European Journal of Human Genetics (1998) 6, 105–113

http://www.stockton-press.co.uk/ejhg

ORIGINAL PAPER

Generation of a transcription map of a 1 Mbase region containing the *HFE* gene (6p22)

Angela Totaro[1], Antonella Roetto[2,3], Johanna M Rommens[4,5], Anna Grifa[1], Massimo Carella[1], Leonardo d'Agruma[1], Maria Assunta Valentino[1], Lucrezia D'Ambrosio[1], Matteo Cicilano[2], Clara Camaschella[2], Brunella Franco[3] and Paolo Gasparini[1]

[1]*Servizio di Genetica Medica, IRCCS-Ospedale 'CSS' San Giovanni Rotondo*
[2]*Dipartimento di Scienze Biomediche ed Oncologia Umana, Ospedale S Luigi, Orbassano (To)*
[3]*Telethon Institute of Genetics and Medicine (TIGEM), San Raffaele Biomedical Science Park, Milan, Italy*
[4]*Department of Genetics, Research Institute, The Hospital for Sick Children, Toronto*
[5]*Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada*

A transcription map was generated of a 1 Mb interval including the *HFE* gene on 6p22. Thirty-seven unique cDNA fragments were characterised following their retrieval from hybridisation of immobilised YACs to primary pools of cDNAs prepared from RNA of foetal brain, human liver, foetal human liver, placenta, and $CaCo_2$ cell line. All cDNA fragments were positioned on the physical map on the basis of presence in aligned and overlapping YACs and cosmid clones of the region. The isolated cDNAs together with established or published sequence tagged sites (STSs) and markers provided sufficient landmark density to cover approximately 90% of the 1 Mb interval with cosmid clones. The precise localisation of two known genes (*NPT*1 and *RING* finger protein) was established. A minimum of 14 additional transcription units has also been integrated. Twenty-eight cDNA fragments showed no similarity with known sequences, but 20 of these detected discrete mRNAs upon northern analysis. Their characterisation is still under investigation. Eleven new polymorphisms were also identified and localised, and the *HFE* genomic structure was better defined. This integrated transcription map considerably extends a recently published map of the *HFE* region. It will be useful for the identification of genetic defects mapping to this region and for providing template resources for genomic sequencing.

Keywords: haemochromatosis; HFE; transcription HAP; 6p22

Introduction

The generation of transcription maps of the human genome is an important step towards the identification of disease genes. During our project for the identification of the hereditary haemochromatosis (*HH*) gene, we focused our attention on the short arm of chromosome 6 (6p22), in the region placed telomeric to marker *D6S105*. Recently, linkage disequilibrium and ancestral haplotype analysis carried out in a large series of patients, provided that the HH gene most likely resides in an interval located further telomeric than previously reported.[1]

Correspondence: Paolo Gasparini, MD, Servizio di Genetica Medica, IRCCS-Ospedale 'CSS', I–71013 San Giovanni Rotondo (Fg), Italy
Received 12 June 1997; revised 13 October 1997; accepted 22 October 1997

This region was poorly investigated and characterised by a low representation of known transcripts. Thus, after cloning the entire 6p22 region in YACs[2] we applied the cDNA selection technique to a 1 Mb region covered by the YAC 790G7, which spans most of the region defined by markers *D6S1016* and *D6S1921*. A novel HLA class I-like gene (*HFE*), mutated in patients affected by HH, has been isolated from this region,[1] and its genomic structure and transcription orientation has been here reported. Location information using established and known markers and the new cDNA fragments was then integrated to achieve a refined map of the region, including a partial cosmid contig.

Materials and Methods

Isolation of YACs, cosmid clones, and STS mapping

YAC clones were provided by the YAC screening centre (DIBIT-HSR, Milan). All the clones with the exception of 935D10 were previously reported in Malaspina *et al*.[2] DNA was obtained following propagation of single colonies in Achilles heel cleavage (AHC) medium at 30°C for 48 hours. Total yeast DNA chromosomes were prepared in 100 μL Seaplaque (FMC, Rockland, ME) agarose plugs.

A chromosome 6 specific cosmid library constructed at the Los Alamos National Laboratory was kindly provided by the Baylor College of Medicine Human Genome Center. High density filters (1536 clones/filter) were prepared for hybridisation screening by plating on to nylon filters from the thawed 384 microtiter dish stocks using the Biomek 2000 workstation. The membranes were transferred to appropriate agar plates for overnight growth, and processed for cell lysis and DNA fixation.[3]

All publicly available and newly identified STS were used as probes to screen the arrayed chromosome 6 specific cosmid filters. In all, 287 cosmid clones were identified. They were retrieved, grown in individual wells of 96 well plates and transferred to filters as for the entire library. All probes used in the hybridisation of either YACs or cosmids were preassociated with a 2×10^4 excess (0.2 mg/ml) of sonicated human placental DNA for 1 hour at 65°C, in order to saturate repetitive DNA sequences. The hybridisation and washing conditions were as previously described.[4]

