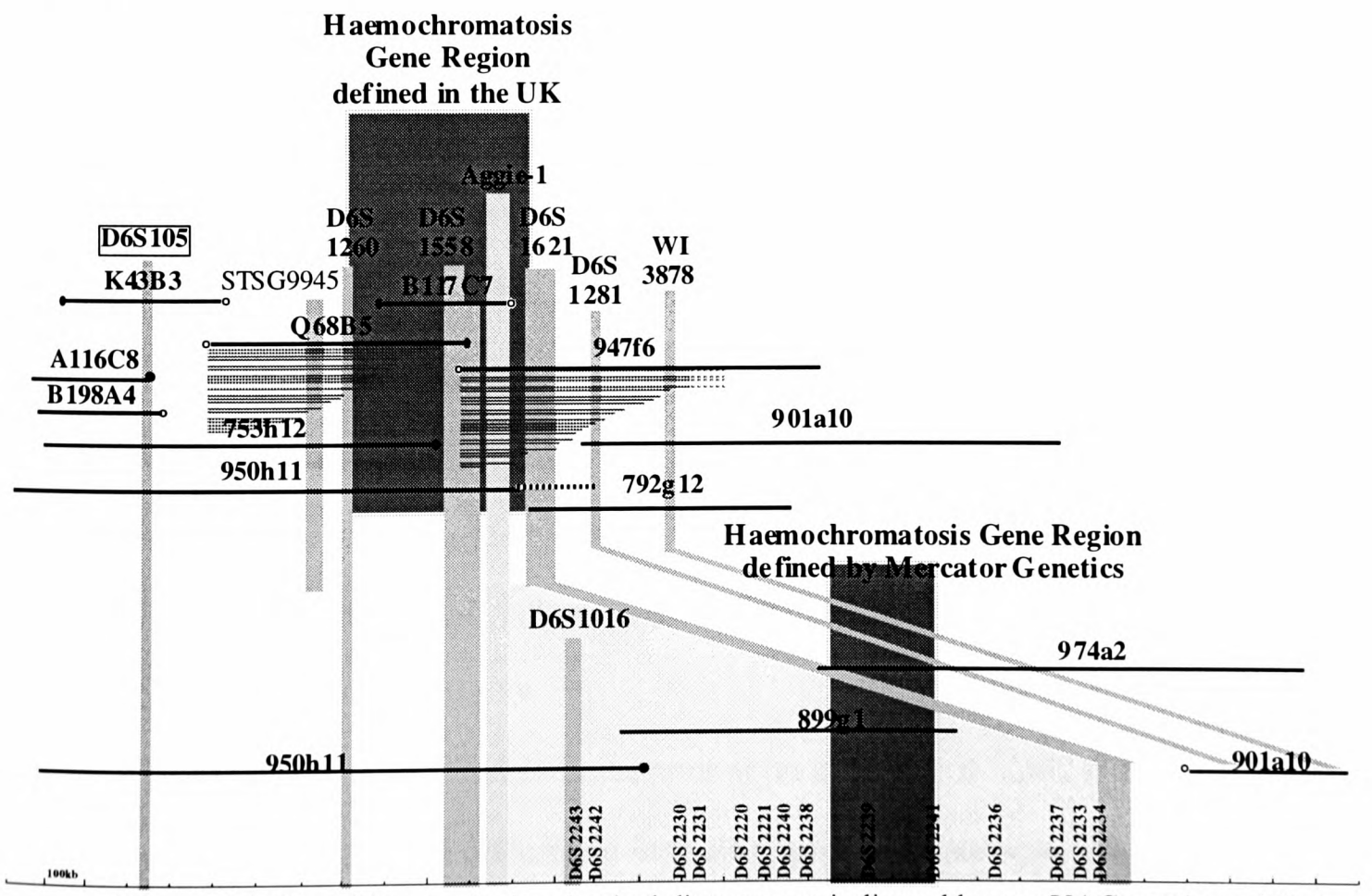


Figure 4.3 The difference in the physical distances as indicated by our YAC contig and that of Mercator could be explained by a deletion of over 1Mb in the CEPH YAC clone 947f6.

At the telomeric end of our YAC contig, an interval of 50-100kb between the markers V40 and D6S1621 "contained" a deletion of over a megabase.

Analysis of the primer sequences of the new markers deposited in the Genome Database (GDB) revealed that the Mercator microsatellite primers for D6S2234 amplify the same CA repeat locus as the Généthon marker D6S1621. This supports the recombination data defining D6S1621 as a telomeric boundary to the gene region (Chapter 2.6, page 107). From the physical map of Mercator (and accepting that 947f6 harboured a large deletion) our "peak" of linkage disequilibrium ($\lambda = 0.72$ for D6S1260 and D6S1558), 1Mb centromeric to the region defined by Mercator is entirely compatible with their claim.

Our identified need to isolate new microsatellite markers telomeric to D6S1260 had been grossly underestimated by the perceived physical distance between D6S105 and D6S1621.

Interpretation

Could we have done other than rely on a single YAC in this region?

Members of the group had been trying to extend the YAC contig telomeric to D6S105 for nearly two years (Jenny Pointon). Despite screening two separate YAC libraries, extending the contig telomeric to the end of Q68B5 had proven impossible due to the consistent finding of YAC clones showing cross-hybridisation with other chromosomes (summarised in Chapter 2.3, page 70). We had identified the need to map more telomeric and made use of all available resources including the publicly accessible YAC databases.

Two aspects of our strategy require critical review. Characterisation of YAC clones by FISH and the use of public access databases.

FISH

The Mercator report makes no comment on analysis of YAC clones by FISH. This represents an important difference in their approach from ours. We considered FISH to be the method of choice for the detection of clone chimaerism. Some form of

chimaerism detection has been considered essential for sequential chromosome walking to prevent the inadvertant walk onto other chromosomes and to maximise efficiency. However this assumes that one is undertaking a sequential walk.

Since this project began the mapping information available through public databases has increased enormously. This is not only true of genomic clones but also STSs. At the outset we aimed to physically link up two points (HLA-F and D6S105) separated by a physical distance of 2.5Mb between which there were no significant STSs. That interval now contains at least 4 microsatellites and two gene loci. This increase in the number of physically mapped STSs makes sequential chromosome walking over large physical distances obsolete. In principle, mapping has switched from the horizontal (clone contigs) to the vertical (STS content mapping). Mapping in this way means that FISH analysis on every clone is no longer necessary. Chimaeric (or "non-consistent") clones can be identified largely on the basis of high density STS content analysis alone.

With their greater resources for library screening, it would appear that Mercator were able to greatly increase the depth of coverage of YAC clones in the region extending telomeric to D6S1260, presumably by re-screening the CEPH library with all new markers (end sequence STSs and new microsatellites). YAC clones with chimaeric inserts and markers not mapping to 6p would be identified by inconsistencies in the STS content analysis and not studied further.

Ironically, if we had not had access to high quality FISH analysis for each of our YAC clones we would not have have identified the region showing cross-hybridisation with chromsome 5 and probably would have walked straight through both it and the "hole". A deletion is the one form of YAC rearrangement that will not be identified by FISH.

The Whitehead Institute Database

Mercator state in their paper that the region immediately surrounding the peak of linkage disequilibrium (i.e. containing the gene) is "not represented on the public databases" citing the Whitehead Institute (Feder *et al* 1996). The development of

publicly accessible databases such as that of the Whitehead Institute has provided a potentially powerful framework for mapping. However it must be remembered that the goal of the Whitehead Institute had been the construction of an STS map of the genome and that the YAC data contained therein is supportive rather than the primary objective (vertical v. horizontal mapping - see discussion above).

We appreciated that data from any database should be verified first-hand. We analysed all clones identified on the Whitehead database both by FISH and by STS content analysis both with our own vectorette STSs and microsatellites. Ideally (given sufficient resources), this could have been extended to re-screening the YAC libraries (CEPH and others) with all STSs appearing on the database. This is how Mercator might have avoided the deletion.

Ultimately the deletion in 947f6 would have been identified by the extension of a bacterial clone contig out to include D6S1621. This might have taken some time.

Other maps of the haemochromatosis gene region

Until Mercator's publication there were very few other sources of physical mapping data from this region. Burt had published a "physical map of the haemochromatosis gene region" (Burt *et al.* 1996). Despite citing evidence that the haemochromatosis gene mapped telomeric to D6S105, their physical map only extended as far as D6S1260 and the data contained within was little more than STS content analysis of CEPH YAC clones such as can be found on the more comprehensive computer databases. As either a physical map or a haemochromatosis gene region this report was incomplete.

More recently Bray-Ward has reported an analysis of FISH-mapped CEPH YAC clones in the region telomeric to the MHC (Bray-Ward *et al.* 1996). Drawing heavily on data from the Whitehead Institute database they too include 947f6 as the only clone bridging the interval between D6S1558 and D6S1621. They report 947f6 to be chimaeric by FISH (presumably seeing the "cross-hybridisation" signal from chromosome 5) but not by STS content analysis.

At the Third International Workshop on the Molecular Genetics of Haemochromatosis held in Rennes in June 1996 other maps of this region were presented more informally. Volz (Berlin, Germany) showed three PAC contigs. One of these was centred around MOG, the second at D6S105 and the last on a histone gene cluster in 6p22.2 which could represent the histone genes identified by Mercator Genetics. Camaschella (Torino, Italy) reported a YAC contig of 8Mb extending across the entire 6p22 band. Although she reported the mapping of 11 known genes and 5 ESTs this group reported no markers other than those in the public database. David (Rennes, France) reported a 4Mb YAC contig extending from HLA-F.

Many of these groups had already committed themselves to methods of gene identification before establishing an accurate gene position. Each of the groups mentioned above (Australia, France, Italy, USA) presented the identification of many cDNAs from the region around, and immediately telomeric to, D6S105. Each had based their gene location on data from a fraction of the number of markers that it had taken Mercator to identify HLA-H. Part of this eagerness to identify candidate genes was due to the extended region of linkage disequilibrium resulting in very high allelic association from markers that are now recognised to be separated form HLA-H by megabases of DNA. Part of the problem was due to genuine difficulties in mapping. Ultimately, some groups had tried to bypass the basic principle of positional cloning - the chance of finding a particular gene increases with more accurate mapping data.

The realisation of the deletion within the YAC clone 947f6 was a huge disappointment to our group. We had defined a region of 250kb on the basis of values of λ that would have been more than high enough to predict the location of any one other single gene disorder (see worked examples in Terwilliger 1995). The fact that these loci lay nearly 1Mb from the position of the best positional candidate yet published is remarkable. The need to identify more microsatellites was underestimated by the perceived physical distance between these markers and our telomeric limit for the haemochromatosis gene region. However, we had been trying to achieve the correct primary objective - the definition of a candidate region with

clear centromeric and telomeric boundaries. In that respect, we were as close (and had been more efficient) than any of the other non-commercial groups in the field.

Chapter 4.3

Analysis of the Cys282Tyr mutation in UK patients with haemochromatosis

Background

The MHC class I-like gene reported by Mercator as harbouring the haemochromatosis mutation provides no immediate explanation of the pathology of the condition. One reservation to accepting the mutations described is the possibility that these sequence variants represent simple sequence polymorphisms in linkage disequilibrium with the disease causing mutation, that might lie in a separate gene. The effect of linkage disequilibrium is likely to be more marked in a new (recently founded) population such as that in North America where there has been less time for recombination to erode the (secondary) founder haplotype of flanking markers.

The prevalence of HLA-H mutations in haemochromatosis patients from the UK would significantly enhance our understanding of the relationship between these sequence variants and the disease.

Methods

Patients

All haemochromatosis patients attending for venesection treatment in the departments of Gastroenterology and Haematology at John Radcliffe Hospital were reviewed. In each case the diagnosis was confirmed to meet the following criteria.

1. An hepatic iron index (HII) > 2 determined by dry weight iron estimation from liver biopsy tissue or by magnetic resonance imaging (MRI).

or 2. A liver biopsy reported to show grade III iron deposition (Scheuer grading)

and 3. Greater than 5g of iron removed by intensive phlebotomy (i.e. more than 20 weekly venesection of 500ml before the development of iron deficiency)

Whole blood was collected into EDTA form each patient and DNA was prepared by phenol chloroform extraction.

A larger study included these patients along with those under the care of clinicians at the other centres comprising the UK Haemochromatosis Consortium

(King's College Hospital and Royal Free Hospitals, London and The University Hospital of Wales, Cardiff). Healthy blood donors from South Wales were used as control subjects for this study (Mark Worwood, Cardiff).

PCR based restriction fragment length polymorphism

Cys282Tyr

The proposed ancestral haemochromatosis mutation in HLA-H is a G to A transition resulting in the proposed cysteine to tyrosine substitution at position 282 of the polypeptide. This nucleotide change creates a new recognition site for the restriction enzymes *Rsa* I and *Sex* AI.

Figure 4.4 Part of the cDNA sequence of HLA-H (accession U60319) demonstrating the new *Rsa I* and *Sex AI* site created by the G to A transition. The primers used to amplify this region, as reported by Feder *et al* 1996 are both intronic and not present in the cDNA sequence.

The PCR primers described by Mercator to amplify this region of the HLA-H gene are intronic and not identifiable within the cDNA sequence deposited in GenBank (accession number U60319). However, amplification of genomic DNA with these primers generates a product which after digestion with *Rsa* I clearly demonstrates two restriction sites in affected individuals and only the one site in normal control samples. Cys282Tyr fragment sizes 203bp, 111bp and 29bp after *Rsa* I digestion, normal allele fragments 203bp and 140bp.

The enzyme *Rsa* I was used in preference to *Sex* AI as the second *Rsa* I site within the amplified product acted as an internal control for the digestion.

His63Asp

The second "mutation" described by Mercator was of less clear genetic or pathological significance in that the amino acid substitution is present in the "binding groove" of the molecule and only appeared to contribute to the phenotype in compound heterozygotes with the Cys282Tyr mutation. The C to G transition causes the loss of *Bcl* I, *Mbo* I and *Sau* 3AI sites.

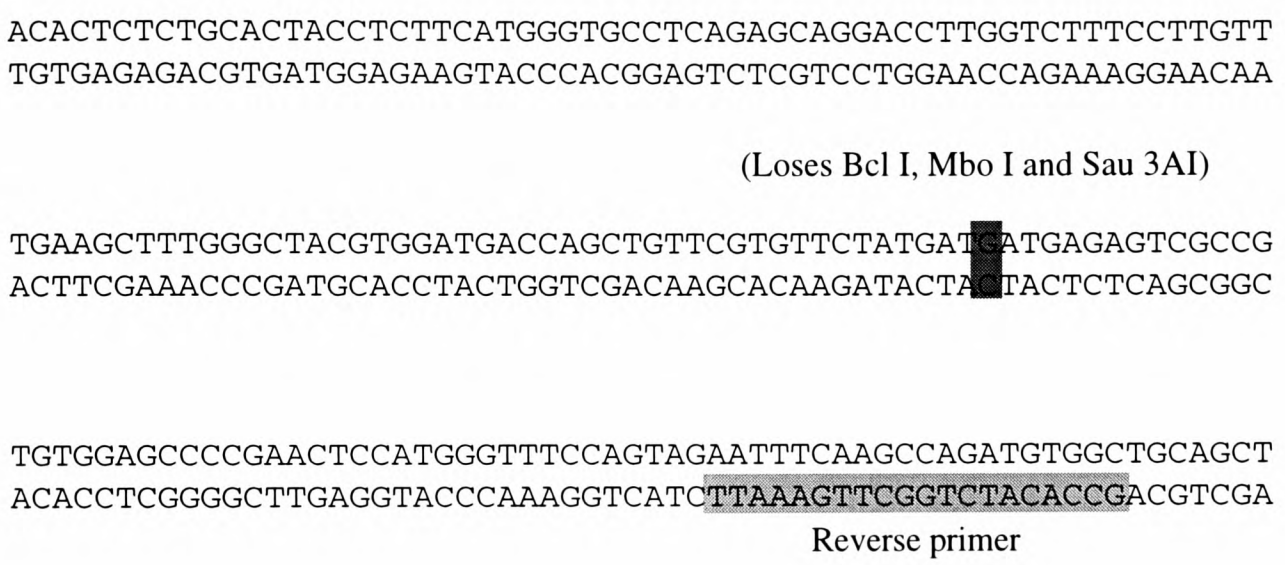


Figure 4.5 Part of the cDNA sequence of HLA-H (accession number U60319). The C to G transition illustrated results in the loss of restriction sites for *Bcl* I and *Sau* 3AI. . The reverse primer to amplify this site is indicated. The forward primer is intronic.

Genotyping for this mutation was done by a PCR based RFLP as for the common mutation.

Results

Oxford patients

Twenty-three patients with an unequivocal diagnosis of haemochromatosis and who attend the John Radcliffe Hospital for venesection treatment were studied for the mutations reported by Mercator. Twenty-two of the 23 were homozygous for the Cys282Tyr mutation. None had the His63Asp variant.

Name	Sex	Age at diagnosis	Haplotype 265, 105, 1260, 1621	Cys282Tyr
N.B.	M	28	**1,1; 8,8; 4,4; 5,5**	+/+
P.G.	M	29	4,5; 6,**8; 4,**5; **5,**10	+/+
A.B.	M	29	3,5; 6,12; **4**,11; **5,5**	+/+
G.H.	M	30	**1,1; 8,8; 4,4; 5,5**	+/+
P.W.	M	35	**1,1; 8,8; 4,4;** 7,10	+/+
A.A.	M	43	**1,1; 8,8; 4,4; 5,5**	+/+
J.S.	M	45	**1**,5; **8**,5; **4,**7; **5,5**	+/+
P.B.	M	47	**1**,3; **8,8; 4**,3; **5,5**	+/+
G.M.	M	50	3,4; **8**,7; **4,4; 5,5**	+/+
B.G.	M		**1,1; 8,8; 4,4; 5**,3	+/+
J.W.	M	54	**1**,4; **8,8; 4**,5; **5,5**	+/+
M.O'K	M	58	5,6; 4,4; **4**,3; **5,5**	+/+
J.Lo.	M	60	**1,**6; **8,8; 4,4; 5,5**	+/+
E.W.	M	60	3,6; 6,7; **4**,8; **5**,8	-/-
J.Ly.	M	63	**1,1; 8,8; 4,4; 5,5**	+/+
B.F.	M	65	**1,1; 8,8; 4,4; 5,5**	+/+
R.S.	M	69	**1,**5; 5,**8; 4,4; 5,5**	+/+
L.W.	F	24	**1,1; 8,8; 4,4; 5,5**	+/+
B.C.	F	28	**1,1; 8,8; 4,4; 5,5**	+/+
A.H.	F	33	**1,1; 8,8; 4,4; 5,5**	+/+
J.B.	F	50	**1,**3; **8,8**; **4**,3; **5,5**	+/+
A.M.	F	57	**1,**3; **8,8; 4,4; 5,**10	+/+
A.S.	F	57	**1,**4; **8,8; 4,4;** 3,**5**	+/+

Table 4.2 Genotyping for the Cys282Tyr mutation in patients with haemochromatosis. The haplotype data refers to alleles at the loci D6S265; D6S105; D6S1260 and D6S1621. The alleles shown in bold are those associated with haemochromatosis and present on the ancestral haplotype as defined in the UK. For further clinical details on the patients see Appendix 3.

A single patient without the Cys282Tyr mutation

A single male patient (E.W.) did not have the Cys282Tyr mutation. The patient was reviewed in the Out-patient department and the clinical history, physical findings and investigations were reviewed.

Mr E.W. is a 60 year old retired builder. He had had a liver biopsy performed in January 1996 for persistently abnormal iron indices that had first been recognised during an episode of flucloxacillin related jaundice in 1995. This showed excess hepatocellular iron and minor portal tract changes in keeping with his recent cholestatic drug reaction. Hepatic iron concentration was estimated by MRI from which his hepatic iron index was calculated to be 5.4. He had been treated by weekly venesection. The only significant finding on clinical examination was swelling of the 2nd and 3rd metocarpo-phalangeal joints, characteristic of haemochromatosis arthropathy.

Although there was no family history to suggest iron overload in his parents, Mr E.W. had an interest in the history of his family and had traced his father's ancestors back to the 17th century. All had come from around Abingdon, Oxfordshire and many had been buried in a parish church in Long Wittenham.

On clinical grounds Mr E.W. has "classic" haemochromatosis.

UK patients and blood donors - Cys282Tyr

Combining the results of the Oxford patients with those from Cardiff, King's College and Royal Free Hospitals, London reveals an remarkable frequency of homozygosity for the common Cys282Tyr mutation in UK patients with haemochromatosis.

Cys282Tyr	Number	+/+	+/-	-/-
Haemochromatosis	115	105 (**91%**)	4	6
Controls	101	1	10 (**10%**)	90

Table 4.3 The prevalence of the Cys282Tyr mutation in HLA-H in UK patients with haemochromatosis.

