Gene mapping and medical genetics

Journal of Medical Genetics 1988, 25, 794–804

The map of chromosome 20

NANCY E SIMPSON
Division of Medical Genetics, Departments of Paediatrics and Biology, Queen's University, Kingston, Ontario, Canada.

SUMMARY The number of gene assignments to human chromosome 20 has increased slowly until recently. Only seven genes and one fragile site were confirmed assignments to chromosome 20 at the Ninth Human Gene Mapping Workshop in September 1987 (HGM9). One fragile site, 13 additional genes, and 10 DNA sequences that identify restriction fragment length polymorphisms (RFLPs), however, were provisionally added to the map at HGM9. Five mutated genes on chromosome 20 have a relation to disease: a mutation in the adenosine deaminase gene results in a deficiency of the enzyme and severe combined immune deficiency; mutations in the gene for the growth hormone releasing factor result in some forms of dwarfism; mutations in the closely linked genes for the hormones arginine vasopressin and oxytocin and their neurophysins are probably responsible for some diabetes insipidus; and mutations in the gene that regulates both α-neuraminidase and β-galactosidase activities determine galactosidosis. The gene for the prion protein is on chromosome 20; it is related to the infectious agent of kuru, Creutzfeld-Jacob disease, and Gersmann-Straussler syndrome, although the nature of the relationship is not completely understood. Two genes that code for tyrosine kinases are on the chromosome, SRC-E, the proto-oncogene and a gene (HCK) coding for haemopoietic kinase (an src-like kinase), but no direct relation to cancer has been shown for either of these kinases. The significance of non-random loss of chromosome 20 in the malignant diseases non-lymphocytic leukaemia and polycythaemia vera is not understood. Twenty-four additional loci are assigned to the chromosome: five genes that code for binding proteins, one for a light chain of ferritin, genes for three enzymes (inosine triphosphatase, s-adenosylhomocysteine hydrolase, and sterol delta 24-reductase), one for each of a secretory protein and an opiate neuropeptide, a cell surface antigen, two fragile sites, and 10 DNA sequences (one satellite and nine unique) that detect RFLPs.

Since 1973, regular human gene mapping (HGM) workshops have been held every few years at which collective decisions have been made and published regarding the nomenclature and assignments of genes and loci to specific chromosomes based on published reports and data brought to the workshop. At the workshops the position of a locus is considered to be confirmed when data from two or more independent laboratories support the same assignment; if the assignments are not the same, the locus is designated inconsistent. When one group of investigators maps a gene or locus it is considered a provisional assignment until confirmed by an independent laboratory. The most recent Human Gene Mapping Workshop (HGM9) was held in September 1987 and until then chromosome 20 had been sparsely populated with mapped genes and DNA sequences in comparison to its sister chromosome 19 (for review of the latter see Shaw et al2).

The usual catalyst for extensive mapping of an autosome has been the chromosome assignment of the locus for an important genetic disease when the gene has no known product, for example, Huntington’s disease on chromosome 4,3 4 cystic fibrosis on chromosome 7,5-7 and myotonic dystrophy on chromosome 19.2 Although a locus for a disease with no known gene product had not been mapped to chromosome 20, at Human Gene Mapping Workshop 6 (HGM6)8 a provisional assignment of the locus for multiple endocrine neoplasia type 2A (MEN 2A), a hereditary curable cancer, sparked interest in the chromosome, even though the assignment proved to be incorrect.9 10 The assignment of genes which code for interesting proteins can also...
The map of chromosome 20

stimulate mapping of a chromosome, such as those in the haemoglobin cluster on 11p[11] and the HLA loci on chromosome 6 together with their associations to a number of chronic diseases.12 At HGM9 there were only eight confirmed assignments to chromosome 20 (table 1). Increasing interest in the chromosome, however, is reflected in that one satellite DNA sequence and nine anonymous unique sequences that detect restriction fragment length polymorphisms (RFLPs) (table 2) were listed and one tentative and 13 provisional new assignments were made at HGM9 in 1987.13

Chromosome 20 is one of the two smallest metacentric chromosomes, the other being chromosome 19. These F group chromosomes can be distinguished quite readily from each other by their banding patterns. Trisomy 20 is not usually viable and only one case has been reported.14 A few cases of partial trisomy 20p have been described and for the most part it is a recognisable syndrome.15-20 Partial trisomy 20q has also been reported but occurs even less frequently than partial trisomy 20p.21 On the other hand, trisomy 20 mosaicism occurs relatively frequently in amniotic fluid samples taken for prenatal diagnosis of genetic disease and more than 50 cases have been reported.22 As with all mosaicisms in culture, it presents a serious genetic counselling problem. The problem is increased as there is no specific birth defect associated with the mosaicism seen in the fetuses in which it has been confirmed.22 From the above one can see that chromosomal abnormalities, in general, have not drawn particular attention to the chromosome and will not be reviewed in detail. The subject of the review will be primarily the genes on the chromosome, with an emphasis on those related to disease.

