THE UNIVERSITY *of* EDINBURGH

Title	Mapping and mutational analysis of chromosome 11q12-13
Author	Coyle, Elizabeth
Qualification	PhD
Year	1996

Digitisation notes:

- Page numbers 219 missing in original

Mapping and mutational analysis of chromosome 11q12-13

Elizabeth Coyle

PhD

The University of Edinburgh

1995

Declaration

I declare:

a) that this thesis was composed by myself and

b) that the work is my own, except where otherwise stated

Abstract

This thesis describes a) the isolation and mapping of new markers from human chromosome 11q12-13 and b) the mutational analysis of a gene from this region as a candidate for an inherited deafness syndrome. Chromosome 11q12-13 is estimated to comprise 28Mb of DNA, is gene rich, and encodes a number of important disease loci, associated with deafness, eye retinopathies, cancer and diabetes. We had at hand a number of somatic cell hybrids containing different sub fragments of this region, which I wished to exploit for the isolation and mapping of new markers. I developed a two step Alu PCR amplification protocol which allowed the majority of products to be used as genomic probes without competitive hybridisation. I compared the efficiency of different cloning methods and demonstrated the value of Turbo cloning, a new method for blunt end cloning. Markers were then ordered along the chromosome using the hybrid panel. A number of markers were concentrated in a sub region of 11q13 which had been linked to several important disorders, and these were used to isolate YACs. Thus the formation of a cloned DNA map of the region was initiated.

In parallel I examined the Olfactory Marker Protein (OMP) gene as a candidate for Usher Syndrome Type IB, by DNA sequencing and mutational analysis of affected individuals. Usher syndrome is characterised by deafness and retinitis pigmentosa. In Type I, vestibular dysfunction is also a feature of the disorder. The disorder is genetically heterogeneous and for each subtype several genetic loci have been linked in individual populations: Type IB has been linked to 11q13. A stereociliary

defect was proposed, and the mapping of OMP to the linked region, its expression in olfactory nasal cilia and the CNS, and the mapping of the mouse Omp gene to the syntenic region on mouse chromosome 7 very close to *Shaker-1*, a possible mouse homologue of Usher syndrome type I, justified close examination of OMP as a candidate for both disorders. A human genomic clone was therefore obtained and sequenced, several different methods of mutation detection compared, and the value of the chemical mismatch cleavage technique demonstrated. At the same time, comparative mapping was carried out by mapping candidates from the *Shaker-1* nonrecombinant region on the somatic cell hybrid panel. Variants of the OMP gene were identifed, but it was formally excluded as candidate when a gene located very close to OMP was shown by others to be mutated in *Shaker-1* mice and USHIB patients.

Acknowledgements

First of all I want to thank the three main people who helped me to put this thesis together: David Porteous for his support and guidance throughout; Chris Boyd for checking all those little details and helping to transfer and hand out each new version; and Andreas Schedl for always being able to look at it from a different angle. My figures are mostly down to Sandy Bruce, who was at one point more anxious than I was to get all the prints done, and Euan Slorach for helping me through the pain of Island Draw. Thanks of course also goes to Sheila Mould for helping to track down those elusive journals.

Getting to work each day was courtesy of John Maule and his bike clinics. Thanks to everyone else in the West Wing for making it worthwhile when I got there. Thanks to Stuart Wyles for all those coffees, lunches, and late night shopping trips. Then of course there are my two other 'mums' Sheila Webb and Heather Davidson who were always willing to give useful, and often conflicting advice on my social life.

Behind everything of course were my family, who always believed in me. Thanks mum.

Abbreviations

ABS	absorbance
BAC	bacterial artificial chromosome
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary DNA
cM	centimorgan
CMGT	chromosome mediated gene transfer
CSC	coincident sequence cloning
°C	degrees centigrade
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleotide triphosphate
dTTP	deoxythymidine triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddTTP	dideoxythymidine triphosphate
ddF	dideoxyfingerprinting
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid disodium salt
EST	expressed sequence tag
FISH	fluorescence in-situ hybridisation
IRS	interspersed repetitive sequence
IPTG	isopropyl β-D-thiogalactopyranoside
IBD	identical by descent

kb	kilobase
LIC	ligation independent cloning
LINES	long interspersed repeated sequences
LMP	low melting point
LOD	logarithm of the odds
MAC	mammalian artificial chromosome
Mb	megabase
mRNA	messenger RNA
OMP	olfactory marker protein
ORN	olfactory receptor neurone
MIM	Mendelian inheritance in man reference number
PAC	P1 artificial chromosome
pBS	pBluescribe
PCR	polymerase chain reaction
PFGE	pulsed field gel electrophoresis
PMSF	phenylmethylsulphonylfluoride
pers. comm.	personal communication
ROMI	rod outer membrane protein 1
RNA	ribonucleic acid
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SINES	short interspersed repeated sequences
SSC	sodium saline citrate
STS	sequence tagged site
UV	ultraviolet
VNTR	variable number of tandem repeat sequences
XGAL	5-bromo-4-chloro-3-indoyl-β-D-galactosidase
YAC	yeast artificial chromosome

Contents

List of figures

List of tables

1. Introduction

1.1 Producing and ordering a library of repeat free clones from chromosome 11q12-13

The initial aims of the Human Genome Mapping Project (HGMP) were to establish physical and genetic maps including features such as polymorphisms, genes, and specific DNA sequences. The intentions were two fold: the development of unique sets of PCR primers that would recognise sites ~100kb apart across the entire genome (STSs); and to begin to assemble large segments (2Mb equivalent to ~1% of the X chromosome) of contiguous cloned DNA at different defined sites in the genome (Hoffman 1994). This has led to large scale mapping projects, with the most impressive results to date being produced by Genethon in the form of dense genetic maps (Gyapay et al., 1994). The availability of increasing numbers of polymorphic markers has greatly aided linkage studies of genetic disorders and consequently led to physical mapping and assembly of cloned DNA maps of linked regions. Large scale physical mapping efforts have also been undertaken leading to the construction of megabase YAC contigs spanning whole chromosomes (Gaston et al., 1991), most of the work has, however, concentrated on the construction of contigs spanning disease associated regions (Todd et al., 1995).

One of the objectives of my PhD was to produce a map of chromosome 11q12-13. This was undertaken for several reasons; a) studies have shown the region to be gene rich and not surprisingly therefore several human genetic disorders had been assigned to the region and b) the region was relatively poorly mapped, but resources were available in the laboratory which would be valuable in generating new markers. To be useful, such markers must contain a minimal amount of repetitive DNA and be large enough to allow efficient labelling and hybridisation. My aim was therefore to develop a method which could generally be applied to the generation of clone-based physical maps of such regions.

Clone-based physical maps are extremely useful as a framework for many types of structural and biological studies and are resources for characterisation and understanding of human genes, diseases and development. Here an Alu-PCR based approach was developed which can be used to assemble clone maps of a specific chromosomal region. The inter-Alu products were cloned and as such could be used directly to produce a map of the region or sequenced and the sequence then released to others as a sequence tagged site (Smith et al., 1993). The method developed here could equally be applied to other less well mapped regions by selection of reduced somatic cell hybrids retaining the appropriate region.

1.1.1 Somatic cell Hybrids

Producing and ordering a library of repeat free clones from chromosome 11q12-13 required a source of human DNA from this region. Conventional human-rodent

somatic cell hybrids have been valuable in their ability to allow the assignment of genes to specific chromosomes (Walter and Goodfellow 1993). The limitation of this approach however is that it does not often allow markers to be located with high resolution. Subchromosomal localisation of genes using human-rodent somatic cell hybrids requires the fragmentation and segregation of chromosomes by γ-irradiation induced chromosome breakage. Hybrids generated in this manner, radiation hybrids, are readily usable in positional cloning experiments, and have facilitated the mapping and cloning of sequences from defined regions of the genome (Cox et al., 1990; James et al., 1994). Radiation hybrids are produced by subjecting human-rodent hybrid cell lines, containing a single intact human chromosome, to lethal X irradiation. This treatment breaks the chromosomes into fragments and results in cell death. The irradiated donor cell line is then fused to a recipient rodent cell line. Human material is usually retained at high frequency in the absence of specific selection, although it is possible to select for the retention of specific regions by using a recipient cell line deficient for a specific marker (Walter and Goodfellow 1993). The human fragments can be distinguished from the rodent background using the interspersed repetitive sequence polymerase chain reaction (IRS-PCR). A single primer directed to human repetitive sequence elements is used to amplify the DNA between two elements juxtaposed in opposing orientations. Thus products are only obtained when a human fragment is present and the resulting pattern can be indicative of a specific chromosome or subchromosomal region (Guzzetta et al., 1991). Radiation hybrids allow the separation of fragments of 0.3-30Mb and thus span the gap between the limits of

cloning and mapping methods such as PFGE, YAC cloning, and conventional somatic cell genetics.

Three different types of somatic cell hybrids were used to create a hybrid mapping panel (figure 1.1): two chromosome 11 translocation cell lines; four X irradiation hybrids specifically selected to contain particular fragments of chromosome 11q; and two independent HRAS 1 selected chromosome mediated gene transformants which also contained several fragments of chromosome 11 (these are described in more detail in section 3.1.3.1). The importance of the latter two types of hybrid is that they contain multiple fragments of chromosome 11 and together comprise a high resolution mapping panel often allowing markers to be assigned to small intervals of 11q with submegabase accuracy. Through mapping of markers on the panel (Slorach et al., 1995; Evans et al., 1995) the region contained in the hybrids has been well defined, and the irradiation hybrid WJX3.4 appeared to contain a single contiguous region from 11p11.2-11q13.5 (Fletcher et al., 1993). Consequently WJX3.4 was used to produce a library of the region 11q12-13, to which a number of disease loci have been linked.

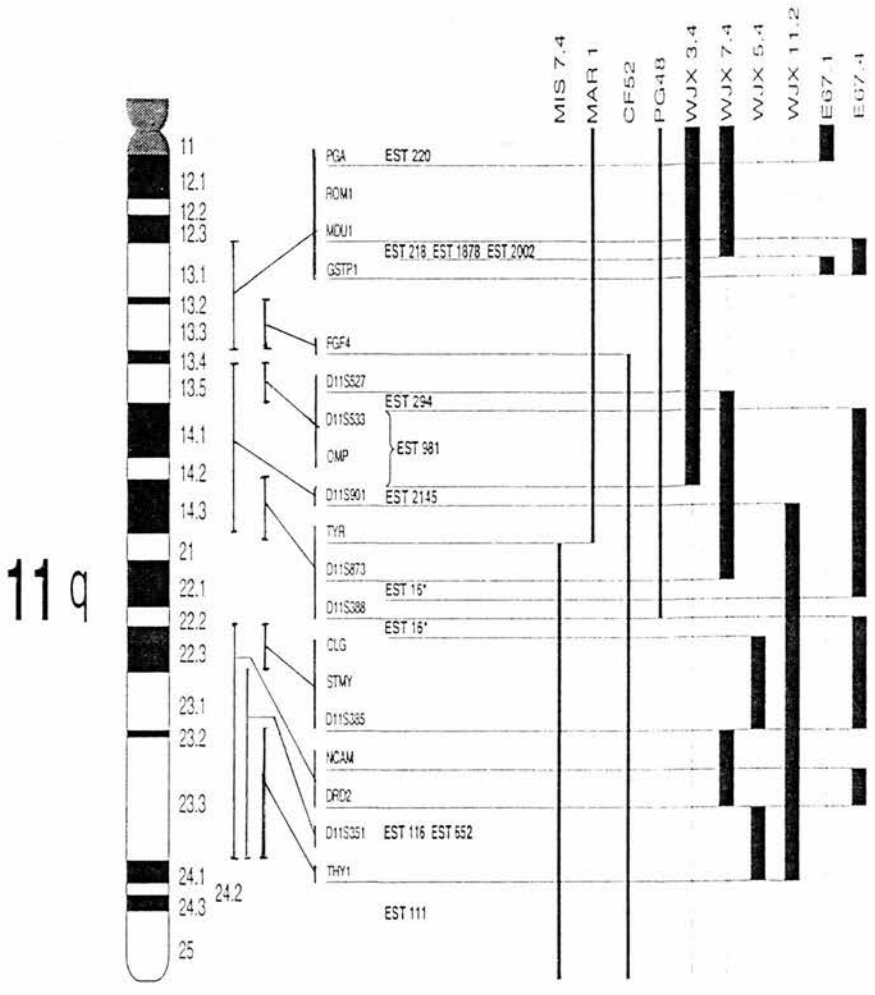


Figure 1.1 Regions of chromosome 11 contained in somatic cell hybrids

Diagrammatic representation of the chromosome 11 content of each of the hybrids is shown on the right. Hybirds MIS7.4 and PG48 were not used in this study. The positions of the chromsome 11 markers (Evans et al., 1995) and the ESTs (Slorach et al., 1995) have been defined by others in the group. Markers are shown in the correct order although evenly spaced for clarity. Regional localisation bars show more accurate mapping positions.

EST 16* is shown in two possible locations since the data available is insufficient to distinguish between the two

1.1.2 Cloned DNA maps

Somatic cell hybrids, which retain different fragments of a chromosome, allow markers to be positioned relative to each other on the chromosome. Estimates can be made of the physical distance between markers, and hence confirm their relative positions, by looking at the frequency of chromosome breakage between them on exposure to a particular dosage of X-rays (Cox et al., 1990). Genetic distance can also be estimated, if markers are polymorphic, by looking at the frequency of recombination between two alleles. Both maps provide information on the distance between markers, although the resolving power of a radiation hybrid map is generally greater than that of a linkage map which is limited by the availability and spacing of polymorphic markers along the chromosome. Neither, however, provides access to the intervening DNA which requires cloned DNA maps.

A cloned DNA map describes a genomic region as a series of overlapping cloned molecules, such that each recombinant vector contains a piece of DNA that partially overlaps with the DNA carried by the adjacent recombinant (Little 1992). Cloned DNA maps are generally constructed using Yeast Artificial Chromosomes (YACs). YAC vectors are capable of carrying several hundred kilobase-pairs of DNA insert, consequently contigs (sets of overlapping clones or sequence) of up to several megabases can be readily assembled. Individual YAC inserts can then be subcloned into cosmids or plasmids, enabling their analysis down to the nucleotide sequence level. Contigs are constructed by screening libraries such as the ICI YAC library with suitable probes (Anand 1992).

Mapping on somatic cell hybrids is an important intermediate stage when assembling a YAC contig. It allows the rationalization of results of the primary YAC library screen, for example if the same YAC is isolated using two markers which are shown to be several megabase apart by mapping on somatic cell hybrids, then this cannot be spanned by a single YAC but will probably be due to the presence of related or repetitive sequence in the markers. It also means that the location of the YAC can be confirmed by *in situ* hybridisation of YACs to somatic cell hybrid chromosome spreads (FISH).

In the next section I will explain why the PCR amplification of human sequences flanked by specific repeated sequences (Alu-PCR) was chosen as a means of generating markers, and how the PCR was optimised to give the maximum complexity of products, and the cloning strategy used. Later sections will explain how clones suitable for mapping on somatic cell hybrids and YACs were selected, and the isolation of CA repeat containing clones. First, however, I will discuss why this region of chromosome 11q is so worthy of study.

1.1.3 Disease loci

A number of disease loci have been linked to this region of 11q. These include genes for cancer, diabetes and hereditary deafness, (Fig. 2 shows their relative positions). In order to facilitate cloning studies of such disease loci, new markers were needed to fill the gaps in the existing map and help in the isolation of candidate genes. These markers would be particularly interesting if they were polymorphic as they would then also be useful in linkage analysis. In this section I describe some of the disorders linked to the region.

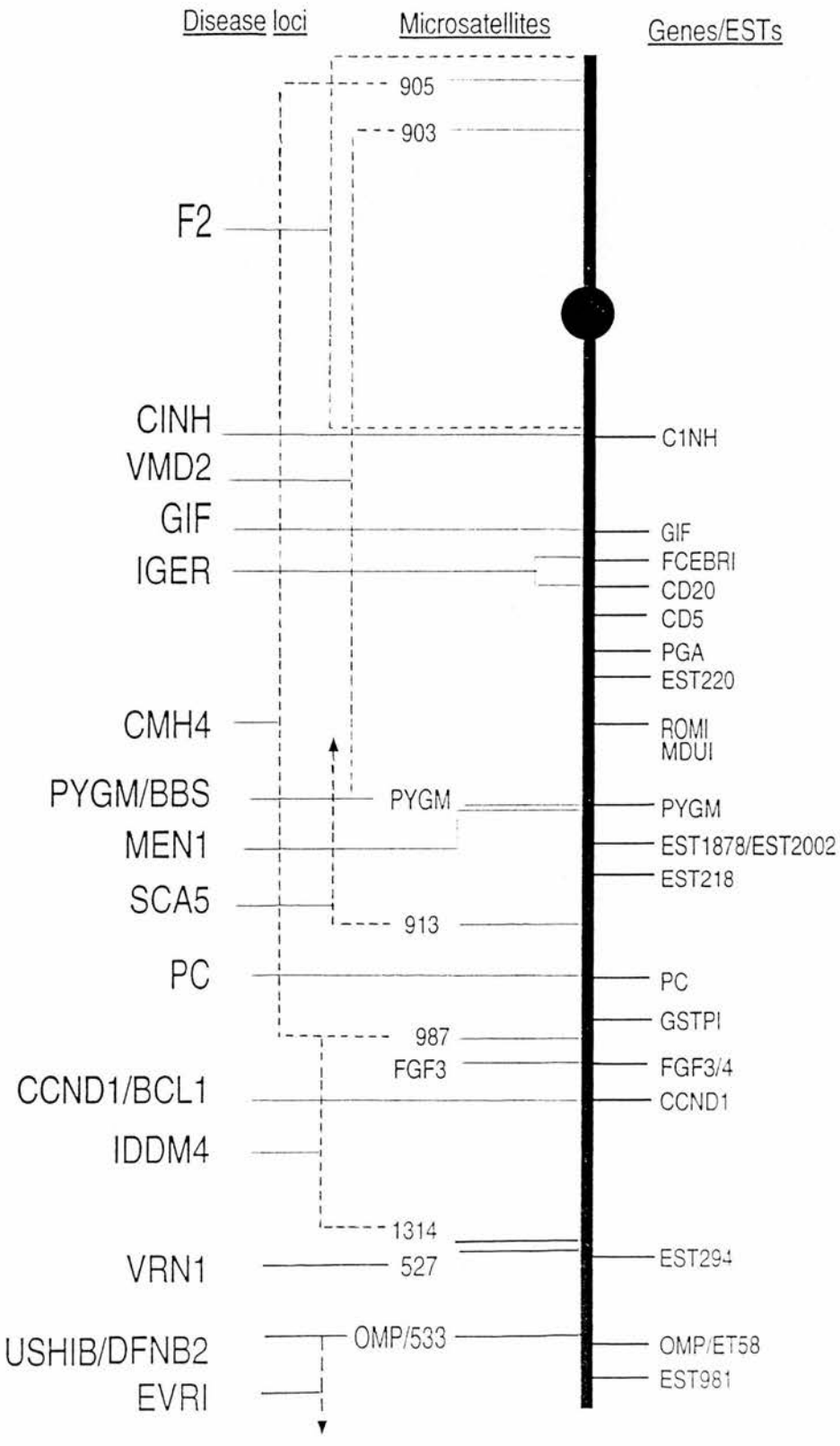

Disease loci
Microsatellites
Genes/ESTs
905
903
F2
CINH
C1NH
VMD2
GIF
GIF
IGER
FCEBRI
CD20
CD5
PGA
EST220
CMH4
ROMI
MDUI
PYGM/BBS
PYGM
PYGM
MEN1
EST1878/EST2002
SCA5
EST218
913
PC
PC
GSTPI
987
FGF3
FGF3/4
CCND1/BCL1
CCND1
IDDM4
1314
VRN1
527
EST294
USHIB/DFNB2
OMP/533
OMP/ET58
EVRI
EST981

Figure 1.2 Map positions of disease loci on 11q

This map is based on that originally formulated by James et al (1994). The gene linked to the prothrombin (F2) locus had only been positioned very roughly (Royle et al., 1987) to 11p11-q12. Some of the polymorphic loci had not been positioned on the radiation hybrids, but an indication of where they map has been given: the SCA5 disease locus lies in an 8cM region between D11S913 and GATA201; the EVRI disease locus has been tightly linked to D11S873 which lies in the region 14cM distal to D11S533; MENI disease locus lies in an 3cM region between PYGM and the more telomeric marker D11S480. The D11S- prefix has been omitted from the microsatellite marker designations.