The following landmarks were integrated in the map:

- anonymous STS: (*D6S1294*);
- polymorphic markers: *D6S2220*, *D6S2221*, *D6S2240*, *D6S2238*, *D6S2241*, *D6S2236*, *D6S2237*, *D6S1621*, *D6S2233*, *D6S1281*. All of them are available at the Genome Database (http://www-genome.wi.mit.edu).
- the IMAGE cDNA clone 307162 (clone zb51g06), used to determine the direction of transcription of *HFE* gene;
- 899G1-1, ALUG7, ALUG12, 792G12-3 are four anonymous probes obtained by Alu-PCR from YAC 899G1, 790G7 and 792G12 using a mix 1:4 of ALUJ/ALUS primers[5] and PCR conditions of: annealing 58°C for 1', and extension 72°C for 2' for 35 cycles. Aliquots of the PCR products or specific ALU bands were isolated from low-melting agarose gel and used as probes.

cDNA Selection and Characterisation

Mammalian genomic DNAs were isolated from human leukocytes, from rodent cell lines and a mouse-human hybrid cell line GM10629 obtained from NIGMS Human Genetic Mutant Cell Repository, Camden, New Jersey. Restriction digestions were carried out according to supplier's recommendations. Electrophoresis, blotting and hybridisation were carried out according to standard procedures. After hybridisation, the blots were washed (0.2XSSC with 0.1% SDS) at 60°C and exposed to autoradiography for 2–72 h.

Randomly primed cDNA was prepared from poly-A^+ RNA of foetal liver, foetal brain, and placenta tissues and from total RNA of adult liver, and of the Caco-2 cell line (ATCC HTB 37). The artificial chromosome of YAC 790G7 was transferred to nylon membrane (Hybond N, Amersham) and immobilised by UV-cross linking following separation by pulsed-field gel electrophoresis. Selection was carried out for two consecutive rounds of hybridisation to the immobilised artificial chromosome DNA. Hybridising cDNA was collected, amplified and cloned as described.[6,7]

We picked and gridded 288 individual colonies on a microtiter format for ordering, prescreening and storage. Initial analysis of cDNA clones included a prescreening for ribosomal sequences. Approximately 8% of the clones were eliminated as they hybridised with radiolabelled cDNA obtained from total RNA. A secondary screening for histone genes was carried out, and an additional 1% of clones was eliminated. Plasmids that did not hybridise in the prescreenings were isolated for further analysis. Of isolated inserts, 52% mapped to the starting YAC clone. Membranes prepared from the gridded clones were also repeatedly hybridised with individual cDNA clones to eliminate overlapping or redundant clones. The sequence of both strands of the retrieved cDNAs was obtained with the dideoxy chain terminator method[8] using the Taq Dye Primer Cycle Sequencing Kit (ABI) and the 373A DNA Sequencer (ABI). The resulting sequences were then analysed by comparison with BLASTX-non redundant protein, BLASTN on GeneBank, GeneBank update, EMBL, EST, SWISSprot databases.

Northern blotting was performed to determine the pattern of expression of each retrieved cDNA clone. Clontech Human Northern Blots (Clontech, USA) containing several tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, leukocytes) were used with the recommended conditions.

Polymorphism Detection

Oligonucleotides for PCR primers corresponding to the first 23 clones isolated (CD30L, CD36, CD64, CD77, CD117, CD140, CD176, CD218, CD235, CD282, CD212, CD123, CD43, CD88, CD104, CD105, CD108, CD110, CD171, CD203, CD207, CD240) were designed to run RNA-SSCP experiments for detecting nucleotide changes, as described.[9,10] Annealing temperatures for amplification and length of each PCR product are reported in (Table 3) (see Results). PCR was performed on 500 ng genomic DNA from 20 unrelated individuals according to standard protocols. After *in vitro* transcription with the incorporation of (α 32)rATP an aliquot of the reaction (4.5 μl) was loaded on

to a 6.5% nondenaturing polyacrylamide gel. Electrophoresis was performed at 30W constant power for 13 h. After electrophoresis, the gel was dried and subjected to autoradiography for 12 h. Fragments with altered migration patterns were then sequenced as described above. When a polymorphism was detected, frequencies of alleles were assessed in a sample of 50 unrelated individuals.

cDNA Library Screening

cDNA-selection products, labelled with a-32P dCTP with Multiprime Kit (Beckman Instruments, Palo Alto, CA) for a concentration of 1×10^6 cpm/ml, were used to screen five human cDNA libraries: NT2 neuronal precursor cell (937230 Stratagene, La Jolla, CA), duodenum (HL1156n Clontech Inc., Palo Alto, CA), liver (937224 Stratagene, La Jolla, CA), placenta (937225 Stratagene, La Jolla, CA) and foetal kidney (HL3028b, Clontech Inc). Prehybridisation and hybridisation and washes were performed as previously described.[2]

