Gene mapping and medical genetics

Molecular genetics of human chromosome 4

JAMES A GUSELLA, T CONRAD GILLIAM, MARCY E MACDONALD, SHIRLEY V CHENG, AND RUDOLPH E TANZI

From the Neurogenetics Laboratory, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, USA.

SUMMARY The recent discovery that the gene causing Huntington’s disease (HD) resides on chromosome 4 has generated increased interest in this autosome. Chromosome 4 contains two of the more informative conventional genetic markers, GC and MNS, but most loci have been assigned to it by recombinant DNA techniques. There are currently more anonymous DNA fragments detecting restriction fragment length polymorphisms (RFLPs) on chromosome 4 than on any other autosome. In addition, most of the cloned genes from this chromosome detect useful RFLPs. A genetic linkage map including both conventional and DNA markers should soon span the entire chromosome and will undoubtedly lead to the localisation of other inherited disorders.

Chromosome 4 comprises 6·5 % of human genomic DNA and therefore encodes perhaps 5000 to 10000 genes. As with all human chromosomes, only a very small minority of the genes on chromosome 4 have been identified. Taking together all genes, pseudogenes, cell surface antigens, fragile sites, and anonymous DNA segments given locus symbols, 69 loci had been officially assigned to the chromosome at the Human Gene Mapping Workshop 8.1 These loci have been mapped by many different techniques including somatic cell genetics, dosage studies, and family studies, but recombinant DNA methods have produced the bulk of the assignments. The chromosome is second only to the X chromosome and chromosome 21 in the number of cloned sequences available.2 At least 16 of the cloned segments represent known genes or pseudogenes (table), while the remainder are anonymous DNA fragments with no known function. The vast majority of DNA segments assigned to chromosome 4 detect restriction fragment length polymorphisms (RFLPs). With the great number of new markers now available, it is expected that a genetic linkage map encompassing the entire chromosome will soon be constructed. This development will set the stage for the systematic pursuit of family studies to identify additional loci causing human genetic disease, as well as facilitate the regional localisation of newly cloned genes.

Genes mapped to chromosome 4

Thirty-one genes or pseudogenes have been mapped to chromosome 4 by a variety of methods and they are listed in the table. Many of the genes were assigned purely by traditional methods, such as family studies, dosage in cells with chromosomal aberrations, and somatic cell genetics. The loci encoding albumin (ALB), vitamin D binding protein (GC), and the fibrinogens (FGA, FGB, FGG) were originally mapped to chromosome 4 by linkage analysis or dosage experiments, but when each of these genes was cloned it was possible to confirm the assignments directly by recombinant DNA techniques.3-15 Several of the genes on chromosome 4, such as interleukin 2 (IL2) and epidermal growth factor (EGF), were mapped only after cloning of cDNAs permitted DNA hybridisation to panels of somatic cell hybrids or in situ hybridisation.

Two of the genes that have not yet been isolated, PGM2 and PEPS, encode the biochemical activities

Received for publication 10 February 1986.
Accepted for publication 12 February 1986.
**TABLE Genes mapped to chromosome 4**.