Spinocerebellar ataxia type 5-*SCA5*

SCA5 is one of a group of clinically and genetically heterogeneous autosomal dominant cerebellar ataxias (Ranum et al., 1994). The disease is characterised by varying degrees of incoordination affecting gait, speech and swallowing, due to progressive multisystem neurodegeneration, Four different SCA loci have previously been identified on chromosomes 6p, 12q, 14q and 16q, one of which (SCA1) has been shown to be caused by expansion of an unstable trinucleotide repeat which may explain the anticipation and clinical variation associated with these disorders (Jodice et al., 1994). A large, ten generation kindred used in multipoint linkage and haplotype analyses provided strong evidence that the gene was located in the 8cM region between GATA2A01 and D11S913.

Familial hypertrophic cardiomyopathy 4-*CMH4*

Hypertrophic cardiomyopathy (MIM - 192600) is a primary cardiac disease that is usually transmitted as an autosomal dominant trait. It is the major cause of sudden death in young athletes. The disorder was shown to be genetically heterogeneous, when less than 50% of published pedigrees were found to be associated with mutations in the β myosin heavy chain gene on chromosome 14q11-q12 (CMH1 - Solomon et al 1990). Two other loci were later identified on chromosome 1q (CMH2 - Watkins et al., 1993) and 15q2 (CMH3 - Thierfelder et al., 1993). Genetic analyses of informative markers in a large family not linked to any of the

previously identified loci, identified two haplotypes cosegregating with the disease. This led to the tentative conclusion that the CMH4 locus lay between markers AFM105xb10 (D11S905) and AFM131ye5 (D11S987) in a 17cM region of chromosome 11 (11p13-q13) (Carrier et al., 1993).

Hypoprothrombinemia; dysprothrombinemia - *F2*

Hypoprothrombinemia and dysprothrombinemia result in an increased bleeding tendency manifested especially in more marked or prolonged posttraumatic and postoperative bleeding (MIM - 176930). Both are inherited in an autosomal dominant manner, and are caused by mutations in the gene for prothrombin (coagulation factor II): in hypoprothrombinemia no identifiable protein is produced; whereas in dysprothrombinemia protein is produced, but is biologically inactive (Lanchantin et al., 1968). The gene for human prothrombin (F2) was assigned to 11p11-q12 by analysis of a panel of somatic cell hybrids and by *in situ* hybridisation, using both cDNA and genomic probes (Royle et al., 1987).

Pyruvate carboxylase deficiency - *PC*

As the name suggests, this autosomal recessive disorder results from a deficiency in pyruvate carboxylase (MIM - 266150), leading to abnormal gluconeogenesis. Clinically this results in growth retardation, mental retardation, and infantile subacute necrotizing encephalomyelopathy. Somatic cell hybrids were used to localise the gene to the long arm of chromosome 11 (Freytag and Collier, 1984), comparative mapping with the mouse suggested that the human PC locus was

located in the proximal part of 11q, perhaps at 11q12-13 (Rochelle et al., 1992). This localisation has since been confirmed (Smith et al., 1993)

Congenital Pernicious anemia - *GIF*

Infantile onset pernicious anaemia is inherited in an autosomal recessive manner, and results from a defect of intrinsic factor secretion (MIM - 26100). Clinically the disorder results in jaundice, anaemia and mental retardation. Somatic cell hybrids were used to localise the gene to chromosome 11 (Hewitt et al., 1991). This localisation was later refined to 11q12-13 by Stafford et al (1994).

Atopy - *IGER*

Atopy is a common condition in man, characterised by familial asthma, hay fever, and eczema (MIM - 147050). It is caused by an excessive allergic response to small quantities of common inhaled allergens such as pollens, animal dander, and house dust mite, and is mediated by Immunoglobin E. Inheritance of this disorder is complex but may involve autosomal dominant factors. Using sibpair analysis which does not require prior assumptions about the mode of inheritance, penetrance of the disease, or the gene frequency, an important locus for atopy in 11q12-13 was identified in British (Young et al., 1992) and Japanese families (Shirakawa et al., 1991). The candidate region was later refined to an area centromeric to the pepsinogen locus (PGA), centred around the CD20 gene (Sandford et al., 1993). Recently Stafford et al (1994) constructed a 2.8 Mb YAC contig of this region, to which they localised two candidate genes for the disorder:

the beta subunit of the high affinity immunoglobulin E receptor (FcεRIβ); and CD20, a molecule involved in B cell differentiation.

Hereditary angioedema - *C1NH*

Hereditary angioedema is an autosomal dominant disorder characterised by episodic local subcutaneous oedema and submucosal oedema involving the upper respiratory and gastrointestinal tracts (MIM - 106100). It is caused by a deficiency of activated C1 esterase inhibitor (C1NH). The gene was assigned to 11p11.2-q13 using somatic cell hybrids (Bock et al., 1986). Theriault et al (1990) used in situ hybridisation to obtain more precise localisation of the gene to 11q11-q13.1.

McArdle disease - *PYGM*

McArdle disease is an autosomal recessive disorder caused by myophosphorylase deficiency (MIM - 232600). Phenotypically this results in exercise intolerance, with premature fatigue, and cramps in exercising muscles. The human muscle glycogen phosphorylase (PYGM) gene has been assigned to chromosome 11 by DNA spot blot analysis of chromosomes (Lebo et al., 1984) and mutations in this have been shown to result in McArdle disease (Tsujino et al., 1993). Lebo et al (1990) later sublocalised the gene to the proximal part of 11q13 by FISH.

Parathyroid adenomatosis 1-*CCND1*

Parathyroid adenomatosis is an autosomal dominant disorder in which parathyroid adenomas result due to overexpression of PRAD1 (MIM - 16841) which encodes a cyclin. The cyclin D1 gene (CCND1) was later assigned to 11q13 by a

combination of somatic cell hybrid studies and FISH (Inaba et al., 1992). CCND1 has been shown to have the ability to transform cultured cells and hence function as an oncogene (Hinds et al., 1994), thus supporting its proposed role as a proto-oncogene for parathyroid adenomas.

Chronic lymphatic B-cell leukemia - *BCL1*

Chronic lymphatic leukemia results in the malignant transformation of B cells (MIM - 151400). The observation of a recurring t(11;14) translocation involving the Ig heavy chain locus at 14q32 and 11q13 (Tsujimoto et al., 1984) led to the 11q13 being designated BCL1, for B-cell leukemia/lymphoma-1 The nature of the gene located at 11q13 which is involved in the malignant transformation of B cells remained unclear until recent work by Komatsu et al (1994) showed that in a particular translocation the PRAD1 gene was the sole candidate known to be located between breakpoints. In addition, PRAD1 mRNA overexpression was shown to be concordant with the presence of a translocation. These results indicated that the putative BCL-1 gene activated by translocation is in fact PRAD1.

Multiple endocrine neoplasia type 1-*MEN1*

Multiple endocrine neoplasia is an autosomal dominant disorder characterised by a high frequency of peptic ulcer disease and primary endocrine abnormalities involving the pituitary, parathyroid, and pancreas (MIM - 131100). Larsson et al (1988) mapped the MEN1 locus to chromosome 11 by demonstrating linkage to a

DNA probe derived from the PYGM locus. In addition to the linkage evidence, loss of heterozygosity (LOH) on chromosome 11q in MEN1 associated tumours suggested the presence of a tumour suppressor gene in that region, and deletion mapping in such tumours placed the MEN1 locus telomeric to PYGM at 11q13 (Byström et al., 1990). The genes PYGM and FAU have been ruled out as candidates for MEN1 by analyses of DNA in MEN1 patients and MEN1 associated tumours (Kas et al., 1993). A new gene (ZFM1) has, however, been isolated from the associated region (Toda et al., 1994) and is currently being analysed.

Familial exudative vitreoretinopathy-*EVR1*

Familial exudative vitreoretinopathy is an autosomal dominant disorder characterised by incomplete vascularization of the peripheral retina. Clinical expression of the disorder varies between lack of any visual symptoms and considerable visual impairment which may lead to blindness at a young age (MIM - 133780). Extended multipoint linkage analyses on two families using 5 markers from 11q13-q23 showed the locus most likely maps between D11S533/D11S527 and D11S35 (Li,Y., 1992). Close linkage without recombination was observed between D11S533 and D11S527. The location of EVR1 has recently been refined to the locus D11S837 which lies between D11S533/D11S527 and D11S35 (Müller et al., 1994)

Bardet-Biedl syndrome-*BBS*

BBS is an uncommon, autosomal recessive disorder characterised by mental retardation, post-axial polydactylia, obesity, hypogenitalism and retinitis pigmentosa (MIM - 209900). A number of other traits have also been shown to be frequently associated with the condition including urogenital malformations, renal diseases and both conductive and sensorineural hearing impairments. Expression of these features is variable both within and between families. Linkage analysis on 31 North American families showed significant linkage to the gene for human muscle glycogen phosphorylase (PYGM) in 17 of the families (Leppert et al., 1994). Consideration of adjacent markers in linked families showed that BBS mapped to a 3cM confidence interval between markers UT5150 and INT2, with the favoured position at the PYGM locus. This locus heterogeneity was supported by previous studies which had shown evidence for linkage to chromosome 16 in an Israeli Bedouin kindred (Kwitek-Black et al., 1993). Here only three families gave positive LOD scores to chromosome 16 markers and subsequently yielded a negative LOD score from PYGM. Heterogeneity in the North American families could not be explained by ethnic origin, and did not readily reflect phenotypic characteristics.

Neovascular inflammatory vitreoretinopathy-*VRN1*

Neovascular inflammatory vitreoretinopathy is an autosomal dominant disorder characterised by retinal and iris neovascularisation, abnormal retinal pigmentation, uveitis, cystoid macular oedema, vitreous haemorrhage, and traction retinal

detachment (MIM - 193235). In linkage analysis using chromosomal loci that had been associated with uveitis, neovascularisation, or retinal degeneration in previous studies, markers at 11q13 which were closely linked to Best's disease locus revealed significant linkage (Sheffield et al., 1992). The most probable location of ADNIV was at D11S527. This was confirmed by Stone et al (1992) who observed no recombinants between the disease phenotype and D11S527. Linkage data suggested that VRN1 and Best's disease, although mapping close to each other, were in fact due to mutations in different genes (Sheffield et al., 1992). Similarly VRN1 and EVR1, also appeared to result from mutations in different genes, since EVR1 lies 7cM from D11S527 (Sheffield et al., 1993). The fact that several eye disorders are linked to the same proximal portion of 11q may indicate the presence of a cluster of genes that have a common function and owe their proximity to a common ancestor. Although until the nature of the genes involved is determined this remains a theory.

Insulin dependent diabetes mellitus-*IDDM4*

Insulin dependent diabetes mellitus is a disorder of glucose homeostasis that is characterised by susceptibility to ketoacidosis in the absence of insulin therapy (MIM - 222100). IDDM is a genetically heterogeneous autoimmune disease and genetic studies have been aimed at the identification of loci associated with increased susceptibility to this multifactorial phenotype. The major locus (IDDM1 which accounts for 42% of familial cases) is encoded by the major histocompatibility complex on chromosome 6p21, other loci, such as the insulin

gene on chromsome 11p15 (IDDM2), are more minor in their contribution (less than 10%). By carrying out linkage tests of microsatellites (average spacing 11cM) Davies et al (1994) confirmed the presence of two other minor loci at 11q (IDDM4) and 6q (IDDM5). The presence of a susceptibility locus on 11q was confirmed by Hashimoto et al (1994).

Vitellliform macular dystrophy/Best's disease-*VMD2*

Best's disease is clinically defined as vitelliform macular dystrophy. It is a juvenile onset, autosomal dominant disorder characterised by an accumulation of yellowish material in the macular area (MIM - 153700). The disease is slowly progressive and eventually leads to atrophy of the retinal pigmented epithelium in photoreceptor cells, severely impairing central vision. The biochemical defect underlying this condition is unknown, but abnormal electroculography (EOG) and histopathological studies, suggest the retinal pigment epithelium as the primary site of the dysfunction. Linkage studies by Stone et al (1993) localised the gene to 11q13, between INT2 and D11S871. Weber et al refined this localisation to a 3.7cM interval between D11S903 and PYGM, which by PCR mapping was shown to span the centromere of chromosome 11. The gene encoding the retina specific protein ROM1 has been isolated and shown to map to the Best's disease region (Bascom et al., 1992a,b). Identification of a recombination event between ROM1 and the disease phenotype excluded ROM1 as the gene responsible for the disorder (Stöhr and Weber 1995).

1.1.4 Physical mapping and human specific repeats

Various different resources can be used as a starting point from which to formulate a physical map of a region. Essentially these depend on how the disorder was mapped to the region in the first place. Various examples are given in the disorders mentioned above; translocations, cytogenetically visible deletions, and linkage disequilibrium. Translocations give a specific physical location to aim for, whereas the area covered by a deletion, or surrounding a linked marker may be quite large. Closer markers may first need to be isolated in each case and a cloned DNA map assembled if there are no obvious candidate genes in the region.

The human genome contains a relatively large proportion of DNA which appears as yet to serve no purpose. In contrast to the DNA of simpler organisms like *E.coli* and yeast, human DNA contains large amounts of sequences present in multiple copies and these form a large proportion of what has been termed 'junk' DNA. While some repetitive sequences are located in discıete chromosomal domains, for example alphoid sequences, most repetitive DNA has been shown to be dispersed apparently at random throughout the entire genome in higher organisms (Moyzis et al., 1989).

Initially it was thought that this prevalence of repeat sequences would complicate the mapping of the human genome, but further investigation of repeat structure revealed that the sequence of such elements was highly conserved. Thus with the advent of the polymerase chain reaction, the frequency and conservation of these sequences could be used to advantage in a technique known as interspersed

repetitive sequence PCR. Interspersed repeat sequence DNAs are now classified as SINES (short interspersed repeated sequences) or LINES (long interspersed repeated sequences). In primates each class is dominated by a single DNA sequence family. The major human LINES family is L1 or Kpn1 with a consensus sequence length of approximately 6.4kb; many family members have serial deletions of the 5' end or internal deletions and rearrangements, but most share the 3' end. The reiteration frequency is about 4,000- 20,000 for the 5' end and increases to 50,000-100,000 for the 3' end (Korenberg and Rykowski 1988). The major human SINE is the Alu sequence family. It has a consensus sequence of about 300 base pairs and is reiterated 300,000- 900,000 times in the genome.

Early work had shown that for particular repeats there were local regions of preference and exclusion (Moyzis et al., 1989). Further investigation revealed that this bimodal distribution correlated with Giemsa dark and light chromosome bands: the Alu family dominates in Reverse bands, and the L1 family dominated in Giemsa/Quinacrine positive bands. Each band type can be characterised by the average G+C content, the stage in the DNA synthetic period when they replicate, and the point at which they condense in mitotic prophase. Giemsa/Quinacrine bands are relatively rich in adenine plus thymine (A+T), replicate their DNA during late S phase, and condense early during mitosis. Reverse bands in contrast are relatively rich in guanine plus cytosine (G+C), replicate their DNA early in S phase, and condense late in mitotic prophase. Alus are composed of 56% G + C and L1s 58% A + T, in fitting with the overall DNA composition, and each may comprise

13% to 18% of the total DNA in a chromosome band. It has been proposed therefore that the distribution of these sequences may account for a large part of human chromosome banding seen with fluorescent dyes (Korenberg and Rykowski 1988).

Alu repeat elements have been the preferred choice for IRS-PCR because of the higher degree of conservation, their high reiteration frequency, the small distance separating adjacent elements of inverted orientation and their location in R bands. Alu repeat elements have been used in mapping in basically two different ways; either a) to provide a reproducible fingerprint for a specific somatic cell hybrid, YAC or cosmid, and consequently provide a means of confirming human specific DNA content; or b) these template specific products can be cloned and mapped back to a specific point in the resource from which they were obtained e.g. FISH on a somatic cell/irradiation hybrid, Southern hybridisation to a restriction fragment of a YAC or a cosmid. The former method is used routinely and presents few technical problems. The latter, while frequently used, suffers from several drawbacks which I have sought to overcome:

1. Alu PCR fragments are largely blunt ended and cloning is therefore inefficient
2. DNA fragments amplified using Alu PCR primers which generate the high density of PCR products required are frequently small (Cole et al., 1992b) and thus label poorly.

3. Alu-PCR products contain at least 100 base pairs of the Alu repeat sequence which must be competed out so they can be used as a unique or low copy hybridisation probes.

Each of these points will be addressed, but first I will discuss the structure and origin of Alu repeat sequences and the nature of the material produced by Alu-PCR.

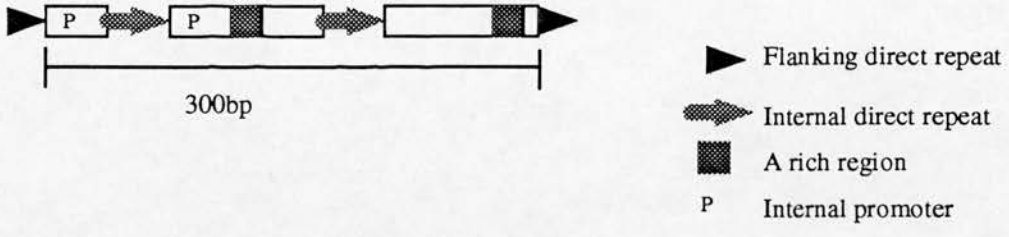


Figure 1.3 The structure of an Alu repeat element

Figure 1.3 shows the structure of an Alu repeat element (Watson et al., 1994). Each Alu element is dimeric in structure, composed of two tandemly arranged halves, with the right half containing an additional 31bp relative to the left half. Alu repeat elements are transcribed by RNA polymerase III, which is directed by a promoter lying within the repeat sequence and therefore is retained in the transcribed sequence. Elements also contain a middle A-rich region and a 3' oligo (dA) tail of variable length, consistent with the belief that these elements have arisen as a result of reverse transcription through an RNA intermediate (Batzer and Deininger 1991). Once transcribed into RNA it is proposed that the 3' oligo dUTP tail formed at the 3' end of all RNA polymerase III products would fold back on the lengthier A-rich tail, forming a primer and allowing reverse transcription of the

repeat element into cDNA. This cDNA would eventually be inserted into the genome via staggered nicks which are repaired to form the direct repeat sequences seen to flank Alu repeat elements.

A theory for the origin of Alu elements is that they originated from a defective 7SL RNA molecule which was reverse-transcribed into DNA and inserted into the genome (retroposition) (Leigh-Brown 1984). 7SL is an abundant small RNA which associates with six proteins to form a small cytoplasmic RNA-protein complex (called SRP) that is essential for the translocation of newly synthesised proteins across membranes of the rough endoplasmic reticulum in higher eukaryotic cells (Watson et al., 1994). Alus are thought to be stable in the genome because a 7SL pseudogene may not possess or respond to any mechanism for excision or transposition. Thus, unless its presence is detrimental to the chromosome in which it inserts, it will be stably maintained throughout replication and cell division. Despite the presence of thousands of copies of Alu elements, the majority of which possess an RNA polymerase III promoter which is functional *in vitro,* there is very little transcription of Alu family members *in vivo.* A number of explanations for this phenomenon have been proposed: absence of the 37bp upstream sequence required for the 7SL gene to be actively transcribed; methylation; and accumulation of inactivating mutations (Deininger et al., 1992, Liu et al., 1994). The fact that some disease related *de novo* mutations have been shown to have resulted from Alu element retroposition indicates that some members are still transcriptionally active (Vidaud et al., 1993). There are two

different schools of thought on the retroposition of Alu repeat elements, each of which is based on specific nucleotide variations in individual Alu elements. In the 'master gene' model all subfamilies are serially related due to their origin from the same master source gene. The subfamilies themselves are not propagated. In the 'transposon model' on the other hand, copies of the source gene may themselves be transcribed and retropose, and hence subfamilies may share some sequence changes, but differ in others. Recent analysis of the Huntington disease region revealed a recently transposed Alu repeat element the sequence of which supported the latter transposon model (Hutchinson et al., 1993) since it was not serially related to previously defined subfamilies.

What is the nature of the material produced by Alu PCR? As previously mentioned, Alus are clustered in R bands. Investigation of R bands (Holmquist 1992) has shown that they are not all homogeneous and some R bands are known as T bands. T bands are particularly GC and Alu rich and although they represent just 15% of all R bands they have been shown so far to accomodate 65% of all mapped genes. This could be related to the fact that two thirds of all genes are associated with GC-rich CpG islands (Craig and Bickmore 1993). This concentration of Alus in gene rich locations makes it possible to postulate that inter Alu PCR products must often span exon sequences and thus can be used as probes to screen cDNA libraries, allowing the isolation of an expressed gene from a specific chromosomal region (Klein et al., 1993). It has, however, been shown that the majority of inter-Alu sequences are derived from within introns and 3'

untranslated regions (Adams et al., 1991) hence this is unlikely to be a very useful technique. Some exceptions do exist, for example the LDL-receptor gene where Alu sequences were located in the nontranslated part of the last exon (Südhof et al., 1985). Analysis of this gene also shed light on the ability of Alu elements to promote genomic rearangements. The gene contains Alu elements in most of its 17 introns, and eight out of nine rearrangements characterised at the sequence level involve Alu elements. Further investigation led to the identification of a conserved 26bp core sequence which it was proposed could function as a recombinational hotspot (Rüdiger et al., 1995). Thus, as well as helping in the identification of candidate genes, Alu elements in a few cases may be responsible for the actual defect. In the main, however, Alu-PCR has been used to increase the density of markers for a specific region. The A rich extensions at the 3' end of Alu repeats are potential VNTRs since it is thought that the majority may be polymorphic (Charlieu et al., 1992). A significant proportion (6.7%) of these extensions consist of microsatellites such as (CAA)n, (CAAAA)n, (TAA)n...in which polymorphisms are easier to detect than in polyA runs (Economou et al., 1990). Thus if a polymorphic marker is required for linkage analysis of a specific region it is advisable to sequence the ends of any Alu repeats which map to that region.