HFE Genomic Structure

The *HFE* intron/exon boundaries have been defined using two strategies:

(1) generation of intron-containing genomic fragments by long-PCR using *HFE* cDNA-derived primers; and

(2) screening of a chromosome 6 specific cosmid library using PCR fragments corresponding to the *HFE* cDNA.[1] PCR was performed with a Perkin Elmer Gene Amp 9600 thermal cycler, in reaction volumes of 25 μl and DNA concentration of 250 ng/ml. The reaction was continued for 30 cycles consisting of 1 min at 94°C, 1 min at 60°C and 2 min at 72°C. The intron/exon boundaries of amplified fragments were sequenced in a 373A automatic sequencer (Applied Biosystems) with the same PCR primers. Simultaneously, a chromosome 6 specific cosmid library, cloned in Lawrist 4 vector, was analysed by hybridisation with different cDNA fragments generated by PCR. Six positive clones (23B12, 11A9, 31H6, 45E12, 24A11, 39K16) were isolated and analysed by PCR with cDNA primer pairs (Table 2, see Results). Five clones (23B12, 11A9, 45E12, 24A11, 39K16) were positive for all six exons of *HFE* gene plus the 3' UTR, while clone 31H6 did not contain the 3' UTR. A primer designed on the first exon of *HFE* gene was used to sequence the 5' UTR region.

Results

We investigated the interval on 6p22 defined centromerically by the markers *D6S1016* and *D6S1921*. We had previously established an extended comprehensive physical map revealing that the YAC clone 790G7 essentially spanned the defined internal and thus it was chosen for cDNA selection experiments. In the meantime, another group reported the isolation of a gene, named *HFE*; from this interval that is mutated in haemochromatosis patients.[1] In addition, they described several new polymorphisms within this region. We incorporated these new polymorphisms and were able to establish the orientation and position of the *HFE* gene with our map.

cDNA Selection Results

Thirty-seven unique cDNA selection products were selected, sequenced, and verified to map to YAC clones 790G7, 792G12, 935D10, 764H9, 905G1, 899G1, and 767G12. To confirm assignment to chromosome 6, each cDNA was hybridised to total human DNA and to a mouse hybrid cell line that contains chromosome 6 as its only human material. Sequence analysis of each clone by alignment to the public databases revealed that 28 corresponded to a new sequence, while the remaining nine shared overlapping sequence or were similar to previously known genes. A complete summary of the results obtained is reported in Table 1.

cDNAs of Known Human Genes

Nine clones shared overlapping sequence, and/or were homologous to known genes, as summarised in Table 1. One clone, CD33, was identical to a sodium-phosphate pump (*NPT1*)[11] and had already been assigned to the 6p23-6p21.3 region. This transcription unit was mapped within the end of YAC 928C4 and was indeed negative on the cosmids of the region. Clone CD235 was identical to a gene encoding a RING finger protein (Y07829). Another clone, CD75, showed similarity with the bovine butyrophilin (BTN) and it was not further investigated. Six additional clones (CD23, CD58, CD134, CD216, CD231, CD296) shared overlapping sequences with known repetitive elements, including LINE 1, and mer sequences (Table 1).

To attempt to get full length cDNAs, groups of retrieved fragments were then used to screen five different human cDNA libraries. Twelve different cDNA clones were isolated and mapped on to the YAC and cosmid clones (data not shown). In particular, two cDNA clones (hDuo11 and hDuo18) were positive with CD30H, three (hNT2 6, hNT2 7b, hNT2 14) were positive with CD115, five (hKi2, hKi1a, hKi5b, hKi9, hKi11) were positive with CD274, one (hLi12) was positive with CD123. Work is in progress to characterise these cDNAs further and to get more clones for the remaining cDNA selection products using additional cDNA libraries.

RNA Hybridisation

cDNA clones isolated by cDNA selection were tested by hybridisation to polyA^+ RNA of a series of tissues. The results are presented in Table 2. Fifteen clones revealed a wide distribution of tissue expression, while five demonstrated more specific expression patterns. CD190 was expressed in only prostate (Figure 1). CD212 and CD235 gave discrete mRNAs only in liver