Major loci related to or possibly related to disease

ADENOSINE DEAMINASE (ADA) AND SEVERE COMBINED IMMUNE DEFICIENCY (SCID)

Since HGM123 the locus for the rare, but very serious, disease severe combined immune deficiency (SCID) has been known to be on chromosome 20 by

### TABLE 1 Genes and fragile sites on chromosome 20.

<table>
<thead>
<tr>
<th>Name and category at HGM9</th>
<th>Regional assignment</th>
<th>Polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confirmed Adenosine deaminase (ADA) deficiency results in severe combined immune deficiency (SCID)</td>
<td>20q13.2—q34 or q12</td>
<td>+</td>
</tr>
<tr>
<td>s-adenosylhomocysteine hydrolase (AHCY); reduced activity in SCID secondary to ADA deficiency</td>
<td>20 cen—q13.1</td>
<td>—</td>
</tr>
<tr>
<td>Prion protein (PRIP), the infectious agent of Creutzfeld-Jacob disease, kuru, and Gertschmann-Strausser syndrome</td>
<td>20p12—pter</td>
<td>+</td>
</tr>
<tr>
<td>Precursor for growth hormone releasing factor (GHRF) or somatocrin; deficiency results in one of the forms of hypothalamic dwarfism</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Avian sarcoma viral (v-src) oncogene homologue (SRC1); the locus is non-randomly lost in some patients with non-lymphocytic leukaemias and polycythaemia vera</td>
<td>20q12—q13</td>
<td>—</td>
</tr>
<tr>
<td>A light polypeptide-like ferritin</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Insolone triphosphatase (ITPA); the biological substrate and function unknown</td>
<td>20p</td>
<td>—</td>
</tr>
<tr>
<td>Fragile site (FRA20A), follic acid induced, rare</td>
<td>20p11.23</td>
<td>—</td>
</tr>
</tbody>
</table>

### Provisional

<table>
<thead>
<tr>
<th>Name and category at HGM9</th>
<th>Regional assignment</th>
<th>Polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arginine vasopressin (AVP), its carrier protein neurophysin II, and diabetes insipidus</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Prepro-oxytocin (OT), its carrier protein neurophysin I, and diabetes insipidus</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Galactosialidosis (GSL), a deficiency of both α-neuraminidase and β-galactosidase</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Haemopoetic cell kinase (HCK), an src-like tyrosine kinase</td>
<td>q11—q12</td>
<td>—</td>
</tr>
<tr>
<td>Desmoprotease to cholesterol enzyme (DCE), probably sterol delta 24-reductase</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Guanine nucleotide binding protein (GNAS), an α-stimulated polypeptide</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>High leucine transport (HTL)</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Metallothionein-like 4 (MTL4), a metal binding protein</td>
<td>20</td>
<td>—</td>
</tr>
</tbody>
</table>

### Polymorphisms

<table>
<thead>
<tr>
<th>Name and category at HGM9</th>
<th>Regional assignment</th>
<th>Polymorphic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prodyphonin (PDYN), an opiate neupeptide</td>
<td>20p12—pter</td>
<td>+</td>
</tr>
<tr>
<td>Ribophorin II (RP2) binds ribosomes for secretory and membrane proteins</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Secretogranin 1 (SCG1) (chromogranin B), a secretory protein</td>
<td>pter—p12</td>
<td>—</td>
</tr>
<tr>
<td>Thrombomodulin (THBD), a cell surface receptor for thrombin</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Msk 38, a cell surface antigen identified by the monoclonal antibody 05</td>
<td>20</td>
<td>—</td>
</tr>
<tr>
<td>Fragile site (FRA20B), aphidicolin type, 'common'</td>
<td>20p12.2</td>
<td>+</td>
</tr>
</tbody>
</table>