1.1.5 Isolation and mapping of repeat free clones

A number of groups (Meese et al., 1992 , Bull et al., 1993) have carried out regional assignment of Alu-PCR products using the primers and methodology suggested by Brooks-Wilson et al (Brooks-Wilson et al., 1990). The primer used,

A1, hybridises to the extreme 3' end of the consensus sequence and can therefore be used to create probes that do not require preannealing. This primer, however, produces fewer products than other primers located more internally in the Alu consensus sequence, probably because it is unable to prime from the more anciently inserted and hence sequence diverged Alu elements (Tagle et al., 1992). Thus to produce the maximum number of products, more internally located primers must be use. This, however, results in products which are flanked by Alu repeat sequence and this has to be efficiently competed out before they can function as unique or low copy probes. Frequently this has meant the use of rigorous competition and stringent washing conditions, which often led to considerable depletion of the amount of functional probe (Monaco et al., 1991). However, methods have been devised which allow extensive competition of pooled Alu -PCR products while minimising this depletion (Cole et al., 1992b). To function as markers Alu-PCR products have to be mapped individually. Thus a method was required which produced the maximum number of products and which involved the least complicated hybridisation protocol. This was achieved by using a two stage amplification procedure: first an internal primer (614) was used under conditions which gave the maximum number of products; then a primer designed to hybridise to the extreme 3' end (Alu18) of the Alu consensus sequence was used to reamplify these products, so the resultant products contained minimal Alu sequence and were available at a high enough density to make them a useful mapping resource. The

next important stage was to optimise the efficiency of cloning these products and hence maintain the number and diversity of Alu-PCR products.

1.1.6 Cloning PCR products

PCR products are largely blunt ended, and the cloning of such DNA is often very inefficient, resulting in a high background of non recombinants. Blunt-end cloning is hampered by the addition of an extra nucleotide, almost exclusively dATP, to the 3' ends of PCR products by Taq polymerase (Clark 1988). This template independent terminal transferase activity of Taq polymerase has been used to advantage in the T-vector technique (Marchuk et al., 1991). Here the vector is digested at a unique site using a blunt end cutter, Taq polymerase is then used to add a single thymidine to the 3' ends of the linearised vector. Thus the single base overhangs on the vector and PCR product are complementary and therefore can be ligated efficiently. Another widely used approach, cohesive end cloning by virtue of restriction sites intrinsic to or designed into the 5' regions of PCR products, is inefficient and tends to generate a high background of non recombinants. Both methods rely on T4 DNA ligase which under some conditions has been shown to aberrantly cause frame-shift mutations in the vector during the cloning process (Wu et al., 1989). This results in lacZα inactivation in the absence of an insert, and consequently produces white colonies which are termed pseudorecombinants. Thus a method which gave a higher cloning efficiency and lower background of pseudorecombinants than those then available was desired. Initially a method for ligation independent cloning of PCR products was devised. The efficiency of this

method was compared to conventional cohesive end cloning, and an 'in-house' method devised by Chris Boyd for the efficient cloning of blunt ended molecules called 'turbo cloning' (Boyd 1993).

Ligation independent cloning (LIC) is a cloning method which both obviates the need for T4 DNA ligase and increases the specificity of the cloning process. Such a method has previously been successfully used by Aslanidis and deJong, although they modified both vector and inter-Alu fragments by PCR amplification using primers incorporating specially adapted ends (Aslanidis and Dejong 1990). No pseudorecombinants produced by rearranged vector were observed. In the method I used, the vector had already been modified, so a stock could be made in the conventional manner alleviating the need for vector amplification. Both methods rely on the creation of long single stranded tails at the end of Alu PCR products and of the plasmid vector. This permits non-covalent bi-molecular associations via base pairing, allowing circulization between two molecules in the absence of DNA ligase. The vector ends are non-complementary, preventing self annealing, and since linear molecules transform *E.coli* poorly, recombinant molecules can be selected by transformation.

Turbo cloning is a method devised to efficiently clone blunt-ended DNA molecules into plasmids (Boyd 1993). It is simple, rapid (1-2 hours), works with minute amounts of insert, and shows lack of selectivity for small inserts. Essentially the method uses any plasmid vector containing the 34 bp site-specific recombination sequence lox and the cognate Cre recombinase system of bacteriophage P1. Here

a derivative of pBluescribe (Stratagene), pBSLox, was used. This vector permits turbo cloning and retains the blue/white screen for recombinants. There are two stages. Firstly, linear vector and fragment DNAs are ligated intermolecularly under conditions of macromolecular crowding (15% polyethylene glycol 6000) which accelerates blunt-end joining 1000 fold. Secondly, circular recombinant molecules are efficiently excised from the ligation products by Cre recombinase acting on pairs of lox sites within directly repeated vector molecules flanking insert DNA.

In addition to isolating and mapping new markers from chromosome 11q12-13, mutational analysis was carried out on a gene which was a candidate for Usher syndrome IB. The olfactory marker protein gene had been mapped to the region linked to the disorder by Kathryn Evans (Evans et al., 1993) using the somatic cell hybrid panel previously described (section 1.1.1). The expression pattern of the gene, and the fact that the mouse Omp gene mapped to the region linked to *Shaker-1,* the putative mouse homologue of Usher syndrome type I, made it a good candidate for the disorder. These factors and the disorder itself are explained in more detail in the following sections.

1.2 Usher syndrome

1.2.1 Historical background and clinical description of Usher syndrome

The disorder was named after C.H.Usher who recognised in 1914 that it an autosomal recessively inherited condition (Usher 1914). Usher syndrome is characterised by deafness and retinitis pigmentosa, which causes progressive night blindness and peripheral vision loss, and is the *sine qua non* for diagnosis of this disorder. Both hearing problems and retinitis pigmentosa are relatively common in the general population (1 in 1000 and 1 in 5000 respectively), thus a certain number of patients can be expected to have both by chance alone. Usher syndrome, however, is a specific constellation of audiologic, vestibular, and ophthalmic findings and can be clearly distinguished from other conditions. Usher syndrome accounts for about 15-20% of patients with retinitis pigmentosa and 50% of patients with combined deafness and blindness (Rosenfield et al., 1994). The prevalence of Usher syndrome in the general population is fairly constant. In Colombia it is estimated at 3.2/100,000 (Tamayo et al., 1991). In Nordic countries the frequency of Usher syndrome is 3-3.6/100,000 (Grøndahl 1987). In the United States estimates are slightly higher at 4.4 per 100,000 (Boughman et al., 1983). These figures are typical for an autosomal recessive single gene disorder. Usher syndrome is clinically and genetically heterogeneous and several different subtypes have been described. Subtypes are distinguished clinically by phenotypic variabilities in auditory acuity and vestibular function as described later. First I will discuss the phenotypic features associated with Usher syndrome.

1.2.1.1 Sensorineural deafness and vestibular dysfunction

Hearing loss is the most common form of sensory impairment (Duyk et al., 1992). About one in every 1000 children is born with hearing impairment, and about half of these are due to genetic factors (Steel and Brown 1994). The vast majority of cases of severe or profound congenital deafness result from inner ear defects. Other causes are middle ear defects and central auditory abnormalities. Late onset hearing loss is also frequently due to genetic factors which presumably make the ear more sensitive to environmental damage. Many different genes appear to be involved in the development and function of the mammalian inner ear. Abnormalities of the inner ear have been classified into three major groups: morphogenetic defects, where gross structural deformities of the labyrinth result from early developmental abnormalities; cochleo-saccular abnormalities, where the primary defect is to the stria vascularis, a secretory epithelium on the lateral wall of the cochlea duct that controls the composition of the endolymph in the lumen of the duct; and neuroepithelial abnormalities, affecting the sensory epithelia of the inner ear, which fail to complete maturation and subsequently degenerate (Steel and Bock 1983).

In contrast to morphogenetic defects, cochleo-saccular abnormalities, and late-onset hearing loss each of which have had a small proportion of genes responsible for individual disorders identified (Steel and Brown 1994), none of the genes resulting in neuroepithelial deafness have been identified. In the mouse, many genes involved in neuroepithelial deafness have been localized, but none identified

(Lyon and Searle, 1989). This type of pathology is usually autosomal recessive in inheritance, and not part of an obvious syndrome. Thus the identification of the responsible genes in humans becomes extremely difficult due to the heterogeneity involved in neuroepithelial deafness and mouse models are used wherever possible. Usher syndrome is one of the few examples of syndromic neurosensory deafness. Recently some major advances have been made in the localisation of nonsyndromic recessive deafness (Guilford et al., 1994) and of course Usher syndrome. Neurosensory defects are likely to represent the major cause of heritable deafness in humans (Steel and Brown 1994). Such defects result in structural and functional abnormalities of the organ of Corti, which consists of neurosensory epithelium, leading to progressive secondary sensory hair cell degeneration. The high susceptibility of this organ to developmental abnormalities might be explained by its particularly complex and precise arrangement of a variety of types of sensory and supporting cells.

Another part of the inner ear is the vestibular apparatus, which comprises the three semicircular canals, the sacculus, and the utriculus. Known also as the labyrinth, it is a complex sense organ stimulated by gravity and rotational movements. It plays an important role in postural activity, giving rise to impulses which reflexly adapt the position of the trunk and limbs to that of the head and enables erect positioning of the head and normal attitude of the body to be maintained. Impulses from the vestibule also reach the cerebral cortex and subserve the recognition of the position and movements of the head (Keele et al., 1993). Vestibular dysfunction is

commonly associated with deafness. In mice vestibular abnormalities are manifest as hyperactivity, head shaking and circling. Humans, in contrast, are able to use other cues, such as well-developed proprioreceptors and stereoscopic vision to maintain their balance (Steel 1991). Thus more complicated tests are required to detect vestibular areflexia in humans.

1.2.1.2 Retinitis pigmentosa

Retinitis pigmentosa has a prevalence of 1 in 5000 in the United Kingdom and is a significant cause of blindness in old age (Kajiwara et al., 1994). Typically retinitis pigmentosa (RP) causes progressive loss of the peripheral visual fields (tunnel vision) and impaired vision under conditions of reduced illumination, generally leading to complete blindness. This is due to the preferential devastation of rod versus cone receptor cells (Heckenlively 1988). The primary defect in RP lies in the rod photoreceptor; rod electroretinograms (ERGs), a measure of the electrical activity of the photoreceptor cells, are invariably reduced or absent in patients with RP (Humphries et al., 1994). Retinitis pigmentosa is genetically heterogeneous, with most cases being due to autosomal dominant, autosomal recessive or X-linked genes (Wright 1992). Already more than 60 molecular mutations are known to cause RP and these result from mutations in four different genes (rhodopsin, peripherin/RDS, the beta subunit of cyclic GMP phosphodiesterase, and the rod cyclic GMP activated channel protein) (Goldberg 1994). A novel genetic transmission pattern has also been identified in RP. Three families have been

isolated in which RP is caused by mutations in the unlinked loci peripherin/RDS and ROM1, indicating digenic inheritance (Kajiwara et al., 1994).

1.2.1.3 Mental retardation

A review of early reports of Usher syndrome (Hallgren 1959) suggests mental retardation is a component of the disorder, subsequent investigations, however, failed to confirm this association (Kimberling et al., 1991). Early diagnoses were not based on standardised modern criteria, but on educational and social performance. In addition, mental retardation was not a consistent feature of sibs concordant for Usher syndrome, suggesting problems with diagnosis, or an unusual association between the two. Vestibular dysfunction leads to a delay in motor functions and this could have led to over diagnosis of mental retardation. MRI and CAT scanning have, however, revealed cerebellar and cerebral abnormalities supporting the hypothesis of a CNS effect of the Usher genes (Piazza et al.,1987, Schaefer et al., 1991). Further study of this potential facet of the disorder is required.

1.2.2 Clinical subtypes of Usher syndrome

In order to identify the affected loci it was important to reliably separate individuals into subtypes. The difficulty in trying to do so has hampered genetic linkage approaches. The fact that USH I and II were consistent within families made it likely that they resulted from separate genetic entities. Miscategorisation across clinical subtypes, which could well result from mutations at separate loci,

could easily prevent identification of true linkage, hence a simple classification scheme was adopted as outlined below (Kimberling et al., 1989).

1.2.2.1 Usher syndrome type I

Type I is clinically the most severe and also the most frequent subtype, accounting for approximately 90% of Usher patients (Kimberling et al., 1991). It has been described in a number of different populations (Acadian: Smith et al., 1992b; Scandinavian: Hallgren 1959; U.S.: Boughman et al.,1983; Colombian: Tamayo et al., 1991; Norwegian- Grøndahl and Mjøen 1986). It is characterised by profound congenital hearing loss and vestibular dysfunction. The latter manifests clinically as a delay in psychomotor development: babies cannot sit without support until around 10 months (normal range 4-5 months), or begin walking before 27 months (normal range 9-12 months). Common motor activities usually learnt later in childhood such as running and cycling are also performed with some difficulty (Smith et al.,1992b). Hearing aids are ineffective and speech is usually unintelligible, so sign language is the normal form of communication

1.2.2.2 Usher syndrome type II

This is the second major type of Usher syndrome, and has also been described in a number of different populations (Acadian: Smith et al., 1992; Scandinavian: Hallgren 1959; U.S.: Boughman et al.,1983; Colombian: Tamayo et al., 1991; Norwegian: Grøndahl and Mjøen 1986). Estimates of the frequency of USH2 vary, but it is probably around 10% of the total (Kimberling et al., 1991). Here

individuals have mild hearing loss in the lower frequencies sloping to a profound loss in the higher frequencies and normal vestibular function.

Studies have shown that although Types I and II cannot strictly be distinguished by the age of onset of retinitis pigmentosa or its rate of progression, although a general pattern is observed. Usher type I individuals have an earlier onset of retinal degeneration, in childhood or early adolescence, and more rapid progression, while Usher type II individuals have an onset in adolescence or early adulthood, and slower progression.

1.2.2.3 Usher syndrome type III

This is a much less frequent subtype representing only 2% of all Usher syndrome affected individuals. It is defined by a congenital or early onset progressive hearing loss that leads to profound deafness. Like the other two subtypes it appears to be mainly found in a particular region: several cases have been reported in studies of Nordic countries (Grøndahl and Mjøen 1986). A few other isolated cases have been identified in other populations (Smith et al., 1992b). A recent study showed that almost 20% of Usher patients in Eastern Finland have USH3, including a number of affected sib pairs (Sankila et al., 1993). A founder effect may explain why this subtype is higher in this particular population. Grøndahl and Mjøen (1986) also proposed that it may be impossible to differentiate between type 2 and 3 without comparing audioradiograms for a series of years, or having data available for other family members. They felt that some studies were carried out over too short a period of time to detect progressive deafness.

1.2.2.4 Usher syndrome type IV

This type is clinically similar to type II, but with X linked inheritance. The classification exists, but no convincing evidence for an X-linked inheritance pattern has so far been shown, in this potentially extremely rare subtype (Kimberling et al., 1989).

1.2.3 Carriers of Usher syndrome

Electrooculographic (EOG) studies of carriers with tapetoretinal dystrophies which had been matched for sex and age with normal controls revealed that for some dystrophies an abnormal EOG may indicate a carrier state (Pinkers et al., 1994). EOG is a non-invasive method of assessing the function of the retinal pigmented epithelium. Results were statistically significant for type I and II, although more pronounced for type II. Overall a subnormal EOG Lp/Dt ratio (light peak and dark trough- these measure the adaptation of the eye to light and dark) was present in 30% of the carriers.

Even if carriers could be identified the overall risk of a child being born with Usher syndrome is so small that a screening programme would be hard to justify. In the US there is estimated to be 1 carrier per 100 inhabitants so that the chance that both parents carry the mutated gene is 1/10,000, and the risk of producing an affected child (25%) is then 1/40,000. By the time an affected child is diagnosed most couples will have completed their families, and even unaffected sibs of the affected individual have a very low chance of themselves producing an affected

child. The problems with an autosomal recessive disorder like Usher syndrome come to the fore in a consanguineous population.

Normally consanguinity occurs for religious and geopolitical reasons. In such a reproductively isolated population, a large number of individuals affected with what is normally a relatively rare disorder, indicating the presence of a recessive deleterious allele, can result from the marriage of normal but not distantly related parents (Bonné-Tamir et al., 1994). This is because when one marries "within the family", for example, a cousin marriage, the probability that the individual and his/her cousin carries the same specific rare gene (and that their child be homozygous) is much higher than if the marriage is "outside the family" (Bittles and Neel 1994). Studies have shown that consanguinity is higher among parents of deaf children than in parents of their normal hearing counterparts (Reardon 1992). In a normal population the preference of a deaf individual for a deaf partner may inadvertently increase the risk of a particular deafness disorder. The risks here, however, are potentially less than in a consanguineous population due to the many different causes of genetic deafness and the fact that half the cases of deafness cannot yet be linked to a genetic cause. Perhaps understandably the deaf are often hostile to the proposition of genetic counselling to prevent them having deaf children (Reardon 1992).

The point at which most help can be given is in the early diagnosis of Usher syndrome. With intact vision, deaf individuals can lead a relatively normal life. However, with visual loss their ability to cope is severely restricted. Timely

identification of affected individuals (e.g. by screening children attending speech and hearing clinics and schools for the deaf) will result in their placement in the proper rehabilitative programme to allow them to acquire skills to prepare for the ensuing loss of vision (Samuelson and Zhan 1990).

Hope also comes from the fact that it is now possible to implant a normal cochlea in USH1 affected children. Initial studies have shown that Usher's patients required very little extra effort in rehabilitation compared to other prelingually deaf patients, and all patients reported considerable advantages in hearing abilities and social life (Hinderink et al., 1994).

1.2.4 Genetic subtypes of Usher Syndrome

1.2.4.1 Genetic heterogeneity

The higher frequency of types 1 and 2, and hence greater availability of families, has meant that gene mapping of the Usher syndromes has concentrated on these two clinical subtypes. The differences in phenotype which are consistent within individual families pointed to heterogeneity. It could not be ruled out, however, that the subtypes themselves could result from mutations in a number of genes. Individuals affected by Charcot-Marie-Tooth type I disease (CMT-I), a common form of genetic neuropathy with slow neural conduction, phenotypically gave no indication that they were in fact the result of mutations of more than two genes. Genetic linkage analysis has however indicated that mutations at a locus on the X chromosome and at least three different autosomal recessive loci can result in CMT-I (Chance et al., 1990). Herein lies the problem with autosomal recessive

disorders. Because each family is unlikely to have more than one or two affected individuals, pooling of data across families is necessary to establish statistical evidence for linkage to a marker. If the disorder is heterogeneous then different families will show linkage to different markers and the end results could be inconclusive. The inability to say with certainty that all families included in a linkage study have mutations at the same locus being analysed means that families showing critical crossovers cannot be accepted with total confidence. The problem can be circumvented by using data from large consanguineous pedigrees, but these are seldom available. Alternatively, as was the case in USH2, as large a sample size as possible can be used in the hope that the likelihood distribution calculated would be less sensitive to effects of heterogeneity (Kimberling et al., 1995). The results of linkage studies to date are given below. Six different genetic loci have been linked to Usher syndrome- one coding for Usher Type III, two coding for Usher Type II, and three coding for Usher Type I.

1.2.4.2 Usher Syndrome Type II

The mapping of Usher syndrome type II to chromosome 1q in 1990 (Kimberling et al., 1990) was a significant breakthrough. USH II was the first autosomal recessive form of retinitis pigmentosa to be localized, as well as being the first autosomal recessive gene causing deafness. Informative members from 8 type II families were typed for three loci previously mapped to chromosome 1. A maximum pairwise LOD score of 6.04 was obtained for the marker THH33 (D1S81), which placed the gene in the distal one third of the long arm of

chromosome 1. Similar analysis of Type I families showed no linkage to any of the three loci. Lewis et al (1990) confirmed the assignment of USH2 to 1q32-q44, and provided further support for heterogeneity of USH1 and 2 (USH1 families again failed to show linkage), using different families. This demonstrated that the observed clinical heterogeneity of Usher syndrome is due to the effects of mutations in two or more separate and distinct loci. Using additional markers from the 1q32-1qter region the map location of USH II was further refined in the original families by Kimberling et al. Significant linkage was observed with the marker D1S48 and the favoured order was USH2-D1S81-D1S48, although this may have been biased by the low level of heterozygosity of surrounding markers and the fact that more information was gained from D1S81 than D1S48 (Weston et al., 1991). Efforts to isolate and clone the gene were complicated when typing and linkage analysis of an USH2 family failed to show linkage to markers from 1q32-44 (including D1S81) suggesting a second different gene was affected in this family (Pieke Dahl et al., 1991, 1993). The locus at 1q32-q44 has been termed USH2A and the new unlinked locus designated USH2B. There was no obvious clinical distinction between this new family and previous families, but clearly it would be useful if some differences could be defined.