Over 90% of haemochromatosis patients in this study were homozygous for the Cys282Tyr mutation. Ten percent of the control population were heterozygotes.

The one control subject found to be homozygous for the Cys282Tyr mutation was subsequently found to have an elevated transferrin saturation and is currently being investigated (Mark Worwood).

Interpretation

For linkage disequilibrium to be informative in determining the position of a disease gene sufficient time must have elapsed to allow this effect to have dissipated at all but the most closely linked markers. The caucasian population of the UK is believed to be ethnically closer to that in which the founder haemochromatosis mutation occurred (Simon *et al.* 1987) and, in population terms, is considerably more ancient than those of the New World (USA, Australia). We have found over 90% of haemochromatosis patients from the UK to be homozygous for the G to A substitution predicted to result in the cysteine mutation in the HLA-H gene product (Cys282Tyr). This is higher than that found in the USA (Beutler *et al.* 1996; Feder *et al.* 1996) and provides overwhelming support for the claim that Mercator Genetics have identified the ancestral haemochromatosis mutation. Furthermore, no one individual has yet been found to be homozygous for this mutation without evidence of iron loading. In the control population heterozygosity for the proposed ancestral mutation is 10% comparing favourably with the predicted carrier rate as derived from disease prevalence figures defined by measurement of iron indices.

Further HLA-H mutations in affected individuals not carrying Cys282Tyr ought to be sought.

Chapter 4.4

HLA-H and haemochromatosis.

Background

On the basis of mapping and mutational analysis, HLA-H carries the ancestral haemochromatosis mutation (Cys282Tyr). A rational explanation of the disease phenotype on the basis of loss of an MHC class I-like product is not apparent. Before considering Mercator's proposed role for HLA-H, the previous evidence supporting the possibility of an aberrant MHC product in haemochromatosis will be reviewed in more detail.

Haemochromatosis, the gastrointestinal tract and MHC gene products

The gastrointestinal tract provides an epithelial barrier to pathogens and toxins in the environment. The classical role of immunological defence in the gut is confined to the specialised mucosa-associated lymphoid tissues (MALT) and the diffuse lymphocyte populations (lamina propria lymphocytes and intraepithelial lymphocytes or IELs). Although the enterocyte constitutes the absorptive lining of the gut there is a precedant for it expressing tissue-specific MHC gene products.

HLA-like molecules

Immunoglobulin (IgG) present in breast milk is absorbed from the neonatal gut after binding to a specific neonatal Fc receptor (FcRn) contributing passive, humoral immunity before the development of a mature immune system (Simister and Mostov 1989). This receptor has strong structural homology with MHC class I genes (and HLA-H). The greatest region of sequence conservation is in the $\alpha 3$ domain. The putative binding groove of FcRn appears to have distinctly less sequence variability than the classical peptide-presenting class I molecules in keeping with its proposed role as a ligand specific receptor. The neonatal Fc receptor represents the tissue specific expression of a class I related gene with restricted function, quite different to that of the classical class I genes.

In mouse, other inherently less polymorphic class I-like genes have been identified at the telomeric end of the H2 complex (Qa, Tla and Hmt). The thymus leukaemia antigen (Tla) is expressed by gut epithelium (Hershberg *et al.* 1990; Wu *et*

al. 1991) and has been proposed to play a role in the presentation of antigens or specific "stress proteins" to the γδ intraepithelial lymphocyte (IEL) population of the mucosa.

A human thymus-leukaemia antigen has in fact already been hypothesised to represent the haemochromatosis gene product (Dorak *et al.* 1994).

> 1. Haemochromatosis has a strong association with HLA-A3 and results in increased intestinal iron absorption,
> 2. The HLA-A3 homozygous genotype is associated with a high relative risk of both haemochromatosis and early-onset leukaemia,
> 3. A thymus-leukaemia antigen-like molecule, TCA, is expressed in leukaemia and recognised by monoclonal antibodies to transferrin receptor,
> 4. In mice, particular thymus-leukaemia antigens are expressed in the intestinal mucosa
>
> (Dorak et al 1994).

Although these observations were indirect, this hypothesis may have now been confirmed.

β_2-microglobulin

One feature common to all class I related products (both classical and non-classical) is their non-covalent association with β_2-microglobulin. In 1994 de Souza reported progressive hepatic iron loading in mice deficient in β_2-microglobulin (de Souza *et al.* 1994). Selective disruption of the β_2-microglobulin gene results in a specific deficiency in the protein which leads to a consequent loss of MHC class I proteins and $CD8^+$ T lymphocytes (Zijlstra *et al.* 1990). Although these animals develop normally (Koller *et al.* 1990), they progressively accumulate tissue iron with a preferential deposition of iron in periportal hepatocytes with almost complete sparing of reticuloendothelial sites such as the spleen (de Souza *et al.* 1994). In the absence of a detectable abnormality in MHC class I antigen expression in haemochromatosis, de Souza concluded that the aberration of iron metabolism in these animals was consequent on the marked reduction in CD8+ T-lymphocytes populations.

Two years later a second report of iron metabolism in β_2-microglobulin-deficient mice attempted to formulate a specific role for class I genes in iron

metabolism (Rothenberg and Voland 1996). These authors had examined 5' regulatory sequences in non-classical class I genes of the mouse and identified a regulatory motif shared by β globin (βGAP). Assuming globin genes to be transcriptionally regulated by iron they made the claim that these elements must also reflect iron regulated transcription of the non-classical class I genes and these in turn must control intestinal iron absorption. Despite citing de Souza's report from two years earlier, Rothenberg and Voland claim to have **predicted** on the basis of their observations that mice deficient in β_2-microglobulin would become iron loaded. Twenty β_2-microglobulin (-/-) mice were studied and found to have increased hepatic parenchymal iron compared with control animals and heterozygous littermates. Nearly 50% had evidence of piecemeal hepatic necrosis and four mice developed primary liver cell tumours. Increased gastrointestinal iron absorption was demonstrated. Rothenberg and Voland claimed that the β_2-microglobulin deficient animals were an attractive model for the study of iron overload diseases and they postulated a role for non-classical class I molecules expressed on the luminal surface of the gut epithelium as a receptor for iron bound to a new class of iron ligand, the product of the HEPH or Hephaestus genes (Rothenberg and Voland 1996). Although it is less than clear from their results and reasoning, Rothenberg and Voland claim that their model explains i) why haemochromatosis maps to the MHC in man, ii) why the condition is so prevalent and iii) why the phenotype in man is so mild.

The only conclusion that can unequivocally be drawn from the observations on β_2-microglobulin deficient mice is that β_2-microglobulin-associated proteins are involved in some aspect of iron metabolism. With regard to comparison to the haemochromatosis phenotype, more information would be necessary before accepting these animals as a true model of the condition (notably transferrin saturations and basic haematological indices).

Proposed role for HLA-H

Although indirect evidence had pointed to the possibilty of MHC genes influencing iron metabolism, the identification of HLA-H as the gene underlying the commonest form of primary iron overload means that a more exact explanation must now be sought. Mercator Genetics accept that the primary defect in intestinal iron absorption is *probably* the transfer of mucosal iron from the basolateral surface of the enterocyte to the plasma and propose three possible mechanisms through which an aberrant MHC class I-like protein could contribute to iron loading.

1. Receptor for an iron binding ligand.

It is proposed that the HLA-H product might act as receptor for an iron binding ligand (Feder *et al.* 1996). In that the common mutation (Cys282Tyr) is predicted to inactivate the gene product, such a receptor is postulated to "normally act in a negative fashion, limiting or controlling the iron absorption process" (Feder *et al.* 1996). Theories as to the nature of the ligand were not expanded on and these are dependant on whether the gene product is expressed at the luminal or basolateral surface of the enterocyte.

a. HLA-H on the luminal side of the cell

If HLA-H is expressed on the luminal surface of the enterocyte (akin to FcRn) then the most attractive associated ligand might be haem. Dietary haem iron is more bioavailable than iron from non-haem sources and control of haem iron absorption appears to be less responsive to body iron stores than that of non-haem iron in iron loaded patients with haemochromatosis (Lynch *et al.* 1989). Uptake of dietary haem iron by the enterocyte is via a specific haem receptor (Grasbeck *et al.* 1982) although the protein structure of this has not been established. Haem would represent a substrate larger than the peptides bound by the classical class I gene products although, by analogy with the Fc receptor, HLA-H could form a homo-dimer in order to accommodate this (Burmeister *et al.* 1994).

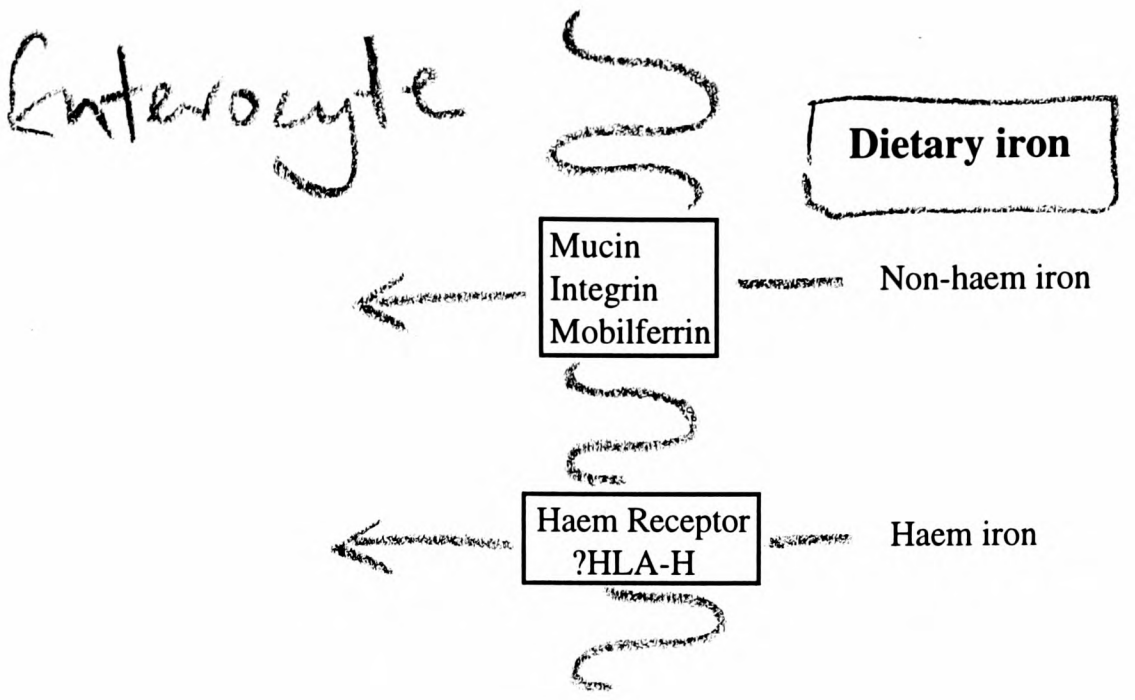


Figure 4.6 Hypothesis: HLA-H as a luminally expressed haem receptor.

There are considerable conceptual reservations to such a model. HLA-H is predicted to normally act in a negative way, which would therefore reduce haem iron absorption and the presence of another pathway for haem uptake would have to be considered. Furthermore, if the mutated HLA-H were to result in increased haem iron absorption this would not explain iron deficiency of the enterocyte in haemochromatosis.

b. HLA-H on the serosal side of the cell

If HLA-H is expressed on the basolateral surface of the enterocyte the theoretical mechanisms are similarly obscure. Transferrin is the recognised iron binding ligand in mammals and it has a well characterised cognate receptor (transferrin receptor, TfR).

The principle of transferrin saturation mediating control of iron absorption has been postulated previously (Bothwell *et al.* 1989) although the molecular mechanisms have never been determined. There is clinical evidence for this in that all conditions that result in low transferrin bound iron (i.e. iron deficiency, atransferrinaemia and acerulplasminaemia) result in increased intestinal iron absorption. The single clinical feature that distinguishes haemochromatosis from other forms of primary iron overload is that increased gastrointestinal iron absorption continues in the presence of

a **high** saturation of circulating transferrin levels. Enterocyte expression of transferrin receptor is upregulated in haemochromatosis (Pietrangelo *et al.* 1992) reflecting the cellular iron-deficient phenotype. For HLA-H to represent an transferrin-independant iron uptake pathway in certain cell types, the form of iron likely to be taken up is unclear.

Could HLA-H interact with the enterocyte transferrin receptor and alter its function or act as a second transferrin receptor? Control of iron uptake by the enterocyte might be postulated to differ from that of cells of a non-iron transporting nature if the relative saturation of circulating transferrin is critical to the regulation of serosal transfer of iron. Whereas most cells can down regulate transferrin receptor expression when intracellular iron stores are repleat, the enterocyte may have an physiological obligation to continue "sampling" circulating transferrin as a measure of body mobilizable iron status.

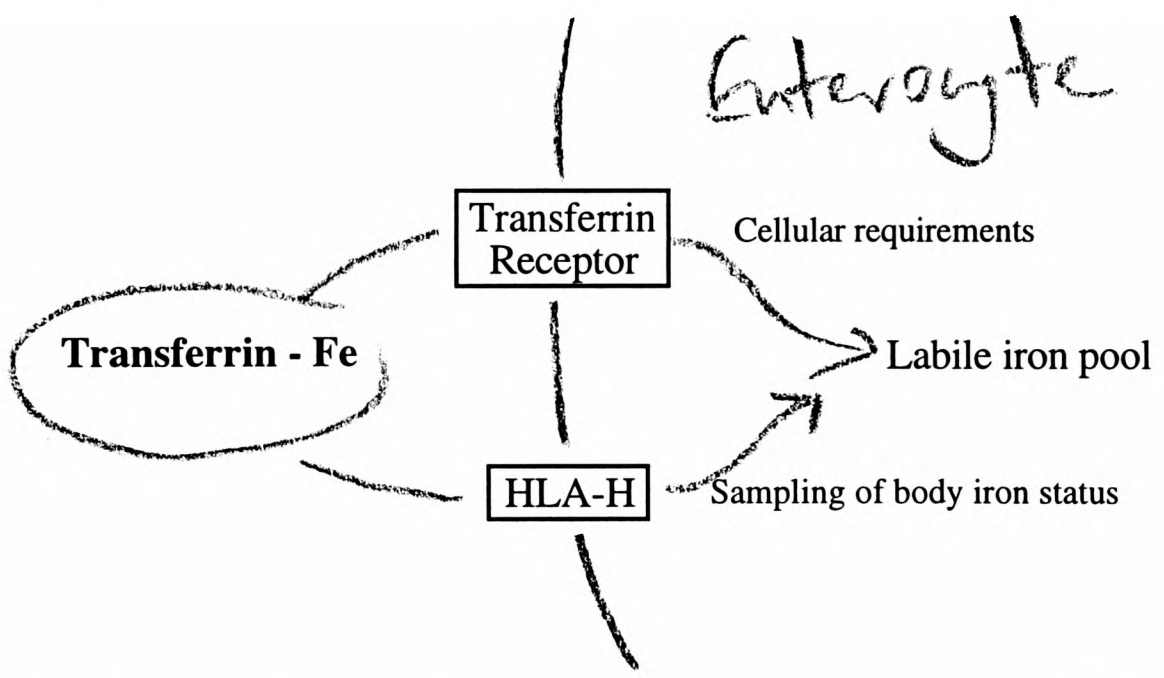


Figure 4.7 Hypothesis: HLA-H expressed on the serosal side of the cell, mediating the "sampling" of circulating transferrin.

2. A role in signal transduction

The second proposal was that HLA-H plays a role in signal transduction, acting as a sensor of circulating iron levels and regulating other genes or gene products that control iron transfer (Feder *et al.* 1996). This proposal is subject to the same conceptual considerations as the iron ligand theory in terms of where on the cell HLA-H is expressed. However, it accepts that HLA-H might not bind and internalise iron but rather signal to other effector molecules or genes. If these effectors control

iron release (by whatever mechanism), and that HLA-H normally acts in an inhibitory way, then its absence would result in increase serosal transfer of iron.

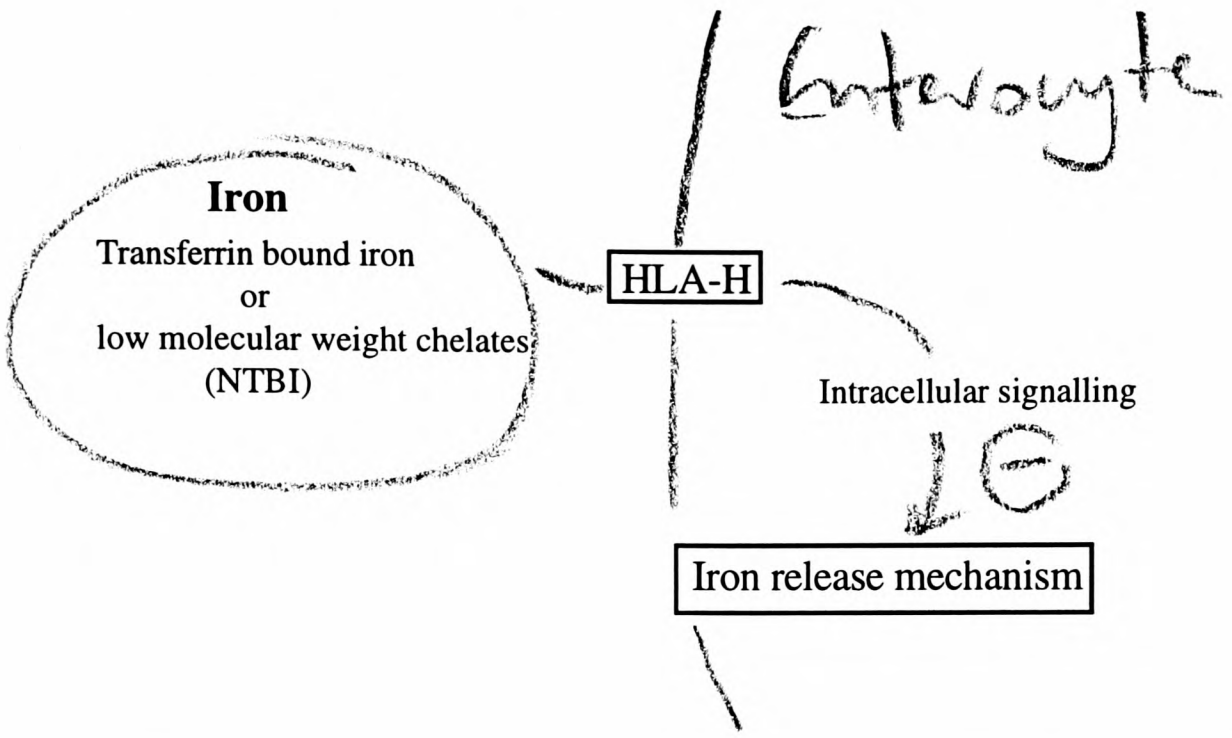


Figure 4.8 Hypothesis: HLA-H as an iron sensor resulting in intracellular signalling resulting in a reduction in serosal transfer of iron.

3. "Indirect role"

The final theory put forward by Mercator is that HLA-H plays an indirect role, interacting with components of the immune system which influence iron metabolism (Feder *et al.* 1996). This is the least explicit of the proposed mechanisms and the most difficult to support with other evidence.

Haemochromatosis, iron overload and the immune response

The effect of pertubations of iron status (both overload and deficiency) on susceptibility to infection is extremely complex (see de Souza 1989, Brock 1993). Iron overload is not associated with either an overt resistance or susceptibility to common bacterial infections, although there have been occasional reports of unusual systemic infections (*Pastuerella pseudotuberculosis, Yersinia enterocolitica*) in heavily iron loaded patients with haemochromatosis. The relationship between iron excess and infections with either viruses (e.g. Hepatitis B and Hepatitis C) or parasites (e.g. *Plasmodium falciparum*) is unclear.

Abnormalities in the cellular arm of the immune response relating to levels of iron loading in haemochromatosis have been reported (Bryan *et al.* 1991), but more importantly some appear to be unrelated to the levels of iron loading.

T lymphocytes

Porto and colleagues have previously reported a reduction in the relative proportion of $CD8^+$ to $CD4^+$ T lymphocytes in haemochromatosis (Porto *et al.* 1994). No control population was included in their study although they claimed that the reduction in $CD8^+$ correlated with severity of iron loading and that this abnormality remained unaltered after venesection treatment. Subsequent studies (presented in abstract only) have demonstrated other $CD8^+$ T cell phenotypic features that appear to correlate with disease severity but which are not corrected by iron removal (Porto *et al.* 1995):

1. diminished Vβ6.7 usage;
2. reduced % in $CD28^+$ and increased % of HLA-DR^+ within the $CD8^+$ pool;
3. reduction in level of autophosphorylation of the CD8-associated p561ck;
4. reduced CTL cytotoxicity

These observations supported de Souza's conviction that the T cell subset abnormalities were responsible for the iron accumulation in the β_2-microglobulin deficient mice (which also lack $CD8^+$ T cells de Souza *et al.* 1994).