(Figure 1). Two other clones were expressed in limited sets of tissues, CD274 in liver and kidney, and CD123 in liver, muscle, kidney and pancreas. All these 20 clones detectedone single band either on chromosome 6 or genomic DNA (data not shown). Five clones (CD36, CD105, CD117, CD203, and CD218) detected less discrete signals on northern blots, indicating either hybridisation to a series of mRNAs or non-specific hybridisation. In this case, all clones detected one band on chromosome 6 DNA, while at genomic level clones CD117 and CD203 detected additional bands (data not shown). For the remaining clones (CD77, CD140, CD152, CD207) expression levels appeared below the sensitivity of our northern assay. Gene fragments CD43, CD115, and CD219, showed identical expression patterns. As they are located in close proximity to each other, they are likely to correspond to a single transcription unit or gene. Four additional transcription units arise from four pairs of closely positioned clones, CD110/CD282, CD88/CD30L, CD240/CD30H, and CD176/CD254. They also showed consistent mRNA sizes and patterns of expression, while RT-PCR, leading to one amplified product, confirmed the presence of only one transcript for each transcription unit (data not shown). Clones CD110/CD282 (transcription unit 4) define a sequence with an homology of 88% at protein level with *NPT3* gene (*NPT3h*. In contrast, the expression pattern and sequence homologies for CD183, CD123, and CD171 clones, that all reside on the same 51M1 cosmid, are consistent with a minimum of three

Table 1 Summary of selected cDNA clones. The first 31 gene fragments correspond to new sequences while the remainder shared overlapping sequence, and/or were homologous to previously known genes

Clone	*Size (bp)*	*Remarks*	*EMBL Acc. No.*
CD 20	578	New sequence	Z83954
CD 30H	514	New sequence	Y10210
CD 30L	362	New sequence	Y10209
CD 36	655	New sequence	Z83936
CD 43	1100	New sequence	Z83934–5
CD 64	637	New sequence	Y10206
CD 67S	317	New sequence	Y10510
CD 77	549	New sequence	Y10204
CD 88	651	New sequence	Y10205
CD 105	635	New sequence	Z83944
CD 110	697	New sequence	Y10506
CD 115	446	New sequence	Y10519
CD 117	689	New sequence	|Z83948
CD 123	800	New sequence	Z83952
CD 140	496	New sequence	Y10208
CD 152	1100	New sequence	Y10514
CD 171	514	New sequence	Y10207
CD 176	1000	New sequence	Y10511
CD 190	1000	New sequence	Y10508
CD 203	480	New sequence	Z83940
CD 207	560	New sequence	Y10202
CD 212	529	New sequence	Z83946
CD 218	1100	New sequence	Z83937
CD 219	442	New sequence	Z83950
CD 240	800	New sequence	Z83942
CD 254	407	New sequence	Z83945
CD 274	570	New sequence	Z83953
CD 282	615	New sequence	Y10512
CD 33	503	Identical to *H. Sap.* mRNA for Na^+-ph	
CD 134	436	Identical to a human line 1 repetitive element	
CD 235	514	Identical to *H. Sap.* gene encoding RING finger protein	
CD 23	403	Similar to human Kpnl repetitive sequence	Z83955
CD 58	900	Similar to mer10 repetitive element	Y10515–6
CD 75	589	Similar to bovine butyrophilin mRNA	Y10203
CD 216	672	Similar to human L1 repetitive element	Y10509
CD 231	613	Similar to human L1 repetitive element	Z83951
CD 296	320	Similar to repetitive elements	Z83949

different transcription units. Combining the data from the physical map with expression results, a minimum of 14 transcription units, in addition to already known genes of the region, were identified (Figure 2).

Identification of New Polymorphisms

The RNA-SSCP technology allowed us to detect 11 new polymorphisms (clones CD30L, CD36, CD64, CD77, CD117, CD140, CD176, CD212, CD218, CD235, CD282) as shown in Table 3. Details on the allele frequencies of each polymorphism are also reported. In each case, Mendelian inheritance was confirmed in 20 large kindreds.

Table 2 RNA analysis of the gene fragments corresponding to new sequences

Clones	*Size (Kb)*	*Widely expressed*	*Specific issue*
CD 20	2	+	–
CD 30L	5	+	–
CD 30H	6.5	+	–
CD 43	5.5	+	–
CD 64	4.5	+	–
CD 67S	2	+	–
CD 88	5	+	–
CD 110	5	+	–
CD 115	5.5	+	–
CD 123	2.8	–	liv., musc., kidn., panc.
CD 171	1.8	+	–
CD 176	2	+	–
CD 190	3	–	prostate
CD 212	5.5	–	liver
CD 219	5.5	+	–
CD 235	4	–	liver
CD 240	6.5	+	–
CD 254	2	+	–
CD 274	2/4	–	liver, kidney
CD 282	5	+	–
CD 36	smear		
CD 105	smear		
CD 117	smear		
CD 203	smear		
CD 218	smear		
CD 77	ND		
CD 140	ND		
CD 152	ND		
CD 207	ND		

mRNA sizes detected in 16 tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas, spleen, thymus, prostate, testis, ovary, small intestine, colon, leukocytes) are indicated. Fragments displaying only limited expression patterns are indicated. Five cDNA clones (CD36, CD109, CD117, CD203, CD218) detected less discrete signals on the blots, indicating a series of RNAs or non-specific hybridisation, whilst for four clones (CD77, CD140, CD152, CD207) expression levels were below the sensitivity of the assay. ND: as data.