There are two main approaches to take when a disorder has been mapped to a specific chromosomal region, either a) mutational analysis of an appropriately expressed gene which has previously been mapped to the region (the candidate gene approach) or b) in the absence of any suitable candidates, refine the linkage

analysis to cover as small a region as possible then look for genes located in this region which would make suitable candidates (the positional cloning approach). Two genes, phosducin and hCHML, which mapped to the appropriate region made suitable candidates for USH2

The gene for human phosducin was mapped to 1q25-q31.1 (Ding et al., 1993). Phosducin is known to have a role in regulating phototransduction in retinal rod cells and is abundantly expressed in the retina and the brain. These two facts made it a good candidate for Usher syndrome type II. The gene is 18kb in length, has four exons and three introns (Abe et al., 1994). Denaturing gradient gel electrophoresis was carried out on PCR amplified exons from Usher patients which had shown linkage to D1S81 (Ara et al., 1992). No mutations in the gene causing an amino acid change were found, but two polymorphisms were identified which in combination produced a specific haplotype (Ara et al., 1993). The frequency of the polymorphisms was different between individuals affected by USH2 and autosomal dominant retinitis pigmentosa, as compared to controls, suggesting that a gene involved in these two disorders was located in the vicinity of the phosducin gene.

Choroideremia is an X-linked disorder characterised by progressive dystrophy of the choroid, retinal pigment epithelium and retina. The gene responsible for this disorder has been identified (hCHM), and in the process of analysing this gene a new gene showing a high degree of homology to hCHM was identified (Cremers et al., 1992). This new gene was called hCHML (choroideremia-like) and although

the deduced amino acid sequence showed 95% similarity to hCHM, hCHML contained no introns. This suggested that hCHML had originated by reverse transcription of hCHM mRNA and subsequent integration into the genome. Analysis of somatic cell hybrids allowed the assignment of hCHML to chromosome 1q31-qter. This chromosomal localisation and the phenotypic similarities between choroideremia and USH2 made hCHML another good candidate for this disorder. The DNA of USH2 patients was analysed by polymerase chain reaction single-strand conformation polymorphism analysis and direct sequencing (van Bokhoven et al., 1994). No mutations were detected.

Thus it would appear that we have gone from two potential candidates to none. It is worth considering however the reliability of the methodology used to identify mutations and extent of each gene which was actually analysed. DGGE is highly reliable and able to identify 95% of mutations. SSCP is less reliable and the conditions used to analyse CHML would miss 20% of all possible mutations. It is also important to note that neither the 5' nor the 3' untranslated regions, which could contain important regulatory sequences, were analysed. Similarly the introns in the phosducin gene were not screened for mutations. The formal exclusion of both genes therefore awaits either more complete mutation analysis; detailed RNA and protein analysis from B-cell lines of patients; more detailed linkage analysis; or of course, the identification and confirmation of a new candidate gene.

Recently more detailed linkage analysis has placed the USH2A gene into a 2.1-cM region between the markers D1S237 and D1S229 on chromosome 1q41

(Kimberling et al., 1995). This localisation rules out phosducin as a candidate, since it was mapped to 1q25-q31.1. Linkage analysis was carried out on a large sample of 68 informative families in order to reduce the effects of locus heterogeneity: approximately 12.5% of the families studied were estimated to be unlinked to 1q41. Cytogenetic analysis of a member from each of 38 families failed to identify any deletions, inversions or translocations. No suitable candidates have been shown to lie within the linked region, hence work has been initiated to construct a YAC contig and physically map the region.

1.2.4.3 Usher Syndrome Type III

It was Sankila et al (1993) who observed an increase in the prevalence of USH3 in Finland. The most recent studies by this group have shown that this proposed enrichment of the USH3 gene is even higher than at first anticipated (Sankila et al., 1995), and may be as much as 42%. Sankila et al excluded the four regions previously linked to Usher Syndrome (Sankila et al in preparation), then carried out a systematic search by genetic linkage analysis in 10 multiple affected Finnish families using polymorphic microsatellite markers. This led to the assignment of USH3 to the 5-cM interval between the markers D3S1555 and D3S1279 in 3q21-25. This confirms the existence of Usher Syndrome Type III as a distinct entity. A number of eye-related genes have been mapped to 3q21-25 (rhodopsin (RHO) in q21-24, cellular retinol binding protein 1 (CRBP1) in q21-22, and CRBP2 in 3p11-qter - Naylor et al., 1995), but further refinement of the region is necessary before candidates can be confidently identified.

1.2.4.4 Usher Syndrome Type I

There was strong evidence that the clinical heterogeneity in Usher syndrome was due to genetic heterogeneity (Kimberling, et al., 1990, Lewis et al., 1990), and although no alternative location for USH1 was suggested, 1q was excluded since USH1 families failed to show any linkage to markers from this region (Weston et al., 1991). The first evidence of linkage to USH1 came from Kaplan et al in 1992 when they showed linkage of 10 out of 15 families to the marker WLJ14 (DS14S13) at chromosome 14q32 (USH1A). Notably 8 of the 10 linked families came from the Poitou-Charentes region in Western France, suggesting a founder effect. There were no obvious candidate genes in the linked region, although a mouse gene controlling heart position has been mapped to the region homologous to human 14q32. This is termed the *iv* locus, which lies 3-cM from the immunoglobulin heavy-chain constant-region complex (Igh-C) on distal chromosome 12 (Brueckner et al., 1989). Normal ciliary function during foetal development is required for the asymmetric left-right position of heart and viscera in higher vertebrates which links in with the fact that the proposed defect affecting the 3 different sensory systems in USH1 individuals is a congenital abnormality of the cilia (see later). Thus a cluster of genes on 14q involved in ciliary motility has been proposed (Kaplan et al., 1992). This proposition is strengthened by reports of two independent cases of Kartagener syndrome (abnormal heart position and deficient ciliary function in respiratory epithelium and sperm) in association with retinitis pigmentosa and deaf-mutism (Lake and Sharma 1973).

In the meantime, extensive linkage analysis (Keats et al., 1992) using affected individuals from 36 families from the United States, England and Sweden, and 206 markers allowed the exclusion of 39% of the genome. No linkage was observed to 14q32 indicating that there were at least two loci responsible for USH1. The next major breakthrough came when Smith et al (1992a) localised two regions linked to USH1 to chromosome 11. The analysis was carried out on British and French-Acadian families, and demonstrated further locus heterogeneity. Each was linked to a separate marker and both failed to show linkage to the region linked to the other geographically defined family set. The French-Acadian families gave a maximum LOD score with the locus D11S419 which had been localised to 11p15-p13 (USH1C). The British families, on the other hand, gave a maximum LOD score for the locus D11S527, which had been localized to 11q13.5 (USH1B). The most common Usher Type I mutation is that found at the 11q13.5 locus (Kimberling et al., 1995).

The localisation of the Acadian Usher syndrome type I locus (USH1C) was further refined to a 2-3cM interval in the p15.5-p14 region of chromosome 11 by Nouri et al in 1994. The results obtained showed evidence for a founder effect. No recombination events were found between USH1C and three microsatellite markers (D11S419, D11S921, and D11S899), and all of the 54 chromosomes with the abnormal allele at the disease locus had a specific haplotype for D11S419/D11S921. The Acadians are the descendants of French immigrants who colonized the Canadian territory known as Acadia (now Nova Scotia and

surrounding areas) in 1604. After being expelled by the British in 1775, many Acadians made their way to Louisiana. They settled on the plains among the Bayous and remained relatively isolated because of linguistic, religious, and cultural cohesiveness, as well as geographic isolation. Thus, a strong founder effect would be anticipated. No obvious candidate genes are known to map to the region delineated by this study. Recently, however, a possible mouse model for USH1C has been identified (Heckenlively et al., 1994). The *tub/tub* strain of mice were determined to have progressive retinitis pigmentosa and deafness. Crossing with known types of retinal degeneration showed that the *tub/tub* strain had a unique retinal degeneration and mapping indicated linkage between *Hbb* and *tub* on mouse chromosome 7. This region is syntenic to human 11p15, making the mouse strain a good candidate for the human Acadian Usher homologue.

1.2.5 Ciliary defect

I have already hinted at the possible involvement of a ciliary defect in Usher syndrome. There are a number of characteristic defects observed in Usher syndrome affected individuals which led to this theory, which I will now discuss. I will then explain how the olfactory marker protein through its expression in olfactory cilia and its map position in mouse and man became a candidate for USH1B.

The presence of a congenital abnormality of the cilia in Usher syndrome affected individuals could explain how a single gene defect could affect three different sensory systems. This is because photoreceptors, auditory hair cells, and vestibular

hair cells all develop from ciliated progenitors. In addition, primary ciliary dyskinesia due to generalised abnormality of the axoneme structure could account for the non progressive nature of the hearing loss, since axonemes are present in auditory hair cells only during their development (Kaplan et al.,1992).

Several lines of evidence indicate that such a generalised abnormality of the axoneme structure has occurred in the cilia of Usher syndrome affected individuals. Many of the studies to date have concentrated on nasal cilia and sperm, which are more accessible sources of human material then the inner ear and the eye, in order to study the basic pathological process. Arden and Fox (1979) found an increased incidence of abnormal ultrastructure of axonemes in nasal cilia. Difficulties in biopsying photoreceptors has also led to the study of sperm in Usher syndrome affected males. It is possible to assess the functional capabilities and three-dimensional structure of cilia in sperm. Hence, Hunter et al (1986) were able to demonstrate that as well as the presence of a number of sperm tail abnormalities, both sperm motility and velocity were decreased in the patient group. Electron microscopic examination revealed abnormal axonemes in not only the sperm, but also the photoreceptors of these patients. Further evidence for the presence of abnormal photoreceptors came from Berson and Adamian's (1992) study of the eye of an USHII affected individual, which led them to conclude that the majority of the connecting cilia of the photoreceptors were abnormal. In an attempt to explain the progressive nature of the Retinitis Pigmentosa they proposed that abnormalities in the cilia might affect the microtubule membrane crosslinker

leading to impairment of the process of outer segment renewal. Alternatively, disruption of the pattern of microtubule doublets could be preventing actin from carrying out its essential role in the process of outer segment renewal. Transport of soluble proteins such as arrestin from the inner to outer segment may also be disrupted due to abnormalities in the structure of the cilia and therefore affect photoreceptor function and viability. The connecting (sensory) cilium of rods and cones is the stalk that separates the outer segment that contains visual pigment in stacks of membrane discs from the inner segment that contains cytoplasmic organelles involved in protein synthesis. Morphological defects in the connecting cilia in photoreceptors belonging to an USH2 affected individual were also observed by Barrong et al (1992).

Shinkawa and Nadol (1986) were able to carry out a histological examination of the inner ear of an USH3 affected patient. They observed significant primary neuronal degeneration throughout the cochlea, leading to a decrease in outer ciliary cells. This finding was supported by earlier work by Buch and Jorgenson (1963) who also observed cochlear degeneration, but this time in USHI affected indviduals. The most recent evidence for a generalised abnormality of cilia in patients with Usher syndrome came from Bonneau et al (1993) who observed bronchiectasis, chronic sinusitis, and reduced nasal mucociliary clearance in two USHI affected sibs.

1.2.6 Olfactory marker protein as a candidate

1.2.6.1 Olfaction

The sense of smell arises from stimulation of receptors in the olfactory mucosa. The olfactory mucosa is above the main respiratory stream and sniffing is often required to stimulate receptors. In man the olfactory mucosa contains about 10-20 million special sensory cells each of which is a bipolar neuron (Keele et al., 1983). The receptor cells lie between supporting cells and their dendrites extend as naked rods which end in fine cilia (6-12 to each rod) which lie in the mucus covering the olfactory mucosa. The supporting cells end in microvilli which secrete mucus. The axons from the olfactory receptor cells are fine unmyelinated fibres, which run in bundles or fascicles containing 20-100 fibres; each fascicle is enclosed in one mesaxon or Schwann cell process. The fascicles run together in the fila olfactoria which pierce the cribriform plate and enter the olfactory bulb. Within the outer layers of the bulb the axons of the olfactory nerves enter the glomeruli and these form synapses with dendrites from mitral and tufted cells. The latter forming the second-order neurons on the olfactory pathway. Olfactory neurons are unusual in that they are continually replaced throughout adult life. Neurons lost as a result of normal neuronal turnover in mature animals or in response to experimentally induced chemical or surgical insult, are replaced by stem cells which are part of the population of basal cells in the neuroepithelium (Graziadei and Monti Graziadei 1978)

1.2.6.2 Olfactory cilia

In many cells e.g. those lining respiratory epithelium the major function of cilia is concentrated in the cytoskeletal core, and cilia appear as motile cellular elements. Olfactory cilia are much more highly specialized and have acquired a secondary function whereby a key role is played by the enclosing lipid bilayer and the axoneme becomes an inert scaffold. This modification of ciliary function and structure has also occurred in other sensory cells i.e. the development of the elaborate membraneous disc structures of retinal rod outer segments. The cilia of both olfactory and visual sensory cells appear to have evolved to give increased receptive membrane area (Lancet 1988)

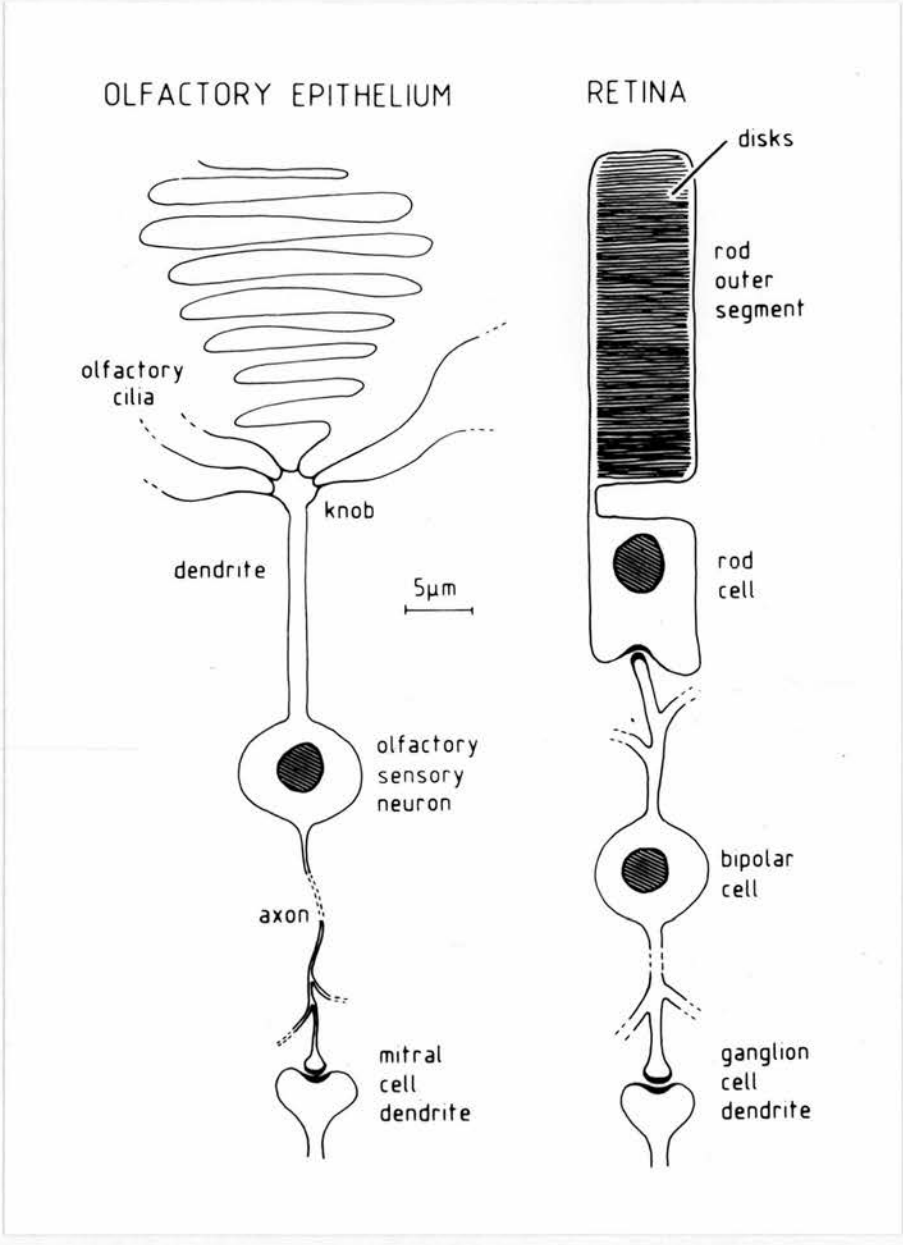


Figure 1.4 Olfactory and visual sensory cells

Schematic representation of the structures involved in transmission of sensory signals in the peripheral olfactory and visual sensory systems, drawn approximately to scale. In both, transduction occurs in modified ciliary structures, olfactory cilia, and retinal rod segments. (Lancet 1988)

1.2.6.3 Odorant induced signal transduction

In the predominant model of the initial events underlying odorant-induced signal transduction the initial event is when odorants bind to receptors on the surface of the cilia that emanate from the extended dendritic process of olfactory neurons. This then activates either one or both of the second messenger pathways in the cilia where either; a) G proteins are activated to interact with membrane bound adenylyl cyclase, resulting in an increase in cAMP which opens a cyclic nucleotide responsive cation channel, depolarizes the cell and initiates an action potential; or b) binding causes the generation of inositol 1,4,5-triphosphate (IP3) which leads to the activation of an IP3 receptor-channel complex in the cilial membrane. Thus the large family of receptor molecules seems to rely on just two secondary messenger systems emphasizing how much the olfactory sensory neuron has evolved to deal with a multitude of diverse ligands while using the components of a simple transduction cascade (Cunningham and Reed 1992). Having outlined the major features of the olfactory system and olfactory cilia, it is necessary now to look at the expression pattern and putative function of OMP to see why it became a candidate for Usher syndrome Type IB.

1.2.6.4 Olfactory marker protein

Olfactory marker protein (OMP) was originally observed as a band on acrylamide gel electrophoresis of native mouse olfactory tissue cytoplasmic extracts. The protein was subsequently purified and characterised from both mouse and rat

olfactory tissue, and specific antibodies prepared. OMP has a molecular mass of 19,000Da and an isoelectric point of approximately 5. Purified OMP protein contains no carbohydrate, or the amino acid cysteine. It represents about 0.1-1% of the olfactory tissue cytoplasmic protein, and antisera directed against rat OMP has been shown to cross react with extracts from olfactory epithelium of many species including humans. In common with other cytoplasmic proteins the amino terminus of OMP is acetylated. No homology has been shown between OMP and any other proteins in the database, and there is no evidence of it exhibiting any common enzymatic activity, suggesting that it plays a unique role in olfactory neuron function (Margolis 1988).

How does olfactory marker protein fit in with the overall picture of odorant-induced signal transduction? In rats immunohistochemistry has been used to show that OMP is restricted to olfactory neurons. It is present in the cytoplasm of the cell body, the peripherally directed dendritic knob, in the proximal portion of the cilia, as well as in the axon and synaptic terminals in the olfactory bulb. It is not expressed in any other cell type in the olfactory mucosa and disappears at the boundary between the olfactory and respiratory mucosa. OMP is synthesised in the mature olfactory receptor perikaryon from which it is transported to the entire cytoplasmic compartment of the neuron in association with the slow component of axoplasmic flow. OMP turnover has a biphasic half-life with a rapid component of about a day and slower component of about a week. Since OMP is absent from the neural precursor basal cells in the olfactory neuroepithelium, it has been used as

a marker for the mature, functioning receptor cells whose axons comprise the first cranial nerve. In rodents, expression of OMP has also been observed in olfactory neurons of the vomeronasal organ (Farbman and Margolis 1980). Less OMP expression is observed in the vomeronasal organ than in cells of the nasal cavity; the major morphological difference between olfactory receptor cells from the two sites is the absence of cilia in the cells of the vomeronasal organ. This indicates that OMP may be involved in sensory transduction in the cilia, although it could also be involved in a more basic function in this specific cell type.