How HLA-H might alter T cell subset numbers is not apparent.

Macrophages

Although the role of macrophages in iron metabolism and the pathogenesis of haemochromatosis has been covered elsewhere in this thesis, evidence exists that the immunological roles of these might also be intrinsically perturbed in haemochromatosis. Gordeuk demonstrated low levels of the cytokine, tumour necrosis factor α (TNFα) in the supernatants of cultured monocytes from patients with haemochromatosis (Gordeuk *et al.* 1992). This was not seen in cells from patients with iron loading anaemia, it persisted despite stimulation of the cells with lipopolysaccharide and was limited to TNFα (i.e. not seen for interleukin 1β).

There are many unexplained features of haemochromatosis that now require renewed study in the light of the identification of HLA-H.

Chapter 4.5

Haemochromatosis and heterogeneity.

Background

Until the necessary biological experiments are completed we are left without even a theoretical explanation as to the role of HLA-H in iron absorption and how the common mutation in HLA-H causes iron loading. Before then there remain important questions regarding the molecular genetics of haemochromatosis and HLA-H.

Genetic heterogeneity

"Haemochromatosis" unrelated to defects in HLA-H

In their study Mercator Genetics found that homozygosity for the Cys282Tyr mutation accounted for 83% of their 178 affected individuals. Only a small number of affected individuals (8) were compound heterozygotes with the second "mutation" (His63Asp). Twenty-one of their patients (12%) had no detectable mutation in the HLA-H gene, an observation that was said to reflect "heterogeneity" in their patient population.

The term heterogeneity in this context could have more than one interpretation. It could imply that their patient population contained individuals with iron loading from other conditions (e.g. porphyria cutanea tarda, alcoholic liver injury). This is unlikely if their diagnostic criteria were adhered to.

Despite the demography of the patient population not being reported, Mercator stated that they found no obvious clinical-genetic correlation to the underlying "heterogeneity", inferring the presence of patients with clinically indistinguishable haemochromatosis unrelated to the HLA-H gene on chromosome 6. In support of the existance of forms of the disease that appear to be unrelated to 6p they erroneously cited Simon's original description of the association with HLA-A3 and B14 (Simon *et al.* 1976) in which no such claim was made.

Evidence for the existance of more than one "haemochromatosis" gene

A precise definition of haemochromatosis is the common, recessively inherited, HLA-linked form of primary iron overload. Other forms of inherited primary iron overload exist worldwide. Most are clearly distinguishable from

haemochromatosis on the basis of the clinical phenotype (e.g. atransferrinaemia, acaeruloplasminaemia) and some occur in distinct ethnic/racial groups (e.g. the dietary iron overload of sub-saharan Africa or Bantu siderosis). An autosomal dominant form of familial iron overload, unlinked to HLA, has been reported in a large Melanesian kindred (Eason *et al.* 1990). These conditions should not be confused with haemochromatosis.

Even after the reported association of haemochromatosis with HLA-A3, the variability in the clinical presentation lead some to propose the existance of more than one gene causing this condition (Bomford *et al.* 1977). Simon refuted this hypothesis by demonstrating the lack of clinical expression of haemochromatosis in 247 individuals classed as being heterozygous for the condition on the basis of sharing a single HLA haplotype with an affected individual (Simon *et al.* 1980). He concluded that there was a single major HLA-linked gene to account for haemochromatosis.

Distinct clinical sub-groups of haemochromatosis

The extent of clinical heterogeneity in haemochromatosis as seen in North America was clearly outlined by Muir (Muir *et al.* 1984). He studied the phenotype of 174 individuals from 9 families and defined four different types of disease expression.

Group I - Classic homozygous haemochromatosis,
Group II - Severe iron overload and accelerated disease,
Group III - Elevated total body iron stores and normal serum values,
Group IV - Elevated serum values and minimal increase in total body iron stores.

Group II, as defined by Muir, consisted of two young women with severe iron overload - one dying at the age of 23 with intractable heart failure. These cases appear similar to other, infrequent, published case reports of young patients (usually male) presenting with very heavy iron loading in which cardiac failure dominates the clinical picture and progressive cardiomyopathy often proves rapidly fatal (Perkins *et al* 1965; Charlton *et al* 1967; Lamon *et al* 1979; Jensen *et al* 1993). Limited family data is available from such cases, although, as in the cases described by Muir, siblings

are often found to be iron loaded. Support for a distinct genetic mutation in cases such as these is the fact that of the small number of individuals identified in the UK **without** the Cys282Tyr mutation, one (M.B., King's College Hospital) was identified at an early age with severe iron loading following the death of a brother with cardiomyopathy. If sequencing of the HLA-H gene were to fail in identifying a distinct mutation, attempts to identify a separate genetic locus in patients such as these might be limited by the fact that these cases are rare.

Muir's Group III consisted of two individuals presenting with characteristic haemochromatosis arthropathy and with tissue iron loading but with normal transferrin saturation and ferritin. Neither was of North-West European extraction (one being an African-American and the other a Greek Orthodox bishop). The mechanism of iron loading in these cases is likely to be different to that in haemochromatosis where iron absorption persists in the presence of saturated circulating transferrin. The recognition of arthropathy in these patients is intriguing as this represents an aspect of haemochromatosis that is independant of iron loading.

Subjects with elevated tissue iron and serum ferritin but with normal transferrin saturation have been reported from Brittany (Moirand *et al.* 1995). In a series of 28 patients with this syndrome the prevalence of HLA-A3 was similar to the control population and iron loading was not seen in HLA-identical siblings inferring a genetic susceptibility locus other than HLA-H. With the diagnostic criteria adopted by Mercator Genetics, patients such as these could account for cases without HLA-H mutations.

The last of Muir's clinical sub-groups (Group IV) might reasonably be considered as being "early" or "mild" haemochromatosis in that the only difference between these and Group I (classic homozygous haemochromatosis) was the relatively lower amounts of storage iron as defined by quantitative phlebotomy. Muir's study was conducted before the hepatic iron index had been described. However, even if we assume that these cases had HLA-linked, primary iron overload,

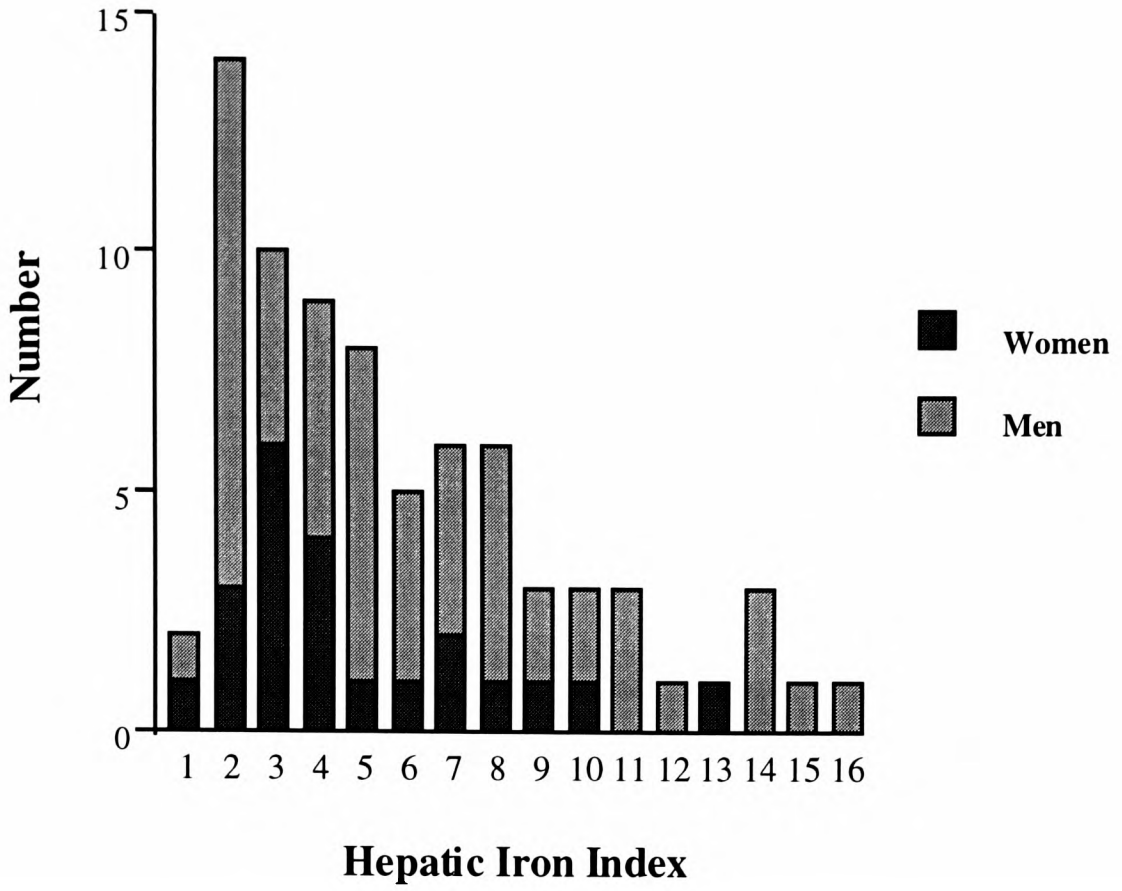
their recognition raises further questions concerning heterogeneity in haemochromatosis.

Phenotypic heterogeneity in a new light

Other than the few patients without an identifiable mutation in HLA-H, the high rate of homozygosity for Cys282Tyr is remarkable for a condition whose phenotype varies considerably. Iron loading in haemochromatosis occurs over many years and by the time sufficient iron has accumulated to cause tissue damage many affected individuals will have concurrent pathologies in affected organs (e.g. ischaemic heart disease, type II diabetes, osteoarthritis). Discussions on the phenotypic variability of haemochromatosis should be restricted to estimations of the **rate** of iron loading. Bassett's original report of the use of the hepatic iron index examined the correlation between hepatic iron concentration and the risk of cirrhosis in haemochromatosis homozygotes (Bassett *et al.* 1986) but the fundamental issue of the huge variation in the essential rate of iron loading as reflected by the hepatic iron index has not been formally addressed. Now that we know that a high proportion of haemochromatosis patients are homozygous for the ancestral Cys282Tyr mutation this variability might be re-evaluated.

Hepatic Iron Index - A measure of phenotype severity

In the original two reports of the measurement of hepatic iron index in haemochromatosis, values ranging from 2.4 to 15.6 were found (Bassett *et al.* 1986; Summers *et al.* 1990).



(Data from Bassett *et al* 1986 and Summers *et al* 1990)

Figure 4.9 Range of hepatic iron index (as reported in Bassett *et al* 1986 and Summers *et al* 1990)

Allowing for the fact that an HII of 2 is considered as diagnostic of haemochromatosis, the values of HII from this group of patients suggests a normal (Gaussian) distribution of severity. The absence of a bimodal distribution might argue against the effect of another dominant genetic factor determining disease severity but the fact that the inferred rate of iron loading amoungst these patients varies nearly 8 fold now suggests that factors other than simple mutation variants at the disease gene influence disease expression.

Three of the Oxford patients illustrate this variability in the severity of the haemochromatosis phenotype.

Name	Sex	Age at diagnosis	Hepatic Iron Index	Cys282Tyr
JS	M	45	2.8	+/+
PB	M	47	10.75	+/+
AH	F	33	10.6	+/+

Table 4.4 Variation in the rate of iron loading in haemochromatosis patients homozygous for the Cys282Tyr mutation.

JS and PB are both males who were diagnosed as having haemochromatosis in their mid-forties. PB was asymptomatic but identified after the diagnosis had been

made in his sister (JB). His hepatic iron index was nearly four times that of JS (10.75 v. 2.8) who was diagnosed following the investigation of malaise and abnormal liver blood tests. Mrs AH was diagnosed a decade earlier (aged 32) with an HII of 10.94 despite having had 4 term pregnancies (a major drain on iron stores). All of these individuals are homozygous for the Cys282Tyr mutation.

Recognition of this concept is of great importance for two reasons. First, prior to its identification, the haemochromatosis gene product might reasonably have been predicted to fulfill the role of controlling iron absorption in mammals. The variation in rates of iron absorption in patients bearing the same mutation would suggest that other factors (environmental or genetic) influence this. Secondly, the motivation for mapping and defining the haemochromatosis gene was the potential to identify individuals at risk of irreversible organ injury from iron loading. If the distribution illustrated above is representative of other haemochromatosis populations then 50% or more individuals identified as being homozygous for the common HLA-H mutation might barely reach established criteria for the clinical diagnosis.

The suggestion that the common haemochromatosis phenotype is very mild is supported by the apparent discrepancy between the prevalence of the condition as determined by measurements of iron status and the number of patients presenting with clinical disease and requiring venesection. The John Radcliffe Hospital serves a population of 1 million people. The prevalence of homozygosity for Cys282Tyr in an unselected UK population is 1 in 250 or 0.4% (Alison Merryweather-Clarke, unpublished data). This predicts that as many as 4000 individuals in the Oxford region are homozygotes for this mutation. At present 21 haemochromatosis patients attend the John Radcliffe Hospital for venesection. One explanation for the discrepancy in these numbers is that with a time related accumulation of iron being the hallmark of this condition a proportion of affected individuals die with other illnesses prior to the presentation of iron overload. This premise can neither be confirmed nor refuted without a prospective screening program of asymptomatic individuals.

An alternative view of haemochromatosis would be to consider that the more common phenotype is one of such subtle iron loading that organ damage and objective indication for venesection treatment (as presently accepted) is the exception. Other factors that predispose a minority of patients to be more severely affected (i.e. in terms of iron loading) must be postulated.

Environmental factors (predominantly diet) have been proposed to account for variation in rates of iron loading although this has never been formally addressed. In the absence of recognised iron excretion pathways a conceptual analysis of potential environmental factors affecting iron balance is made more straight forward. Iron is only lost through the daily desquamation of gut epithelium and physiological demands such as pregnancy or chronic haemorrhage. The first of these accounts for only minor loses (in the order of 1mg/day) and the other two are usually clinically detectable. Other than by transfusion, iron can only enter the body through the gastrointestinal absorption of dietary iron. Dietary composition is recognised to influence rates of gut mucosal uptake of iron but as discussed earlier in this thesis there is evidence to suggest that the critical level of control of net iron absorption is the serosal transfer of this iron into the circulation.

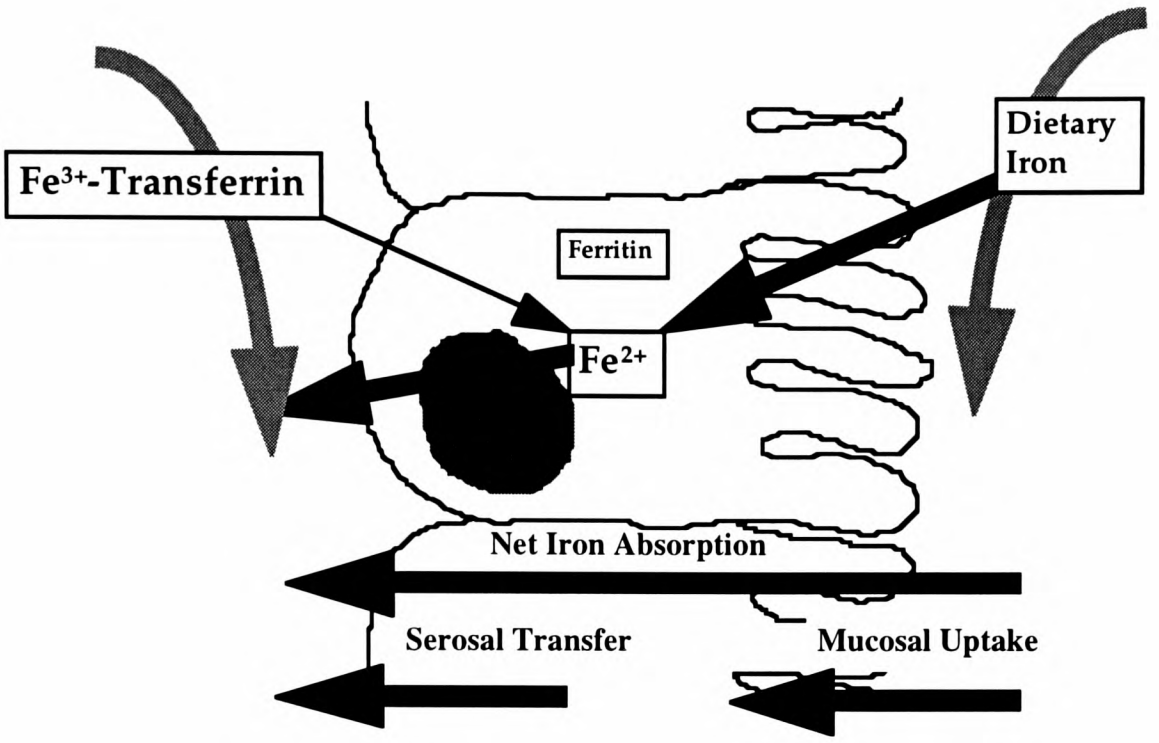


Figure 4.10 The point of control for intestinaliron absorption lies at the serosal surface of the enterocye.

In individuals without haemochromatosis, body iron homeostasis is maintained by controlled release of iron from the enterocyte into the circulation, largely independant of the influence of dietary composition. In patients with haemochromatosis this rate of serosal transfer appears to be variable despite a remarkable consistency in the underlying genetic mutation.

Although the enterocyte release of iron might be influenced by environmental factors (acting from the gut lumen) the recognition of quantitative differences in iron release by reticuloendothelial cells provides support for the existence of further genetic variability in iron release mechanisms. Clinical quantification of reticuloendothelial iron stores is possible by magnetic resonance imaging of the spleen and lumbar vertebral marrow at the time of non-invasive determination of hepatic iron concentration. Such data could generate an "hepatic/spleen iron index" (HSII cf HII) which would account not only for the time related accumulation of iron but also reflect quantitative differences in internal iron distribution.

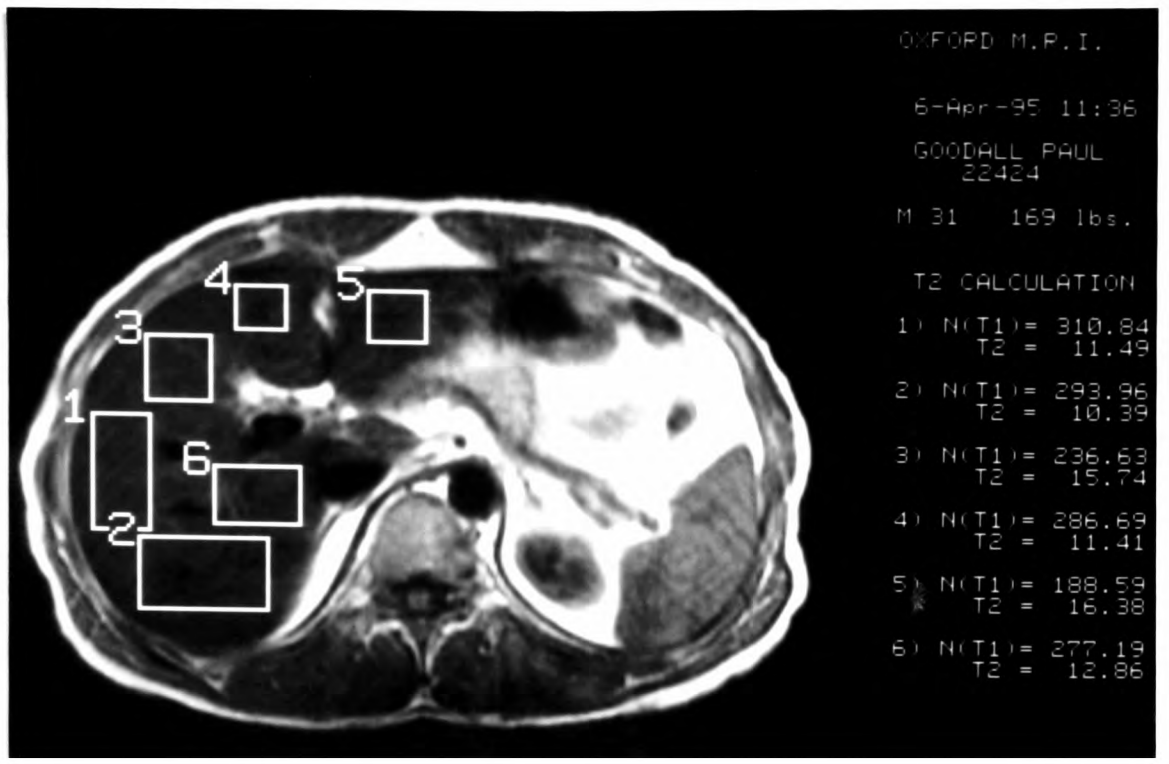


Figure 4.11 Magnetic resonance imaging (MRI) scan of a patient with haemochromatosis. From a scan such as this the splenic iron concentration can be derived from T2 relaxation data. Combined with estimation of hepatic iron concentration the relative distribution of iron between the two sites can be calculated

If the same molecular mechanism controls release of iron from both enterocyte and macrophage this would provide strong support for other genetic factors controlling the rate of iron release and hence absorption.