Isolation of Cosmid Clones

All the available fragments together with Alu-PCR probes derived from the YAC clones were used to screen a chromosome 6 specific cosmid library to detect a total of 287 cosmid clones. The cosmid clones allowed for the ordering and precise mapping of the different STSs. The compiled results of these experiments are shown in Figure 2.

Genomic Structure and Orientation of HFE gene

The human *HFE* gene spans approximately 9.5 Kb and the open reading frame (ORF) is contained in 6 exons. An additional non-coding exon is located within the 3' UTR. The size of the 7 introns ranges from 160 bp (intron 4) to approximately 3 Kb (intron 1). All exon/intron splice junctions conform to the eukaryotic 5' donor 3' acceptor consensus splice junctions GT/AG rule (Table 4). Junctions at all introns are type 1 (splicing occurring after the first based of the codon). Two cosmid clones (23B12 and 31H6) were further analysed to characterise the upstream promoter region. Among the 448 bp of the putative regulatory regions, we noted a region relatively GC rich (64% from base 1 to –78) and the absence of classical TATA and CAAT boxes. Thus, the human *HFE* promoter has structural features in common with those of several housekeeping genes and is consistent with the ubiquitous expression pattern of the mRNA in all tested human tissues.[1] Apart from the noted increase in GC content, 80–90 bp before the initiation codon, no significant homology of the putative regulatory sequence of the *HFE* gene has been found to promoter regions of other *HLA* class I genes. A more precise characterisation of the promoter by functional analysis with a reporter gene, and identification of the trans-acting factors will provide insights into the control of expression of the *HFE* gene.

Using primers designed from genomic sequence established by analysis of *HFE* genomic structure, two probes were generated corresponding to the 5' end of the *HFE* gene, and to the middle part of the gene, together with the EST 307162 which corresponds to the 3' end of the *HFE* gene. The three probes were used in hybridisation experiments on YAC and cosmid filters. The presence or absence of the three probes on cosmid clones mapped in the region and the simultaneous presence of the 3' end of the gene and marker *D6S2238*, that has been mapped centromerically to *HFE*,[1] on cosmid 39K16 allows us to determine that *HFE* is transcribed from telomere to centromere. Furthermore

our results showed that most of the *HFE* gene is contained in a portion of YAC790G7 which is deleted in our clone.

Discussion

The generation of transcription maps is an important step towards the definition of the human genome and the identification of functioning genes. We focused our attention on a 1 MB region of 6p22, where the *HH* locus was positioned. Using a large YAC from our previously assembled contig,[2] cDNA selection experiments were carried out.

Thirty-seven gene segments have been isolated; 28 of them corresponded to new sequences, while the remaining nine shared overlapping sequence or were similar to previously known genes. The isolated cDNAs together with established or published STSs and markers provided sufficient landmark density to cover approximately 90% of the 1 Mb interval with cosmid clones. mRNAs sizes and expression patterns were used to establish the number of transcription units identified by the 37 cDNA fragments analysed. In most cases, for the fragments that map close together different expression patterns or different RNA sizes were detected. Combining the data from the physical map with expression results, at least 14 transcription units in total were detected in addition to known genes of the region such as *HFE*, *NPT1*, *BTN* etc.

The absence of integration between either a detailed genetic or a medium-range physical map of the region does not allow comparisons for more precise locations of the transcription units identified. Our data clearly suggests the presence of a high gene density in this region of 6p22 with an average gene frequency of a gene approximately every 80 Kb. Our data are in agreement with those recently reported in a 1 Mb transcription map of the *HFE* region, that describes 20 novel transcripts.[12] This latter map extends more centromerically, but a minimum overlap of approximately 300 Kb with our map occurs in the area of the