Expression of OMP is first observed in olfactory neurons early in the last trimester of gestation at about the time when innervation of the olfactory bulb begins. In the mouse, expression of OMP in receptor cells is observed one day after synaptogenesis, and evidence in rodents is consistent with the theory that formation of a synapse induces receptor cells to synthesize OMP. In humans, however, OMP is not synthesized until approximately two months after synaptogenesis has occurred (Meng and Zheng 1992). Another difference between humans and rodents is that unlike the latter, where OMP expression is seen as an indication of mature olfactory neurons, human OMP is found in selected groups of cells of various ages and more likely serves as an indicator of the metabolic state of the cell.

The function of OMP remains unknown although study of its expression throughout vertebrate classes could led to fundamental insights regarding its role, as indicated above. In addition, identification of OMP expression in amphibians,

which are extensively utilised as a model system for innovative studies of sensory transduction in isolated olfactory cell preparations, could be very useful. Thus the demonstration of OMP expression in olfactory receptor neurons of frogs and salamanders was seen as a major breakthrough (Krishna et al., 1992). The distribution of OMP in olfactory receptor neurons of amphibians is characteristic of that already reported in mammalian species. The presence of OMP in the primary olfactory system of a wide variety of species that range from fish to human (Margolis 1988), and the phylogenetic conservation of cellular localization to olfactory receptor neurons strongly supports the supposition that this protein plays an important role in the function of ORNs. The evidence presented so far indicates that OMP is selectively expressed in olfactory receptor cells, but no olfactory defects have been observed in Usher syndrome patients. There is however immunohistochemical evidence for low levels of OMP in discrete neurons in the cerebellum, hypothalamus, and spinal cord of rat and hamster, indicating that OMP might be a more ubiquitously distributed neural protein (Baker et al.,1989). This supports the hypothesis of a CNS effect of the Usher genes due to the observed cerebellar and cerebral abnormalities (Piazza et al.,1987, Schaefer et al., 1991).

Equally as important as the ability to observe the expression of OMP in its normal cellular context is the ability to manipulate the expression of OMP both temporally and spacially. Thus it became essential to clone and characterise the mRNA and gene for OMP in order to determine how expression of the gene was controlled. This was done initially in the rat. Rogers et al (1987) isolated a very large

polyadenylated mRNA and were able to show that the primary translation product was indistinguishable from OMP isolated from olfactory tissues. The isolated mRNA and resultant cDNA allowed confirmation of the amino acid sequence of the protein, and showed the gene to be composed of a 486 nucleotide coding region followed by 1630 nucleotides of the 3' untranslated region. The absence of the initiator codon and the cap site from the cDNA indicated that a portion of the 5' noncoding region was missing from the cDNA, although the presence of a poly(A) tail suggested that the cDNA insert was complete at the 3' end. Danciger et al (1989) then isolated a phage clone containing the whole gene including the 5' untranslated region. They were able to show that the coding region was colinear to that previously observed by Rogers in the cDNA clone, indicating that OMP is intronless. The gene lacked both CAAT and TATA boxes in the flanking region. These features are normally characteristic of a pseudogene so it was necessary to prove that the isolated gene was capable of being expressed. This was done by making an 11kb chimeric gene where the coding region was replaced by Thy1.1. Immunohistochemical localisation of THY1.1 transgenic mice showed the expression of the chimeric gene to be restricted to olfactory receptor neurons and their axons, consistent with the presence of regulatory sequences controlling the olfactory specific expression of the OMP gene. The position had now been reached where it was possible to identify the minimal sequence information necessary to achieve biological specificity of OMP expression and determine the method by which this is achieved.

The rodent OMP promoter was analysed by electrophoretic mobility shift assay (EMSA). Here radiolabelled OMP gene DNA fragments were incubated with nuclear protein extracts and the extent of gel retardation in the presence and absence of inhibitors assessed. DNaseI footprinting was also carried out. Together these methods allowed the identification of two elements that interact with an olfactory neuron specific factor(s) (Kudrycki et al., 1993) and an additional region (upstream binding element-UBE) involved in binding of nuclear proteins in extracts from all tissues tested. The former two elements were shown to be located approximately 0.5Kb apart and share a common motif that binds the novel olfactory neuron specific factor Olf-1 (Kudrycki et al., 1993). A cDNA clone encoding the Olf-1 binding activity has recently been identified by genetic selection in yeast (Wang and Reed 1993). The deduced amino acid sequence of Olf-1 is predicted to contain two helical repeats with similarity to the helix-loop-helix class of transcription factors, and may represent the first identified member of a novel subclass containing this protein motif. Danciger *et al* identified two transcription initiation sites at 55 and 58bp upstream of the ATG and using this information Kudrycki et al (1993) designed an OMP-*lacZ* transgene, where the lacZ gene was under the control of a fragment of the rat OMP gene spanning nucleotides -239 to +55 (here 0 was the transcription start point and hence +55 marked the translation initiation codon). This fragment included the proximal Olf-1 binding site and was sufficient to confer olfactory tissue specific expression of a reporter gene in transgenic mice (Grillo et al., 1992; Servenius et al., 1994).

1.2.6.5 A mouse model for Usher syndrome Type IB

Mice have been called the experimental surrogate of humans. The physiological similarities between mouse and man mean that mutations in a mouse gene affecting a specific biological process such as hearing may either give insights into development and function of the ear or serve as candidates for specific human hearing disorders (Nadeau et al., 1991). Support that human and mouse disorders which are physiologically similar may result from mutations in a homologous gene can also be gained on the genetic level. Comparative mapping has demonstrated that loci that are closely linked in one species tend to be closely linked in the other, whereas loci that are loosely linked in one species tend to be unlinked in the other (Lalley et al., 1978). These linked loci therefore mark homologous chromosomal segments, also known as syntenic groups of genes, and have been used to identify candidate homologues between mouse mutations and human hereditary hearing disorders (Nadeau et al., 1991).

These similarities between mouse and man are particularly advantageous when it comes to hearing disorders. Analysis of the development of the human inner ear is not possible in humans, and temporal bone specimens on the rare occasions when they are obtained have undergone further degeneration due to environmental factors such as noise and drugs, over and above the initial genetic defect (Steel 1991). Most functional development of the ear occurs postnatally making it relatively easy to draw conclusions about normal auditory function by observing defects in the mouse (Steel and Harvey). The normal structure of the human and

mouse ear differ in only a few minor features, even the structure of the highly complex organ of Corti is preserved. Thus it is likely that there will be a similar array of genes controlling its development in the two species. Cloning genes for deafness, even autosomal recessive deafness, is a realisable goal in mice because large mouse backcrosses can be constructed that are known to carry the same mutation and that give a high density of recombination breakpoints around the gene of interest. Indeed several groups have assembled large backcrosses of selected mouse strains to facilitate high-resolution gene mapping. For example the European Backcross Consortium has a mapping panel of 1000 mice typed for 78 anchor loci, which provides an informative panel of recombinants.

One of the many mouse mutations implicated in hearing impairment is *shaker-1 (sh-1)*, a spontaneous mutation which has a characteristic phenotype consisting of sensorineural deafness, hyperactivity, head shaking, and circling (Lord and Gates 1929). These behavioural abnormalities are attributed to vestibular dysfunction because of associated structural anomalies in the vestibular part of the labyrinth (Steel 1991). Considerable electrophysiological research has been carried out to characterise the nature of the deafness in *sh-1* mice and this implicated either the inner hair cells or their innervation as the primary site of the lesion (Steel and Harvey 1992).

The mutation is autosomal recessive and shows complete penetrance. This disorder was localised to the distal region of mouse chromosome 7 (Lyon and Searle 1989) where it is flanked proximally by the gene tyrosinase (*Tyr* at the *c*

locus) and distally by the gene for β-globin (*Hbb*). This mapping result was fortuitous, since the locus *c* is a well-known coat-colour locus, 'null' mutations at which produce albino (or chinchilla etc.) mice. The fact that it is a visible marker led to the *c* locus being used in specific locus mutation-rate experiments and as a result a large number of mutations were induced at this locus (Doolittle et al., 1990). Many of these mutations were chromosomal deletions of varying lengths that affected neighbouring genes in addition to the primary selected locus and as such facilitated mapping of other mutations, in this case *sh-1*, linked to this region. Thus Rinchik et al (1991) were able to localise the *sh-1* locus within the albino (*c*) deletion complex by looking for deletions which could elicit pseudodominance over *sh-1* i.e. crossing mice heterozygous for the lethal albino deletions to *sh-1* homozygous mice resulted in animals exhibiting the *sh-1* phenotype hence the deletion must include the *sh-1* locus. The proximal border of the *sh-1* interval was refined by the mapping of an N-ethyl-N-nitrosurea (ENU) induced lethal mutation within the deletion complex. In addition, two loci defined by molecular clones, *Omp* and *D70R1* (an anonymous clone) were mapped to the distal region of non overlap. An intraspecific backcross segregating for the *sh-1* mutation [(C57BL/10+/+ × sh-1/sh-1) × (sh-1/sh-1)] was then set up and yielded a total of 1066 progeny mice, 67 of which were identified as recombinant in the vicinity of the *sh-1* mutation (Brown et al., 1992). This panel was then used to generate a detailed genetic map of the region and identify markers closely linked to the *sh-1* locus. Thus only one recombinant was identified between *Omp* and the *sh-1* locus.

Therefore the two loci were separated by a genetic distance of 0.1cM which in the mouse genome represents an approximate physical distance of 200kb. Subsequent genetic mapping of a number of other genetic markers through the recombinant panel indicated that the *Omp/sh-1* recombinant mouse had been mis-scored (Brown,K.A.,Steel,K.P., and Brown,S.D.M., unpublished data).

At this point the OMP gene was mapped in humans (Evans et al., 1993) to 11q13.5, identifying a new region of homology between mouse chromosome 7 and human 11. The gene was mapped using a subfragment of a human cosmid, which encompassed the OMP gene, as a hybridisation probe on a somatic cell hybrid panel carrying various portions of chromosome 11. This allowed the gene to be placed on the long arm of chromosome 11 centromeric to TYR. This position was further refined using fluorescently labelled cosmid as an *in situ* hybridisation probe on banded human metaphase chromosomes. Hence human OMP was mapped to the region linked to Usher syndrome type Ib. The similarities in the mode of inheritance, penetrance, and the basic pathology between Usher syndrome and *sh-1* led to the conclusion that the two disorders may be homologous. Retinitis pigmentosa had not been observed in *sh-1* mice, but examples do exist where mutations in homologous genes result in similar but not identical phenotypes in mouse and man. Mutations in the human PAX3 gene causes varying degrees of deafness in association with dystopia canthorum and variable degrees of pigmentary disturbance. In contrast, mutations in the homologous mouse gene Pax-3, which is located in a region syntenic to that wherein the human gene is

located, cause pigmentary anomalies, cranofacial abnormalities, but no auditory defects (Steel and Smith 1992). It was proposed that the difference in phenotype may be due to different parts of the gene being mutated, or variations in modifying influences yet to be identified. In the same manner, therefore, USH1B and *sh-1* could be true homologues, and result from mutations in the olfactory marker protein gene. Thus the comparative mapping results and the observed ciliary defects in USH1B patients made the OMP gene a strong candidate for both disorders.

The possibility that mutations in Omp resulted in the *sh-1* phenotype was investigated by direct sequencing of the coding region in the original *sh-1* mutation and five new mutations recovered following ENU mutagenesis (Rinchick et al., 1990). No sequence variants were detected in the coding region of any of the *sh-1* mutants, and expression of Omp in the original spontaneous *sh-1* mutant was apparently normal. Therefore it was concluded Omp was not the shaker-1 gene (Brown et al., 1994). While this result lessened the confidence in the OMP gene as the locus responsible for USH1B there were a number of reasons why OMP remained a candidate. It was possible that *sh-1* and USH1B were not true homologues and mutations in OMP result in USH1B. Alternatively neither disorder could involve OMP directly, but OMP could occasionally be involved in USH1B as a result of a contiguous gene deletion. The third possibility of course was that OMP is not involved in either defect. The only way to confidently resolve

this issue was to characterise the OMP gene in normal individuals and families carrying the Usher Type 1B defect.

In the meantime, Steve Brown and his group, who had formally excluded *Omp* as a candidate in *sh-1*, began to construct a 1.4Mb YAC contig across the *sh-1* region using Omp as a start point. Microsatellites were isolated from two YAC clones which contained *Omp*, and used to analyse backcross progeny from the original *sh-1* cross in order to delimit the *sh-1* non-recombinant region to approximately 500kb. Exon trap products were then recovered from one of the YACs which lay within the non-recombinant region (Gibson et al., 1994). Those products with open reading frames in the correct orientation and showing no homology to repetitive sequences were hybridised to zoo blots. Two exon trap clones-ET17 and ET58- clearly detected sequences conserved across a wide variety of species, which were mapped in humans during the course of this project. If they mapped to 11q13.5 then the gene from which they had originated would be a strong candidate for both disorders. In parallel the Brown group also mapped a potassium channel gene to the non-recombinant region, and an attempt was also made to map this in human.

1.2.7 Mutational analysis of OMP

To confirm that a candidate gene is responsible for a disorder it is necessary to carry out mutational analysis. Once mutations have been observed in a candidate gene, these or other mutations in the same gene must be seen in a number of other affected individuals for their role in the disease phenotype to be confirmed and

understood. The method chosen to carry out mutational analysis on a particular gene depends on a number of factors, and no single method is applicable for all situations. Primarily the procedure selected depends on whether one is screening for known mutations (e.g. in prenatal diagnosis of a common genetic disorder such as cystic fibrosis), or scanning a candidate gene for previously unknown mutations. The latter was the situation applicable to analysis of the OMP gene, hence I will only consider methods relevant to this. The method chosen depends on: the expected nature of the mutation; the size and structure of the locus being analysed; the availability of mRNA; the degree of sensitivity needed; and the equipment available. Candidate genes can be anything from a few hundred to hundreds of thousands of bases long, and may be composed of many exons. Thus the search for a disease causing mutation may require screening thousands of nucleotides. A number of different techniques have been developed which show varying degrees of accuracy and ease of use. If a gross chromosomal structural alteration has occurred this can be readily identified using cytogenetic techniques. If a disorder is caused by a large deletion, insertion, duplication, or inversion in the gene under analysis this may be detectable by Southern blot hybridisation and restriction mapping or multiplex PCR. These, however, are relatively rare compared to alterations in one or a few bases. The most common methods used to detect single base changes are SSCP, DGGE, HOT, and direct sequencing; I will discuss each of these in turn. All of the methods require prior amplification of the gene from genomic DNA and as such are PCR based.

1.2.8 Comparison of methods to detect single base changes

1.2.8.1 Single-strand conformation polymorphism (SSCP)

Overall, this is the fastest and simplest technique. Wild type and mutant DNA are amplified by PCR, denatured, then electrophoresed adjacent to each other in a non-denaturing polyacrylamide gel containing glycerol. The single stranded DNA molecules from each denatured product assume a 3-D conformation dependent on their primary sequence. If there is a difference between wild type and mutant DNA differential migration may result. The major disadvantages of this technique are that it only detects up to 90% of mutations and screening must be carried out on small fragments (less than 400bp) (Condie et al., 1993). It has nevertheless been used very successfully to identify mutations (Berger et al., 1992) and has the advantage that due to its simplicity a large number of samples can be screened in parallel.

1.2.8.2 Denaturing gradient gel electrophoresis (DGGE)

Once the correct conditions have been established, DGGE is also relatively simple and can be carried out on a large number of samples in parallel. Here PCR products are electrophoresed through an increasingly stringent gradient of the denaturing agents urea and formamide. When the denaturation stringency becomes high enough strands of DNA dissociate in discrete sequence dependent domains of low melting temperature. Again sequence differences between wild type and mutant DNA can cause each to migrate differently due to different degrees of 'melting'. The sensitivity of the method is greatly enhanced if heteroduplex DNA

formed between wild type and mutant DNA is analysed. To ensure the sequence heterogeneity lies within a domain of relatively low melting temperature a 30-50bp high melting temperature GC rich sequence (GC-clamp) is attached to one of the primers, and primers are designed using a computer program to predict theoretical melting profiles. The fragment sizes screened depend on the melting profile of the DNA, but are normally in the range of 500bp. This technique has also been used very successfully in the identification of mutations (Fodde et al., 1992). The denaturing concentration which gives maximal separation between the wild-type and mutant DNA fragments can be selected experimentally, once a particular mutation is known, and this can then be used instead of a denaturing gradient (CDGE- constant denaturing gradient gel electrophoresis). Theoretically DGGE/CDGE has the potential to pick up 100% of mutations, although in practice the figure is more like 90% (Condie et al., 1993). DGGE/CDGE requires quite a lot of effort to set up although commercial equipment and software are now available.

1.2.8.3 Chemical mismatch or hydroxylamine and osmium tetroxide (HOT)

The chemical mismatch or HOT (because of the use of the chemicals osmium tetroxide and hydroxylamine) technique is the most accurate, but most demanding of the mutation detection techniques. A heteroduplex is created between wild type and mutant PCR products by boiling and reannealing. Mismatched bases in the heteroduplex molecules are then modified using Maxam-Gilbert sequencing chemistry; osmium tetroxide is used to modify mispaired thymines and

hydroxylamine for mismatched cytosines. The DNA is cleaved at modified bases by piperidine. In order to detect the point at which cleavage has occurred, either wild type or mutant DNA is labelled, and the products run out on a denaturing polyacrylamide gel which is autoradiographed. The precise distance from the end of the molecule where the sequence alteration has occurred is indicated by the size of the cleavage band and the cleaving reagent. Only a small proportion of the mutant gene needs therefore to be sequenced. The HOT technique has also been used very successfully (Prosser et al., 1990) and is capable of picking up all mutations (Condie et al., 1993). Moreover this method has the lowest size constraints of the three mutation scanning techniques and products of up to 1.7kb have been accurately screened (Saleeba et al., 1992).

Both SSCP and DGGE/CDGE are simple to carry out, and can be used to screen a large number of samples. The HOT method is very accurate and can be used to analyse large fragments, but involves the use of hazardous chemicals, and is much more labour intensive than the other two. In all cases the final step is to sequence the fragment under analysis to confirm the presence of a mutation. The HOT technique has the added advantage that the position and nature of the mutation is produced and therefore only a defined region needs to be sequenced.

1.2.8.4 Recent advances

A number of advances have recently been made to the DGGE, SSCP, and HOT techniques. These were made after I had developed an experimental strategy for

OMP mutation analysis, but are worthy of consideration for future candidate gene analysis.

DGGE

A technique has been developed which overcomes the size limitation of DGGE. Guldberg and Güttler (1994) were able to show that multiple DNA fragments can be scanned for mutations by heteroduplex analysis in a 0-80% denaturing gradient gel, where the running time is determined as the time it takes for the shortest migrating fragment to reach its final position in the gel. Thus by using one 'broad-range' gel and protocol they were able to analyse simultaneously all 13 PCR products covering the human phenylalanine hydroxylase gene and detected all of the 75 expected point mutations.

SSCP

SSCP has also undergone recent improvements due to the advent of automated fluorescent sequencing technology. A new technique called the multiple fluorescence-based polymerase chain reaction single-strand conformation polymorphism (MF-PCR-SSCP) has been developed (Iwahana et al., 1994). It takes advantage of the fact that the ABI Model 373A DNA sequencer can detect up to four types of fluorescence in one lane, so each sample can include an internal control and lane to lane differences in mobility can be standardised. In addition, a gel temperature controlling apparatus has been developed which allows strict temperature control. This in combination with a higher percentage gel, than used

in standard SSCP analysis, enables better separation and can result in 100% detection of mutations often under one set of electrophoresis conditions.

HOT

Similarly the HOT technique has been adapted to make use of fluorescence technology, and a new technique called FAMA (fluorescence assisted mismatch analysis) has evolved (Verpy et al., 1994). The wild type and putative mutant allele which form the heteroduplex are end labelled with strand specific fluorophores prior to the formation of heteroduplex. This further increases the sensitivity of detection, because fluorescent differential labelling allows the identification and measurement of strand-specific background cleavages at matched cytosine or thymine residues. The technique is less time consuming than conventional HOTs and produces a high signal-to-noise ratio indicating that it may be applicable to the detection of somatic mutations.