Haemochromatosis as a "polygenic" disorder

Although limited by the available measures of disease severity, a significantly higher hepatic iron index has been reported in haemochromatosis patients with two copies of an ancestral haemochromatosis haplotype, as defined by markers lying centromeric to HLA-H (Crawford *et al.* 1995; Piperno *et al.* 1996). Given that the long sought haemochromatosis gene has now been defined as an MHC class I-like molecule and that the majority of patients are homozygous for a single point mutation, this raises the possibility that HLA-H might interact with other polymorphic MHC products to generate the marked heterogeneity in the clinical phenotype. Other genetic loci, such as those accounting for primary iron overload conditions not linked to HLA-H, may provide the critical controlling element for intestinal iron absorption. The interest in identifying these other genes must now be all the greater.

Historically the impetus for a positional cloning strategy for the identification of the haemochromatosis gene was driven by the early observation of linkage to HLA-

A. Although this observation lead to the clarification of haemochromatosis being an inherited condition it could be argued that the strength of this linkage actually held back the cloning of the gene due to the effect of linkage disequilibrium beyond the MHC. If no HLA association had been described and the inherited nature of haemochromatosis was still a subject of controversy the condition might have been considered suitable for analysis by a genome-wide screen with polymorphic markers spanning all the chromosomes as has recently been reported for typically "polygenic" conditions such as diabetes (Davies *et al.* 1994), Crohn's disease (Hugot *et al.* 1996) and multiple sclerosis (Sawcer *et al.* 1996). Such an approach would clearly identify strong association at markers flanking HLA-H, 4Mb telomeric to the MHC. It would be fascinating to know whether significant associations might lie elsewhere.

HLA-H is not the functional haemochromatosis candidate that everyone had been expecting and its cloning has raised as many questions concerning mammalian iron metabolism and the haemochromatosis phenotype it has answered. If these issues are not clearly resolved by biological experiments in the near future the search for genetic determinants of iron absorption in man should be continued.

Genotyping for Cys282Tyr in the UK

The motivation behind mapping the haemochromatosis gene was the prospect of identifying asymptomatic individuals at risk of iron loading. Genotyping for the Cys282Tyr mutation in UK haemochromatosis patients suggests that the identification of this mutation will far exceed expectations in this respect. Haemochromatosis can now be defined as the presence of two haemochromatosis alleles in HLA-H regardless of iron status. Genotyping for Cys282Tyr now provides a diagnostic test that is over 90% sensitive and 100% specific for haemochromatosis. If applied to the population of the UK as a screening test this would identify over 200,000 people at risk of developing iron overload. A small proportion of these people might accumulate sufficient iron to develop cirrhosis that could be prevented by the early institution of venesection. Lower levels of iron loading than that which results in cirrhosis has been

associated with coronary heart disease (Salonen *et al.* 1992) and malignancy (Stevens *et al.* 1994). These might only represent potential risks but the proposed therapeutic intervention (venesection) is both safe and relatively cheap. Even if venesected blood from haemochromatosis homozygotes was not taken up by the Blood Transfusion Service (as is presently the case), the potential for generating soluble blood products for either clinical practice or research is enormous.

We may have to wait for some time yet before we understand the molecular mechanisms controlling intestinal iron absorption in man and how this is disrupted in haemochromatosis. However, we now have a simple molecular genetic test for the early identification of this condition which will contribute significantly to the understanding of its clinical course.

Chapter 5

General methods and materials

Contents - Chapter 5

Chapter 5

Pages 220-240

General methods and materials

Suppliers

Unless otherwise stated materials were obtained from the following suppliers:

Biochemicals	Sigma or BDH
Radiochemicals and hybridisation membranes	Amersham
Restriction enzymes	New England Biolabs (NEB)
	NBL
Media	Difco
Sequencing reagents	USB/Amersham

Basic methods

Measurement of DNA concentration

DNA concentrations were calculated from the optical density reading taken at 260nm (OD_{260}) on a Beckman DU-62 spectrophotometer compared to water. DNA concentration was calculated by multiplying the OD260 by the conversion factor 50 for double stranded DNA and by 35 for olignucleotides as described in (Sambrook *et al.* 1989).

Polymerase chain reaction (PCR)

Much of the experimental work presented in this thesis uses the polymerase chain reaction (PCR). Unless otherwise stated PCR reactions were carried out in a total volume of 50µl comprising the following components:

Template DNA @ 100ng/µl	1µl
dNTPs (dATP, dGTP, dCTP, dTTP - 2.5mM of each)	2µl
10x PCR Buffer (Mg^{2+})	5µl
Primer Oligos @ 100ng/µl or 7.5pmol	1µl each
Taq Polymerase (AmpliTaq, Perkin Elmer)	0.25µl (1U)
Autoclaved "MilliQ" water	39.75µl
	50µl

The final magnesium concentration in the reaction mix was 1.5mM unless stated otherwise.

Thermal cycling was performed in a Hybaid Omnigene thermal cycler. Unless otherwise stated cycling conditions were as follows:

(95°Cx2min; ATx30sec; 72°Cx30sec) x1
(95°Cx1min; ATx30sec; 72°Cx30sec) x38

where AT= annealing temperature defined for specific primer pairs.

PCR primer pairs used in this thesis are included in Appendix 2

Restriction enzyme digestion of DNA

Restriction digestion of DNA was conducted in the buffer supplied with the enzyme (NEB). All digestions were carried out at 37°C in a dry heating block except for *BssH* II which was incubated at 65°C as per the manufacturer's instructions.

Electrophoresis

Agarose gel electrophoresis

DNA was separated by electrophoresis through agarose. PCR products were resolved in 2% agarose and digested DNA fragments in 1% agarose. In each case the sample was loaded into the gel in a ten fold dilution of loading buffer (30% Ficoll, 10x TBE, 10% bromophenol blue and 10% xylene cyanol) and run with appropriate size standards (1kilobase ladder, λHindIII, GIBCO). Each gel was run in 1x TBE under a potential of 100 Volts. After electrophoresis gels were stained with ethidium bromide and destined for 10 minutes in distilled water before illumination under ultraviolet light. Where necessary permanent record was captured on Polaroid film.

Pulsed-field gel electrophoresis

High molecular weight DNA, such as yeast chromosomes, were separated by pulsed-field gel electrophoresis (PFGE). DNA (in agarose blocks) was loaded into a 1% agarose gel in 0.5x TBE. Pulsed-field gels were run in either an LKB Pulsaphor or a Biorad CHEF DRIII apparatus. Lambda ladder and/or low range marker (NEB) were run concurrently as size standards. Switching times varied depending on the range of sized fragments to be resolved. Where appropriate these and the total run times are indicated in the legend of the relevant illustration.

Polyacrylamide gel electrophoresis

Sequence reactions and CA repeat analyses were resolved in 4-6% denaturing polyacrylamide gels (39:2 acrylamide:bis-acrylamide; 7M urea). Gels were run in a vertical electrophoresis tank (BRL S2), in 1xTBE at constant power (25 Watts for a small gel, 60 Watts for a large gel).

After electrophoresis the gel was fixed in 10% methanol/10% acetic acid before being dried on Whatman 3MM paper under vacuum at 80°C for 1 hour (Biorad Gel drier 583)

Southern blotting

DNA was transferred from agarose gels to nylon membranes by Southern blotting. After visualisation by ethidium bromide staining gels with high molecular weight DNA (PFGE) were depurinated in acid before denaturation.

Depurination (0.25M HCl)	20mins
Denaturation (0.5M NaOH, 1.5M NaCl)	2x20mins
Neutralisation (1M Tris-HCl pH7.5)	2x30mins.

The gel was placed, inverted, on a platform covered by a wick of Whatman 3MM paper soaked in 10x SSC. The denatured DNA was transferred to Hybond™ N nylon membrane (Amersham) under 500g of pressure overnight.

After blotting, the membrane was air-dried and the DNA was fixed to the membrane UV irradiation (Stratalinker, Stratagene) and baked for 2hrs at 80°C.

Hybridisation proceedures

Labelling of DNA probes

DNA fragments used as hybridisation probes (PCR products, Plasmids etc) were labelled with [^{32}P α]dATP by random hexamer labelling (Multiprime, Amersham). Twenty-five nanograms of DNA was labelled with 30-50μCi of label and purified by column chromatography through a Sephadex (G50-80) column. The DNA probe was denatured at 100°C for 3min before being added to the hybridisation solution.

Hybridisation conditions

Hybridisations were conducted in Church and Gilbert's hybridisation solution at 65°C after at least 1 hour of pre-hybridisation of the membrane. Hybridisations were carried out in a rotating Hybaid hybridisation oven at 65°C overnight.

Post-hybridisation washes

After hybridisation filters were washed under continuous monitoring with a Geiger counter to stringency dependant on the received signal. The wash protocol was:

6xSSC/0.1%SDS	room temperature	30mins
6xSSC/0.1%SDS	room temperature	30mins
2xSSC/0.1%SDS	room temperature	30mins
2xSSC/0.1%SDS	65°C	30mins
0.1xSSC/0.1%SDS	65°C	30mins

After washing filters were patted dry and sealed in cling-film and exposed to radiographic film.

Removal of labelled DNA probe

After hybridisation each filter was stripped of the probe by washing in 0.1% SDS at 96°C for at least 30min. Removal of the labelled DNA probe was confirmed by autoradiography.

Autoradiography

Membranes hybridised with probes incorporating [^{32}P α]dATP were exposed to X-OMAT LS film (Kodak) with intensifying screens at -70°C. Sequence gels ([^{35}S α]dATP] and microsatellite analysis gels ([^{33}Pγ]dATP) were exposed to either X-OMAT LS or BIOMAX MR film (Kodak) without screens at room temperature.

Films were developed in an XOGRAPH X150.

YAC Methods

Screening the YAC Libraries

Two human YAC libraries were held in the Institute of Molecular Medicine. The Washington University library (also known locally as the St. Louis or Olson library - (Brownstein *et al.* 1989)) and the ICRF library (often referred to as 4X, Larin *et al.* 1991). Both libraries were constructed in the vector pYAC4 transformed into the host yeast strain AB1380. Each has at least 2x coverage of the autosomes.

The two YAC libraries were screened by PCR (Green and Olson 1990). In the case of the Washington Library two rounds of PCR on pooled Yeast clones identify a specific 96 plate containing the target clone (see Figure 2.2, page 42). One round of PCR on pooled yeast DNA identifies the plate in the ICRF library.

Having identified the target plate, the individual clone containing the target sequence must be identified. This is achieved by "rows and columns' PCR. The target plate is replicated onto AHC-agar and grown at 30°C. A sample of each colony in a specific row on the plate is picked with a sterile pipette tip and "pooled" into 100µl of MilliQ water in a sterile Eppendorf tube. This process is repeated for each row and each column of the plate in turn. The twenty resulting pooled samples are boiled for 10 minutes to lyse the yeast, centrifuged at 13000 rpm to pellet the cellular debris and then 1µl of the supernatant is used in a PCR reaction. The resulting PCR identifies a single positive row and column defining the grid position of the clone containing the target sequence.

The positive colony is then picked and streaked on AHC agar to obtain single colonies. Five individual colonies and a pool sample are picked from this plate for further characterisation.

Preparation of YAC DNA - Blocks

High molecular weight DNA from whole yeast was prepared in agarose blocks for pulsed-field gel electrophoresis in addition to the preparation of DNA in solution for PCR analysis. A single yeast colony was picked from an AHC plate and

innoculated into 100mls of YPD supplemented with glucose and ampicillin (50µg/ml). This was grown at 30°C using gentle rotation overnight.

A 500µl sample of thie overnight culture was mixed with 200µl of sterile 80% glycerol and stored at -70°C.

Fifty millilitres of the culture was taken for preparation in agarose blocks using the following method:

1. Centrifuge cells at 1000rpm for 10mins
2. Resuspend in 1.6ml 1M sorbitol; 20mM EDTA; 14mM β-mercaptoethanol and 1mg/ml lysing enzymes (Sigma). Incubate at 37°C for 2 hours.
3. Melt low-melting point (LMP) agarose (Gibco) in 1M sorbitol; 20mM EDTA to a final concentration of 1-1.5%. Add β-mercaptoethanol to a final concentration of 14mM (1µl/ml).
4. When cooled to 50°C, add 2ml of molten agarose to the 1.6ml of cells, mixed and dispense into plug moulds (Biorad). Cool on ice for 1 hour.
5. When set push agarose blocks out into 15ml Yeast Resuspension Solution (1M sorbitol; 20mM EDTA; 10mM Tris HCl pH 7.5, lysing enzymes at 1mg/ml). Add proteinase K (50µl of 10mg/ml) and incubate at 37°C for 2hours.
6. Replace resuspension solution with 10ml Yeast Lysis Solution (100mM EDTA; 10mM Tris HCl pH8.0; 1% lithium dodecylsulphate - LDS). Incubate at 37°C for 30mins.
7. Replace Yeast Lysis Solution and incubate cells overnight to lyse.
8. Wash blocks 3-4 times in 50ml of T.E. buffer pH7.5 at 40-50°C to remove all of the LDS.

Blocks stored for more than 1 month are kept in 0.5M EDTA. If such stored blocks were subsequently used the washing steps must be repeated.

Preparation of YAC DNA - Liquid DNA

Twenty-five millilitres of overnight culture was prepared as DNA in solution for PCR analyses using the following method.

1. Centrifuge cells at 1000rpm for 10min.
2. Resuspend cells in 1ml of 1.2M sorbitol; 0.2M Tris HCl pH7.5; 0.02M EDTA and transferred to an Eppendorf tube. Add 10µl of Novozyme (100mg/ml stock, Novo Biolabs) and 20µl β-mercaptoethanol. Incubate at 37°C for 1hour.
3. Centrifuge cells at 13 000rpm for 2min.
4. Wash pellet in 1ml of 1M sorbitol and pellet again.
5. Gently resuspend pellet in 1ml Yeast Lysis Solution (50mM Tris pH 8.0; 100mM NaCl; 100mM EDTA; 0.5% sodium dodecyl sulphate). Add 40µl proteinase K (10mg/ml) and 10K units of RNase A (100µg, Sigma). Incubate at 65°C for 2 hours.

6. Extract DNA with an equal volume of phenol chloroform and precipitate with 2 volumes of absolute alcohol at -20°C for at least 1 hour.
7. Centrifuge the DNA at 13000rpm for 15minutes, resuspend in 400µl of TE buffer and repeat the extraction and precipitation.
8. Resuspended the DNA pellet in 100µl of TE buffer.

Characterisation of YAC Clones

DNA from five individual colonies and a pool sample for each clone were characterised to establish integrity of the YAC. Each was studied by PCR with the primer pair used to screen the library.

Yeast chromosomes were separated by PFGE of high molecular weight DNA in agarose blocks. The separated DNA was transferred to nylon membranes and sequentially hybridised with DNA probes to the YAC vector right and left arms and C_0t 1 DNA to identify the YAC. The size of the YACs were calculated by determining the distance of the YAC band from the sample well on the autoradiograph and plotting this onto a size/distance graph generated for each gel from size standards (λ ladder, low range marker, NEB) run concurrently.

STS Content Mapping

STS content of YAC clones was determined by PCR with STS specific primers using 100ng of total yeast DNA from each clone.

Restriction mapping

Each YAC was restriction mapped using infrequently cutting restriction endonucleases.

Complete digestion was carried out using 40units of the enzyme per half agarose block that had been pre-equilibrated in the appropriate buffer for at least an hour. The digestion proceeded overnight at 37°C.

Partial restriction mapping was conducted with serial dilutions of enzyme. A single agarose block was divided into three and each fragment was equilibrated with the appropriate enzyme buffer. The enzyme was diluted to give a final activity of 20u; 10u and 1u (except for *BssH* II - 1u; 0.5u and 0.05u) and added to the block on ice.

The digestion was allowed to proceed for 30 minutes only at 37°C and the activity of the enzyme was quenched with 5µl 0.5M EDTA.

YAC digestion fragments were resolved by PFGE, the DNA transferred to nylon membranes and hybridised with YAC left and right arm DNA probes. Restriction sites were calculated from the size of restriction fragments identified with each YAC arm DNA probe. Clones from the Washington and ICRF libraries were mapped by both complete and partial digestion.

Isolation of YAC insert end sequence

YAC insert ends were sequenced to generate STS tags for chromosome walking and for STS content analysis. The chosen strategy was that of a vectorette mini method of James and Ogilvie (unpublished). This was adapted from the method of Riley (Riley *et al.* 1990). Vectorette units and primers were obtained from Cambridge Research Biochemicals.

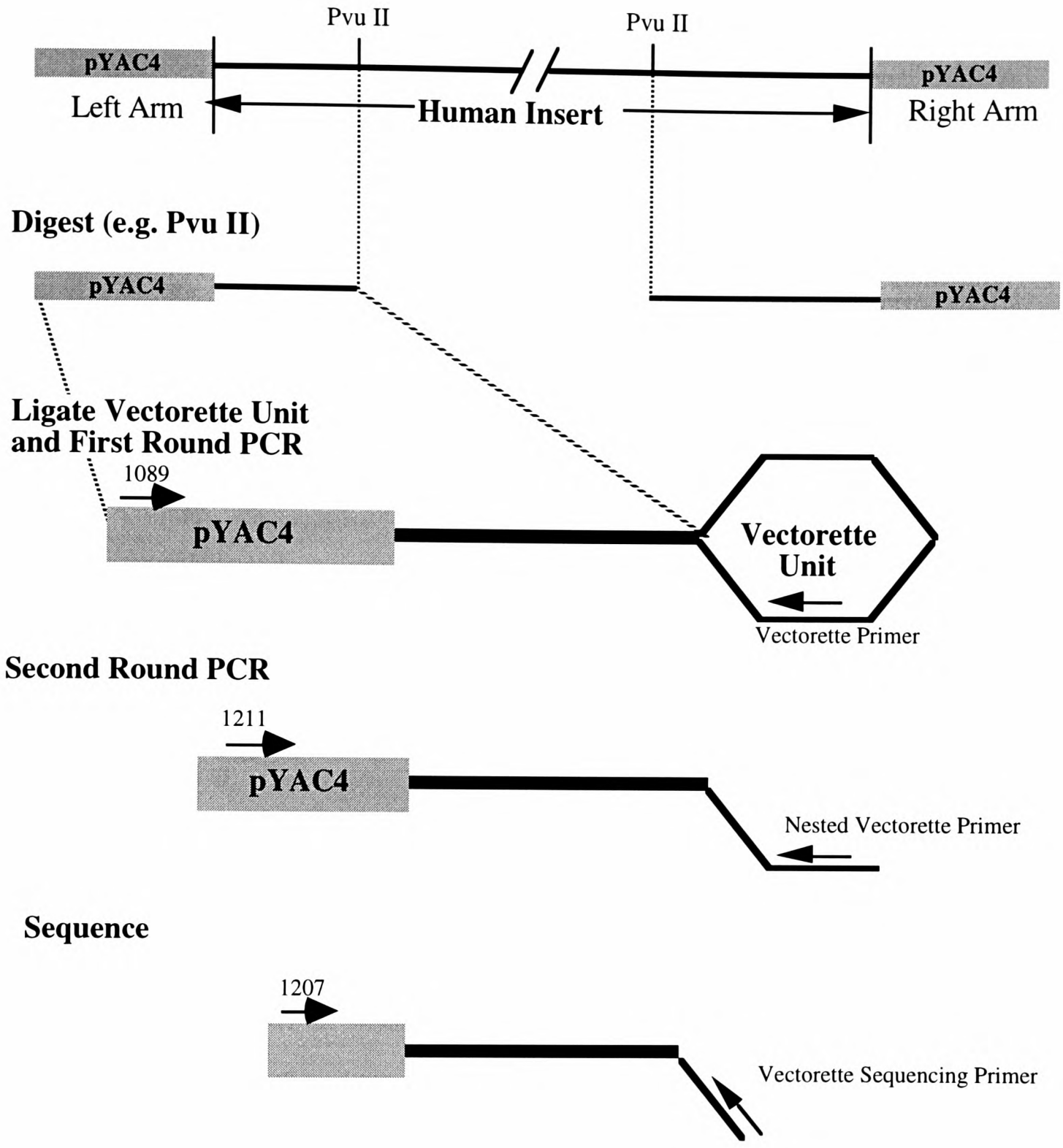


Figure 5.1 Vectorette strategy for isolating YAC insert end sequence (see text for protocol)

Vectorette Library Construction

For each YAC, vectorette libraries were after digestion with 4 separate restriction enzymes (*Alu* I, *Pvu* II, *Rsa* I and *Eco* RV) using the blunt vectorette unit (Cambridge Research Biochemicals). If clean second round PCR product was not obtained with any of these libraries then 3 further libraries were made using the *Bam* HI vectorette unit (*Bcl* I, *Bgl* II and *Bam* HI).