<table>
<thead>
<tr>
<th>Gene symbol</th>
<th>Description</th>
<th>Location</th>
<th>Cloned DN</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD</td>
<td>Huntington's disease</td>
<td>p16, linked to D4S10</td>
<td></td>
<td>60-68</td>
</tr>
<tr>
<td>KLF2</td>
<td>Processed pseudogene of raf oncogene</td>
<td>p15-pter</td>
<td>+</td>
<td>32</td>
</tr>
<tr>
<td>PGM2</td>
<td>Phosphoglucomutase 2</td>
<td>p14-q12</td>
<td></td>
<td>16-18</td>
</tr>
<tr>
<td>P1PN</td>
<td>Peptidase S</td>
<td>p11-q12</td>
<td></td>
<td>18-21</td>
</tr>
<tr>
<td>PP1AT</td>
<td>Phosphoribosylpyrophosphate amidotransferase</td>
<td>pter-q21</td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>MT2P1</td>
<td>Metallocarboxypeptidase 2</td>
<td>p11-q21</td>
<td>+</td>
<td>33-35</td>
</tr>
<tr>
<td>TAP</td>
<td>Alpha-1 protein</td>
<td>q11-q13</td>
<td>+</td>
<td>6-23</td>
</tr>
<tr>
<td>HP15P</td>
<td>Hemidextrin persistence of alpha-1 protein</td>
<td>q11-q13</td>
<td>+</td>
<td>3-7-23</td>
</tr>
<tr>
<td>GC</td>
<td>Group-specific component; vitamin D binding protein</td>
<td>q12-q13</td>
<td>+</td>
<td>3-8-10</td>
</tr>
<tr>
<td>ASSP8</td>
<td>Argininosuccinate synthetase pseudogene</td>
<td>q21-qter</td>
<td>+</td>
<td>36</td>
</tr>
<tr>
<td>DGI1</td>
<td>Dominogenicity imperfect</td>
<td>Linked to GC</td>
<td></td>
<td>57-59</td>
</tr>
<tr>
<td>ADH1</td>
<td>Alcohol dehydrogenase, class 1</td>
<td>q21-q25</td>
<td>+</td>
<td>25</td>
</tr>
<tr>
<td>ADH2</td>
<td>Alcohol dehydrogenase, class 1</td>
<td>q21-q25</td>
<td>+</td>
<td>24-25</td>
</tr>
<tr>
<td>ADH3</td>
<td>Alcohol dehydrogenase, class 2</td>
<td>q21-q25</td>
<td>+</td>
<td>24</td>
</tr>
<tr>
<td>ADH5</td>
<td>Alcohol dehydrogenase, class III</td>
<td>q21-q25</td>
<td>+</td>
<td>26</td>
</tr>
<tr>
<td>IDH</td>
<td>Pyruvate dehydrogenase</td>
<td>q21-q25</td>
<td>+</td>
<td>73-76</td>
</tr>
<tr>
<td>ASG1</td>
<td>Aspartateglutaminidase</td>
<td>q21-qter</td>
<td>+</td>
<td>54-55</td>
</tr>
<tr>
<td>FG1</td>
<td>Epidermal growth factor</td>
<td>q25-q27</td>
<td>+</td>
<td>40-42</td>
</tr>
<tr>
<td>IL2</td>
<td>T cell growth factor, interleukin 2</td>
<td>q26-q28</td>
<td>+</td>
<td>38-39</td>
</tr>
<tr>
<td>FG2</td>
<td>Fibrinogen alpha chain</td>
<td>q26-q28</td>
<td>+</td>
<td>13-14</td>
</tr>
<tr>
<td>FGB</td>
<td>Fibrinogen beta chain</td>
<td>q26-q28</td>
<td>+</td>
<td>13-15</td>
</tr>
<tr>
<td>FGG</td>
<td>Fibrinogen gamma chain</td>
<td>q26-q28</td>
<td>+</td>
<td>11-13-15</td>
</tr>
<tr>
<td>MNS</td>
<td>MNS blood group; glycoporphins A &amp; B</td>
<td>q28-q31</td>
<td>+</td>
<td>27-30-45</td>
</tr>
<tr>
<td>SF</td>
<td>Steatosis blood group</td>
<td>Linked to MNS</td>
<td></td>
<td>46-47</td>
</tr>
<tr>
<td>TYS</td>
<td>Sclerotinosis</td>
<td>Linked to MNS</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>CD2D</td>
<td>Leucocyte surface antigen 5 (CD2p60)</td>
<td>Not regionally mapped</td>
<td></td>
<td>77</td>
</tr>
<tr>
<td>MD11</td>
<td>T cell antigen detected by monoclonal Ab</td>
<td>Not regionally mapped</td>
<td></td>
<td>78</td>
</tr>
<tr>
<td>MIC16</td>
<td>ALL antigen detected by monoclonal Ab</td>
<td>Not regionally mapped</td>
<td></td>
<td>79</td>
</tr>
<tr>
<td>QDPR</td>
<td>Quinolinate oxidoreductase</td>
<td>Not regionally mapped</td>
<td>+</td>
<td>49-51</td>
</tr>
<tr>
<td>NA</td>
<td>e-kit proto-oncogene</td>
<td>Not regionally mapped</td>
<td>+</td>
<td>37</td>
</tr>
</tbody>
</table>

NA = not assigned.