1.2.8.5 Direct sequencing

Direct sequencing is the final step in mutational analysis of candidate genes, but for a number of reasons it is used to confirm rather than initially identify mutations. Sequencing is more effective when the template is made single stranded. This may be achieved by carrying out an initial PCR in which one of the primers is biotinylated. The DNA fragment is then selectively immobilised to streptavidin coated magnetic beads and after strand-specific elution, the eluted strand as well as

the remaining immobilised strand, is used for bidirectional dideoxy sequencing (Hultman et al., 1991). An alternative method is cycle sequencing which involves the simultaneous amplification and sequencing of template by the addition of dideoxy terminators to a secondary PCR reaction. This has the advantage of requiring minute amounts of template, and the use of vast excess of sequencing primer and the high temperature of the reaction permit analysis without prior preparation of single stranded template (Ruano and Kidd 1991). The recent availability of automated sequencing apparatus for analysis of fluorescently labelled sequencing products has greatly increased the efficiency of direct sequencing. Expensive specialised equipment is required so this technology is not at present available to all laboratories, but after an initial investment the returns are great since a large number of samples can be processed and directly analysed by a computer. It is however still time-consuming to detect mutations by comparing large numbers of sequencing reactions, and sequence ambiguities may be falsely mistaken for mutations. Consequently new methods have been designed which allow more direct comparison of sequencing reactions. An example of one of these is orphan peak analysis.

1.2.8.6 Orphan peak analysis

Orphan peak analysis is a fluorescent adaptation of dideoxy fingerprinting (Sarkar et al., 1992) where Sanger sequencing reactions of multiple samples using one dideoxynucleotide are run out alongside each other on a non-denaturing gel. A sequencing ladder results and mutations are detectable as shifts of individual bands

in the ladder. Similarly in Orphan Peak analysis, PCR amplified DNAs from four different individuals are subjected to a single base-specific sequencing reaction and the products are applied to a set of four lanes of a Pharmacia automated DNA sequencer. The computer software then applies a different colour to each lane, and a base substitution in an individual is clearly visible as an individual-specific peak with a colour specific for the individual. The intensity of the resultant signal is high enough that a mutation can be detected even when several individuals are applied to the same lane (Hattori et al., 1993). Obviously for both ddF and Orphan Peak analysis to search for unknown mutations the patterns of all four (A,C,G,T) bases must be analysed. The method may also be applicable to an ABI-type automated sequencer, where four primers labelled with a different fluorochrome would be employed.

1.2.8.7 Enzyme mismatch cleavage

A new method has been devised for the detection of mismatches in heteroduplex molecules formed between wild type and mutant DNA (Dean 1995). It takes advantage of the ability of bacteriophage resolvases, whose function *in vivo* is to cleave branched DNA, to recognise and cut DNA at mismatches. This has been tried successfully by two groups, and while there are still some problems with the method it has great potential (Youil et al., 1995; Mashal et al., 1995). Basically, PCR products from the two DNA resources are combined to form heteroduplex. Youil et al (1995) end labelled wild type DNA and combined it with 10 × excess unlabelled mutant DNA, Mashal et al (1995) on the other hand combined the two

DNAs in equimolar proportions then carried out end labelling after the formation of heteroduplex or after incubation with resolvase. Products are then incubated with resolvase. Youil et al (1995) used T4 endonuclease VII, whereas Mashal et al (1995) used T4 endonuclease VII and T7 endonuclease I, both in combination and separately. Fragments are finally resolved on a polyacrylamide gel. Fragments of up to 1kb can be screened, no toxic chemicals are involved, and it can be set up in any laboratory without the need for expensive equipment. Reactions are identical for all DNA fragments, and resolvases recognise all four classes of nucleotide mismatch. Overall the method is simple and rapid, but does require some fine tuning. In some sequence contexts certain mismatches are not cleaved, and there is a high degree of nonspecific background cleavage. Both problems may be solved by the use of highly purified commercially available resolvases. Alternatively, resolvases from other viruses or organisms may be discovered which show a preference for presently uncleaved mismatches and indeed give a lower background.

1.3 Experimental strategy

The aims of this project were to develop a rapid and simple method of cloning inter Alu PCR products which were readily mappable. The cloning resource was a somatic cell hybrid which contained the chromosomal subregion 11q12-13. PCR conditions were optimised to give the maximum number of inter ALu products from the hybrid, and the capabilities of three different cloning methods were then compared. Finally clones were mapped on a somatic cell hybrid panel which

subdivided 11q12-13, and used to isolate YACs for the region. In parallel, the human olfactory marker protein gene, which maps to 11q12-13 and was a candidate one of the subtypes of the sensorineural deafness syndrome Ushers IB, was cloned and sequenced. The sequence was then used to design primers and the best method of mutation analysis selected before carrying out mutation analysis on affected individuals.

2. Materials and Methods

2.1 Cell culture

2.1.1 Bacterial cells

2.1.1.1 Media and additives

All media was sterilised by autoclaving.

L-broth and L-agar

2.46g $MgSO_4$, 10g tryptone (Difco), 5g yeast extract (Difco) and 10g NaCl were added per litre of distilled water. To make plates 15g agar (Oxoid) was added per litre of L-broth.

Terrific broth

12g tryptone, 24g yeast extract, 4g glycerol were added per 900ml of distilled water. After autoclaving, 100ml autoclaved phosphate buffer (0.1M KH_2PO_4 and 0.72M K_2HPO_4) was added.

2 YT medium

16g bacto-tryptone, 10g bacto-yeast extract, and 5g NaCl were added per litre of distilled water. The pH was adjusted to 7.0 with 5N NaOH

Ampicillin

Ampicillin was added to agar and broth in order to select for bacteria carrying plasmids which confer resistance to this antibiotic. A stock solution of ampicillin (Sigma) was made up at a concentration of 50mg/ml in dH_2O. This solution was filter sterilised and stored at -20°C. It was added to melted broth and agar (cooled to 60°C) to give a final concentration of 50μg/ml.

5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-gal)

X-gal (Sigma) acts as a substrate for β-galactosidase. It was added to agar to enable blue/white colour screening for recombinants. A stock solution of X-gal was made up at a concentration of 20mg/ml in dimethylformamide (Sigma) and stored, protected from light at -20°C. It was added to agar to give a final concentration of 40μg/ml.

Isopropyl β-D-Thiogalactopyranoside (IPTG)

IPTG is a derepressor of the Lac operon. It was added in addition to X-gal to agar to enable blue/white colour screening for recombinants. A stock solution of IPTG (Sigma) was made up at a concentration of 100mM in dH_2O and stored at -20°C. It was added to give a final concentration of 0.5mM.

2.1.1.2 Bacterial cell culture

Escherichia coli (*E.coli*) is the most widely used host organism for molecular cloning of DNA and a diverse set of assays involving cloned genes. There are two major parameters involved in efficiently transforming a bacterial organism: the method used to induce competence; and the genetic constitution of the host strain of the organism being transformed.

2.1.1.3 Bacterial strains

The *recA* gene is the master regulator of recombination in *E.coli* (Hanahan et al., 1991) It is important that the cloned DNA is not rearranged within the bacterial cell, hence *recA* mutants are used for the propagation of vectors containing cloned inserts. Another necessary genetic component of the host strain is the *lacZΔM15* mutation. The multiple cloning site (MCS) in appropriate plasmid cloning vectors is incorporated in-frame into the amino terminal DNA fragment of *lacZ* carried by the vector. When such a vector is transformed into a host strain carrying the *lacZΔM15* mutation, active β-galactosidase is produced and colonies turn blue on indicator plates containing X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactosidase). If DNA fragments cloned into the MCS of the vector cause the α subunit to be disrupted such that it cannot combine with the β subunit to restore β-galactosidase activity, white colonies develop. Thus blue-white colony screening can be performed to identify vectors containing cloned inserts. Strains can carry the *lacZΔM15* by two different means : on an F' (JM101); or on the chromosome as part of the ϕ80dlacZΔM15 transducing phage (JM83,DH5α, DH5αMCR, DH10B). The genotype of the three strains used is given below:-

DH5α - *supE*44 Δ*lac*U169 (ϕ80 *lacZ*ΔM15) *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1 *rel*A1

XL1-Blue - *sup*E44 *hsdR*17 *rec*A1 *end*A1 *gyrA*46 *thi rel*A1 *lac*$^-$ F' [*pro*AB$^+$ *lac*I^q *lacZ*ΔM15 Tn*10*(*tet*^{r})]

DH10B - *araD139* Δ(*ara-leu*7697) *galU galK mcrA* Δ(*mcr-hsdRMS-mcrBC) rpsl deoR* ϕ80*d lacZV* Δ(*M15) endA1 nupG recA1*

2.1.1.4 Plasmid vectors

The cloning vectors used were pBS (formerly Bluescribe) and pBS-SK, both of which are produced by Stratagene™.

2.1.2 Yeast artificial chromosomes

YACs are currently the cloning system of choice for long range physical libraries. This is because of their ability to clone large inserts (up to about 2Mb), which means that fewer markers are required to assemble a contiguous array of YACs representing a multimegabase genomic region. Yield of insert DNA from YACs is typically low. YACs are carried primarily as single copy chromosomes in a yeast host, and yeast cell densities in overnight cultures are significantly lower than *E.coli* cell densities. About 10μg of yeast DNA may be obtained from a 5ml overnight culture of a YAC clone, but only 2% to 8% of that is YAC DNA. The rest is yeast chromosomal DNA, which so resembles the YAC in physical properties that neither alkaline lysis procedures nor cesium gradients can separate them. Preparative pulsed field gel electrophoresis (PFGE) can be used to separate the YAC from the endogenous yeast chromosomes on a size basis, if the YAC does not comigrate with one or other of the yeast chromosomes, but yields are low by comparison with *E.coli* plasmid DNA preparations. The most common way to use YACs in mapping is to use ordered overlapping YACs or single copy markers ordered using those YACs to identify overlapping clones from a smaller insert, chimera free library e.g. in cosmids, for a region of interest (Dracopoli et al., 1995).

2.1.2.1 Media and additives

AHC broth and agar

AHC (Acid Hydrolysate of Casein) is a rich selective medium which lacks uracil and trytophan. It was used for selective growth of YAC recombinants prior to production of plugs and isolation of DNA. Combine :-

1.7 g yeast nitrogen base (without amino acids and ammonium sulphate- DIFCO), 5 g ammonium sulphate, and 10 g acid casein hydrolysate (low salt) were added to 1 litre of distilled water and the pH adjusted to 5.8. For AHC agar 17-20 g of Bacto agar (Difco) were then added. Autoclave, and when cooled to 60°C add 50ml filter sterilised 40% glucose and 10ml 2mg/ml sterile adenine sulphate

2.1.3 Somatic cell hybrids

2.1.3.1 Cell lines

The following cell lines were analysed. The translocation chromosome MAR1 which carries the derived 11 translocation chromosome, in the absence of any normal chromosome 11 material. Four X irradiation hybrids, WJX3.4, WJX5.4, WJX7.4, and WJX11.2, which contain fragments of chromosome 11 in the background of a Chinese hamster ovary cell line WG3H, were also analysed. These fragmentation hybrids were produced by X irradiation of a chromosome 11-only hybrid J1 C14 (Jones et al., 1984), followed by rescue of the fragments produced, by fusion with WG3H cells, and finally, cell surface marker selection to isolate those cell lines which contained fragments of chromosome 11 (Fletcher et al., 1993). Two additional chromosome 11 fragmentation hybrids, E67.1 and E67.4, were studied. These were derived by transfer of metaphase chromosomes from a human EJ bladder carcinoma cell line to the murine C127 cell line. This was followed by selection for HRAS1 mediated transformed growth, in addition to selection for the HRAS locus and other markers on 11p (Porteous et al., 1986). E67.1 and E67.4 had been analysed previously for 11p markers (Porteous et al., 1987) and a limited number of 11q markers (Spurr et al., 1988). Two other translocation cell lines were also analysed. CF52, which has a well characterised t(11;16)(q13;p11) translocation chromosome as its sole human component in the background of a murine cell line, A9 (Koeffler et al., 1981). MAFLI, which is a permanent B lymphoblastoid cell line established from a patient with schizophrenia and bearing the t(1;11)(q42.2;q21) translocation (Fletcher et al., 1993). J1 Cl4 (Jones et al., 1984) was used as a positive control and WG3H and RAG (murine background) were negative controls. Cells were available as a cell pellet from which DNA was extracted.

2.2 Transformation and isolation of DNA

2.2.1 Reagents and protocols for the transformation of competent bacteria

Bacterial cultures should always be initiated from a fresh agar plate derived from a stock stored at -70°C, or cells maintained in slow anaerobic growth in an agar stab stored in the dark at room temperature. These methods will always produce reliably competent cells whereas cells maintained in suspension in glycerol or on plates at 4°C produce erratic competence. There are two major methods used to induce competence in cells (Hanahan 1983; Hanahan et al., 1991). Several different methods using chemicals exist to do this, with the degree of simplicity of the method directly reflected in the degree of competence achieved. These methods all ultimately involve a heat shock step and hence are also referred to as heat shock transformation. The other major method is electroshock transformation. Transformation efficiency can be roughly

measured as the number of colonies formed per unit mass of DNA. Not every cell is competent so a limited number of transformed cells will be obtained irrespective of how much DNA is added. Electrocompetent cells have 10 fold higher transformation efficiency and the fraction of competent cell is also 10 fold higher, so they can be expected to produce 10-100 times more colonies than chemical methods. The method chosen is determined by the nature of the DNA to be transformed i.e. whether it just retransformation of a plasmid clone, transformation of an efficient or inefficient ligation, or transformation of cloned high complexity cDNA and genomic cloning. The other major parameter, which also requires consideration of the nature of the DNA to be transformed, is the genetic constitution of the host cells as discussed above (section 2.1.1.3).

2.2.1.1.1 Preparation of bacteria for heat transformation

Standard laboratory method

- Inoculate 100ml of pre-warmed L-broth (containing 10mM $MgSO_4$, 10 mM $MgCl_2$, 10mM NaCl, 25mM KCl) with 2ml of an overnight culture
- Incubate cells in a 37°C shaking incubator until the $A_{500nM} = 0.48$
- Incubate 15min on ice
- Centrifuge 5000rpm 15min, in sterile Falcon tubes at 4°C
- Take up pellet in 33ml of RF1 (100mM RbCl, 50mM $MnCl_2.4H_20$, 30mM KOAc, 10mM $CaCl_2.2H_2O$ 15% w/v Glycerol pH to 5.8 with 0.2M acetic acid. Filter sterilise.)
- Incubate 15min on ice
- Centrifuge 5000rpm 5min, in sterile Falcon tubes at 4°C
- Take up pellet in 25ml of RF2 (10mM MOPS, 10mM $MgCl_2$, 75mM $CaCl_2.2H_2O$, 15% w/v Glycerol pH6.8 with NaOH. Filter sterilise.)
- Incubate 20 min on ice
- Freeze in 200µl aliquots in dry ice/methanol bath
- Store at -70°C for up to 6 months

Hanahan method (Hanahan 1991)

- Inoculate 50 ml of prewarmed L-broth with 3ml of a fresh overnight culture
- Incubate in a 37°C shaking incubator until the $A_{500nM} = 0.48$-0.55
- Incubate 30ml of cells 15min on ice
- Centrifuge 2500×gmax 10 min, in sterile Falcon tubes at 4°C

- Resuspend pellet in 10 ml of TFB (10mM K-MES pH6.2, 100mM RbCl, 45mM $MnCl_2.4H2O$, 10mM $CaCl_2.2H_2O$, 3mM Hexamine Cobalt (III) Chloride) by gentle vortexing
- Incubate 15min on ice
- Centrifuge 2500×gmax 10min at 4°C
- Resuspend pellet in 2.4ml of TFB
- Add 97.5µl of fresh DMSO
- Incubate on ice 5 min mixing occasionally
- Add 97.5µl of 2.25M DTT 4mM potassium acetate
- Incubate on ice 10 min
- Add 97.5µl of fresh DMSO
- Incubate on ice for 5 min, then transform immediately

Calcium/Manganese-based method (CCMB - Hanahan 1991)

- Inoculate 50 ml of prewarmed L-broth with 3ml of a fresh overnight culture
- Incubate cells in a 37°C shaking incubator until the absorbance at 550nM reads 0.3
- Incubate 15-30min on ice
- Centrifuge 4000×gmax 15min, in sterile Falcon tubes at 4°C
- Resuspend cells in 1/3 volume of CCMB80 (80mM $CaCl_2$, 20mM Mn Cl_2, 10mM Mg Cl_2, 10mM KOAc, 10% w/v glycerol)
- Incubate on ice for 20 min
- Centrifuge 4000×gmax 15min, in sterile Falcon tubes at 4°C
- Resuspend cells in 1/12 volume of CCMB80
- Freeze in 1ml aliquots in dry ice/methanol bath
- Store at -70°C for up to 6 months

2.2.1.2 Preparation of bacteria for electro-transformation

- Inoculate 1 litre of L-broth with 10ml of a fresh overnight culture
- Incubate cells in a 37°C shaking incubator until the absorbance at 600nM reads 0.5-1
- Incubate 15-30min on ice
- Centrifuge 4000×gmax 15min, in sterile Falcon tubes at 4°C
- Resuspend pellets in 1 litre cold, low ionic strength buffer, usually dH_2O
- Centrifuge 4000rpm 15min, in sterile Falcon tubes at 4°C
- Resuspend pellets in 500ml cold, low ionic strength buffer, usually dH_2O
- Centrifuge 4000rpm 15min, in sterile Falcon tubes at 4°C

- Resuspend pellet in ≈20ml cold 10% glycerol
- Centrifuge 4000rpm 15min, in a sterile Falcon tube at 4°C
- Resuspend to a final volume of 2-3ml in 10% glycerol
- Freeze in 40µl aliquots in dry ice/methanol bath
- Store at -70°C for up to 6 months

2.2.1.3 Heat shock transformation of competent bacteria

- Thaw competent cells on ice
- Mix up to 1.5µl of DNA with 100µl of competent cells
- Heat shock by incubation at 42°C for 1 min
- Incubate on ice 1 min
- Add 1ml of 2YT medium
- Incubate in a shaking 37°C incubator for 1 hour
- Plate out 200µl on a selective plate

2.2.1.4 Electro-transformation of competent bacteria

- Thaw elctrocompetent cells at room temperature and place on ice
- In a cold 1.5ml eppendorf, mix 40 µl of cell suspension with 1-2µl of DNA
- Mix well, incubate on ice 1 min
- Set the Gene Pulser (Biorad) apparatus to 25µF and 2.5 Kv. Set the pulse controller to 200Ω.
- Transfer the DNA/cells to a cold 0.2cm electroporation cuvette, shake suspension to. the bottom
- Place cuvette in chilled safety chamber slide. Pulse. The time constant should be 4.5-5msecs
- Add 1ml of L-broth to the cuvette immediately
- Quickly transfer the cells to a 1.5ml eppendorf and incubate in a shaking 37°C incubator for 1 hour
- Plate out on selective media

2.2.2 Frozen stocks of bacterial strains

Frozen stocks were made by freezing down to -70°C an aliquot of an overnight culture in 20% glycerol.

2.3 Isolation of plasmid DNA from bacterial cells

All of these methods can be found in Sambrook et al., (1989).

2.3.1 Small scale plasmid DNA purification

Set up an overnight culture by picking a single colony from a fresh plate into 5ml terrific broth containing a selective antibiotic for the plasmid which is to be purified, generally ampicillin. Incubate at 37°C overnight in a shaking water bath Either spin down the full 5ml (5 min) or a 1.5 ml(1 min) aliquot. The larger volume gives a higher yield of plasmid without the need to adjust the volume of other solutions used in the method.

- Add 250µl lysozyme (10mg/ml) in GTE (50mM Glucose,25mM Tris,50mM EDTA) to pellet, and incubate 10 min room temperature. This breaks down the cell wall and outer membrane.
- Add 500 µl alkaline/SDS (1% SDS,0.2M NaOH), and incubate 10 min on ice. SDS lyses the spheroplasts.
- Add 250µl high salt (3M Potassium Acetate, pH to 4.8 with acetic acid),and incubate on ice 10 min. Here the disrupted of the plasmid base pairing is returned to the native superhelical structure, but the chromosomal DNA remains irreversibly denatured.
- Centrifuge in the cold room 15 min
- Transfer 900µl supernatant to a fresh tube
- Repeat centrifuge and transfer to fresh tube
- Add 540µl of isopropanol
- Centrifuge at room temperature 15 min
- Wash pellet with 70% ETOH
- Dry 3 min under vacuum
- Resuspend in 30µl TE, and run 3µl on a gel to estimate the recovery (typically 3-5µg).

2.3.2 Large scale plasmid DNA preparation

Set up an overnight culture shaking, at 37°C, by picking a single colony from a fresh plate into 5ml of terrific broth containing a selective antibiotic for the plasmid which is to be purified, generally ampicillin. Use 500µl of the overnight culture to inoculate 500ml of L-broth containing ampicillin (100µg/µl). This must also be incubated overnight as before.

The DNA is then extracted as described below.