The following method was used for library construction:

1. Divide a single, well washed YAC block into 4 equal portions with a clean blade
2. Equilibrate each quarter block in 500µl of T4 buffer on ice for 1 hour.

3. Replace equilibration buffer with 50µl of buffer containing 20-25u of enzyme and incubate at 37°C overnight
4. Add a further 75µl of T4 buffer and melt the block at 65°C for 10min.
5. Mix carefully and remove 40µl to a clean tube at 37°C
6. Add 10µl of Ligation Mix (85µl 1xT4 buffer; 4µl 100mM ATP; 4µl Vectorette Unit @ 1pmol/ml; 40U T4 ligase @ 2.5U/µl) and incubate at 37°C for 2 hours.
7. Add 200µl of sterile water and store at -20°C until required.

Vectorette PCR

Two rounds of PCR with nested primer pairs specific to the vectorette unit and each of the YAC vector arms (see Appendix 1)

First round PCR

DNA (Vectorette Library)	2µl
dNTPs	2µl
10x PCR Buffer (1.0mM Mg^{2+})	5µl
Vectorette Primer (33pmol)	1µl
YAC arm primer (1090 right arm or 1089 left arm)	1µl
Water	37µl

(96°Cx10min; 92°Cx10min) x1 cycle
Add 1U Taq polymerase to each reaction
(92°Cx2min; 60°Cx2min; 72°Cx5min) x39 cycles
(72°Cx10min) x1 cycle

Two microlitres of the primary PCR product is diluted in 400µl of sterile water and 2µl used as substrate for the second round PCR.

Second Round PCR

DNA (diluted primary PCR product)	2µl
dNTPs	4µl
10x PCR buffer (1.0mM Mg^{2+})	10µl
Nested Vectorette Primer (100pmol)	1µl
Nested YAC arm Primer (20359 right or 1211 left)	1µl
Water	74µl

(96°Cx10min; 92°Cx10min) x1 cycle
Add 1U Taq polymerase to each reaction
(92°Cx2min; 60°Cx2min; 72°Cx5min) x39 cycles
(72°Cx10min) x1 cycle

Ten microlitres of the second round PCR product is run in 2% agarose gel and visualised by staining with ethidium bromide. The "cleanest" (i.e. most specific) PCR products are chosen for sequencing either directly or after cloning the PCR product into pCR Script SK (+) (Stratagene).

Sub-cloning into λ phage

The Washington library YAC B117C7 was sub-cloned into λ phage (Lambda DASH® II, Statagene). The method used for this and the problems encountered are dealt with in Chapter 2.4.

Phage methods

Transferring phage clones to filters

Screening the phage library for recombinant clones containing human DNA was performed by hybridisation to phage DNA on nylon membranes. The phage library plates were overlaid with Hybond™ N nylon membrane discs. The plate was then marked with a sterile needle and india ink before the membrane was removed after 5 minutes. The membrane was air-dried and then denatured (1.5M NaCl, 0.5M NaOH) for 2 minutes, neutralised (1.5M NaCl, 0.5M Tris-HCl, pH8.0) and rinsed (0.2M Tris-HCl, pH 7.5, 2xSSC). The DNA was fixed to the membrane by baking (80°C for 1 hour).

Filters were hybridised with radiolabelled probes of Cot 1 DNA and positive colonies identified by alignment of the resulting autoradiograph with the original plate.

Preparation of phage DNA

Individual phage clones identified by a C_0t 1 probe were picked from the secondary screening plates and eluted in SM buffer with chloroform.

For phage DNA preparation each clones was plated in top agarose on NZY agar to near confluence. 50-100µl of stock phage in SM buffer was adsorbed at 37°C for 15 minutes to 200µl of XL1 Blue MRA (P2) diluted in 10mM $MgSO_4$ to an OD_{600} of 0.5. Once the top agarose was set the plates were incubated overnight at

37°C. Phage were eluted from the agarose with SM buffer, gently rotating at room temperature for 4-6hrs.

The SM buffer was removed and "mini-preps" were prepared using the Qiagen <lambda> miniprep kit. This provided sufficient DNA for screening clones for CA repeats, sub-cloning insert DNA into pUC 18 and cycle sequencing of insert ends.

Phage clone 46 bridged a small gap in the bacterial clone contig established between D6S1260 and D6S1558. DNA from this clone was prepared in greater quantity by liquid culture in 500mls of NZY medium. Phage particles were precipitated with polyethylene glycol (PEG) and sodium chloride and purified on a caesium chloride gradient centrifugation as described in Sambrook (Sambrook *et al.* 1989).

Screening for CA Repeats

Phage clones containing CA repeats were identified in two ways. Firstly the filters generated from the screening of the phage library for human recombinant clones were hybridised with a CA probe and clones containing both human insert and CA were identified by comparison of the resulting autoradiographs. Furthermore, mini-prep DNA from clones identified to contain human inserts were digested with a combination of *Eco* RI and *Hin*d III. The digested DNA was resolved on a 1% agarose gel and after visualisation of the fragments with ethidium bromide staining the gel was Southern blotted onto a nylon membrane. This blot was hybridised with a CA probe to identify clones containing CA repeats. The resulting autoradiograph also allowed identification of clones appearing to contain different CA repeats on the basis of different sized *Eco* RI/*Hin*d III fragments. Subsequent hybridisation of the same membranes with a human C_0t 1 probe generated *Eco* RI/*Hin*d III fingerprints of phage clones which allowed prediction of likely overlap between clones.

Sub-cloning into pUC

Phage clones containing CA repeats were subcloned into pUC 18 (Life Technologies). Phage DNA (10µl) was digested to completion with *Sau* 3A. After phenol-chloroform extraction and ethanol precipitation, DNA was resuspended in

sterile water. 100ng of this digested DNA was ligated to 100ng of *Bam*HI digested vector with 0.5µl of T4 ligase (400u/µl, NEB) at 4°C overnight. One microlitre of a 1:5 dilution of the ligation was used to transform 100µl of MAX Efficiency DH5α competent cells (Life Technologies) by heat shock (42°C for 45 seconds). After the addition of 0.9ml of SOC, the transformed cells were plated onto Hybond™ N discs laid onto LB agar. After incubation at 37°C overnight copy filters were produced. The replica filters were laid onto Whatman 3MM paper soaked in denature solution, neutralising solution and finally 2xSSC (5 minutes on each). The filters were air dried and the DNA fixed by baking for at least 1 hour at 80°C before hybridisation with a CA probe.

Plasmid Methods

Plasmid Mini-prep

Plasmid DNA was prepared by alkaline lysis.

1. Centrifuge 10ml of cells at 3000rpm for 5-10mins
2. Add 100µl (50mM Glucose; 25mM Tris-HCl pH8.0; 10mM EDTA), vortex, transfer to an Eppendorf tube.
3. Add 200µl (0.2M NaOH; 1% SDS), invert and leave on ice for 2mins.
4. Add 150µl 5M KOAc, vortex, leave on ice for 2 mins. Centrifuge at 13000rpm for 5 mins.
5. Remove supernatant to fresh tube and add RNase A (1-2µl of 10mg/ml stock). Incubate at 37°C for 30mins
6. Extract with equal volume of phenol-chloroform. Centrifuge 5 mins.
7. Precipitate with 2 volumes ethanol. Centrifuge 5mins.
8. Wash pellet in 70% ethanol
9. Resuspend in 100µl sterile water

Screening pUC Libraries for CA repeats

Plasmid clones clones identified as containing CA repeats were picked and grown in LB overnight at 37°C. DNA from these clones was diluted to a concentration of 100ng/µl and sequenced with M13 primers.

M13 forward primer **GAT ATC GAA TTC CTG CAG CCC**

M13 reverse primer **CCC TCA CTA AAG GGA ACA AAA GC**

Cloning YAC Ends

Large YAC end Vectorette PCR products were cloned into pCR Script SK(+) (Stratagene) before sequencing.

Bacterial Strains

The following bacterial strains were used:

Bacterial strain	Genotype
XL1-Blue MRA (Stratagene)	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 gyrA96 relA1 lac*
XL1-Blue MRA(P2) (Stratagene)	XL1-Blue MRA (P2 lysogen)
DH5α™ (Life Technologies)	F^-ϕ80d*lac*ZΔM15 Δ(*lac*ZYA-*arg*F)U169 *deo*R *rec*A1 *end*A1 *hsd*R17(r_K^-,m_K^+) *pho*A *sup*E44 λ^-*thi*-1 *gyr*A96 *rel*A1
Epicurian Coli® XL1-Blue MRF' Kan (Stratagene)	*Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F' *proAB lacI*^{q}Z*ΔM15* Tn*5* (Kan^r)]

DNA Sequencing

Direct Sequencing of Vectorette Products

Vectorette PCR products were sequenced directly. A biotinylated second round PCR primer (1211 left arm, 20359 right arm - see Appendix 1) allowed separation of PCR product strands using streptavidin coated magnetic beads (Dynabeads, Dynal). After incubation of the PCR product with the beads at room temperature the DNA strands were denatured with alkali. The biotinylated strand was captured on the side of the Eppendorf tube within an magnetic particle concentrator (MPC Dynal). The non-biotinylated strand was collected and transferred to a clean tube and neutralised with 0.2M Hcl and 1.0M Tris-HCl (pH 7.5). The biotinylated strand was washed with further alkali and TE buffer and resuspended in sterile water.

The two separate strands were sequenced using Sequenase® Version 2.0. Biotinylated strands were sequenced using the Vectorette sequencing primer and the

non-biotinylated strands with primers 1207 (YAC left arm) and 1208 (YAC right arm, Appendix 1)

Sequencing Plasmids

Plasmids were sequenced after alkali denaturation and ethanol precipitation.

Cycle Sequencing of Phage Insert Ends

Lambda phage clones were sequenced directly using primers designed to the T3 and T7 promotor sequences of the vector. The primers were end-labelled with [^{33}Pγ]dATP using T4 polynucleotide kinase (NEB) and cycle sequencing was performed using ThermoSequenase (Amersham) as per the manufacturer's protocol.

Samples were denatured at 96°C for 3minutes before electrophoresis on a 6% polyacrylamide gel.

Patient analysis

Extraction of DNA from whole blood

Whole blood (10-20mls) was collected into sterile EDTA tubes. A wash in sterile distilled water was followed by lysis with 0.1% Nonidet P40. The nuclear pellet was then resuspended in lysis buffer (100mM NaCl, 25mM EDTA) and 10% SDS added to a final concentration of 0.5%. Proteinase K was added (500µg) and the sample incubated at 37°C overnight. One fifth volume 5xANE was added and after two phenol chloroform isoamyl alcohol (24:25:1) extractions, DNA precipitated by adding one tenth volume of 4M NaCl and 2 volumes of absolut ethanol. Precipitated DNA hooked out with a sterile Pasteur pipette, washed in 70% ethanol and air-dried. The DNA was resuspended in 1ml of sterile MilliQ water.

Microsatellite amplification

One of the microsatellite primers was end-labelled with [^{33}Pγ]dATP using T4 polynucleotide kinase (NEB).

Primer (300ng)	3µl
T4 polynucleotide kinase (10U)	1µl
10x buffer (NEB)	1.5µl
Sterile water	8.5µl
[^{33}Pγ]dATP (10µCi)	1µl

Incubate at 37°C for 45 minutes and then heat inactivate the kinase at 95°C for 5 minutes.

PCR amplification of the microsatellite.

Master mix

Labelling mix (as above)	15µl
Other primer (300ng)	3µl
10x PCR buffer	30µl
dNTP (2.5mM of each)	6µl
Taq polymerase (4U)	1µl
Sterile water	245µl

To 10µl of the master mix was added 1µl of genomic DNA (@ 100ng/µl) and overlayed with mineral oil. After PCR amplification, 10µl of Stop solution (USB) was added, the samples were denatured at 95°C for 3 minutes before being loaded onto a 6% denaturing polyacrylamide gel.

Solutions

Chemicals were all obtained from either BDH or Sigma and were of AnalaR grade where available.

TE Buffer

10mM Tris-HCl (pH 7.5); 1mM EDTA

1x TBE Buffer (per 20 litres)

Tris	216g
EDTA	18.6g
Orthoboric Acid	110g

20x SSC Buffer (per 5 litres)

NaCl	876.5g
Sodium Citrate	441g

5x ANE (per litre)

1M NaAc3H_20	10ml
5M NaCl	20ml
0.5M EDTA	2ml
20% SDS	25ml

Church and Gilbert's Hybridisation Buffer (per litre)

SDS	35g
NaH_2PO_4	39g
0.5M EDTA	1ml

pH to 7.2 with NaOH

SM Buffer (per litre)

NaCl	5.8g
$MgSO_4 \cdot H_2O$	2g
1M Tris-HCl (pH 7.5)	50ml
2% gelatin	5ml

10x PCR Buffer (per 10ml)

2.5M Kcl	2mls
Tris-HCl pH8.3	1ml
2M $MgCl_2$	75µl

Made up to 10mls with sterile MilliQ H_20

Filter sterilised and stored at -20^{0}C

Media

All media were made using MilliQ water and autoclaved. Media components (yeast extracts, peptone, tryptone and agar) were obtained from DIFCO.

YPD (per litre)

Yeast Extract	10g
Peptone	20g

pH 5.8 with HCl

50ml filter sterilised 40% glucose and Ampicillin (50µg/ml) added after autoclaving.

AHC Plates (per litre)

Yeast Nitrogen Base	6.7g
Casein Acid Hydrolysate	10g
0.5% adenine hemisulphate	2ml

pH 5.8 with HCl

50ml filter sterilised 40% glucose and Ampicillin (50µg/ml) added after autoclaving.

Luria Broth (LB per litre)

NaCl	10g
Tryptone	10g
Yeast Extract	5g

pH 7.5 with NaOH

LB Plates

As above with 15g of agar per litre of broth

NZY Broth (per litre)

NaCl	5g
$MgSO_4 \cdot 7H_2O$	2g
Yeast Extract	5g
Casein Hydrolysate	10g

pH 7.5

NZY Plates

As above with 15g of agar per litre of broth

NZY top agarose

As above with 0.7% (w/v) of agarose

SOB (per litre)

Tryptone	20g
Yeast extract	5g
NaCl	0.5g

Autoclave

Add filter sterilised 10ml 1M $MgCl_2$ and 10ml 1M $MgSO_4$

2ml of 20% filter sterilised glucose prior to use.

References

Abderrahim, H., Sambucy, J.-L., Iris, F., Ougen, P., Billault, A., Chumakov, I. M., Dausset, J., Cohen, D., and Le Paslier, D. (1994). Cloning the human major histocompatibility complex in YACs. Genomics, **23**, 520-527.

Adams, P. C., Chau, L. A., White, M., and Lazarovits, A. (1991). Expression of transferrin receptors on monocytes in hemochromatosis. American Journal of Hematology, **37**, 247-52.

Adams, P. C., Gregor, J. C., Kertesz, A. E., and Valberg, L. S. (1995). Screening blood donors for hereditary hemochromatosis: Decision analysis model based on a 30-year database. Gastroenterology, **109**, 177-188.

Albig, W., Drabent, B., Kunz, J., Kalff-Suske, M., Grzeschik, K.-H., and Doenecke, D. (1993). All known human H1 histone genes except the H1° gene are clustered on chromosome 6. Genomics, **16**, 649-654.

Altschul, S. F., Boguski, M. S., Gish, W., and Wootton, J. C. (1994). Issues in searching molecular sequence databases. Nature Genetics, **6**, 119-129.

Amadou, C., Ribouchon, M. T., Mattei, M. G., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., Avoustin, P., and Pontarotti, P. (1995). Localization of new genes and markers to the distal part of the major histocompatibility complex (MHC) region and comparison with the mouse: new insights into the evolution of mammalian genomes. Genomics, **26**, 9-20.

Askwith, C., Aide, D., VanHo, A., Bernard, P. S., Li, L., Davis-Kaplan, S., Sipe, D. M., and Kaplan, J. (1994). The FET3 gene of *S.cerevisiae* encodes a multicopper oxidase required for ferrous iron uptake. Cell, **76**, 403-410.

Askwith, C. C., deSilva, D., and Kaplan, J. (1996). Molecular Biology of Iron Acquisition in *Saccharomyces cerevisiae*. Molecular Microbiology, **20**, 27-34.

Bacon, B. R., and Britton, R. S. (1990). The pathology of hepatic iron overload: A free radical-mediated process? Hepatology, **11**, 127-137.

Ballabio, A. (1993). The rise and fall of positional cloning. Nature Genetics, **3**, 277-279.

Barry, M. (1974). Liver iron concentration, stainable iron, and total body storage iron. Gut, **15**, 411-415.

Barton, J. C., Edwards, C. Q., Bertoli, L. F., Shroyer, T. W., and Hudson, S. L. (1995). Iron overload in african americans. American Journal of Medicine, **99**, 616-623.

Bassett, D. E., Boguski, M. S., and Hieter, P. (1996). Yeast genes and human disease. Nature, **379**, 589-590.

Bassett, D. E., Boguski, M. S., and Spencer, F. (1995). Comparative genomics, genome cross-referencing and XREFdb. Trends in Genetics, **11**, 372-373.

Bassett, M. L., Halliday, J. W., and Powell, L. W. (1982). Ferritin synthesis in peripheral blood monocytes in idiopathic hemochromatosis. Journal of Laboratory and Clinical Medicine, **100**, 137-145.

Bassett, M. L., Halliday, J. W., and Powell, L. W. (1986). Value of hepatic iron measurements in early hemochromatosis and determination of the critical iron level associated with fibrosis. Hepatology, **6**, 24-29.

Baynes, R. D., Bukofzer, G., Bothwell, T. H., Meyer, T. E., Friedman, B. M., Macfarlane, B. J., and Lamparelli, R. D. (1989). Iron metabolism in normal and hemochromatotic macrophages. American Journal of Hematology, **31**, 21-5.

Baynes, R. D., Meyer, T. E., Bothwell, T. H., and Lamparelli, R. D. (1991). Maternal and fetal iron measurements in a hemochromatotic pregnancy. American Journal of Hematology, **36**, 48-9.

Beutler, E., Gelbart, T., West, C., Lee, P., Adams, M., Blackstone, R., Pockros, P., Kosty, M., Venditti, C. P., Phatak, P. D., Seese, N. K., Chorney, K. A., Elshof, A. E. T., Gerhard, G. S., and Chorney, M. (1996). Mutation analysis in hereditary hemochromatosis. Blood Cells, Molecules, and Diseases, **22**, 187-194.

Bjorn-Rasmussen, E., Hageman, J., vandenDungen, P., Prowit-Ksiazek, A., and Biberfeld, P. (1985). Transferrin receptors on circulating monocytes in hereditary haemochromatosis. Scandinavian Journal of Haematology, **34**, 308-311.

Boguski, M. S. (1994). Bioinformatics. Current Opinion in Genetics and Development, **4**, 383-388.

Boguski, M. S., Lowe, T. M. J., and Tolstoshev, C. M. (1993). dbEST - database for "expressed sequence tags". Nature Genetics, **4**, 332-333.

Boguski, M. S., and Schuler, G. D. (1995). ESTablishing a human transcript map. Nature Genetics, **10**, 369-371.

Bomford, A., Eddleston, A. L. W. F., Kennedy, L. A., Batchelor, J. R., and Williams, R. (1977). Histocompatibility antigens as markers of abnormal iron metabolism in patients with idopathic haemochromatosis and their relatives. Lancet, 327-329.

Bomford, A., and Williams, R. (1976). Long term results of venesection therapy in idiopathic haemochromatosis. Quarterly Journal of Medicine, **45**, 611-623.

Boretto, J., Jouanelle, A.-M., Yaouanq, J., ElKahloun, A., Mauvieux, V., Blayau, M., Perichon, M., LeTreut, A., Clayton, J., Borot, N., LeGall, J.-Y., Pontarotti, P., and David, V. (1992). Anonymous markers located on chromosome 6 in the HLA-A class I region: allelic distribution in genetic haemochromatosis. Human Genetics, **89**, 33-36.

Bothwell, T. H., Charlton, R. W., and Motulsky, A. G. (1989). Hemochromatosis. The Molecular Basis of Inherited Disease, C. R. Scriver, A. L. Beaudet, and W. S. e. a. Sly, eds., McGraw Hill, New York, 1433-1454.

Bray-Ward, P., Bowlus, C., Choi, J., LePaslier, D., Weissenbach, J., and Gruen, J. R. (1996). FISH-mapped CEPH YACs spanning 0 to 46cM on human chromosome 6. Genomics, **36**, 104-111.

Brink, P., Disler, P., Lynch, S., Jacobs, P., Charlton, R., and Bothwell, T. (1976). Patterns of iron storage in dietary iron overload and idiopathic haemochromatosis. Journal of Laboratory and Clinical Medicine, **88**, 725-731.

Brock, J. H. (1993). Iron and immunity. Journal of Nutritional Immunology, **2**, 47-106.

Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, J., Schlessinger, D., and Olson, M. V. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. Science, **244**, 1348-1351.

Bryan, C. F., Leech, S. H., Kumar, P., Gaumer, R., Bozelka, B., and Morgan, J. (1991). The immune system in hereditary hemochromatosis: a quantitative and functional assessment of the cellular arm. American Journal of Medical Sciences, **301**, 55-61.

Buckler, A. J., Chang, D. D., Graw, S. L., Brook, J. D., Haber, D. A., Sharp, P. A., and Housman, D. E. (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc. Natl. Acad. Sci. USA, **88**, 4005-4009.

Burke, D. T., Carle, G. F., and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science, **236**, 806-812.

Burmeister, W. P., Gastinel, L. N., Simister, N. E., Blum, M. L., and Bjorkman, P. J. (1994). Crystal structure at 2.2Å resolution of the MHC-related neonatal Fc receptor. Nature, **372**, 339-343.

Burt, M. J., Smit, D. J., Pyper, W. R., Powell, L. W., and Jazwinska, E. C. (1996). A 4.5-megabase YAC contig and physical map over the hemochromatosis gene region. Genomics, **33**, 153-158.

Cairo, G., Recalcati, S., Montosi, G., Castrusini, E., Conte, D., and Pietrangelo, A. (1996). Monocytic iron regulatory protein activity is up-regulated in genetic haemochromatosis. European Iron Club, Jaca, Spain.

Calandro, L. M., Baer, D. M., and Sensabaugh, G. F. (1995). Characterisation of a recombinant that locates the hereditary hemochromatosis gene telomeric to HLA-F. Human Genetics, **96**, 339-342.

Camaschella, C., Roetto, A., Gasparini, P., Piperno, A., Fortina, R., Surrey, S., and Rappaport, E. (1996). Allelic association of microsatellites of 6p in Italian haemochromatosis patients. Human Genetics, **97**, 476-481.

Campuzano, V., Montermini, L., Molto, M. D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Trouillas, P., DeMichele, G., Filla, A., DeFrutos, R., Palau, F., Patel, P. I., DiDonato, S., Mandel, J.-L., Cocozza, S., Koenig, M., and Pandolfo, M. (1996). Friedrich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science, **271**, 1423-1427.

Carozzi, N., Marashi, F., Plumb, M., Zimmerman, S., Zimmerman, A., Coles, L. S., Wells, J. R. E., Stein, G., and Stein, J. (1984). Clustering of human H1 and core histone genes. Science, **224**, 1115-1117.

Cartwright, G. E., Edwards, C. Q., Kravitz, K., Skolnick, M., Amos, D. B., Johnson, A., and Buskjaer, L. (1979). Hereditary hemochromatosis: phenotypic expression of the disease. New England Journal of Medicine, **301**, 175-179.

Chandler, M. E., Kedes, L. H., Cohn, R. H., and Yunis, J. J. (1979). Genes coding for histone proteins in man are located on the distal end of the long arm of chromosome 7. Science, **205**, 908-910.

Charlton, R. W., Abrahams, C., and Bothwell, T. H. (1967). Idiopathic hemochromatosis in young subjects. Archives of Pathology, **83**, 132-140.

Chen, E., d'Urso, M., and Schlessinger, D. (1994). Functional mapping of the human genome by cDNA localisation versus sequencing. BioEssays, **16**, 693-698.

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. Nature Genetics, **9**, 347-350.

Conrad, M. E., and Umbreit, J. N. (1993). A concise review: iron absorption-the mucin-mobilferrin-integrin pathway. A competitive pathway for metal absorption. American Journal of Hematology, **42**, 67-73.

Crawford, D. H. G., Powell, L. W., Leggett, B. A., Francis, J. S., Fletcher, L. M., Webb, S. I., Halliday, J. W., and Jazwinska, E. C. (1995). Evidence that the ancestral haplotype in Australian hemochromatosis patients may be associated with a common mutation in the gene. American Journal of Human Genetics, **57**, 362-367.

Crosby, W. H. (1963). The control of iron balance by the intestinal mucosa. Blood, **22**, 441-449.

Cross, S. H., Charlton, J. A., Nan, X., and Bird, A. P. (1994). Purification of CpG islands using a methylated DNA binding column. Nature Genetics, **6**, 236-244.

Cystic Fibrosis Genetic Analysis Consortium (1994a). Population variation of common cystic fibrosis mutations. Human Mutation, **4**, 167-177.

Dabkowski, P. L., Angus, P. W., Smallwood, R. A., Ireton, J., and Jones, R. M. (1993). Site of principal metabolic defect in idiopathic haemochromatosis: insights from transplantation of an affected organ. British Medical Journal, **306**, 1726.

Dancis, A., Klausner, R. D., Hinnebusch, A. G., and Barriocanal, J. G. (1990). Genetic evidence that ferric reductase is required for iron uptake in *Saccharomyces cerevisiae*. Molecular and Cellular Biology, **10**, 2294-2301.

Dancis, A., Roman, D. G., Anderson, G. J., Hinnebusch, A. G., and Klausner, R. D. (1992). Ferric reductase of *Saccharomyces cerevisiae*: molecular characterisation, role in iron uptake and transcriptional control by iron. Proc. Natl. Acad. Sci. USA, **89**, 3869-3873.

Davies, J. L., Kawaguchi, Y., Bennett, S. T., Copeman, J. B., Cordell, H. J., Pritchard, L. E., Gough, S. C. L., Jenkins, S. C., Palmer, S. M., Balfour, K. M., Rowe, B. R., Farrall, M., Barnett, A. H., Bain, S. C., and Todd, J. A. (1994). A genome-wide search for human type 1 diabetes susceptibility genes. Nature, **371**, 130-136.

Davis, W. D., and Arrowsmith, W. R. (1950). The effect of repeated bleeding in hemochromatosis. Journal of Laboratory and Clinical Medicine, **36**, 814-815.

Davis, W. D., and Arrowsmith, W. R. (1953). The treatment of hemochromatosis by massive venisection. Annals of Internal Medicine, **39**, 723-734.

de Souza, M. (1989). Immune cell functions in iron overload. Clin. Exp. Immunol., **75**, 1-6.

de Souza, M., Reimao, R., Lacerda, R., Hugo, P., Kaufmann, S. H. E., and Porto, G. (1994). Iron overload in β_2-microglobulin-deficient mice. Immunology Letters, **39**, 105-111.

Deugnier, Y. M., Guyader, D., Crantock, L., Lopez, J. M., Turlin, B., Yaouanq, J., Jouanolle, H., Campion, J. P., Launois, B., and Halliday, J. W. (1993). Primary liver cancer in genetic hemochromatosis: a clinical, pathological, and pathogenetic study of 54 cases. Gastroenterology, **104**, 228-34.

Deugnier, Y. M., Loreal, O., Turlin, B., Guyader, D., Jouanolle, H., Moirand, R., Jacquelinet, C., and Brissot, P. (1992). Liver pathology in genetic haemochromatosis: a review of 135 homozygous cases and their bioclinical correlations. Gastroenterology, **102**, 2050-2059.

Dixon, R. M., Styles, P., Al-Refaie, F., Kemp, G., Donohue, S. M., Wonke, B., Hoffbrand, A. V., Radda, G. K., and Rajagopolan, B. (1994). Assessment of hepatic iron overload in thalassemic patients by magnetic resonance spectroscopy. Hepatology, **19**, 904-910.

Dorak, T. M., Burnett, A. K., and Worwood, M. (1994). Thymus-leukaemia antigens: the haemochromatosis gene product? Immunology and Cell Biology, **72**, 435-439.

Dujon, B. (1996). The yeast genome project: what did we learn? Trends in Genetics, **12**, 263-270.

Eason, R. J., Aston, C. E., Adams, P. C., and Searle, J. (1990). Familial iron overload with possible autosomal dominant inheritance. Australian and New Zealand Journal of Medicine, **29**, 226-230.

Edwards, C. Q., Griffen, L. M., Dadone, M. M., Skolnick, M. H., and Kushner, J. P. (1986). Mapping the locus for hereditary hemochromatosis: localisation between HLA-B and HLA-A. American Journal of Human Genetics, **38**, 805-811.

Edwards, C. Q., Griffen, L. M., Goldgar, D., Drummond, C., Skolnick, M. H., and Kushner, J. P. (1988). Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. New England Journal of Medicine, **318**, 1355-1362.

ElKahloun, A., Chauvel, B., Mauvieux, V., Dorval, I., Jouanolle, A. M., Gicquel, I., Le, G. J., and David, V. (1993). Localization of seven new genes around the HLA-A locus. Human Molecular Genetics, **2**, 55-60.

European Chromosome 16 Tuberous Sclerosis Consortium (1993). Identification and characterisation of the tuberous sclerosis gene on chromosome 16. Cell, **75**, 1305-1315.

European Polycystic Kidney Disease Consortium (1994b). The polycystic kidney disease 1 gene encodes a 14 kb transcript and lies within a duplicated region on chromosome 16. Cell, **77**, 881-94.

Fargion, S., Fracanzani, A. L., Piperno, A., Braga, M., D'Alba, R., Ronchi, G., and Fiorelli, G. (1994). Prognostic factors for hepatocellular carcinoma in genetic hemochromatosis. Hepatology, **20**, 1426-1431.

Feder, J. N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D. A., Basava, A., Dormishian, F., Domingo, R., Ellis, M. C., Fullan, A., Hinton, L. M., Jones, N. L., Kimmel, B. E., Kronmal, G. S., Lauer, P., Lee, V. K., Loeb, D. B., Mapa, F. A., McClelland, E., Meyer, N. C., Mintier, G. A., Moeller, N., Moore, T., Morikang, E.,

Prass, C. E., Quintana, L., Starnes, S. M., Schatzman, R. C., Brunke, K. J., Drayna, D. T., Risch, N. J., Bacon, B. R., and Wolff, R. K. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nature Genetics, **13**, 399-409.

Fillet, G., Beguin, Y., and Baldelli, L. (1989). Model of reticuloendothelial iron metabolism in humans: abnormal behavior in idiopathic hemochromatosis and in inflammation. Blood, **74**, 844-51.

Finch, S. C., and Finch, C. A. (1955). Idiopathic hemochromatosis, an iron storage disease. Medicine (Baltimore), **34**, 381-430.

Flanagan, P. R., Hajdu, A., and Adams, P. C. (1995). Iron-responsive element-binding protein in hemochromatosis liver and intestine. Hepatology, **22**, 828-832.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J. M., Weidman, J. F., Phillips, C. A., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudeck, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O., and Venter, J. C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Science, **269**, 496-512.

Francanzani, A. L., Fargion, S., Romano, R., Piperno, A., Arosio, P., Ruggeri, G., Ronchi, G., and Fiorelli, G. (1989). Immunohistochemical evidence for a lack of ferritin in duodenal absorptive epithelial cells in idiopathic hemochromatosis. Gastroenterology, **96**, 1071-8.

Gandon, G., Jouanolle, A.-M., Chauvel, B., Mauvieux, V., LeTreut, A., Feingold, J., LeGall, L.-Y., David, V., and Yaouanq, J. (1996). Linkage disequilibrium and extended haplotypes in the HLA-A to D6S105 region: implications for mapping the hemochromatosis gene (HFE). Human Genetics, **97**, 103-113.

Gandon, Y., Guyader, D., Heautot, J. F., Reda, M.-I., Yaouanq, J., Buhe, T., Brissot, P., Carsin, M., and Deugnier, Y. (1994). Hemochromatosis: diagnosis and quantification of liver iron with gradient-echo MR imaging. Radiology, **193**, 533-538.

Gasparini, P., Borgato, L., Piperno, A., Girelli, D., Olivieri, O., Gottardi, E., Roetto, A., Dianzani, I., Fargion, S., Schinaia, G., Cappellini, M. D., Gandini, G., Pignatti, P.-F., Fiorelli, G., de Sandra, G., and Camaschella, C. (1993). Linkage analysis of 6p21 polymorphic markers and the hereditary haemochromatosis: localisation of the gene centromeric to HLA-F. Human Molecular Genetics, **2**, 571-576.

Georgatsou, E., and Alexandraki, D. (1994). Two distinctly regulated genes are required for ferric reduction, the first step of iron uptake in *Saccharomyces cerevisiae*. Molecular and Cellular Biology, **14**, 3065-3073.

Geraghty, D. E., Pei, J., Lipsky, B., Hansen, J. A., Taillon-Miller, P., Bronson, S. K., and Chaplin, D. D. (1992). Cloning and physical mapping of the HLA class I region spanning the HLA-E-to-HLA-F interval by using yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA, **89**, 2669-2673.

Goldwurm, S., Smit, D. J., Kun, J. E., Halliday, J. W., Powell, L. W., and Jazwinska, E. C. (1996). Isolation of hemochromatosis candidate genes. Hepatology, **23**, I-22.

Goodfellow, P. N., Banting, G., Trwosdale, J., Chambers, S., and Solomon, E. (1982). Introduction of a human X-6 translocation chromosome into a mouse teratocarcinoma: investigation of control of HLA-A, B, C expression. Proc. Natl. Acad. Sci. USA, **79**, 1190-1194.

Gordeuk, V., Mukiibi, J., Hasstedt, S. J., Samowitz, W., Edwards, C. Q., West, G., Ndambire, S., Emmanual, J., Nkanza, N., Chapanduka, Z., Randall, M., Boone, P., Romano, P., Martell, R. W., Yamashita, T., Effler, P., and Brittenham, G. (1992a). Iron overload in Africa: Interaction between a gene and dietary iron content. New England Journal of Medicine, **326**, 95-100.

Gordeuk, V. R., Ballou, S., Lozanski, G., and Brittenham, G. M. (1992b). Decreased concentrations of tumor necrosis factor-alpha in supernatants of monocytes from homozygotes for hereditary hemochromatosis. Blood, **79**, 1855-60.

Goya, N., Miyazaki, S., Kodate, S., and Ushio, B. (1972). A family with congenital atransferrinaemia. Blood, **40**, 239-245.

Grasbeck, R., Majuri, R., Kuovonen, I., and Tenhunen, R. (1982). Spectral and other studies on the intestinal haem receptor of the pig. Biochem. Biophys. Acta, **700**, 137-142.

Green, E. D., and Olson, M. V. (1990). Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc. Natl. Acad. Sci. USA, **87**, 1213-1217.

Greene, J. R., Brown, N. H., DiDomenico, B. J., Kaplan, J., and Eide, D. J. (1993). The GEF1 gene of *Saccharomyces cerevisiae* encodes an integral membrane protein; mutations in which have effects on respiration and iron-limited growth. Mol Gen Genet, **241**, 542-553.

Group, The Huntington's Disease Collaborative Research Group (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell, **72**, 971-983.

Gruen, J. R., Goei, V. L., Summers, K. M., Capossela, A., Powell, L. W., Halliday, J., Zoghbi, H., Shukla, H., and Weissman, S. (1992). Physical and genetic mapping of the telomeric major histocompatibility complex region in man and relevance to the primary hemochromatosis gene (HFE). Genomics, **14**, 232-240.

Gruen, J. R., Nalabolu, S. R., Chu, T. W., Bowlus, C., Fan, W. F., Goei, V. L., Wei, H., Sivakamasundari, R., Liu, Y.-C., Xu, H. X., Parimoo, S., Nallur, G., Ajioka, R., Shukla, H., Bray-Ward, P., Pan, J., and Weissman, S. M. (1996). A transcription map of the major histocompatibilty complex (MHC) class I region. Genomics, **36**, 70-85.

Harris, Z. L., Migas, M. C., Hughes, A. E., Logan, J. I., and Gitlin, J. D. (1996). Familial dementia due to a frameshift mutation in the caeruloplasmin gene. Quarterly Journal of Medicine, **89**, 355-359.

Harris, Z. L., Takahashi, Y., Miyajima, H., Serizawa, M., MacGillivray, R. T. A., and Gitlin, J. D. (1995). Aceruloplasminaemia: molecular characterisation of this disorder of iron metabolism. Proc. Natl. Acad. Sci. USA, **92**, 2539-2543.

Hastbacka, J., de la Chapelle, A., Kaitila, I., Sistonen, P., Weaver, A., and Lander, E. (1992). Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genetics, **2**, 204-211.

Hastbacka, J., de la Chapelle, A., Mahtani, M. M., Clines, G., Reeve-Daly, M. P., Daly, M., Hamilton, B. A., Kusumi, K., Trivedi, B., Weaver, A., Coloma, A., Lovett, M., Buckler, A., Kaitila, I., and Lander, E. S. (1994). The diastrophic dysplasia gene encodes a novel sulfate transporter: positional cloning by fine-structure linkage disequilibrium mapping. Cell, **78**, 1073-1087.

Hayashi, A., Wada, Y., Suzuki, T., and Shimizu, A. (1993). Studies on familial hypotransferrinaemia: unique clinical course and molecular pathology. American Journal of Human Genetics, **53**, 201-213.

Hearne, C. M., Ghosh, S., and Todd, J. A. (1992). Microsatellites for linkage analysis of genetic traits. Trends in Genetics, **8**, 288-94.

Hershberg, R., Eghtesady, P., Sydora, B., Brorson, K., Cheroutre, H., Modlin, R., and Kronenberg, M. (1990). Expression of the thymus leukemia antigen in mouse intestinal epithelium. Proc. Natl. Acad. Sci. USA, **87**, 9727-9731.

Hochgeschwender, U. (1992). Toward a transcriptional map of the human genome. Trends in Genetics, **8**, 41-44.

Hoglund, P., Sistonen, P., Norio, R., Holmberg, C., Dimberg, A., Gustavson, K.-H., de la Chapelle, A., and Kere, J. (1995). Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. American Journal of Human Genetics, **57**, 95-102.

Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S.-H., Hu, X., Colbert, A. M. E., Rosenberg, C., Reeve-Daly, M. P., Rozen, S., Hui, L., Wu, X., Vestergaard, C., Wilson, K. M., Bae, J. S., Maitra, S., Ganiatsas, S., Evans, C. A., DeAngelis, M. M., Ingalls, K. A., Nahf, R. W., Horton, L. T., Anderson, M. O., Collymore, A. J., Ye, W., Kouyoumjian, V., Zemsteva, I. S., Tam, J., Devine, R., Courtney, D. F., Renaud, M. T., Nguyen, H., O'Connor, T. J., Fizames, C., Fauré, S., Gyapay, G., Dib, C., Morisette, J., Orlin, J. B., Birren, B. W., Goodman, N., Weissenbach, J., Hawkins, T. L., Foote, S., Page, D. C., and Lander, E. S. (1995). An STS-Based Map of the Human Genome. Science, **270**, 1945-1954.

Hugot, J.-P., Laurent-Puig, P., Gower-Rousseau, C., Olson, J. M., Lee, J. C., Beaugerie, L., Naom, I., Dupas, J.-L., VanGossum, A., Orholm, M., Bonaiti-Pellie, C., Weissenach, J., Mathew, C. G., Lennard-Jones, J. E., Cortot, A., Colombel, J.-F., and Thomas, G. (1996). Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature, **379**, 821-823.

Hunkapiller, T., Kaiser, R. J., Koop, B. F., and Hood, L. (1991). Large-scale and automated DNA sequence determination. Science, **254**, 59-67.

Iancu, T. C. (1993). Animal models in liver research: iron overload. Advances in Veterinary Science and Comparative Medicine, **37**, 379-401.

Jankovic, G. M., Petrovic, M. D., Colovic, M. D., Milosevic, R. A., Janosevic, S., and Trpinac, D. P. (1991). Increased transplacental and breast milk iron delivery in hereditary hemochromatosis: an evolutionary enigma resolved? American Journal of Hematology, **38**, 153-5.

Jazwinska, E. C., Burt, M. J., Smit, D. J., Goldwurm, S. J., Cullen, L. M., Zournazi, A., Banyer, J. L., Francis, J. S., Busfield, F., and Powell, L. W. (1996) Detailed physical mapping of the haemochromatosis gene region. Molecular Genetics of Hemochromatosis - 3rd International Workshop, Rennes.

Jazwinska, E. C., Lee, S. C., Webb, S. I., Halliday, J. W., and Powell, L. W. (1993). Localisation of the hemochromatosis gene close to D6S105. American Journal of Human Genetics, **53**, 347-352.

Jazwinska, E. C., Pyper, W. R., Burt, M. J., Francis, J. L., Goldwurm, S., Webb, S. I., Lee, S. C., Halliday, J. W., and Powell, L. W. (1995). Haplotype analysis in Australian hemochromatosis patients: evidence for a predominant ancestral haplotype exclusively associated with hemochromatosis. American Journal of Human Genetics, **56**, 428-433.

Jeffreys, A. J., Wilson, V., and Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature, **314**, 67-73.

Jensen, P. D., Bagger, J. P., Jensen, F. T., Baandrup, U., Christensen, T., and Ellegaard, J. (1993). Heart transplantation in a case of juvenile hereditary haemochromatosis followed up by MRI and endomyocardial biopsies. European Journal of Haematology, **51**, 199-205.