### TABLE 2 Unique anonymous DNA sequences from chromosome 20 that detect restriction fragment length polymorphisms.

<table>
<thead>
<tr>
<th>Locus name</th>
<th>No of alleles</th>
<th>Location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>D2055</td>
<td>4</td>
<td>20p12</td>
<td>102 103</td>
</tr>
<tr>
<td>D2056</td>
<td>2</td>
<td>20p12</td>
<td>103</td>
</tr>
<tr>
<td>D2058</td>
<td>5</td>
<td>20q13</td>
<td>141</td>
</tr>
<tr>
<td>D2054</td>
<td>3</td>
<td>20q13.2</td>
<td>103 142</td>
</tr>
<tr>
<td>D20514</td>
<td>2</td>
<td>20</td>
<td>143</td>
</tr>
<tr>
<td>D20515</td>
<td>23</td>
<td>20</td>
<td>144</td>
</tr>
<tr>
<td>D20516</td>
<td>17</td>
<td>20</td>
<td>144</td>
</tr>
<tr>
<td>D20517</td>
<td>2</td>
<td>20</td>
<td>145</td>
</tr>
<tr>
<td>D20518</td>
<td>2</td>
<td>20</td>
<td>145</td>
</tr>
<tr>
<td>D2021</td>
<td>?</td>
<td>20c1-2a</td>
<td>146</td>
</tr>
</tbody>
</table>

virtue of the fact that it is caused by a deficiency of the enzyme adenosine deaminase (ADA) which had been assigned to the chromosome. The causal relationship between this enzyme deficiency and the disease was first recognised by serendipity. Giblett et al. found two unrelated patients with SCID and virtually no detectable ADA. Levels of the enzyme in both pairs of parents were reduced to 50 to 60% of normal. Furthermore, one of the pairs of parents was consanguineous. This suggested that the patients were homozygous and the parents heterozygous for a null allele for ADA. The mechanism whereby the enzyme deficiency in the purine pathway caused the immune deficiency characterised by T and B cell depletion, impaired lymphocyte response to mitogens, and declining immunoglobulins in the first few months of life was obscure at that time. Giblett et al. suggested that the inborn error might be in the metabolic pathway involved in the salvage of lymphoid cells at some point in the immune response, or that the mutation was a cytologically undetectable deletion that included the gene for ADA and an adjacent or closely linked immune response gene. This made sense then as it was thought that the HLA locus was linked to the ADA locus in man.

ADA had been well characterised; three isozymic forms had been described by Hopkinson et al., one of which had low activity but not reduced to the same extent as in SCID. Tischfield et al. assigned the locus for ADA to chromosome 20 at HGM1 by somatic cell hybridisation and this in turn assigned the locus for the disease. The locus was regionalised to 20q13.1→qter mainly by dosage studies of the enzyme in the cells of patients with various deletions or rearrangements of chromosome 20 and by use of somatic cell hybrids containing translocations involving chromosome 20. More recently, three groups have suggested that the ADA gene is at 20q12. These latter data make the sublocalisation of the locus inconsistent.

Both genomic and cDNA for ADA are available and the gene has been sequenced. It has been shown that most patients with SCID have normal amounts of mRNA by northern analysis, suggesting that the mutation for the disease is usually a point one. In some cases the mutation may be in a splicing region of the gene. The ADA gene in DNA from two patients with SCID was sequenced and the mutations were single amino acid changes. There is, however, one report of a deletion of the exon 1 sequence of the ADA gene.

S-ADENOSYLMETHIONINE HYDROLASE (AHCY) AND SCID
The enzyme s-adenosylhomocysteine hydrolase (AHCY) converts s-adenosylhomocysteine to adenosine and homocysteine and catalyses the reverse reaction. Its activity is dependent on the activity of ADA and it is impaired when ADA activity is reduced probably from the inhibition of 2'- deoxyadenosine, one of the substrates for ADA. The AHCY activity is reduced in patients with SCID but is obviously not the primary defect, as shown by the molecular studies of ADA in patients with SCID.

The locus that codes for AHCY has also been mapped to chromosome 20; it was first shown to be syntenic with that for ADA using a monoclonal antibody to identify which human/hamster hybrids expressed the enzyme. Later the gene for the enzyme was localised to cen→q13.1. It is of interest that the loci of the two functionally related enzymes ADA and AHCY may be even closer together than indicated by their regional assignments. This has been suggested by linkage studies, but more data are needed for a firm conclusion. The two enzymes have considerable homology and it is possible that AHCY arose from a tandem reduplication of or recombination with a portion of the ADA gene. The evidence points to ADA being the older gene; AHCY occurs only in eukaryotes, whereas ADA occurs in both eukaryotes and prokaryotes.