- Centrifuge 5min at 6000rpm
- Resuspend pellet in 20ml GTE + 200mg lysosyme (10mg/ml)
- Add 40ml alkaline/SDS, and incubate 10 min on ice
- Add 30ml high salt, and incubate on ice 5-30 min
- Centrifuge 30min at 12000rpm

- Filter through muslin to remove bacterial debris
- Add 55ml isopropanol, incubate 5 min room temperature
- Centrifuge 15min at 8000rpm
- Wash pellet with 70% ETOH
- Centrifuge 5min 800rpm to repellet cells
- Resuspend pellet in 5ml TE
- Add 5g caesium chloride
- Add 5ml EtBr (10mg/ml)
- Transfer to Quick-seal centrifuge tubes (Beckman) tubes, balance tubes using 1g/ml CsCl, then seal the tops
- Ultracentrifuge at 80,000rpm at 20°C for 4-16 hours
- Plasmid DNA forms a band, just below the bacterial chromosomal DNA near the top of the tube. RNA and protein are deposited in the bottom of the tube. The DNA band is removed by inserting a needle in the top of the tube to relieve any vacuum formed, then inserting another needle attached to a syringe just below the plasmid DNA band which can then be used to draw off the band
- Remove the EtBr by butanol extraction: add the same volume of butanol, invert several times by hand, allow the two phases to separate then remove and discard the upper phase. This process is repeated until the lower layer becomes clear
- The DNA can then be precipitated: add two volumes of dH_2O to dilute the salt, then six volumes of cold absolute ETOH to precipitate the DNA. Incubate -20°C 20min. Centrifuge at 8000rpm (Sorvall SM24) for 10 min
- Wash pellet with 70% ETOH, centrifuge as above for 5 min
- Resuspend the DNA pellet in 500µl-1ml dH_2O (yield typically 0.5-2mg)

2.3.3 Sequencing standard DNA preparation

This method yields DNA suitable for use with the ABI cycle sequencing kit i.e. high quality supercoiled plasmid DNA which is relatively free of contaminating chromosomal DNA and RNA. The initial steps are very like conventional alkaline lysis mini preps, but there are additional RNase A and PEG (polyethylene glycol) precipitation steps.

- Incubate 5ml cultures overnight 37°C in terrific broth in a 50ml tube to allow adequate aeration
- Centrifuge 5 min to pellet the cells
- Remove the supernatant by aspiration, resuspend pellet in 200µl GTE
- Add 500µl freshly prepared 0.2M NaOH/1% SDS
- Vortex and incubate on ice 5 min

- Neutralise by adding 250µl 3M Potassium Acetate pH4.8
- Vortex and incubate on ice 10 min
- Centrifuge 10 min room temperature
- Transfer supernatant to a fresh tube leaving behind cellular debris
- Repeat the last two steps
- Add RNase A to a final concentration of 20µg/µl
- Incubate 37°C 30 min
- Extract supernatant twice with 400µl chloroform
- Mix layers by hand for 30 seconds after each extraction
- Centrifuge 1 min to separate phases
- Remove aqueous layer to a fresh tube
- Precipitate DNA by adding an equal volume of 100% isopropanol and immediately centrifuging 10 min room temperature
- Wash pellet with 500µl 70% ETOH
- Dry 3 min under vacuum
- Resuspend pellet in 33.6µl dH_2O
- Precipitate by adding 6.4µl NaCl, 40µl 13% PEG8000
- Mix thoroughly, incubate 20 min on ice
- Pellet plasmid DNA by centrifuging 15 min cold room
- Wash pellet with 500µl 70% ETOH
- Dry 3 min under vacuum
- Resuspend in 20µl dH_20
- Store at -20°C

2.3.4 Yeast DNA preparation

- A single colony, freshly grown on an AHC plate, was grown overnight in 5ml of AHC medium, at 30°C, shaking at 200rpm
- centrifuge 3000rpm for 4 min
- resuspend cells in 0.5ml 1M Sorbitol; 0.1M EDTA pH7.5
- add 20µl of 2.5mg/ml zymolyase 6000 and incubate at 37°C for 1.5-2 hours
- centrifuge 1 minute
- resuspend pellet in 0.5ml 50mM Tris pH7.4; 20mM EDTA
- add 50µl 10% SDS and mix carefully but thoroughly
- incubate 65°C for 30 min

- add 200µl 5M potassium acetate
- incubate on ice 1 hour, then centrifuge for 5 min
- transfer supernatant and combine with 1 volume of isopropanol
- mix, incubate room temperature 5 min
- spin 10 secs, take off supernatant with a pastette and air dry pellet
- resuspend in 300µl 10mM Tris; 1mM EDTA pH7.4 (TE)
- add 1.5µl 10mg/ml pancreatic RNase
- incubate 37°C for 30 min
- add 15µl NaCl; 900µl 100% ETOH
- centrifuge 15 min, wash pellet with 70% ETOH
- resuspend dried pellet in 100µl TE
- before digesting DNA centrifuge 15 min to remove any insoluble material which may inhibit digestion

2.3.5 Preparation of agarose plugs for pulsed field gel electrophoresis

This method allows the extraction of intact yeast chromosomes, including yeast artificial chromosomes (YACs), from cells. It was adapted from Maule (1994). The volume can be adjusted to suit the number of plugs required.

- A single colony, freshly grown on an AHC plate, was grown overnight in 10ml of AHC medium, at 30°C, shaking at 200rpm
- Increase the volume of AHC medium to 100ml, and reincubate until cell count reaches ~ 1×10^8 cells/ml (~16 hours)
- Harvest the cells by spinning at 2000g for 10min at 4°C
- Discard the supernatant, and gently disrupt the pellet with a sterile loop before adding 50ml of chilled 50mM EDTA pH7.5
- Centrifuge 2000g, 5 min, at 4°C
- Repeat the last two steps
- Discard supernatant, and take pellet up in 3ml of ice-cold 50mM EDTA pH7.5
- Transfer cells to a 20ml universal container, with a fine tipped sterile Pastette, and heat to 37°C
- Add 6ml of 1% low melting temperature agarose (in 0.125M EDTA pH7.5) which has been cooled to 50°C
- Add 1.2ml of freshly prepared cell wall digestion solution (2ml SCE [1M sorbitol, 0.1M trisodium citrate, 60mM EDTA pH7- autoclave], 0.1ml 2-mercaptoethanol (Sigma), 2mg zymolyase (ICN Biochemicals)

- Immediately mix thoroughly and dispense in 100µl aliquots into plug moulds and allow to set on ice
- Eject the plugs into a 50ml Falcon tube containing 25ml of ETM solution ([0.45M EDTA, 10mM Tris-HCl pH8 - autoclave], 7.5% 2-mercaptoethanol), and incubate at 37°C overnight
- Replace ETM with 20ml 1% NDS solution (0.45M EDTA, 10mM Tris.HCl pH9, 1% SDS) containing proteinase K (Boehringer Mannheim)at 1mg/ml, and incubate at 50°C overnight
- Repeat the last step
- Store plugs in 20ml ETM at 4°C
- Plugs should be equilibrated for at least one hour in gel running buffer before use

2.3.6 Isolation of genomic DNA from cultured cells

- Human genomic DNA was extracted from of a pellet of cultured cells as described below
- Resuspend pellet in 0.5ml of TE buffer (10mM Tris pH7.5; 1mM EDTA), and transfer to a phenol resistant tube
- Vortex while adding 2.5ml of lysis buffer (0.5% SDS; 150mM NaCl; 100mM Tris pH8; 100mM EDTA)
- Add 10µl/ml RNase
- Incubate 37°C for 15 - 60 min
- Add proteinase K to final concentration of 0.5mg/ml
- Incubate 50°C overnight
- Add and equal volume of water saturated phenol
- Centrifuge at 2500 rpm for 5 min
- Remove upper aqueous layer and add an equal volume of 1:1 phenol/chloroform
- Centrifuge at 2500 rpm for 5 min
- Remove upper aqueous layer and add an equal volume of 24:1 chloroform/isoamyl alcohol
- Remove upper aqueous layer and add half volume of 7.5M Ammonium Acetate
- Add two volumes of absolute ethanol and spool out DNA on a glass pasteur with sealed end
- Air dry then dip in 66% ETOH 0.8M ammonium acetate
- Resuspend DNA in 200µl TE

2.4 Amplification of DNA by the polymerase chain reaction (PCR)

The PCR (Saiki et al., 1985; Mullis and Faloona 1987) is an enzymatic technique which permits the cyclic and exponential amplification of a specific ds DNA sequence. The sequence specificity is provided by a pair of oligonucleotide primers which direct amplification. The primers are

complementary in sequence to the 5' ends of the sequence to be amplified and are annealed to the template under conditions where they hybridise only to their exact complement. The PCR is a series of cycles, each of which consist of three steps. The first is a high temperature step (93-94°C) which causes denaturation of the template. This is followed by a reduction in temperature (to that which is optimum for precise annealing of the primers, usually 45-65°C) and finally an extension step at 72°C (which is the optimum temperature of the thermostable polymerase).

2.4.1 Constituents and conditions

Generally PCRs were carried out in a 50µl volume which contained 5µl of 10× buffer, 3µl 25mM $MgCl_2$, 1µl of each primer at 300ng/µl, 1µl of a 50× dNTP stock where each dNTP (Pharmacia) is at 10mM, 1 unit of thermostable DNA polymerase, and DNA template the concentration of which depended on the complexity of the resource being amplified. Initially reactions were carried out using Promega Taq polymerase enzyme, buffer (10× buffer= 100mM Tris-HCl pH8.8, 500mM KCl, 15mM $MgCl_2$, 1% Triton X-100) and magnesium chloride. Licensing restrictions, however, meant that a switch to Cetus Amplitaq, and the corresponding buffer (10× buffer= 100mM Tris-HCl pH8.3, 500mM KCl, 15mM $MgCl_2$, 0.01% (w/v) gelatin), had to be made. It was found that amplification with Cetus enzyme could be improved to the same level as Promega by the addition of 5µl of 1% Triton X-100 to the reaction mix. In each case reactions were overlaid with mineral oil (Sigma).

PCRs were carried out on various different machines; a Perkin Elmer Cetus DNA Thermal Cycler, Hybaid Thermal Reactors and a Hybaid Omnigene. The temperature and the length of the different steps was varied according to the primer pair, the machine and the length of the fragment to be amplified. Conditions for the various sets of primers used were determined as follows. Each set of conditions always had an initial cycle with a longer denaturation step at a higher temperature (generally, 94°C for 3 minutes- Perkin Elmer Cetus DNA Thermal Cycler, 93°C for 2 minutes- Hybaid machines). This was then followed by a series of cycles, the number of which was dependent on the efficiency of amplification (generally 30 cycles were performed initially). This series of cycles consisted of a denaturation step (generally, 92°C for 45 seconds- Perkin Elmer Cetus DNA Thermal Cycler, 91°C for 30 seconds- Hybaid machines) followed by an annealing step. The annealing temperature (T_a) was 5°C lower than the melting temperature (T_m) of the primers. This was determined by the equation T_m (at 1M concentration) = 4(G+C) + 2(A+T). Annealing was generally carried out for 1 minute- (Perkin Elmer Cetus DNA Thermal Cycler) or for 30 seconds- (Hybaid machines). The annealing step was followed by extension at 72°C for approximately 1 minute per kb to be amplified. The final cycle had an increased extension step of 5 or 10 minutes depending on expected size of the product. In a number of

cases these conditions required modification, in order to promote maximum amplification. Generally changes were made to the annealing temperature and the number of cycles of amplification. 'Touchdown' PCR (Don et al., 1991) has also been reported to enhance the specificity of PCR reactions. Here the annealing temperature of the first cycle is approximately 10°C higher than the theoretical annealing temperature. Over subsequent cycles the temperature is lowered by 1 or 2°C every one or two cycles until the theoretical temperature is reached. The cycles which are carried out under the stringent conditions provided by higher temperatures should promote amplification of the correct template molecules, in preference to aberrant amplifications from mismatched template. Thus by the time the temperature has been lowered to that which prevents mispriming, the correctly amplified product has an advantage over any misprimed products.

2.4.2 Template

2.4.2.1 PCR from plasmid isolates

Only a very small amount of template is needed for a plasmid in pure form since its small size means that there are many copies present in a small mass of DNA. Thus generally only 1-100pg were used as template. For amplification of the OMP cosmid, however , I found it necessary to amplify 100pg-1ng, due to the increase in complexity.

2.4.2.2 PCR from bacterial colonies

Template can be prepared by either of two simple methods I have found the two stage method gives more product.

- Touch the colony lightly with a toothpick then place directly in a tube containing the rest of PCR components
- Swirl, then remove toothpick
- Simply add oil then amplify as normal

Or

- Touch the colony lightly with a toothpick, then place in an eppendorf containing 100µl of dH2O
- Swirl, then remove toothpick
- Incubate at 94°C 5 min (to denature the DNA)
- Incubate on ice 2 min
- Centrifuge 2 min
- Amplify 5µl of the supernatant

2.4.2.3 PCR from human genomic DNA

Generally 100ng of genomic DNA was amplified in a standard 50µl PCR. The only difference between setting up genomic and standard plasmid PCRs is that you must ensure that the DNA is properly in solution before removing an aliquot to PCR, and that the aliquot is evenly dispersed in the PCR mix. This is done by thoroughly vortexing the PCR mix before the addition of oil.

2.4.3 Oligonucleotides

Primers were designed using either the Oligo4 programme (Hybaid) or the Primer programme-(Whitehead Institute for Biomedical Research, available through HGMP)

Oligonucleotides were synthesised, in the form of ammonium stocks, on an Applied Biosystem 381A oligonucleotide synthesiser. HPLC purification was not performed.

2.4.3.1 Purification

Oligonucleotide stocks consist of about 1ml 30% NH_4OH and a sediment of beads on which the oligonucleotides are synthesised. The DNA must therefore be precipitated from the NH_4OH supernatant. Two methods were used from this although only the second method is now in use.

Conventional method:-

To 350µl NH_4OH stock

- Add 3M NaOAc pH5.5
- Add 770µl cold absolute EtOH
- Vortex briefly
- Incubate -20°C at least 30 min
- Centrifuge at 1400rpm for 30 min at 4°C
- Wash pellet 2×70% ETOH
- Air dry
- Resuspend in 200µl TE
- Store -20°C

Rapid/butanol method (Sawadogo & van Dyke 1991)

- 100µl NH_4OH stock
- Add 1ml n-butanol (ACS reagent grade)

- Vortex vigorously 15 secs
- Centrifuge 1 min 12,000rpm
- Remove single H_2O-containing n-butanol phase and discard
- In some cases to complete removal of contaminants- resuspend pellet in 100µl dH_2O and repeat the butanol extraction
- Dry pellet under vacuum
- Resuspend pellet in 50µl d H_2O, or TE
- store at -20°C

2.4.3.2 Annealing Oligonucleotides

This was used to create insert for the ligation independent cloning vectors. Combine an equal volume of the two complementary oligonucleotides in 10mM Tris HCl, 5mM MgCl2.

Heat to 70°C for 10 min, then cool slowly to room temperature. The newly annealed oligonucleotides can then be stored at -20°C until ready to use

2.4.4 Details of primers and amplification conditions

Table 2.1 OMP PCR primers

PRIMER NAME	DESCRIPTION	SEQUENCE	PCR CONDITIONS
5un Cod5	OMP PCR primer pair for amplification of 5' coding and untranslated region	5' AGG CTT TGG TGA TTG GGT GTG 3' 5' CGA GGT GCC TGT GAT GGT GAC 3'	T_a=57°C 10% DMSO
Codd3 3uun	OMP PCR primer pair for amplification of 3' coding and untranslated region	5' CTA CGC GTG GAG AGC CTG AAG 3' 5' AAC CCT GCC CCA TAT CCT GAG 3'	T_a=53°C
5uun2	OMP external 5' primer used with 3uun to amplify gene and 3' & 5' untranslated regions	5' TTG GTG ATT GGG TGT GGA 3'	T_a=45°C 13% DMSO
C C'	OMP rat derived primer pair used to amplify the 5' section of the coding region.	5' GCA GAG GAC GGG CCA CAG AAG C 3' 5' CAG TTC TGC GAG GTG CCC GTG 3'	T_a=56°C
S	OMP mouse equivalent of primer C	5' GCA GAG GAT GGG CCG CAG AAG C 3'	T_a=56°C
A	OMP rat derived primer, nested within primer R.	5' GCT GGT TAA ACA CCA CAG AGG C 3'	T_a=56°C
R	OMP rat derived primer which overlaps the stop codon.	5' GCC ATC AGA GCT GGT TAA ACA C 3'	T_a=56°C

Table 2.2 OMP sequencing primers

PRIMER NAME	DESCRIPTION	SEQUENCE	PURPOSE
IntAf	OMP internal forward primer A	5' ATC TCC AAT CAC CAA GCC CTC 3'	ABI sequencing
IntAr	OMP internal reverse primer A	5' TAC TTT CTG CCC TTC CCT CTT C 3'	ABI sequencing
IntBf	OMP internal forward primer B	5' AAA GGG AGG GGA AGA GGG A 3'	ABI sequencing
IntBr	OMP internal reverse primer B	5' GCC AAA GGT GAC GAG GAA GTA C 3'	ABI sequencing
IntCf	OMP internal forward primer C	5' GAC CCC ACT GCC ATC TTC TG 3'	ABI sequencing
IntCr	OMP internal reverse primer C	5' CTC TCT TGG TCT TTC TCA GTC TCT 3'	ABI sequencing
X	OMP internal reverse primer	5' GGG TGA AGT TGA GGC GGT A 3'	Used as probe to identify 5' end of gene
Y	OMP internal sequencing of Pst clone	5' CCC CAT CTC TGT CTC C 3'	Conventional sequencing
Yint	OMP sequencing internal to Y primer	5' GCA GCC GGA GGA NAG GTG TG 3'	Conventional sequencing

Table 2.3 Universal PCR primers

The M13 sequencing primers are single-stranded DNA primers suitable for sequencing and amplification of inserts cloned into all M13 *lac* cloning phages and plasmids (e.g. pUC) beginning at opposite sides of the polylinker (Sanger et al., 1980). Recently the design of the M13 reverse primer has been altered to increase its performance in cycle sequencing (Kretz et al., 1994).

PRIMER	PURPOSE	SEQUENCE	CONDITIONS
M13 forward	sequencing + PCR	5' GTA AAA CGA CGG CCA GT 3'	Ta =55°C
M13 reverse	sequencing	5'GGA AAC AGC TAT GACCAT G3'	Ta =55°C
M13 reverse	PCR	5' CAG GAA ACA GCT ATG AC 3'	Ta =52°C

Table 2.4 Alu PCR primers

PRIMER NAME	SEQUENCE	CONDITIONS
614	5' G TGA GCC GAG ATC GCG CCA CTG CAC T 3'	Ta =65°C
C324	5' CCC CCG CGG CCC GAG CCG AGA TCG CGC CAC 3'	Ta =55°C
C325	5' GGG GGC GCC CCC GAG CCG AGA TCG CGC CAC 3'	Ta =55°C
Alu_{18}	5' CC/TA/G TCT CAA AAA AAA AAA 3'	Ta =46°C

Table 2.5 CA repeat primers

The sequence derived from the clone containing each repeat was used to design primers flanking the CA dinucleotides. Where possible primers were designed to give a product of 150 -250 bp. In each case the shortest primer was fluorescently labelled.

PRIMER NAME	PURPOSE	SEQUENCE	CONDITIONS
CA6f	Primer designed for amplification of 13mer CA repeat, used in combination with Alu_{18}	5' TCA GGA AGT GGG GGG GTT 3'	Ta = 56°
A4f A4r	Primers designed for the amplification of the 13mer CA repeat A4.	5' CTA TGG GGA TGG TAA GTA GCC3' 5' GCC ACC TCT CTT TCT GCT CT3'	Touchdown to Ta = 48°C
F5f F5r	Primers designed for the amplification of the 14mer CA repeat F5.	5' GCA GCC TCC ACC TTC CAG 3' 5' GAG AGG CCG AGG CAG ACA 3'	Touchdown to Ta = 52°C
CA12r	Primer designed for amplification of 14mer CA repeat, used in combination with Alu_{18}	5' TGA ACT GGT ATG ATC TCA 3'	Ta = 55°C
CA8r	Primer designed for amplification of 12mer CA repeat, used in combination with Alu_{18}	5' ATA TGA AAT CGT GAG TGG3'	Ta = 55°C

Table 2.6.1 and 2.6.2 PCR primers for chromosome 11 markers

These are the primers and PCR conditions for markers (Evans et al., 1995) and ESTs (Slorach et al., 1995) previously mapped to 11q12-13 by other members of the group.