* The list does not include anonymous DNA sequences or fragile sites which can be found in references 1 and 2.

most frequently used as markers for the presence of chromosome 4 in somatic cell hybrid experiments.16-21 The PPAT locus, also not yet cloned, was mapped by complementation in auxotrophic mutant hamster cells deficient in the enzyme phosphoribosylpyrophosphate amidotransferase.22 This makes PPAT a potential selectable marker for directed somatic cell genetic studies of chromosome 4.

The approximate locations of the known genes that have been regionally mapped on chromosome 4 are shown in the figure. Several of the genes appear to cluster into families of related sequences. Below the centromere, ALB and AFP encode related proteins and are separated by only a few thousand base pairs.23 Moreover, the cloning of GC which is tightly linked to the ALB locus has yielded sequence data suggesting that it is a member of the same gene family.10 Further down the long arm, there is a set of genes encoding alcohol dehydrogenases.24-26 It is likely that when these are fully sequenced and characterised they will consist of related sequences tightly clustered on the chromosome. The MNS blood group locus results from polymorphisms in glycoporphin A (MN) and glycoporphin B (SS), two related proteins encoded by tightly linked DNA sequences.27-30 A third glycoprotein from red blood cell membranes, glycoporphin C, might well be encoded in the same region.31 Finally, towards the bottom of the long arm there is a cluster of three fibrinogen genes which are contiguous in a 50 kilobase segment.15

In contrast to the clustering of the functional gene families, there are three loci on chromosome 4 that
Molecular genetics of human chromosome 4

![Diagram](image)

**FIGURE** Regionally localised genes on chromosome 4. The order and regional localisations are only approximate.

Each represent a processed pseudogene, whose functional counterpart is located on an entirely different chromosome. RAF2, MT2PI, and ASSP8 are processed pseudogenes homologous to the RAF1 protooncogene (RAFI) on chromosome 3, the metallothionein-2 gene (MT2) on chromosome 16, and the argininosuccinate synthetase gene (ASS) on chromosome 9, respectively.

The recent mapping of the human homologue of v-kit, the viral oncogene from the Hardy-Zuckerman 4 feline sarcoma virus (HZ4-FeSV), to chromosome 4 is the first indication of a potential oncogene on this autosome. The only previous implication of chromosome 4 in carcinogenesis has been the observation of a translocation of the long arms of chromosomes 4 and 11 in some cases of acute lymphocytic leukaemia. This translocation joins the ETS1 oncogene locus from chromosome 11 and a chromosome 4 sequence that has not yet been identified.

**Conventional genetic markers, GC and MNS**

Two of the loci mapped to chromosome 4 by conventional methods, GC and MNS, show considerable polymorphism, placing them among the more informative expressed genetic markers. GC, group specific component or vitamin D binding protein, displays polymorphism that is detectable as differences in electrophoretic mobility of the protein. It has a polymorphism information content (pic) of 0.33.\(^{44}\)

The MNS blood group marker actually results from two tightly linked genes.\(^{27-30}\) The first, encoding glycoporphin A, the major red cell surface protein, displays a sequence polymorphism at its 5' end that determines the M or N specificities. This protein is highly glycosylated, with the presence of the appropriate sugar residues being essential for detecting the appropriate specificities.\(^{30}\) Glycoporphin B, a less abundant red cell surface protein, has an identical 5' amino acid sequence to the N form of glycoporphin A. The remaining sequence diverges, however, and a polymorphism in this region results in either S or s alleles, again detected as antigenic differences. Since the sequences encoding the two polymorphisms are close together in the genome, the resultant haplotypes are used, in practice, as alleles for a single locus. Less frequent variants at these two loci have produced proteins that lack either M or N activity, or that appear to be fusions between glycoporphins A and B.\(^{45}\) The MNS marker has a pic of 0.64, making it second only to HLA in informativeness among the expressed markers.\(^{44}\)