PRION PROTEINS (PRNP) AND DEGENERATIVE ENCEPHALOPATHIES
Perhaps the most interesting gene that has been mapped recently to chromosome 20 is a unique sequence that codes for the human prion protein (PrP 27-30), a 27 to 30 kDa sialoglycoprotein. The biological significance of this protein is still controversial but it was discovered during the attempts by Prusiner to identify the infectious agent responsible for Creutzfeld-Jakob disease (CJD), kuru, and Gerstmann-Strassler syndrome in humans, and scrapie in sheep and goats. The cause of this group of degenerative encephalopathies has been the subject of intense investigation for more than 20 years since Gajdusek et al. showed that kuru could be transmitted from diseased brain tissue to chimpanzees. The term prion was coined by Prusiner to describe a proteinaceous agent capable of giving hamsters scrapie but distinguishable from viruses, plasmids, and viroids since no nucleic acid could be found. The scrapie agent was shown primarily to be a protein that aggregates into amyloid-like structures; it is not infectious in the purified form. The availability of the purified form of the protein has made it possible to use reverse genetics to isolate cDNA probes that code for the prion protein from both hamster and human DNA. It is clear, however, that the protein itself is not the infectious agent since normal as well as affected brains contain its mRNA. The proteins...
The map of chromosome 20

from normal and scrapie infected brains differ immunologically\(^{55}\) and the isoforms are reviewed by Diener.\(^{59}\) In the last few years several hypotheses have emerged to explain the observations. Brown\(^{60}\) has suggested three hypotheses. (1) The infectious agent may yet prove to be a virus although nobody has found nucleic acid in the prion protein. (2) The agent may be a virino with a minimal amount of nucleic acid, difficult to detect, but able to induce disease by disrupting the regulatory functions of the nucleic acids of the host. (3) The prion protein is self replicating. To accept the latter hypothesis by current dogma the demonstration of nucleic acid is still needed. Bandea\(^{61}\) suggests that the pathogen is an inherited endogenous virus which he calls a prionic virus. The prionic virus produces the prion analogous to a pathogenic bacterium producing a toxin. Prions produced by prionic viruses may in turn stimulate other viruses to form different prions with pathological characteristics. This hypothesis fits much of the data but is not yet proven. The relation of the protein to the encephalopathies may be found within the next few years now that CDNA, mRNA, and monoclonal antibodies to the protein are available (reviewed in Prusiner et al\(^{62}\)).

Although the significance of the prion protein and its relation to the degenerative encephalopathies are still not clear, the assignment of its gene (PRIP) to chromosome 20 is confirmed. With a human cDNA probe that codes for the prion protein, the gene was localised to chromosome 20 using filter bound chromosome DNA after chromosome sorting;\(^{58}\) localised by in situ hybridisation to 20p by Robakis et al\(^{63}\) and to the smaller region of 20p12pter by Sparkes et al.\(^{64}\) The protein is well conserved in mammals and the PRIP gene is syntenic on the murine chromosome 2 with three other genes which are on human chromosome 20, ITPA, ADA, and SRC1. The DNA sequence for the human prion protein detects a PvuII polymorphism with gene frequencies of 0.1 and 0.9.

The peptides have been clinically useful in establishing the basic defect in some forms of pituitary dwarfism.

The cDNA for the GRF hypothalamic peptide (GRF-44) or somatocrinin was isolated and sequenced by Gubler et al.\(^{68}\) The clone was made from pancreatic tumour tissue mRNA (for G-40) after biological, immunological, and physico-chemical evidence indicated that the tumour peptide and the hypothalamic peptides were identical. Gubler’s cDNA codes for a 44 amino acid peptide in the precursor of growth hormone releasing factor and the gene was designated GHRF at HGM9.\(^{69}\) This cDNA gene GHRF was used by Mayo et al\(^{70}\) to isolate and sequence genomic clones for the entire precursor of the hormone. Lebo et al\(^{71}\) assigned the GHRF gene to chromosome 20. Mayo et al\(^{70}\) showed that the gene contained five exons and spanned 10 kbp of human genomic DNA; they mapped the cDNA gene for the GHRF-44 peptide to chromosome 20 by dot blot analysis of DNA from high resolution dual laser sorted chromosomes. In the same year, Riddell et al\(^{72}\) mapped the cDNA for the peptide to chromosome 20 using a human/rodent somatic cell hybrid panel.