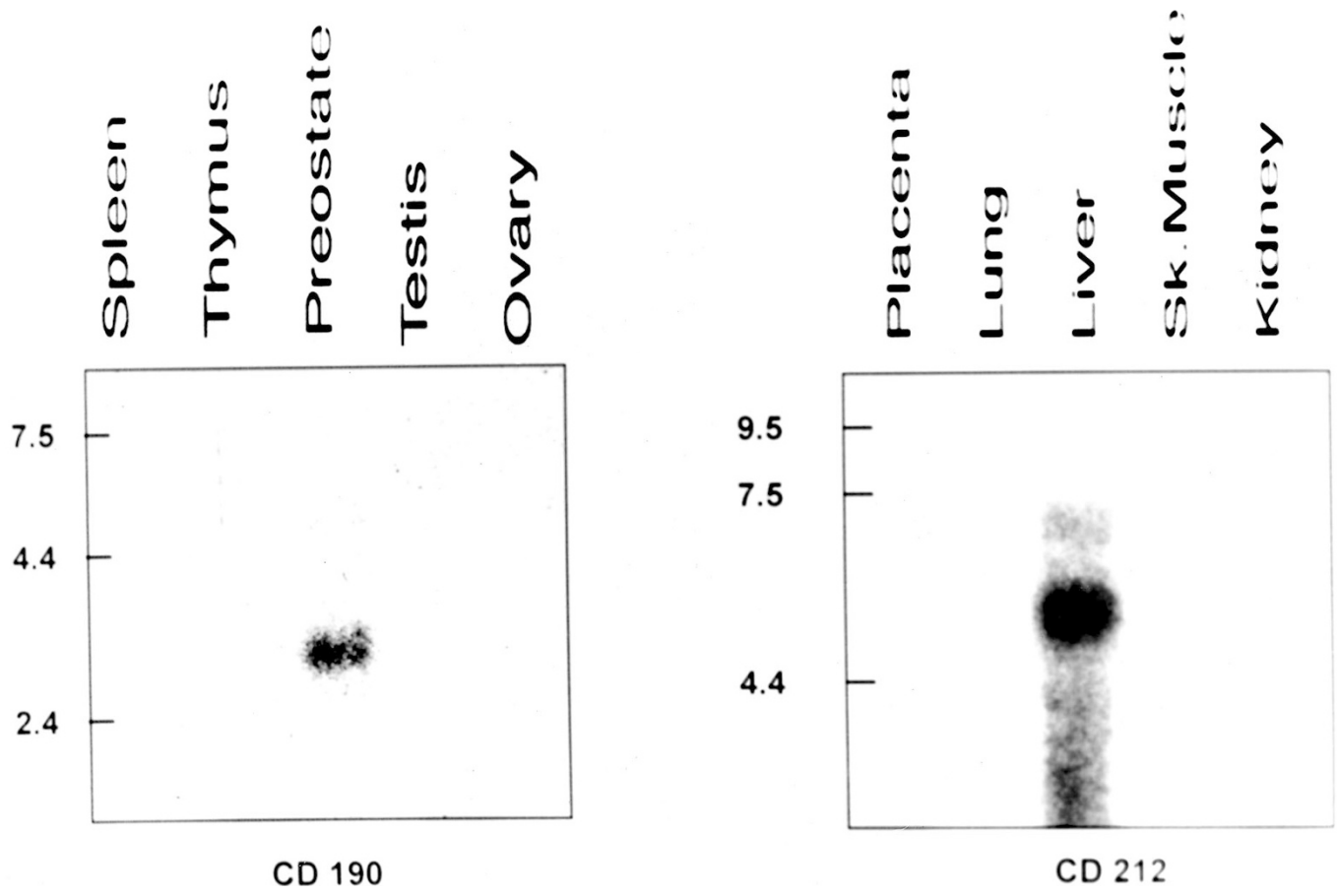

Figure 1 *Northern results of clones CD190 and CD212 that show specific tissue expression in prostate and in liver respectively*