Jordan, I., and Kaplan, J. (1994). The mammalian transferrin-independant iron transport system may involve surface ferrireductase activity. Biochemical Journal, **302**, 875-879.

Jouanolle, A.-M., Yaouanq, J., Blayau, M., Périchon, M., Fauchet, R., Font, M.-P., LeGall, J.-Y., and David, V. (1990). HLA class I gene polymorphism in genetic hemochromatosis. Human Genetics, **85**, 279-282.

Jouanolle, A. M., Mosser, J., Gandon, G., Andrieux, N., Gicquel, I., Thenie, A., LeGall, J. Y., and David, V. (1996) Genetic studies and physical map of the HFE critical region. Molecular Genetics of Haemochromatosis - 3rd International Workshop, Rennes.

Kaplan, N. L., Hill, W. G., and Weir, B. S. (1995). Likelihood methods for locating disease genes in nonequilibrium populations. American Journal of Human Genetics, **56**, 18-32.

Karlsson, M., Ikkala, E., and Reunanen, A. (1988). Prevalence of hemochromatosis in Finland. Acta Medica Scandinavia, **224**, 385-390.

Kerem, B.-S., Rommens, J. M., Buchanan, J. A., Markiewicz, D., Cox, T. K., Chakravarti, A., Buchwald, M., and Tsui, L.-C. (1989). Identification of the cystic firosis gene: genetic analysis. Science, **245**, 1073-1080.

Klausner, R. D., Rouault, T. A., and Harford, J. B. (1993). Regulating the fate of mRNA: the control of cellular iron metabolism. Cell, **72**, 19-28.

Koller, B. H., Geraghty, D. E., DeMars, R., Duvick, L., Rich, S. S., and Orr, H. T. (1989). Chromosomal organisation of the human major histocompatibility complex class I gene family. Journal of Experimental Medicine, **169**, 469-480.

Koller, B. H., Marrack, P., Kappler, J. W., and Smithies, O. (1990). Normal development of mice deficient in β_2M, MHC class I proteins, and $CD8^+$ T cells. Science, **248**, 1227-1230.

Koskinas, J., Portmann, B., Lombard, M., Smith, T., and Williams, R. (1992). Persistent iron overload 4 years after inadvertant transplantation of a haemochromatotic liver in a patient with primary biliary cirrhosis. Journal of Hepatology, **16**, 351-354.

Kuhn, L. C. (1994). Molecular regulation of iron proteins. Balliere's Clinical Haematology, **7**, 763-785.

Lamon, J. M., Marynick, S. P., Rosenblatt, R., and Donnelly, S. (1979). Idiopathic hemochromatosis in a young female. Gastroenterology, **76**, 178-183.

Larin, Z., Monaco, A. P., and Lehrach, H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA, **88**, 4123-4127.

Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. Genomics, **13**, 1095-107.

Lee, G. R., Nacht, S., Lukens, J. N., and Cartwright, G. E. (1968). Iron metabolism in copper-deficient swine. Journal of Clinical Investigation, **47**, 2058-2069.

Leggett, B. A., Halliday, J. W., Brown, N. N., Bryant, S., and Powell, L. W. (1990). Prevalence of haemochromatosis amongst asymptomatic Australians. British Journal of Haematology, **74**, 525-530.

Lehesjoki, A. E., Koskiniemi, M., Norio, R., Tirrito, S., Sistonen, P., Lander, E., and de la Chapelle, A. (1993). Localisation of the EPM1 gene for progressive myoclonus epilepsy on chromosome 21: linkage disequilibrium allows high resolution mapping. Human Molecular Genetics, **2**, 1229-1234.

Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996). The I.M.A.G.E. consortium: an integrated molecular analysis of genomes and their expression. Genomics, **33**, 151-152.

Litt, M., and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actinn gene. American Journal of Human Genetics, **44**, 397-401.

Logan, J. I., Harveyson, K. B., Wisdom, G. B., Hughes, A. E., and Archbold, G. P. R. (1994). Hereditary caeruloplasmin deficiency, dementia and diabetes mellitus. Quarterly Journal of Medicine, **87**, 663-670.

Lord, D. K., Dunham, I., Campbell, R. D., Bomford, A., Strachan, T., and Cox, T. M. (1990). Molecular analysis of the human MHC class I region in hereditary haemochromatosis. A study by pulsed-field gel electrophoresis. Human Genetics, **1990**, 531-536.

Lynch, S. R., Skikne, B. S., and Cook, J. D. (1989). Food iron absorption in idiopathic hemochromatosis. Blood, **74**, 2187-93.

Macdonald, R. A. (1965). Primary haemochromatosis: inherited or acquired? Prog. Haematol., **5**, 324-357.

Malfroy, L., Coppin, H., Calandro, L. M., Borot, N., Baer, D. M., Sensabaugh, G. F., and Roth, M.-P. (1996). A new microsatellite marker at the RFP locus on chromosome 6p22 definitely locates the haemochromatosis gene at least one megabase telomeric to HLA-A. European Journal of Human Genetics, **4**, 5.105.

Martin, M., Mann, D., and Carrington, M. (1995). Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. Human Molecular Genetics, **4**, 423-428.

McLaren, G. D. (1989). Reticuloendothelial iron stores and hereditary hemochromatosis: a paradox [editorial]. Journal of Laboratory and Clinical Medicine, **113**, 137-8.

McLaren, G. D., Nathanson, M. H., Jacobs, A., Trevett, D., and Thomson, W. (1991). Regulation of intestinal iron absorption and mucosal iron kinetics in hereditary hemochromatosis. Journal of Laboratory and Clinical Medicine, **117**, 390-401.

Moirand, R., Mortaji, A., Loreal, O., Gandon, Y., Guyader, D., Brissot, P., and Deugnier, Y. (1995) Unexplained isolated hyperferritinaemia: a new entity of non-HLA-A3 linked and overweight-related iron overload. Fifth Conference of the International Association for the Study of Disorders of Iron Metabolism, Boston.

Muir, W. A., McLaren, G. D., Braun, W., and Askari, A. (1984). Evidence for heterogeneity in hereditary hemochromatosis: evaluation of 174 persons in nine families. American Journal of Medicine, **76**, 806-814.

Mullner, E. W., Neupert, B., and Kuhn, L. C. (1989). A specific mRNA binding factor regulates the iron-dependent stability of cytoplasmic transferrin receptor mRNA. Cell, **58**, 373-82.

Nemeth, A. H., Hunter, N., Coleman, M. P., Borts, R. H., Louis, E. J., and Davies, K. E. (1993). Rescue of a single yeast artificial chromosome from a cotransformation event utilizing segregation at meiosis. GATA, **10**, 123-127.

Niederau, C., Fischer, R., Purschel, A., Stremmel, W., Haussinger, D., and Strohmeyer, G. (1996). Long-term survival in patients with hereditary hemochromatosis. Gastroenterology, **110**, 1107-1119.

Niederau, C., Fischer, R., Sonnenberg, A., Stremmel, W., Trampisch, H. J., and Strohmeyer, G. (1985). Survival and causes of death in cirrhotic and noncirrhotic patients with primary haemochromatosis. New England Journal of Medicine, **313**, 1256-1262.

Noyes, W. D., Bothwell, T. H., and Finch, C. A. (1960). The role of the reticulo-endothelial cell in iron metabolism. British Journal of Haematology, **6**, 43-55.

Nunez, M.-T., Gaete, V., Watkins, J. A., and Glass, J. (1990). Mobilization of iron from endocytic vesicles. Journal of Biological Chemistry, **265**, 6688-6692.

Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. Science, **245**, 1434-1435.

Olson, M. V. (1995). A time to sequence. Science, **270**, 394-396.

Osaki, S., Johnson, D. A., and Frieden, E. (1971). The mobilization of iron from the perfused mammalian liver by a serum copper enzyme, ferroxidase I. Journal of Biological Chemistry, **246**, 3018-3023.

Perkins, K. W., McInnes, I. W. S., Blackburn, I. W. S., and Beal, R. W. (1965). Idiopathic haemochromatosis in children. American Journal of Medicine, **39**, 118-126.

Pietrangelo, A., Casalgrandi, G., Quaglino, D., Gualdi, R., Conte, D., Milani, S., Montosi, G., Cesarini, L., Ventura, E., and Cairo, G. (1995a). Duodenal ferritin synthesis in genetic hemochromatosis. Gastroenterology, **108**, 208-217.

Pietrangelo, A., Gualdi, R., Casalgrandi, G., Montosi, G., and Ventura, E. (1995b). Molecular and cellular aspects of iron-induced hepatic cirrhosis in rodents. Journal of Clinical Investigation, **95**, 1824-1831.

Pietrangelo, A., Rocchi, E., Casalgrandi, G., Rigo, G., Ferrari, A., Perini, M., Ventura, E., and Cairo, G. (1992). Regulation of transferrin, transferrin receptor, and ferritin genes in human duodenum. Gastroenterology, **102**, 802-809.

Pietrangelo, A., Rocchi, E., Ferrari, A., Ventura, E., and Cairo, G. (1991). Regulation of hepatic transferrin, transferrin receptor and ferritin genes in human siderosis. Hepatology, **14**, 1083-1089.

Piperno, A., Arosio, C., Fargion, S., Roetto, A., Nicoli, C., Girelli, D., Sbaiz, L., Gasparini, P., Boari, G., Sampietro, M., and Camaschella, C. (1996). The ancestral hemochromatosis haplotype is associated with a severe phenotype expression in Italian patients. Hepatology, **24**, 43-46.

Pointon, J. J. (1995). Isolation of microsatellites in the haemochromatosis candidate region, Msc Thesis, University College, London.

Porto, G., Arosa, F., Cabeda, J. M., Lacerda, R., and de Souza, M. (1995) Review of CD8+ T lymphocyte anomalies in GH; correlation with the clinical expression. European Iron Club, Hamburg.

Porto, G., Reimao, R., Goncalves, C., Vicente, C., Justica, B., and DeSouza, M. (1994). Haemochromatosis as a window into the study of the immunological system: a novel correlation between $CD8^+$ lymphocytes and iron overload. European Journal of Haematology, **52**, 283-290.

Pountney, D. J., Simpson, R. J., and Wrigglesworth, J. M. (1994). Duodenal feric reductase: purification and characterisation. Biochemical Society Transactions, **22**, 284S.

Powell, L. W. (1992). Does transplantation of the liver cure genetic hemochromatosis? Journal of Hepatology, **16**, 259-261.

Powell, L. W., Campbell, C. B., and Wilson, E. (1970). Intestinal mucosal uptake of iron and iron retention in idiopathic haemochromatosis as evidence for a mucosal abnormality. Gut, **11**, 727-731.

Radisky, E. S., Ajioka, R. S., Edwards, C. Q., Griffen, L. M., and Kushner, J. P. (1994). Mapping recombinant events with molecular markers in hemochromatosis pedigrees. Cytogenet Cell Genet, **67**, 126-8.

Raha-Chowdhury, R., Tigue, N., J., and Worwood, M. (1994). Trinucleotide repeat microsatellite in the 5' untranslated region of HLA-F. Human Molecular Genetics, **3**, 2084.

Raha-Chowdhury, R., Bowen, D. J., Stone, C., Pointon, J. J., Terwilliger, J. D., Shearman, J. D., Robson, K. J. H., Bomford, A., and Worwood, M. (1995). New polymorphic microsatellite markers place the haemochromatosis gene telomeric to D6S105. Human Molecular Genetics, **4**, 1869-1874.

Raja, K. B., Pountney, D., Bomford, A., Przemioslo, R., Sherman, D., Simpson, R. J., Williams, R., and Peters, T. J. (1996). A duodenal mucosal abnormality in the reduction of Fe(III) in patients with genetic haemochromatosis. Gut, **38**, 765-769.

Raja, K. B., Simpson, R. J., and Peters, T. J. (1992). Investigation of a role for reduction in ferric iron uptake by mouse duodenum. Biochemica et Biohysica Acta, **1135**, 141-146.

Raja, K. B., Simpson, R. J., and Peters, T. J. (1993). Corrigendum, Biochemica et Biohysica Acta, **1176**, 197.

Raja, K. B., Simpson, R. J., and Peters, T. J. (1994). Intestinal iron absorption studies in mouse models of iron-overload. British Journal of Haematology, **86**, 156-162.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., R. Anand, Smith, J. C., and Markham, A. F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Research, **18**, 2887-2890.

Riordan, J. R., Rommens, J. M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.-L., Drumm, M. L., Iannuzzi, M. C., Collins, F. S., and Tsui, L.-C. (1989). Identification of the Cystic Fibrosis Gene: Cloning and Characterisation of Complementary DNA. Science, **245**, 1066-1073.

Ross, C. E., Muir, W. A., Ng, A. B. P., Graham, R. G., and Kellermeyer, R. W. (1975). Hemochromatosis. Pathophysiologic and genetic considerations. American Journal of Clinical Pathology, **63**, 179-191.

Roth, M.-P., Dolbois, L., Borot, N., Amadou, C., Clanet, M., Pontarotti, P., and Coppin, H. (1995). Three highly polymorphic microsatellites at the human myelin oligodendrocyte glycoprotein locus, 100kb telomeric to HLA-F. Human Immunology, **43**, 276-282.

Rothenberg, B. E., and Voland, J. R. (1996). 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.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., and Arnheim, N. (1985). Enzymatic amplification of b-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science, **230**, 1350-1354.

Sallie, R. W., Reed, W. D., and Shilkin, K. B. (1991). Confirmation of the efficacy of hepatic tissue iron index in differentiating genetic haemochromatosis from alcoholic liver disease complicated by alcoholic haemosiderosis. Gut, **32**, 207-210.

Salonen, J. T., Nyyssonen, K., Korpela, H., Tuomilehto, J., and Salonen, R. (1992). High stored iron levels are associated with excess risk of myocardial infarction in eastern finnish men. Circulation, **86**, 803-811.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). *Molecular cloning: a laboratory manual*, Cold Spring Harbour Laboratory Press.

Sargent, C. A., Chalmers, I. J., Leversha, M., and Affara, N. A. (1994). A rearrangement on Chromosome 5 of an expressed human beta-glucuronidase pseudogene. Mammalian Genome, **5**, 791-796.

Sawcer, S., Jones, H. B., Feakes, R., Gray, J., Smaldon, N., Chataway, J., Robertson, N., Clayton, D., Goodfellow, P. N., and Compston, A. (1996). A genome screen in multiple sclerosis reveals susceptibility loci on chromosome 6p21 and 17q22. Nature Genetics, **13**, 464-468.

Scheuer, P. J., Williams, R., and Muir, A. R. (1962). Hepatic pathology in relatives of patients with haemochromatosis. Journal of Pathology and Bacteriology, **84**, 53-64.

Sheldon, J. H. (1935). *Haemochromatosis*, Oxford University Press, London.

Simister, N. E., and Mostov, K. E. (1989). An Fc receptor structurally related to MHC class I antigens. Nature, **337**, 184-187.

Simon, M., Bourel, M., Fauchet, R., and Genetet, B. (1976). Association of HLA-A3 and HLA-B14 antigens with idiopathic haemochromatosis. Gut, **17**, 332-334.

Simon, M., Bourel, M., Genetet, B., and Fauchet, R. (1977a). Idiopathic hemochromatosis: demonstration of recessive transmission and early detection by family HLA typing. New England Journal of Medicine, **297**, 1017-1021.

Simon, M., Bourel, M., Genetet, B., Fauchet, R., Edan, G., and Brissot, P. (1977b). Idiopathic haemochromatosis and iron overload in alcoholic liver disease: differentiation by HLA phenotype. Gastroenterology, **73**, 655-658.

Simon, M., Fauchet, R., Hespel, J. P., Beaumont, C., Brissot, P., Hery, B., Hita de Nercy, Y., Genetet, B., and Bourel, M. (1980). Idiopathic hemochromatosis: a study of biochemical expression in 247 heterozygous members of 63 families: evidence for a single major HLA-linked gene. Gastroenterology, **78**, 703-708.

Simon, M., LeMignon, L., Fauchet, R., Yaouanq, J., David, V., Edan, G., and Bourel, M. (1987). A study of 609 HLA haplotypes marking for the hemochromatosis gene: (1) Mapping of the gene near the HLA locus and characters required to define a heterozygous population and (2) Hypothesis concerning the underlying cause of hemochromatosis-HLA association. American Journal of Human Genetics, **41**, 89-105.

Simpson, R. J., Konijn, A. M., Lombard, M., Raja, K. B., Salisbury, J. R., and Peters, T. J. (1993). Tissue iron loading and histopathological changes in hypotransferrinaemic mice. Journal of Pathology, **171**, 237-44.

Sizemore, D. J., and Bassett, M. L. (1984). Monocyte transferrin-iron uptake in hereditary hemochromatosis. American Journal of Haematology, **16**, 347-354.

Skikne, B., and Baynes, R. D. (1994). Iron absorption. In Iron metabolism in health and disease, J. H. Brock, J. W. Halliday, M. J. Pippard, and L. W. Powell, eds., W.B. Saunders, London.

Smith, B. N., Kantrowitz, W., Grace, N. D., Greenberg, M. S., Patton, T. J., Sorger, K., Doyle, J., Semeraro, J., Cooper, A., Kamat, B., Maregni, L., and Ookubo, R.(1995) Is British-Irish ancestry a risk factor for genetic hemochromatosis? International Association for the Study of Disorders of Iron Metabolism, Boston, Massachusetts.

Stal, P., Broome, U., Scheynius, A., Befrits, R., and Hultcrantz, R. (1995). Kupffer cell iron overload induces intercellular adhesion molecule-1 expression on hepatocytes in genetic hemochromatosis. Hepatology, **21**, 1308-1316.

Stearman, R., Yuan, D. S., Yamaguchi-Iwai, Y., Klausner, R. D., and Dancis, A. (1996). A permease-oxidase complex involved in high-affinity iron uptake in yeast. Science, **271**, 1552-1557.

Stevens, R. G., Graubard, B. I., Micozzi, M. S., Neriishi, K., and Blumberg, B. S. (1994). Moderate elevation of body iron level and increased risk of cancer occurrence and death. International Journal of Cancer, **56**, 364-369.

Stone, C. (1995). Molecular and genetic mapping of the haemochromatosis locus, Dphil Thesis, St.Catherine's College, Oxford.

Stone, C., Pointon, J. J., Jazwinska, E. C., Halliday, J. W., Powell, L. W., Robson, K. J. H., Monaco, A. P., and Weatherall, D. J. (1994). Isolation of CA dinucleotide repeats close to D6S105; linkage disequilibrium with haemochromatosis. Human Molecular Genetics, **3**, 2043-2046.

Stremmel, W., Arweiler, D., Vierbuchen, M., Teichmann, R., Diede, H. E., and Strohmeyer, G. (1991). The membrane iron binding protein is enriched in the liver of patients with hemochromatosis. Hepatology, **14**, 142A.

Summers, K. M., Halliday, J. W., and Powell, L. W. (1990). Identification of homozygous haemochromatosis subjects by measurement of hepatic iron index. Hepatology, **12**, 20-25.

Takahashi, Y., Miyajima, H., Shirabe, S., Nagataki, S., Suenaga, A., and Gitlin, J. D. (1996). Characterisation of a nonsense mutation in the ceruloplasmin gene resulting in diabetes and neurodegenerative disease. Human Molecular Genetics, **5**, 81-84.

Tazi, J., and Bird, A. (1990). Alternative chromatin structure at CpG islands. Cell, **60**, 909-920.

Teichmann, R., Strohmeyer, G., and Stremmel, W. (1991). Iron absorption in man involves a membrane carrier protein. Gastroenterology, **98**, A670.

Terwilliger, J. D. (1995). A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. American Journal of Human Genetics, **56**, 777-787.

Totaro, A., Rommens, J. M., Grifa, A., Lunardi, C., Carella, M., Huizenga, J. J., Roetto, A., Camaschella, C., DeSandre, G., and Gasparini, P. (1996). Hereditary hemochromatosis: generation of a transcript map within a refined and extended map of the HLA class I region. Genomics, **31**, 319-326.

Tugendreich, S., Bassett, D. E., McKusick, V. A., Boguski, M. S., and Hieter, P. (1994). Genes conserved in yeast and humans. Human Molecular Genetics, **3**, 1509-1517.

Valberg, L. S. (1978). Tissue Iron Distribution in Idiopathic Hemochromatosis. Gastroenterology, **75**, 915-916.

Volz, A., Boyle, J. M., Cann, H. M., Cottingham, R. W., Orr, H. T., and Ziegler, A. (1994). Report of the Second International Workshop on Human Chromosome 6. Genomics, **21**, 464-72.

Volz, A., and Ziegler, A. (1996) Physical mapping of a 6Mbp region directly telomeric of the HLA-complex. Molecular Genetics of Hemochromatosis - 3rd International Workshop, Rennes, France.