McKusick\(^{73}\) lists four forms of pituitary dwarfism (types I to IV) and two forms with sella turcica defects leading to pituitary dwarfism. Within these groups further heterogeneity exists. As will be seen below, not all dwarfism in these former classifications are the result of defects of pituitary factors. Type I has been known as isolated growth hormone (GH) deficiency in contrast to type III which is a result of deficiencies of several pituitary hormones including GH. Type I has been subdivided into IA with an absolute lack of growth hormone and IB with reduced amounts of growth hormone according to their response to the peptide GRF-40; an ability to produce GH when stimulated with the peptide suggested a deficiency of GHRF.\(^{74}\) Attempts were also made to discriminate between GH and GHRF dwarfism\(^{75}\) using the peptide GRF-44 and insulin provocative tests. Patients who were non-responders to both tests were considered to have growth hormone deficiency and those who responded by an increased production of growth hormone were considered to have lack of the hypothalamic growth hormone releasing factor. As yet it has not been possible to diagnose those patients with a GHRF deficiency using molecular techniques. Using the cDNA for the GHRF gene Parks et al\(^{76}\) were unsuccessful in finding differences in the DNA from patients with isolated type I (five families) or multiple pituitary type III dwarfism (three families). Their lack of success may have been the result of chance selection of inappropriate patients or non-
informative restriction enzymes, or the differences do not occur in the exons coding for the peptide. To our knowledge, no one has probed the DNA from patients with a known positive response to the peptides GRF-40 or GRF-44. In contrast, a deletion of the gene coding for human growth hormone (GH), known to be on chromosome 17, has been shown in members of a family using cDNA probes.77 These patients, as might be expected, form antibodies to the hormone and are thereby difficult to treat. Use of the probes for GHRF and GR to identify the biochemical defect in patients with clinical pituitary dwarfism would avoid the unpleasant GRF-44 and insulin tests.

THE PROTO-ONCOGENE SRC1 AND A GENE FOR HAEMOPOIETIC CELL KINASE (HCK)
The SRC1 locus on 20q codes for the human proto-oncogene c-src which is a cellular homologue of the avian retrovirus known as v-scr, the oncogene of the Rous sarcoma virus. The cellular genes are well conserved in vertebrates and the c-scr oncogene is no exception. It has been found in Drosophila which makes it probably older than 800 million years.78 The c-scr human proto-oncogene was mapped to 20cen→q13.1 by hybridisation of a v-src probe to a somatic cell hybrid panel.79 Furthermore, the c-src gene in mice mapped to chromosome 2. More recently, Le Beau et al80 assigned c-src by in situ hybridisation of a genomic c-scr probe to two sites, 20q12→q13 and 1p34→p36. Parker et al81 isolated two sequences from human libraries whose kinase domain and adjacent regions at the 5’ and 3’ ends were homologous to those of the c-src gene. When they sequenced the two DNAs they found that one sequence, which they named c-src1 (human) was more similar to c-src (chicken) than was the other that they designated c-src2 (human). The former proved to be the sequence that maps to chromosome 20 and is the functional gene that recognises a polyadenylated transcript of the protein kinase. The latter mapped to the human chromosome 1p as does a similar gene to the murine homologue chromosome 4. The sequences in the two probes appeared to be unique and rearrangements were not observed in either gene in a tumour line (A431, Kelly). Parker’s c-src2 gene is now thought to be c-fgr, the feline sarcoma viral oncogene, and does not appear to be an independent gene on 1p (A Y Sakaguchi, 1988, personal communication); the fgr gene had been mapped to 1p36.1→p36.2.82 The c-src1 gene was designated SRC1 and the former SRC2 as FGR (Gardner-Rasheed feline sarcoma viral oncogene homologue) at HGM9.83 Parker et al81 suggested that there are other DNA fragments in the human genome that are homologous to v-scr.

It is of considerable interest that recently a gene that codes for a human src-like tyrosine kinase primarily expressed in haemopoietic cells, known as haemopoietic cell kinase (HCK), has been assigned to the position 20q11→q12 in close proximity to the SRC1 locus.83 84 Further studies will tell us how close the genes for the two kinases are to each other and their relation to cancer, if any.

LOSS OF CHROMOSOME 20 AND CANCER
The significance of loss or rearrangements of chromosomes in cancer is for the most part not well understood. Loss of whole or partial chromosomes in apparently dominantly inherited cancers resulting in the loss of heterozygosity in the tumour cells has been reported, for example, in retinoblastoma,85-87 Wilms' tumour,88-91 and recently in familial adenomatous polyposis.92 93 An interstitial deletion of 20p12.2 associated with the rare, dominantly inherited cancer multiple endocrine neoplasia (MEN 2A) was reported by VanDyke et al94 95 and Babu et al.96 They observed the deletion in patients with MEN 2A from seven families and thus the locus for the disease gene was provisionally assigned to 20p12.2 at HGM6.9 Several investigators were unable to identify the deletion in their patients,97-100 although Butler et al101 were able to confirm the association between the deletion and the disease.