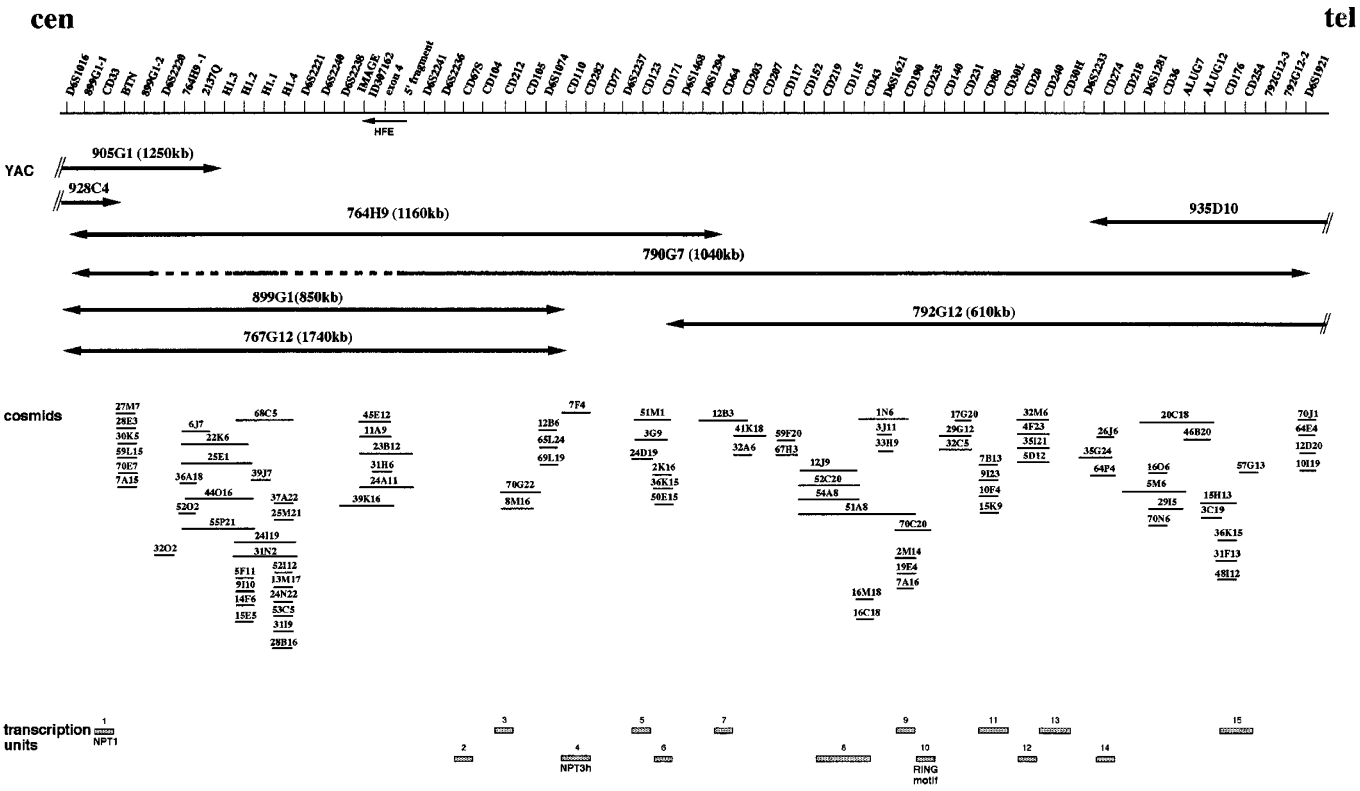

Figure 2 *The transcription/physical map of the HFE gene region. The isolated cDNAs and the class I genes were positioned by bins on cosmids. The relative position of each gene fragment within a given cosmid is not known*

Table 3 Description of the polymorphisms identified by 12 gene fragments

Clones	*Polymorphisms*	*Oligonucleotide sequences*	*Annealing temp. (°C)*	*Amplification product size (bp)*	*Frequency Allele 1*	*Allele 2*
CD 30L	Mae III	F-TCTAAGGAAGTCAGCCCGCC R-TGCTGAAATTGGAAAGGTGA	60	376	75%	15%
CD 36	Nla III	F-GACAGCACAAATGTTGAAGT R-CCACTTCAGAGAGAGTAACA	60	327	67%	33%
CD 64	Dde I	F-GATTCAGGCAAATTCTCTAA R-TCTGACAGTTGTGAAAGGAC	60	185	65%	35%
CD 77	Rna-SSCP	F-GAATGGACAGCATGGTTTC R-ATCCATAGTTTGTCTTCCTT	58	415	82%	18%
CD 117	Rna-SSCP	F-CACAACTGTCCAACAGAATC R-TGAGGCTATAGATGAGGATG	58	696	89%	11%
CD 140	Rna-SSCP	F-AGCGTACACTCTTGAGGGAT R-GCTTCAGGTTACCTCAGC	58	448	77%	23%
CD 176	Rna-SSCP	F-GCAGAGCAGCAACAACT R-ACAGTTAGACATTGAAGCT	58	265	93%	7%
CD 212	Rna-SSCP	F-GAGCAAGCCTCCTTGCAATGT R-CTAAATCCAGACTCTGGTAGG	54	383	84%	16%
CD 218	Nsi I	F-CAGCACAGTGTACCTCATAG R-CCAACAACTGGTACAAATGC	55	266	86%	14%
CD 235	Dde I	F-ATTATAGCACAGTGTGGTGG R-ATGGCACACATCTGCCCCTAA	60	266	73%	27%
CD 282	Rna-SSCP	F-CTACAATGCTGAGTTGCACAG R-ATTTGGGTCAAATGGGCT	58	467	71%	29%

histone genes. Transcription units 1 (*NPT1*), and *BTN* (butyrophilin gene) are contained in both maps. For the sodium-phosphate transporter *NPT1* gene, recently described,[11] we detected two hybridisation signals, one very strong in the area reported in Figure 1 and one somewhat weaker in the region defined by Ruddy *et al*.[12] In addition, another sodium-phosphate pump gene, highly homogolous to *NPT3* gene and thus named *NPT3h*, was isolated and placed telomeric to *NPT3* and to marker *D6S2236* of the Ruddy map. Combining the data from the two maps, it is possible to suggest that this region arose from several duplications of genes such as *butyrophilin*, *histone*, *HLA* class I, *NPT* and RING finger protein, yielding a high gene density.

During the construction of our map, the *HFE* gene was isolated from this interval.[1] By defining its genomic structure and integration with the map the transcriptional orientation of the *HFE* gene from telomere to centromere was clearly determined. The *HFE* genomic structure will facilitate the search for mutations from genomic DNA in all patients. A consistent proportion of our HH chromosomes of Southern Europe are negative for the presence of the *C282Y* mutation.[13,14] The possibility of an extensive analysis of the *HFE* gene at DNA level will help to define the exact proportion of HH patients carrying a mutated allele, and further support or exclude the possibility of genetic heterogeneity.