A third expressed polymorphism mapping to chromosome 4 is the Stoltzfus blood group marker.\(^{46-47}\) Like MNS, Stoltzfus (SF) is detected as differences in the antigenic determinants on red cell membranes. It is genetically linked to the MNS marker and is not frequently used in genetic linkage studies.

Regional localisation of the GC and MNS markers by dosage studies in families with rearranged chromosomes places them towards opposite ends of the long arm. Despite this apparently great physical separation, very extensive family studies have demonstrated statistically significant but distant linkage between the two loci. This linkage is detectable in male meioses, presumably reflecting a greater frequency of recombination on the long arm of chromosome 4 in females than in males.\(^{29-48}\)

**Disease genes mapped to chromosome 4**

In general, genetic diseases can be mapped to chromosomes either by family studies to detect coinheritance with a polymorphic marker or by...
demonstrating an alteration in a gene product from a known locus. The latter is more common, since infrequent variants have been described for many well characterised proteins. For chromosome 4, both approaches have contributed only a few genetic defects to the map.

The QDPR locus encoding dihydropteridine reductase is the presumed site of the defect in type II phenylketonuria where patients do not respond to a phenylalanine free diet.\(^49^-\text{51}\) This enzyme is involved in the production of biopterin, a necessary cofactor for the enzyme phenylalanine hydroxylase, which is defective in type I phenylketonuria. Analbuminaemia is a rare condition in which subjects fail to produce albumin from the ALB locus on the long arm of chromosome 4.\(^52\) A second condition, hereditary persistence of alphafetoprotein (HPAFP), is closely linked to the AFP and ALB loci and may represent a mutation in regulatory sequences at the alphafetoprotein locus. Recently, the AGA locus encoding aspartylglucosaminidase, the enzyme deficient in the lysosomal storage disorder aspartyl-glucosaminuria, has been assigned to the long arm of chromosome 4.\(^54^-\text{55}\)

Genetic linkage studies have been modestly successful on chromosome 4, although the high quality of GC and MNS as markers makes it somewhat surprising that this approach has not identified more disease loci. In fact, only two disease genes have been located using these markers. In 1968, scleroptilosis (TYS), an autosomal dominant skin disorder, was provisionally linked to MNS, although this assignment has never been confirmed.\(^56\) In 1982, Ball et al showed linkage of dentinogenesis imperfecta (DGI1), a defect in dentition, to GC.\(^57^-\text{58}\) This linkage has been confirmed by Conneally et al.\(^59\)

**DNA markers on chromosome 4**

The advent of recombinant DNA techniques for directly detecting polymorphisms in DNA has produced the potential for detecting loci by linkage anywhere in the genome. The discovery of a DNA marker linked to the Huntington's disease (HD) locus on chromosome 4 was the first successful use of this approach for an autosomal disease.\(^60\) HD is a late onset neurodegenerative disorder that shows autosomal dominant transmission. The initial linkage has now been confirmed in several families, although the possibility of non-allelic heterogeneity in HD has not yet been excluded.\(^61^-\text{63}\) The linked marker, D4S10, is detected using the anonymous single copy DNA probe G8. Although originally two polymorphic HindIII sites were used to generate four haplotypes for this locus, the discovery of several additional RFLPs with other enzymes has made this a highly informative locus displaying greater than 90% heterozygosity.\(^64\) The DNA marker is linked to HD at a distance of 4 to 5% recombination (1-loci confidence interval 1–8%).\(^1\)