We decided to test the assignment by linkage studies using unique DNA sequences (D10S5 and D10S6) that detected RFLPs and that mapped to 20p12 by in situ hybridisation and were found to be polymorphic.102 103 By analogy with familial retinoblastoma and Wilms' tumour, a gene in the putative deletion accompanied by a loss of the normal allele in tumour cells was thought to be a possible hypothesis for MEN 2A. The family studies, however, excluded close linkage between the disease locus and the DNA sequences and subsequently excluded the disease locus from most of chromosome 20, indicating that the MEN 2A locus was not at the site of the deletion.104 The above data led to the designation of the assignment of the locus for MEN 2A on chromosome 20 as inconsistent at HGM8.105 The final proof of the exclusion of the MEN 2A locus from chromosome 20 came with the establishment of positive linkage on chromosome 10 by two independent groups9 10 and so the locus for MEN 2A was removed from chromosome 20 and assigned to chromosome 10.11 At HGM9 a ‘common’ aphidicolin enhanced fragile site was reported at the same site as the deletion associated with MEN 2A. Is the product of the gene for MEN 2A affecting this fragile site?

Non-random losses of chromosomes in some patients with malignant disease, non-lymphocytic
leukaemia or myelodysplastic syndrome, and polycythaemia vera have been observed, but the relation between the loss and the neoplasias remains obscure. Most of the losses involve a deletion of 20q106-108 and the deletion always includes the SRC1 locus. The comprehensive study by Davis et al107 of 20 patients with haematological malignant disease did not show any clinical differences between those with or without 20q-. The presence of clonal chromosome abnormalities of any kind does not seem to have prognostic value for polycythaemia vera.106 In patients with post-polycythaemia vera and idiopathic myelofibrosis, Miller et al108 suggested that the abnormal karyotypes including 20q— may be related to cytotoxicity therapy rather than to the disease itself. It therefore appears that the SRC1 locus does not have a direct causal relation, at least in the above neoplasias.

ARGININE VASOPRESSIN (ARVP), OXYTOCIN (OT), THEIR NEUROPHYSINS, AND DIABETES INSIPIDUS

Genes for the two hormones, arginine vasopressin (ARVP) and oxytocin (OT), and their carrier proteins, the neurophysins, have been provisionally mapped to chromosome 20. They are related to diabetes insipidus. Hereditary diabetes insipidus is rare and has been classically divided into two forms by its aetiology. One type has a central nervous system (pituitary or hypothalamic) and the other a renal (nephrogenic) basis. Both result in a defect in the ability to concentrate urine. The central nervous system form is usually inherited in an autosomal and the renal in an X linked fashion.73 Both vasopressin and oxytocin regulate renal tubular resorption and mutations in either of the genes for these hormones might be expected to produce diabetes insipidus. It has indeed been shown in the Brattleboro rat that their recessively inherited diabetes insipidus is the result of defective translation of the message from a deletion of a G residue in the coding domain for the neurophysin carrier protein of vasopressin.109 Molecular descriptions of mutations in these genes for the two hormones and their carrier proteins from patients with human diabetes insipidus will no doubt be forthcoming now that the genes have been isolated and sequenced.

The two hormones, arginine vasopressin and oxytocin, are synthesised together in the hypothalamus as a large precursor protein along with a glycoprotein.110 In the large precursor molecule there are two carrier proteins known as neurophysin I, the carrier protein for oxytocin, and neurophysin II for vasopressin. The hormones and the carrier proteins are packaged in granules, transported along the axons to the neurohypophysis, where they are released into the blood, possibly stored, and then cleaved from their carriers for their respective functions.73 The genes for both the hormones ARVP and OT have been isolated from human libraries and their nucleotide sequences determined.111 The two hormones and their carrier proteins are each encoded by three exons and two introns; the prepropeptide coding sequence of the hormones is in the first exon of each gene and parts of the coding sequences for the neurophysins are in all three exons of each gene.111 The sequences for the functional peptides have opposite transcriptional orientations and the genes were shown to have only 12 kb between them, indicating that they are syntenic and closely linked. Using a bovine cDNA probe from Land et al112 and a panel of human/hamster somatic cell hybrids, Riddell et al72 mapped the arginine vasopressin gene (ARVP) to chromosome 20 and the assignment of the closely linked gene for prepro-oxytocin (OT) can thereby be inferred.