Walsh, S., and Barrell, B. (1996). The Saccharomyces cerevisiae genome on the World Wide Web. Trends in Genetics, **12**, 276-277.

Walters, G. O., Jacobs, A., Worwood, M., Trevett, D., and Thomson, W. (1975). Iron absorption in normal subjects and patients with idiopathic haemochromatosis: Relationship with serum ferritin concentration. Gut, **16**, 188-192.

Watkins, J. A., Altazan, J. D., Elder, P., Li, C.-Y., Nunez, M.-T., Cui, X.-X., and Glass, J. (1992). Kinetic characterisation of reductant dependant processes of iron mobilization from endocytic vesicles. Biochemistry, **31**, 5820-5830.

Weber, J. L., Kwitek, A. E., May, P. E., and Zoghbi, H. Y. (1991). Dinucleotide repeat polymorphism at the D6S105 locus. Nucleic Acids Research, **19**, 968.

Weber, J. L., and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics, **44**, 388-396.

Whittaker, P., Skikne, B. S., Covell, A. M., Flowers, C., Cooke, A., Lynch, S. R., and Cook, J. D. (1989). Duodenal iron proteins in idiopathic hemochromatosis. Journal of Clinical Investigation, **83**, 261-7.

Williams, R., Smith, P. M., Spicer, E. J. F., Barry, M., and Sherlock, S. (1969). Venisection therapyin idiopathic haemochromatosis. Quarterly Journal of Medicine, **38**, 1-16.

Witte, D. L., Crosby, W. H., Edwards, C. Q., Fairbanks, V. F., and Mitros, F. A. (1996). Practice guideline development task force of the College of American Pathologists - Hereditary hemochromatosis. Clinica Chimica Acta, **245**, 139-200.

Worwood, M. (1994). Laboratory determination of iron status. In Iron Metabolism in Health and Disease, J. H. Brock, J. W. Halliday, M. J. Pippard, and L. W. Powell, eds., W.B. Saunders Company Ltd, London.

Worwood, M., and Darke, C. (1993). Serum ferritin, blood donation, iron stores and haemochromatosis. Transfusion Medicine, **3**, 21-8.

Worwood, M., Raha-Chowdhury, R., and Darke, C. (1994). Distribution of alleles at D6S105 and D6S265 with possible HLA haplotype associations. Tissue Antigens, **44**, 322-325.

Wu, M., VanKaer, L., Itohara, S., and Tonegawa, S. (1991). Highly restricted expression of the thymus leukaemia antigens on intestinal epithelial cells. Journal of Experimental Medicine, **174**, 213-218.

Yamaguchi-Iwai, Y., Stearman, R., Dancis, A., and Klausner, R. D. (1996). Iron-regulated DNA binding by the *AFT1* protein controls the iron regulon in yeast. The EMBO Journal, **15**, 3377-3384.

Yip, R., Johnson, C., and Dallman, P. R. (1984). Age-related changes in laboratory values used in the diagnosis of anaemia and iron deficiency. The American Journal of Clinical Nutrition, **39**, 427-436.

Yoshida, K., Furihata, K., Takeda, S., Nakamura, A., Yamamoto, K., Morita, H., Hiyamuta, S., Ikeda, S., Shimizu, N., and Yanagisawa, N. (1995). A mutation in the ceruloplasmin gene is associated with systemic hemosiderosis in humans. Nature Genetics, **9**, 267-272.

Zijlstra, M., Bix, M., Simister, N. E., Loring, J. M., Raulet, D. H., and Jaenisch, R. (1990). β_2-microglobulin deficient mice lack $CD4^{-}8^{+}$ cytolytic T cells. Nature, **344**, 742-746.

Zoghbi, H. J., and Chinault, A. C. (1994). Generation of YAC contigs by walking. YAC Libraries: a user's guide, Nelson, D., L., and Brownstein, B., H., eds., W.H. Freeman and Co., New York.

Left Arm

```
....AGCATC  GCCAGTCACT  ATGGCGTGCT  GCTAGCGCTA  TATGCGTTGA
....TCGTAG  CGGTCAGTGA  TACCGCACGA  CGATCGCGAT  ATACGCAACT
                                1089
TGCAATTTCT  ATGCGCACCC  GTTCTCGGAG  CACTGTCCGA  CCGCTTTGGC
ACGTTAAAGA  TACGCGTGGG  CAAGAGCCTC  GTGACAGGCT  GGCGAAACCG

CGCCGCCCAG  TCCTGCTCGC  TTCGCTACTT  GGAGCCACTA  TCGACTACGC
GCGGCGGGTC  AGGACGAGCG  AAGCGATGAA  CCTCGGTGAT  AGCTGATGCG

GATCATGGCG  ACCACACCCG  TCCTGTGGAT  CGCGGCCGCS  GCGGCCGCCT
CTAGTACCGC  TGGTGTGGGC  AGGACACCTA  GCGCCGGCGS  CGCCGGCGGA

TTAGTATAAA  TTTCACTCTG  AACCATCTTG  GAAGGACCGG  ATAATTATTT
AATCATATTT  AAAGTGAGTC  TTGGTACAAC  CTTCCTGGCC  TATTAATAAA

GAAATGTGTT  TTTCAATTGT  ATATGTGTTA  TGTAGTATAC  TCTTTCTTCA
CTTTACACAA  AAAGTTAACA  TATACACAAT  ACATCATATG  AGAAAGAAGT
      1211                              1207
ACAATTAAAT  ACTCTCGGTA  GCCAAGTTGG  TTTAAGGCGG  AAGACTTTAA
TGTTAATTTA  TGAGAGCCAT  CGGTTCAACC  AAATTCCGCC  TTCTGAAATT
            Eco RI
TTTATCACTA  CGGAATTCCG  TAAATCTTGA  GATCGGGCGT  TCGACTCGCC
AAATAGTGAT  GCCTTAAGGC  ATTTAGAACT  CTAGCCCGCA  AGCTGAGCGG
                                1208
CCCGGGAGAT  TTTTTTGTTT  TTTATGTCTC  CATTCACTTC  CCAGACTTGC
GGGCCCTCTA  AAAAAACAAA  AAATACAGAG  GTAAGTGAAG  GGTCTGAACG
                                20359
AAGTTGAAAT  ATTTCTTTCA  AGGCGGCCGC  SGCGGCCGCG  ATCCTCTACG
TTCAACTTTA  TAAAGAAAGT  TCCGCCGGCG  SCGCCGGCGC  TAGGAGATGC

CCGGACGCAT  CGTGGCCGGC  ATCACCGGCG  CCACAGGTGC  GGTTGCTGGC
GGCCTGCGTA  GCACCGGCCG  TAGTGGCCGC  GGTGTCCACG  CCAACGACCG
                        1090
GCCTATATCG  CCGACATCAC  CGATGGGGAA  GATCGGGCTC  GCCACTTCGG
CGGATATAGC  GGCTGTAGTG  GCTACCCCTT  CTAGCCCGAG  CGGTGAAGCC

GCTCATGAGC  GCTTGTTTCG  GCGTGGGTAT  GGTGGCAGGC  CCCGTG....
CGAGTACTCG  CGAACAAAGC  CGCACCCATA  CCACCGTCCG  GGGCAC....
                                    Right Arm
```

Appendix 1 Abbreviated map of the cloning site of pYAC4

The Eco RI site of the vector is shown along with the primers designed to left and right arm sequence used to amplify insert ends by the vectorette method (Chapter 5, page 230).

Probes specific to each arm of the vector were amplified by PCR using the following primers:

Left arm

Forward GGT AGT TTA TCA CAG TTA AA
Reverse GTG TGG TCG CCA TGA TCG CG
Product 325bp

Right arm

Forward CTC GCC ACT TCG GGC TCA T
Reverse GAT CAT GTC GCG CTC CAG CGA
Product 356bp

Name	Forward/CA Primer	Reverse/GT Primer	Product Size (bp)	Other names	Source
D6S265	ACG TTC GTA CCC ATT AAC CT	ATC GAG GTA AAC AGC AGA AA	125	AFM101XA1	Généthon
HLA-F	CTG TCC TAT TTC ATA TGC TCA GG	ATG AAC TTG TCC TGA GAA TGA AG	168-340	RF	Raha-Chowdhury *et al* 1994
D6S1683	CTG CAC ATG TAT CCG AGA A	TTT NAA GTA GAG ACA GGA TTT CTT G	186	AFM351TA9	Généthon
D6S131	TCT CTG GAC ATT TCC CTG AAG	TCA ACT CAT GGG TCT GGC ACT	271		Amadou *et al* 1995
D6S258	GCA AAT CAA GAA TGT AAT TTC CC	CTT CCA ATC CAT AAG CAT GG	196	AFM031YH12	Généthon
D6S306	TTT ACT TCT GTT GCC TTA ATG	TGA GAG TTT CAG TGA GCC	237	AFM248XH1	Généthon
D6S105	GCC CTA TAA AAT CCT AAT TAA C	GAA GGA GAA TTG TAA TTC CG	130		Weber *et al* 1991
D6S464	TGC TCC ATT GCA CTC C	CTG ATC ACC CTC GAT ATT TTA C	206	AFM323VB5	Généthon
JJP-1	ACT TCC AAA GGA CAC AAA GAA	GTT GTG ATA TGT GTC TCT GAA	240		Jenny Pointon
D6S1260	ACT GCT CCT GGG CAT GGT TG	GTA CAT GCC TTT GTT AAC ATC	150	CS-5	Caroline Stone
AM-2	ATG TCC TCT GAT CCG CCG	GAG TTC TTT CAC TTT CCT AAG G	332		Alison Merryweather-Clarke
Aggie-3	CTA ATC TCC AAA TGC CTA AG	GAT TTA GAA ATG TAG GCC ACT	142	D6S2252	This Thesis
D6S1558	GCT ACT TGG GAG GCT GGA C	CTG GCA GGA GGG CTA GTG	251	AFMA192WG9	Généthon
Aggie-1	TGT GGT TAT CAG GAT GTA AC	ATA GAA ATG GTA TTT TAT TCC TT	178	D6S2251	This Thesis
D6S1016	GCT TAA AAT TTA AAA GTG AGT TTC C	CCT GTC AGC TAG AGA GGC AG	231	CHLC.GGAA10G12	CHLC
D6S1621	AAA GAT TTA GAG TAA ATG CTG ATG A	ACC ACA GAT GAG AAT GCC TT	286	AFM207WH2	Généthon
D6S1281	GAT GCC ACG TTT TAA AAT GC	AGA AGC AGC TGT GCT TTG TT	197	GATA P19326	CHLC
D6S1571	GGA CCT ACG CAT CTG GTG	TGG CTC TAA TGG TTA CTT TTT ACA	168	AFMA223XD9	Généthon
D6S1545	AAT CTA TGC TCC TGG GTT G	GAA GTT CTG GAA ATA CAG CCT C	225	AFMA116ZE1	Généthon
D6S1554	CAA CAA TAG AAA CAG TCC TTG ATG	CCA GAC AGA ATA TAA GGC AAT TAC	168	AFMA183WB5	Généthon
D6S276	TCA ATC AAA TCA TCC CCA GAA G	GGG TGC AAC TTG TTC CTC CT	205	AFM158YE9	Généthon
D6S461	TAT GAC TTC TGG ACA GTT AGG GG	ACA ACC CAT CAG CCC ACT	245	AFM316ZG5	Généthon
D6S299	AGG TCA TTG TGC CAG G	TGT CTA TGT ATA CTC CTG AAT GTC T	223	AFM217XG7	Généthon

Appendix 2.1 Microsatellite primer sequences

Sequences for microsatellites referred to in this thesis. Microsatellites are ordered from centromere (top) to telomere (bottom) as determined by STS content analysis on YAC clones from the region.

Name	Forward Primer	Reverse Primer	Product Size (bp)	Comment
V13	CGA TCA CAT ACT TCT CTC CT	TAG ACA TAG AGA TCT GCA GC	170	
V12	GAA GGG AGC TGA GGA GGT AA	ACA ACA CCT GTT GAC AGC AC	520	
V10	CCC TCT TTA GCA GTA AGC	GAT AAC CCA GGC ACC AG	350	
V16	TGC ATG CTT ACA TGG TTG TC	CCA CAG CTG AGT CAA ACC TA	~300	
V6	CTT AAG ACC TAG TCT CGT CGG	AAG CCC CCA ACT CCA AGG GTT	208	Jenny Pointon / Kathryn Robson
V11	GTG CCG GAC AGA GAC TGC TA	GAA TTC TAC TAA CCC AGG	269	
V15	TCC ACA GGT ACC TTG TAA	CAG AGC ATT TTA AGA CCA	181	
V26	CCC AAA GTG TGT TAG GAG AG	CTA TGG TCA AGA CCT ACT TGC	140	
V17	GCC TCA AGA AAC CTT GTC AC	CAG CCA GCA CTA GGA AAA TG	727	
V34	ATA AAC AGT GGA ATT GGA AA	ATG AAC ACG AAT TAT CTC CC	293	
V39	ACC ACT GCC AGT TGT CTC TG	GTG GCT GTG CTC ACC AAC TT	133	
V30	AAT CTG TTG TGC ATC TAA TG	GAG CAC CTG TGC TTA TGT CA	550	
V51	TCC TAA CTG GTA CTA CTC CTT C	ATT AAG GAG ACT GAA GAT CCT TC	143	Jenny Pointon
V4	AGC ATT TCA ATC ACT CAT AGA CTT TGG AG	GGG TTA GGA ATG GCT CTA A	146	Jenny Pointon
V40	CCT GGA TAC CTT GTC TTC TA	GGA AGA ATC CTA AGA GTC AT	209	

Appendix 2.2 YAC Insert End STSs

Sequence tagged sites generated from YAC insert end sequence. STSs are ordered from centromere (top) to telomere (bottom).

Name	Forward Primer	Reverse Primer	Product Size (bp)
C7p13T3	GCT GGT GGA CCA AAT AAA C	GAA AGC AAA GAG CTG CAC	113
C7p3T3	CGA GTC TGG CTA AGG AGA AGC	CGT TAG AAC GCC ATT ACA TGG	213
C7p3T7	GGG ACA CCA GCA AGA TGA AA	GAA CCC AGC TAA TTC TCC TG	98
C7p1T3	GCC ACA TTC CTG CAA AGA G	GGT GAC ATG ATG GTG ATG ATG	109
C7p50T7	TGG TTC CAG TCT TCT ATG TC	CCT CAT GTA TCT GCT TTT G	100
C7p1T7	GCA GAT CAA CTC TAT CCA AC	GAT GCT TCC TCA TAT ACC C	103
C7p26T7	AAG AAA GAG AAG AAA CCA GC	TTT TCC CAA GGT CAT CAG	131
C7p50T3	CCT CCA CAT CTT CTC ATT CC	GCC ATA AGT ACA TAA ACA AGG G	108
C7p46T7	TGC CCT CAA AAA ATG ACC	TGA AGA CCA GTG ACC AGA AG	143
C7p54T3	GCT TCA ATG ACA CAA CCA C	CTT TCT CTC TCT CCC TTC C	128
C7p19T3	TAG TGA GGC AGG AGA ACA G	AGG GGT ACA AGT GAC ATG G	176

Appendix 2.3 Phage Clone Insert End STSs

Sequence tagged sites generated from insert ends of λ phage clones derived from the YAC B117C7 (St. Louis).

Name	Forward Primer	Reverse Primer	Product Size (bp)	Source
STSs				
STSG 9945	GCT CAA CAA GTC AGG CCT TC	GAA TCT GGG TTT TAA AAG TCA TGG	137	Whitehead Insititute
WI 3111	TTG ATG AAA TTT GTC CTT GTT AAT T	GGC ACC ATC CCT CAA GTT AA	199	Whitehead Insititute, aka D6S1918
WI 3878	ACA CAA TTT TGT TCT CCT TAG TTG G	ATG TAC AGA CCA CAT TTG GAT CC	250	Whitehead Insititute, aka D6S1921
EST				
N63208	GGG CAC AAG TAA AAA CCG	GCA CCA AGG ACG ATA TGA AG	142	See Chapter 3.2
Genes				
MOG	ATC TCC CAG GAA AGA ACG CT	CAC ATT CCG GAT CCT GAG C	187	Amadou *et al* 1995
RFP	AGC CTG ATC CTC TCT GAT	CTC ACC AGC ATC ATA GTC C	356	Amadou *et al* 1995
BT	TCT GTC CGA CTG GAA GAT TCA CGT	AGC CAG TGG GAG AGG GGT CCG CAG TG	278	Amadou *et al* 1995
Prolactin	CAT CTG GTC ACG GAA GTA CG	GCT GAC TAT CAG CTC CAT GC	117	Amadou *et al* 1995
HLA-H				
Cys282Tyr	TGG CAA GGG TAA ACA GAT CC	CTC AGG CAC TCC TCT CAA CC		Feder *et al* 1996
His63Asp	ACA TGG TTA AGG CCT GTT GC	GCC ACA TCT GGC TTG AAA TT		Feder *et al* 1996

Appendix 2.4 Other STSs

Other sequence tagged site primers.

Name	Hosp No	Sex	DOB	Age at Δ	Family History	Liver Biopsy	HII	HLA-A Serotype	Haplotype (D6S265, 105, 1260, 1621)	Cys282Tyr mutation
N.B.	0860758B	M	23.11.56	28	+ Mother (JB)	"Large amount" of iron		**A3, A3**	**1,1; 8,8; 4,4; 5,5**	+/+
A.B.	0225462R	M	22.4.55	29		"Probable" haemochromatosis			3,5; 6,12; **4,**11; **5,5**	+/+
P.G.	0352994S	M	29.6.66	29		Grade II/III siderosis	9	A11, A32	4,5; 6,**8**; **4**,5; **5**,10	+/+
G.H.	0498789S	M	4.2.53	30		Grade IV siderosis		**A3, A3**	**1,1; 8,8; 4,4; 5,5**	+/+
P.W.	1164335	M	20.7.55	35	+ Father (died)	Grade III siderosis Severe fibrosis		**A3, A3**	**1,1; 8,8; 4,4;** 7,10	+/+
A.A.	0108471K	M	31.8.48	43		Grade II/III siderosis	4		**1,1; 8,8; 4,4; 5,5**	+/+
J.S.	0392900	M	11.3.47	45		Grade II/III siderosis	2.8	A2, **A3**	**1,**5; 5,**8; 4,**7; **5,5**	+/+
P.B.	0845831K	M	27.10.37	47	+ Sister (JB)	Grade IV siderosis	10.75	**A3,** A1	**1,**3; **8,8;** 3,**4; 5,5**	+/+
G.M.	0754144	M	10.3.44	50		Grade IV siderosis		A1, A26	3,4; 7,**8; 4,4; 5,5**	+/+
J.W.	0548125D	M	11.7.32	54		Grade III siderosis		**A3**, A11	**1,**4; **8,8; 4,5; 5,5**	
M.O'K	0428574	M	20.9.35	58	+ Brother (died)	Grade III siderosis Minor portal tract fibrosis	5.6	A2, A2	5,6; 4,4; 3,**4; 5,5**	+/+
E.W.	1168647K	M	28.2.36	60		Grade II siderosis	5.6		3,6; 6,7; **4,**8; **5,**8	-/-
B.G.	0620716S	M	13.9.40			Alcoholic liver disease with iron overload			**1,1; 8,8; 4,4;** 3,**5**	+/+
J.Lo.	0445269D	M	18.7.22	60	+ Brother	"Early haemochromatosis"	5.6	A2, **A3**	**1,**6; **8,8; 4,4; 5,5**	+/+
J.Ly.	0228817	M	17.12.30	63		Grade III/IV siderosis Severe fibrosis		**A3, A3**	**1,1; 8,8; 4,4; 5,5**	+/+
B.F.	0335371H	M	3.12.28	65		Grade IV siderosis Cirrhosis			**1,1; 8,8; 4,4; 5,5**	+/+
R.S.	0828697	M	9.12.22	69		Grade III siderosis			**1,**5; 5,**8; 4,4; 5,5**	+/+
L.W.	1450073L	F	6.4.69	25		(Wigan)			**1,1; 8,8; 4,4; 5,5**	+/+
B.C.	1250688H	F	15.2.58	28		(Australia) **HCV+**	4	**A3, A3**	**1,1; 8,8; 4,4; 5,5**	+/+
A.H.	0881349L	F	4.2.53	33	+ Brother (GH)	Grade IV siderosis	10.6	**A3, A3**	**1,1; 8,8; 4,4; 5,5**	+/+
J.B.	0742636V	F	15.6.34	50		(King's College)		**A3**, A1	**1,**3; **8,8;** 3,**4; 5,5**	+/+
A.M.	0845863S	F	6.5.37	57		Alcoholic liver disease and severe iron overload			**1,**3; **8,8; 4,4; 5,**10	+/+
A.S.	0341505E	F	6.9.39	57		No Biopsy (clotting disorder)		**A3**, A25	**1,**4; **8,8**; **4,4**; 3,**5**	+/+

Appendix 3 Haemochromatosis patient clinical details