The two loci have recently been mapped to the terminal band of the short arm, 4p16.\(^65^-\text{68}\) Gusella et al\(^65\) showed that in patients with Wolf-Hirschhorn syndrome, a congenital defect caused by the heterozygous deletion of this region, the D4S10 locus is also deleted. This result has been confirmed by in situ hybridisation by several groups.\(^66^-\text{68}\) In one study it was possible using patients with interstitial deletions of the short arm to narrow the assignment of D4S10 to the terminal half of 4p16.1, one of three sub-bands seen in prometaphase banding experiments.\(^68\) The location of D4S10, and by inference HD, to the terminal part of the short arm explains the failure to detect the disease locus by linkage to either GC or MNS on the long arm.\(^60\)

The mapping of HD to chromosome 4 has provided the impetus to generate large numbers of anonymous DNA markers for this autosome. These now represent more than half of all the assigned loci on chromosome 4. A few markers have been mapped using somatic cell hybrid panels after they were identified in searches for RFLPs using DNA probes chosen at random from genomic DNA.\(^60\) This mapping is being confirmed by in situ hybridisation by several groups.\(^69^-\text{71}\) The majority of the anonymous DNA sequences, however, have come from libraries enriched for chromosome 4 sequences by flow sorting of metaphase chromosomes.\(^71^-\text{72}\) Subsequent screening for RFLPs has revealed that most of these can be successfully used as genetic markers in linkage studies. A total of 37 anonymous DNA loci had been described for chromosome 4 by Human Gene Mapping Workshop 8, and 30 of these represent useful polymorphic markers.\(^1\) Many additional anonymous DNA loci are currently being characterised (Gusella et al, unpublished results).

Of the cloned genes and pseudogenes from chromosome 4, 12 detect RFLPs useful in linkage analysis, although this number is slightly misleading due to the tight clustering of some related genes\(^3\) (Gusella et al, unpublished results). Overall, including both anonymous sequences and cloned genes, there are currently more DNA markers for chromosome 4 than for any other autosome.\(^3\)

No additional markers linked to HD have yet been identified from this pool of DNA loci, perhaps because of its terminal location on the short arm. An intensive effort is now underway to saturate 4p16 with new DNA markers. These are needed to increase the accuracy and applicability of any presymptomatic or prenatal diagnostic test procedures in HD which could potentially be based on linkage analysis. They will also be required to refine
Molecular genetics of human chromosome 4

genetic mapping of the HD locus in anticipation of cloning and characterising the defect.

A detailed linkage map of 4p16 will not be the only byproduct of the search for the HD gene. This effort has already produced a large number of anonymous DNA markers from both the long and short arms. These RFLPS, along with those detected by cloned probes from chromosome 4, are currently being traced through large reference pedigrees by several groups. This combined effort will soon generate a complete linkage map spanning all of chromosome 4. The vast array of as yet unidentified genes on chromosome 4 undoubtedly include several involved in genetic disease. The availability of a selection of spaced polymorphic markers for this autosomal will permit directed family studies with more efficient and powerful data analysis leading to the eventual localisation of these defects.

Preparation of this manuscript was supported by NIH grants NS16367 and NS20012. JFG is supported by the Searle Scholar Program/Chicago Community Trust. TCG and MEM are recipients of the Anna Mitchell Fellowship and the Grace E. Neuman Fellowship, respectively, from the Hereditary Disease Foundation. SVC received a fellowship from the National Huntington's Disease Association.

References
James F Gusella, T Conrad Gilliam, Marcy E MacDonald, Shirley V Cheng, and Rudolph E Tanzi


Van der Goes R, Geurts van Kessel A, Hagemeijer A, Wijnen
Molecular genetics of human chromosome 4


Correspondence and requests for reprints to Dr J F Gusella, Neurogenetics Laboratory, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
Molecular genetics of human chromosome 4.

J A Gusella, T C Gilliam, M E MacDonald, S V Cheng and R E Tanzi

doi: 10.1136/jmg.23.3.193