GALACTOSIALIDOSIS (GSL)

Although not confirmed as yet, the locus for the interesting metabolic disease galactosialidosis (GSL) has been mapped to chromosome 20. The disease is characterised biochemically by its deficiency of two enzymes, α-neuraminidase and β-galactosidase, and has been known as sialidosis II113 to distinguish it from sialidosis I (mucolipidosis I) where only neuraminidase is deficient. The disease was first clinically delineated by Goldberg et al114 and has been known by the eponym, the Goldberg syndrome. The syndrome has also been variously called GM1 gangliosidosis type 4, the cherry-red-spot-myoclonus syndrome with dementia, and the juvenile onset form of sialidosis type II. The age of onset of GSL is most commonly at about 10 years, but it can also present in infancy and adult life. The infantile and adult forms of GSL were distinguished from two forms of sialidosis I, but not from each other, by complementation studies.115 Hoogeveen et al115 suggested that the mutation for sialidosis I was in the structural gene for neuraminidase and the mutation for GSL involved a regulatory factor. More recently, the molecular and complementation studies of Mueller et al116 have supported the hypothesis that the deficiency of neuraminidase is not the result of a mutation in the structural gene for the enzyme (which they have shown maps to chromosome 10 and determines sialidosis type I), but rather one in a gene on chromosome 20 that is necessary for the expression of both neuraminidase and β-galactosidase.

Mueller et al116 assigned the gene for neuraminidase expression by measuring the enzyme activity in
somatic cell hybrid cell lines from a mapping panel and complementation studies. Hybrids were made from two selectable mouse mutant cell lines (RAG and LM/TK−) and normal human fibroblasts. The normal human/RAG mouse hybrid required chromosome 10 for the expression of neuraminidase and a LM/TK− hybrid required chromosome 20. To explain the discrepancy they made RAG hybrids with fibroblasts from a patient with early onset galactosialidosis and found that the hybrids required only chromosome 10; they concluded that the mouse line was providing the missing gene for neuraminidase expression. Hybrids made from the LM/TK− mouse and human cells from a sialidosis I patient, however, needed both chromosomes 10 and 20 for the expression of the enzyme. Parental cells from the RAG mice complemented cells from the patient with galactosialidosis but not from the patient with sialidosis. From these results they suggested that the structural gene for neuraminidase was on chromosome 10 and the gene for the expression of the enzyme in the normal human/RAG mouse hybrid came from the mouse genome.

**Locis not related to disease**

**Light chain gene for ferritin (FTL1)**

One of the family of light chain genes for ferritin (FTL1) has a confirmed assignment to 20q12→q13.31.117,118; genes for other light chains are on chromosomes 19 and X. Ferritin is an iron storage protein consisting of 24 subunits made up of two chains, heavy (H) and light (L) chains, forming a protein spherical shell through which ferrous iron enters and is oxidised to ferric iron.119,120 Both chains are coded for by two distinct families of genes.121 The ratio of heavy and light chains in ferritin varies in the different tissues and iron accumulates more in the L rich and is released more in the H rich ferritins.120 The differences in uptake, storage, and release of iron from known isoferritins has been explained at the molecular level.120 The ratios change during development and disease. The light chain ferritin is increased in iron overload and the H chain in fetal development and in cancer. The cDNAs have been sequenced for both the heavy and light chains and indicate that they are derived from a common ancestor, but probably diverged about 200 million years ago.120 Their homology is extensive in the coding but not the non-coding regions. The specific function of the chromosome 20 gene is not known. As a member of a family of at least three genes for the light chain it may be important for cell survival and therefore there are more than one in the genome, or one or more of them are pseudogenes.

**Inosine triphosphatase (ITPA)**

The locus for the enzyme inosine triphosphatase (ITPA) has been assigned to chromosome 20,122,123 sublocalised to 20p.124 and is on the homologous chromosome 13 of the gorilla.123 The enzyme hydrolyses three nucleotides, inosine, deoxyinosine, and xanthine triphosphate.124 The physiological substrate for the enzyme is unknown although there has been speculation regarding a possible role in purine metabolism.125,126 In any case, there is a very broad range of activity and the enzyme appears to be in all tissues although it is usually measured in erythrocytes. True deficiency (absence in both red and white cells) is rare and is not known to be associated with any disease or abnormal state.124

**Fragile sites**

There are two fragile sites on chromosome 20. One is a rare fragile site (FRA20A) that is folic acid dependent and has been observed at 20p11.23,127,128 The other is the 'common' aphidicolin enhanced fragile site (FRA20B) observed at 20p12.2 (HGM9),129 already mentioned in connection with the putative deletion in MEN 2A.