Finally, the new polymorphisms here described could be useful for linkage analysis and for the construction of at-risk HH haplotypes.

In conclusion, we describe the isolation and mapping of several new gene fragments on 6p22. Our map extends the recently published map increasing the number of transcripts in the area of the *HFE* gene. Additional genes, both known and novel, could be isolated either by increasing the number of unique clones characterised in detail, or by additional cDNA selection experiments with cDNAs from tissues not used in the present study. The novel fragments have already been useful as single-copy probes for screening larger-insert cDNA libraries, in order to complete their respective gene characterisation and for refining the physical map. This material forms a useful source for completion of the full transcript map of 6p22, for the future identification of defects and genes located to this region, and for developing template resources for genomic sequencing.

Acknowledgements

The first two named authors of this paper contributed equally to this work. JM Rommens is a scholar of the Medical Research Council of Canada, and BIOMED. This work was supported by grants from the Italian Ministry of Health, AIRC (Associazione Italiana Ricerca Cancro), Telethon (Grant E149 and E537).

Table 4 Brief description of *HFE* intro-exon boundaries and primers useful for amplifying at genomic level

Exon	*Position*[a]	*Donor sequence*	*Acceptor sequence*	*Primers (5′ to 3′)*
1	1–297		GCTGCgtgagt	F-TTCACAAGCAGGTACCTTCT
				R-ggtcctccaaagttagcaaac
2	298–561	ctccagGTTCA	CAAGGgtatgt	F-acatggttaaggcctgttgc
				R-agctgtttccttcaagatgc
3	562–837	ttccagAGTCC	ACAAGgtatgg	F-aatagggacctattcctttgg
				R-gtagaaaagctctgacaacct
4	838–1113	gtcaagTGCCT	CTGGGgtatgt	F-tggcaagggtaaacagatcc
				R-ctcaggcactcctctcaacc
5	1114–1227	ttttagAGCCC	TTCAAgtgagt	F-gtatgtgactgatgagagcca
				R-cagaggtactaagagacttc
6	1228–2304	ccacagGAGGA	AGCGGgtggat	F-tagtgcccaggtctaaattg
				R-gttagccaggatggtctcga
7	2305–2746	ccaaagTGCTG		F-accatgttggccaggctgg
				R-CAATACAATGTACATTAT

EMBL accessior numbers for the entire genomic structure are: Y09801 (promoter and exon 1), Y09800 (exon 2 and 3), Y09803 (exon 4 and 5), Y09799 (exon 6), Y09802 (exon 7).

[a]Exon sequences are in capital letters; intron sequences are in lower case letters; exon positions are in coding sequence;

References

1 Feder JN *et al*: A novel *MHC* class I-like gene is mutated in patients with hereditary haemochromatosis. Nature Genetics 1996; **13**: 399–408.
2 Malaspina P *et al*: Construction of a YAC contig covering human chromosome 6p22. *Genomics* 1996; **36**: 399–407.
3 Franco B *et al*: A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* 1991; **353**: 529–536.
4 Wapenaar MC *et al*: A YAC based binning strategy facilitating assembly of cosmid contig: 1.6 megabases of overlapping cosmid in Xp22. *Hum Mol Genet* 1994; **3**: 1155–1161.
5 Ferrero GB *et al*: An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3–Xp21.3. *Hum Mol Genet* 1995; **10**: 1821–1827.
6 Rommens J *et al*: A transcription map of the region containing the Huntington disease gene. *Hum Mol Genet* 1993; **2**: 901–907.
7 Rommens J *et al*: Towards a transcriptional map of the q21-q22 region of chromosome 7. In: Hochgeshwender W, Gardiner K (eds). *Identification of Transcribed Sequences*. Plenum Press: New York, 1994, pp65–79.
8 Sanger F, Nicklen S, Coulson A: DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 1977; **74**: 5463–5467.
9 Sarkar G, Yoon H-S, Sommer S: Screening for mutations by RNA single-strand conformation polymorphisms (rSSCP): comparison with DNA-SSCP.. *Nucleic Acids Res* 1992; **20**: 871–878.
10 Bisceglia L, Grifa A, Zelante L, Gasparini P: Development of RNA SSCP protocols for the identification and screening of CFTR mutations: identification of two new mutations. *Hum Mut* 1994; **4**: 136–140.
11 Chong SS, Kristjansson K, Zoghby HY, Hughes MR: Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23. *Genomics* 1993; **18**: 355–359.
12 Ruddy DA *et al*: A 1.1 transcript map of the hereditary hemochromatosis locus. *Gen Research* 1997; **7**: 441–456.
13 Borot N *et al*: Mutation in the *MHC* class I-like candidate gene for hemochromatosis in French patients. *Immunogenetics* 1997; **45**: 320–324.
14 Carella M *et al*: Mutation analysis of *HFE* gene in Italian hemocromatosis patients. *Amer J Hum Gen* 1997; **60**: 828–832.