**Binding proteins**

Five genes for proteins with a binding function have been provisionally assigned to the chromosomes (table 1). Guanine nucleotide binding protein (GNAS), one of the polypeptides that mediate intracellular responses to a variety of extracellular stimuli, was mapped to the chromosome at HGM9.130 The locus for leucine (L) amino acid transport activity (HTL) has been mapped to chromosome 20.131 This is one of three transport mechanisms. The L system is not Na+ dependent and reacts toward branched chain and aromatic amino acids. The locus for leucine transport activity was mapped making use of a temperature resistant human/hamster hybrid line with an increased L system transport. One of the family of metallothionein genes (MTL4) has been tentatively assigned to chromosome 20.132 These metal binding proteins come in a number of forms and the main cluster of genes map to chromosome 16. Others were assigned to chromosomes 1, 4, 18, and 20. Schmidt et al.132 mapped the genes in this family while searching for a mutated one on the X chromosome that might be responsible for Menkes' disease; the patients with this disease are known to accumulate an excess of copper metallothioen in some tissues and cultured cells. One gene for a ribophorin (RPN2) was mapped to chromosome 20 by the panel method and reported at HGM9.133 Two ribophorins have been characterised and sequenced.134 They are glycoproteins in the rough endoplasmic reticulum and are
thought to be involved in the binding of ribosomes related to the synthesis of secretory and membrane proteins. The cDNA for thrombomodulin (THBD) has been isolated, sequenced, and mapped to chromosome 20 by hybridisation to flow sorted chromosomes.\textsuperscript{135} Thrombomodulin is a cell surface receptor which converts thrombin to a physiological anticoagulant by forming a complex with thrombin that activates a protein C, which in turn degrades two of the clotting factors Va and VIIIa. The thrombin-thrombomodulin complex is internalised and the thrombomodulin is recycled, probably as a mechanism of regulation of plasma thrombin levels.\textsuperscript{135} Domains of the precursor of thrombomodulin are organised in a similar way to those described by Sudhof et al\textsuperscript{136} for the low density lipoprotein receptor (LDLR); the locus for LDL is on chromosome 19.\textsuperscript{2} The LDL receptor is also internalised in endothelial cells, in this case as a regulator of plasma cholesterol.\textsuperscript{137}

\textbf{OTHER PROTEINS AND POLYMORPHIC LOCi}

Finally, loci for three additional proteins, one of which is polymorphic, 18 unique anonymous sequences, nine of which are polymorphic (table 2), and one cell surface antigen identified by a monoclonal antibody have been provisionally assigned to chromosome 20 (HGM9). The gene for secretogranin 1 (SCG1) is one of a gene family and has been localised to 20p12.2.\textsuperscript{138} It is another secretion associated protein as is the ribophorin mentioned above. The locus for what is probably an enzyme for the conversion of desmosterol to cholesterol (DCE) was mapped to chromosome 20 as early as 1974,\textsuperscript{139} but has never been confirmed. Desmosterol is converted to cholesterol by the enzyme sterol delta 24-reductase and is presumed to be missing or defective in mouse L cells in which desmosterol but not cholesterol can be found. The hybrid cells were made with several selectable lines of mouse and human cell lines and the presence of the reductase, as measured by the conversion to cholesterol in the hybrids, was correlated with chromosome 20 in the cell clones.\textsuperscript{139} The gene for an opiate neupeptide, prodynorphin (PDYN), has been localised to 20pter-p12 using both a somatic cell hybrid panel and in situ hybridisation.\textsuperscript{140} This gene detects a useful TaqI polymorphism with gene frequencies of 0.7 and 0.3.

I thank Michael W Partington and Alessandra M V Duncan for careful reading of this manuscript. This work was supported in part by Medical Research Council of Canada grant MT-5783.

\textbf{References}

25. Hopkinson DA, Cook PJL, Harris H. Further data on the


47. Honig J, Martiniuk F, d’Eustachio P, et al. Confirmation of the regional localization of the genes for human acid phosphogluco-


The map of chromosome 20


100 Butler MG, Rames LJ, Joseph GM. Cytogenetic studies of individuals from four kindreds with multiple endocrine neoplasia type II syndrome. Cancer Genet Cytogenet 1987;28:253-60.


108 Schmale H, Ivell R, Breindl M, Darmer D, Richter D. The mutant vasopressin gene from diabetes insipidus (Brattleboro) rats is transcribed but the message is not efficiently translated. EMBO J 1986;5:3289-93.


110 Sausville E, Carney D, Battey J. The human vasopressin gene


133 Barton DE, Cimaudo C, Hortsch M, Francke U. The genes for ribophorins I and II are on human chromosomes 3p and 20 and mouse chromosomes 6 and 2, respectively. *HG3M. Cytogenet Cell Genet* 1987;46:577A.


Correspondence and requests for reprints to Professor Nancy E Simpson, Department of Paediatrics, Queen's University, Kingston, Ontario K7L 3N6, Canada.