LOCUS	SEQUENCE	CONDITIONS	REFERENCES
PGA	5'GCA TCT CTG ACA CCA ATC AG 3' 5'TGG AGA AGA GAC AGA TGG AG3'	Ta = 54°C	W Cookson pers. comm. Evers et al., 1989
ROMI	5'AGG CCT GGA GCT TGG GGT GA3' 5' CCA TCC AAA AAC TTT ATT GAG TCT 3'	Ta = 55°C	Primers designed from Bascom et al., 1992 by Evans et al (1995)
MDUI	5'TCT TCA AAG CCT CTG TCT AAC C 3' 5' CTC ATC TCC AAC CTG TCT AAC C 3'	Ta = 57°C	Primers designed from Gottesdiener et al., 1988 by Evans et al (1995)
GSTPI	5' GGG AGG GAT GAG AGT AGG ATG 3' 5' GGA GGT TCA CGT ACT CAG GGG 3'	Touchdown to Ta = 60°C	Primers designed from Cowell et al., 1988 by Evans et al (1995)
FGF4	5' GAT GAG TGC ACG TTC AAG GAG 3' 5'CAG AGA TGC TCC ACG CCA TAC 3'	Ta = 62°C	Primers designed from Yoshida et al., 1987 by Evans et al (1995)
D11S527	5' ATG CGC CTC CAA GAC AAG TTC 3' 5'GCC CCT CTA CTT GTC TGG AG 3'	Ta = 55°C	Brown et al 1991

LOCUS	SEQUENCE	CONDITIONS	REFERENCES
EST01878	5' TTT TCC AAT AAT GTG ACT TC 3' 5'CAA TCC CAA CCG TAA CAG GC 3'	Touchdown to Ta = 50°C	Slorach et al (1995)
EST02002	5' TGA AGG TGA CTT AAA CAG CT 3' 5' CGC AAT TCT CCC TGC GTC AT 3'	Touchdown to Ta = 50°C	Slorach et al (1995)
EST00220	5' TTT CCG AAG AAG GCA GTT TG 3' 5' ATC AGC CTA GAG GCC TGA CT 3'	Touchdown to Ta = 55°C	Slorach et al (1995)
EST00294	5' GTT TGA AGG AAG TGA TTT CC 3' 5' TAG GGC CAC CTC CAG TTC AT 3'	Touchdown to Ta = 55°C	Slorach et al (1995)
EST00218	5' TAC CCC GTC AGA TCA GTT ATG 3' 5' CCA ACA ACG TTT GCCT CAC AT 3'	Ta = 55°C	Slorach et al (1995)

2.5 Enzymatic manipulation of DNA

2.5.1 Restriction endonuclease digestion of DNA

Digestions with restriction endonucleases (Boehringer Mannheim) were carried out in the appropriate buffer (supplied by the manufacturer) at the recommended temperature. 1µg DNA was digested in 10µl, containing 5mM spermidine (Sigma), with 1-2 units of restriction enzyme for between 1 hour (Plasmids and low complexity DNA) and overnight (YACs, somatic cell hybrids, and genomic DNA). When two different enzymes, both of which required the same buffer, were used the reactions were carried out simultaneously. Otherwise, digestion with the enzyme requiring the lower salt buffer was carried out first. The salt concentration was then increased by addition of the appropriate amount of sodium chloride solution and the second reaction carried out. Alternatively KGB was used. Reactions were terminated by heating to 68°C or 80°C for 10 minutes, according to the heat sensitivity of the enzyme, or if the enzyme was heat stable, by phenol/chloroform extraction of the reaction. Restriction digests which were run on agarose gels were terminated by the addition of $1/10^{th}$ of their volume of 'stop mix' (100mM EDTA pH8, 20% Ficoll and orange G)

2.5.2 Dephosphorylation of linear plasmid DNA

This is used to remove the 5' phosphate and prevent the self-ligation of DNA fragments. Usually a stock of dephophorylated vector was prepared.

- To 1µg of linearised plasmid DNA
- Add 2µl of 10× calf intestinal phosphatase buffer (BM)
- Add dH_20 to 19µl
- Add 1µl (1.5 units) calf intestinal phosphatase (BM)
- Incubate for 30min at 37°C (cohesive ended molecules) or 50°C (blunt ended molecules)
- Add an additional 1µl calf intestinal phosphatase (BM)
- Incubate for 30min at 37°C (cohesive ended molecules) or 50°C (blunt ended molecules)
- Stop the reaction by adding 0.4µl 0.5M EDTA (final concentration 10mM)
- Phenol/chloroform extract the DNA

2.5.3 Ligation of blunt ended molecules

According to Cobianchi and Wilson (Cobianchi and Wilson 1987) the optimal ratio of ends for such a ligation event is 3:1 vector to insert. Typically, between 10 and 100ng of vector was used.

To vector and insert DNA add

- 1µl 10× T4 DNA ligase buffer (50mm Tris-HCl pH7.5, 7mM $MgCl_2$, 1mM DTT, 1mM rATP:BM)
- Potassium chloride to 30mM final concentration
- ATP to 1mM final concentration
- dH_2O to 10µl
- 0.1 unit of T4 DNA ligase (BM)
- Incubate overnight at 4°C
- Incubate 68°C 20 min to inactivate the enzyme

2.5.4 Ligation of cohesive ended molecules

According to Cobianchi and Wilson (Cobianchi and Wilson 1987) the optimal ratio of ends for such a ligation event is 1:2 vector to insert. Typically, between 10 and 100ng of vector was used.

- To vector and insert DNA add
- 1µl 10× T4 DNA ligase buffer (BM)
- dH_2O to 10µl
- 0.1 unit of T4 DNA ligase (BM)
- Incubate overnight at 16°C
- Incubate 68°C for 20 min to inactivate the enzyme

2.5.5 Ligation Independent cloning

Alu-PCR products need to be purified to remove excess nucleotides (which would prevent T4 from producing single stranded tails) and primers (which could interfere with the cloning process). This was either done by gel purification and DNA isolation using a Spin-X column, or directly using a Wizard PCR column. The optimal conditions are given below:-

Production of single stranded tails

Vector or insert DNA was combined with:-

- 1.5× KGB buffer
- 0.5mM dATP
- 0.5mM dTTP
- 1 unit of T4 DNA polymerase (BM)
- Incubate 15 min 37°C
- Incubate 68°C

Non-covalent linkage of ends

Combine vector and insert DNA in a 3:1 insert to vector ratio and

- Incubate 10 min at 60°C
- Immediately transfer to 4°C

2.5.6 Turbo cloning

This is a rapid, efficient method for the cloning of blunt-ended DNA into plasmids. It relies on coupling efficient intermolecular ligation under conditions of macromolecular crowding (15% PEG 6000), with subsequent Cre-mediated recircularization at vector-borne *lox* sites. Again inserts were prepared by a two stage PCR protocol, initial amplification with 614 primer was followed by reamplification of an aliquot with the Alu_{18} end primer. A stock of *SmaI* cut dephosphorylated pBSlox vector was kindly provided by Dr Chris Boyd.

Primer phosphorylation

A 10µg aliquot of Alu_{18} was kinased by combining it with-

- 1× Polynucleotide kinase buffer (BM)
- 10 units of Polynucleotide kinase (BM)
- ATP at 0.5mM final concentration
- Incubate 40 min 37°C
- Incubate 10 min 68°C to heat inactivate the kinase
- Precipitate in 0.3M Sodium Acetate and 2 volumes of cold ethanol
- Resuspend in 10µl dH_2O

This was then used to amplify an aliquot of 614 PCR products. T4 DNA polymerase was then used to ensure that the ends of the PCR products were blunt.

Combine an aliquot of the Alu18 PCR reaction with

- dNTPs at 200µm final concentration
- 1 unit of T4 DNA polymerase (BM)
- Incubate 30 min 37°C
- Incubate 10 min 68°C
- Use a Wizard PCR column to purify the DNA and reassess the concentration

Turbo cloning

This method was adapted from Boyd (1993). Typically a 16µl reaction was set up, in the order given-

- 8µl vector (dephosphorylated) and insert (at a 10:1 vector to insert molar ratio)
- 1.6µl 10× buffer (0.5M Tris-HCl, 50mM $MgCl_2$, 5mM ATP, 5mM DTE pH8.0)
- 6µl 40% (w/v) PEG 6000 (BDH)
- 0.4µl T4 DNA ligase (Boehringer: 1U/µl)
- Incubate room temperature 15 min
- Incubate 75°C 5 min, to heat kill ligase

- Add 64µl (4 volumes) of M buffer (medium salt restriction enzyme buffer- Boehringer Mannheim) containing 0.1-0.3µg Cre protein (Dupont)
- Incubate 30°C 30 min
- Incubate 75°C 10 min, to heat kill Cre
- Desalt the reaction prior to transformation by drop dialysis (section 2.7.4)

2.5.7 Restriction endonuclease digestion of DNA embedded in agarose plugs

This method was also adopted from Maule (1994)

- soak plugs for 10 min in a large excess of sterile TE, inverting tube frequently
- immerse plugs in 5ml TE containing phenylmethylsulfonylfluoride (PMSF - a protease inhibitor) at 40µg/ml
- incubate at 50°C for 30 min
- repeat previous two steps
- soak plugs for 2 hours in 10 volumes of 1X restriction enzyme buffer at room temperature, inverting frequently
- transfer single plugs to eppendorfs containing 1X restriction enzyme buffer, 0.1% Triton X-100, and 200µg/ml BSA. Add 20U of restriction enzyme. The final volume should be 100µl., and plug should be fully immersed in the absence of air bubbles.
- incubate overnight in waterbath at recommended temperature for digestion
- cool tube on ice 15 min then remove supernatant with a fine tipped pastette
- fill tubes with ice-cold TE

2.6 Electrophoresis of DNA

2.6.1 Electrophoresis solutions

20×TBE: 1M Tris, 20mM EDTA and 1M Boric acid pH8.3. Agarose gels were run in 0.5XTBE buffer.

20×TAE: 0.8M Tris, 20mM EDTA,pH8 and 0.4M acetic acid. Agarose gels were run in 1XTAE buffer.

10× DNA loading buffer/ 'stop mix': 20% ficoll, 100mM EDTA, orange G

6% Denaturing polyacrylamide (Severn Biotech LTD 6% polyacrylamide ratio 19:1 7M Urea, 1XTBE)

2.6.2 Agarose gel electrophoresis

DNA molecules were separated, according to size, in horizontal agarose gels by electrophoresis. The percentage of agarose (Sigma Type II medium EEO) in the gel was varied according to the range of size separation required. Restricted genomic DNA was run in a 0.8-1% agarose gel. Smaller DNA fragments, such as plasmids or PCR products were run on 1-2% gels, with fragments smaller than 200bp being run on either 2% low melting point agarose (Gibco BRL, ultrapure LMP agarose) or 4% Nusieve agarose (FMC Bioproducts). Ethidium bromide was added to all agarose gels and buffer at a concentration of 1µg/ml buffer. 1/10 volume of 10X stop mix was added to the sample prior to loading on the gel. The size marker used was 1kb ladder (Gibco BRL) or ϕX 174 HaeIII marker (Promega), for large and small fragments respectively. Gels were run in Electro-4 gel boxes (Hybaid) generally at 75-120 volts or at 10-30 volts overnight.

2.6.2.1 Preparative gel electrophoresis

DNA fragments were run in low melting point agarose gels (Ultrapure LMP agarose-Gibco BRL). Gels were viewed on a transilluminator at 305nm , and the required fragment was cut out of the gel using a sterile scalpel blade. Care was taken to ensure that the minimum size of gel slice was excised and that exposure to UV was minimised.

DNA was isolated from the gel slice using one of the methods in section 3.7.

DNA fragments which were to be used only as hybridisation probes were not isolated, but prepared as follows:- the gel slice was weighed and 1-3× the weight of water was added, such that 13µl of diluted gel slice contained 25-50 ng of DNA.

2.6.3 Polyacrylamide gel electrophoresis

Denaturing polyacrylamide gels were used to visualise DNA sequencing reactions

Preparation of gel plates

The gel plates are initially scrubbed using a powder detergent such as Flash, to remove any residual gel. After rinsing thoroughly with water they are laid flat on a sheet of 3MM in a fume hood. The plates are then rubbed down thoroughly with absolute alcohol, and once dry each plate is coated appropriately. The backplate is coated in a solution of 10ml 100% ethanol; 30µl acetic acid; and 30µl methacryoxypropyltrimethylsilane. The frontplate is coated with about 10ml of dimethyldichlorosilane solution. In each case the coating is spread across the plate lightly with a

tissue and left to dry for about 30 minutes, after which time the plates are rubbed dry very lightly with a tissue.

Pouring the gel

The plates are placed together, coated side inwards separated by two spacers at either side. They were then either taped together and the gel poured at an angle or laid flat and the gel dispersed between the two plates by capillary action in the absence of tape. The gel itself consisted of 40ml of 6% acrylamide mix which was polymerised by the addition of 240µl 10% ammonium persulphate and 40µl Temed. The gel takes about 40 minutes to polymerise.

Running the gel

The gel is run vertically (Scotlab vertical slab gel apparatus) in 1×TBE buffer at 3000V, 300mA with a constant 30W power. A metal plate is placed across the front of the glass plates to ensure even heat distribution and the gel is often preheated before loading by running for about half an hour. The urea is washed out of the wells just prior to loading samples. The wells are formed by inserting a sharks tooth comb.

Processing the gel

It is very important for the method used here to ensure that the gel sticks completely to the backplate. It was discovered that this could be facilitated by first cooling the plates in the cold room before attempting to separate them. After separation the backplate is soaked in 10% methanol and 10% acetic acid for 20 min. This fixes the gel and removes the urea. The gel is then washed under flowing water for another 20 min, before drying in a 68°C oven. Although it is not necessary for the development of [α-S^{35}] ATP to expose the gel to film at -70°C, it does ensure that the gel doesn't stick to the film and thus alleviates the need to wrap the gel in Saran wrap which reduces the signal . The gel was incubate at -70°C overnight, or longer if required.

2.6.4 Pulsed field gel electrophoresis

Large undigested YACs were resolved using the CHEF. The sizes of inserts in the ICI YAC library range between 200 and 500 Kb so appropriate conditions to allow separation across this size range were selected, i.e. 200 volts for 24 hours at a pulse time of 50 seconds. Markers were chromosomes of the yeast strains *AB970* (Link and Olson 1991), and strain *YP148* (Jones et al., 1989).

Size of chromosomes in YP148 (Kb) : 90; 220; 280; 360; 445; 555; 610; 690; 760; 800; 830; 920; 960; 1010; 1100; 1600; and 2500.

Size of chromosomes in AB970 (Kb) : 240; 280; 350; 440; 590; 680; 755; 810; 840; 950; 980; 1095; 1120; 1130; and 1640.

2.7 Isolation of DNA from agarose blocks

2.7.1 Gene clean

Here the Geneclean® II kit (BIO 101 Inc.) was used. This method gave the lowest yield of DNA, typically only 10-15%, but the DNA was very clean.

- For DNA in solution add 3 volumes NaI
- For DNA in agarose add 4.5 volumes NaI, if TBE buffer was used also add 0.5 volumes TBE modifier. Mix well and melt agarose at 45-55°C for 5-10 min
- Add 5µl glass milk suspension to 5µg DNA (an additional 2µl should be added for each additional 1µg)
- Vortex, and incubate on ice for 5 min, vortexing after every 2 min
- Centrifuge 5 secs, and discard supernatant
- Resuspend pellet in 250µl New Wash
- Centrifuge 5 secs, and discard supernatant
- Repeat the last two steps 3 times
- Resuspend pellet in 10µl TE
- Incubate 2-3 min at 45-55°C
- Centrifuge 30 secs and remove supernatant which contains the DNA
- Repeat the last two steps

2.7.2 Wizard DNA preparation

These kits are produced by Promega and were found to be easy to use and gave consistently high recovery of DNA (60-100%).

Wizard™ DNA clean up system

- Melt agar block at 70°C
- Add 1ml of resin and gently invert several times
- Pipette mixture into a 2ml syringe, and use plunger to gently push solution through the mini column
- Detach syringe, remove plunger, reattach syringe

- Wash column by pipetting 2ml 80% isopropanol into syringe, and push through
- Centrifuge 20 secs, and leave at room temperature for 5-15 min to remove isopropanol
- Transfer column to a fresh eppendorf and add 50µl prewarmed TE (70°C)
- Incubate 1-30 min
- Centrifuge 20 secs to elute the DNA

Wizard™ PCR clean up sytem

- Transfer reaction to a fresh tube, avoiding carry over of mineral oil
- Add 100µl direct purification buffer
- Vortex briefly
- Add 1ml of resin and vortex 3 times over a 1 min period
- Pipette mixture into a 2ml syringe, and use plunger to gently push solution through the mini column
- Detach syringe, remove plunger, reattach syringe
- Wash column by pipetting 2ml 80% isopropanol into syringe, and push through
- Centrifuge 20 secs, and leave at room temperature for 5-15 min to remove isopropanol
- Add 30µl of TE to the column, and incubate 1-30 min
- Centrifuge 20 secs to elute the DNA

2.7.3 Spin X

These were simple columns produced by Costar. This is the quickest method of purification of DNA from agarose, but the yield (40%) is not as high as with Wizard™ kits nor is it is clean as with Gene Clean. Good for fast crude recovery of DNA. DNA must be in regular agarose (Seakem GTG) , LMP agarose does not work

- Chop up agarose block into small pieces
- Place agarose on spin column
- Spin 15 min at room temperature
- Rotate column through 180°
- Spin 15 min at room temperature
- Discard column and precipitate DNA- 0.3M sodium acetate and 2 volumes cold ethanol
- Purification of DNA

2.7.4 Drop dialysis

This is a very simple and quick method which can be used to purify DNA away from salts, buffers, and enzymes. It can also be used to change the buffer which the DNA is in simply by dialysing the DNA against the buffer instead of distilled water.

- Half fill a petri dish with dH_2O or buffer
- Drop a filter (0.025μM Millipore VS), glossy side up, on the surface
- Leave to equilibrate for 15 min
- Drop ≥5μl of the DNA solution on top
- Dialyse for ≥15 min
- Take off the dialysed DNA

2.8 Transfer of DNA to membranes

The transfer of DNA fragments from an electrophoresed agarose gel to a membrane support results in immobilisation of the DNA fragments, so that the membrane carries a semi permanent replica of the banding pattern of the gel. After immobilisation, the DNA can be subjected to hybridisation analysis, enabling bands with sequence similarity to a labeled probe to be identified (Ausubel et al., 1995).

2.8.1 Southern transfer

- Soak gel in denature for 30 min
- Rinse with dH_2O
- Soak gel in neutraliser (pH to 5.5 with HCl) for 40 min
- Rinse with dH_2O
- Place gel in 2×SSC
- Place gel, wells down, on 3MM wick
- Cover gel with a sheet of Hybond membrane
- Place two sheets of 3MM (Whatman) on top of membrane
- Place one sheet of 15MM on top of 3MM
- Cover evenly with paper towels
- Place a plastic board on top of towels to disperse evenly the 0.5 - 1Kg weight which is then placed on top
- Leave transfer 4 hours - overnight
- Remove towels, weights & paper
- Mark wells on membrane before lifting off gel
- Wash gel briefly in 2×SSC
- Dry on bench 30 min

- Bake 80°C Vacuum oven 30 min

When transferring molecules over 20kb eg YACs the gel was initially depurinated by incubation in 250mM HCl for 20 min, prior to denaturation, and the gel was blotted for 48 hours.

2.8.2 Colony lifts

Bacterial colonies containing a particular insert can be identified directly by hybridisation. The first step in the nucleic acid hybridisation screening procedure is to grow large numbers of colonies on agar plates. Replica copies of these colonies are transferred to nitrocellulose filters where they can be screened.

- Lay filter down gently on plate, being careful not to trap any air bubbles
- To allow identification of colonies mark the plate and membrane at corresponding positions
- The filter is then placed ,DNA side up, on a series of sheets of 3MM paper soaked in various different solutions as indicated below
- 1min 10% SDS (sharpens signal)
- 2-5min 1.5M NaOH (denatures DNA), 1.5M NaCl(sharpens signal)
- 2-5min 1.5M NaCl, 0.15M Tris pH7.4 (neutralises)
- Rub down the filters using a tissue under 2×SSC (removes bacterial cells and salt)
- Dry filters on 3MM paper then bake under vacuum in an 80°C oven for about half an hour

2.9 Hybridisation protocols

2.9.1 Hybridisation solutions

Hybridisation buffer (1 litre) commonly used for Random Prime labelled DNA:

- 5×SSC
- 5×Denhardts (BSA, Ficoll, Polyvinylpyrolidone all at 1g/l)
- Add ≈100ml dH_2O
- Dissolve stirring on low heat
- Add detergent, the amount and type depends on what the buffer will be used for :-
- Hybond membranes -0.5% SDS
- Nitrocellulose membranes -0.1% SDS
- YAC DNA -2.5% Lauroyl Sarcosine
- Add 0.1% Pyrophosphate
- Add 10% Dextran Sulphate

- Make up to 1litre with dH_2O
- Filter through a 8μM filter under suction

Quick oligonucleotide hybridisation mix (1 litre):

- 5× SSC
- 2.5×Denhardts (BSA, Ficoll, Polyvinylpyrolidone all at 0.5g/l)
- 0.5% SDS
- Add dH_2O to 1 litre

2.9.2 The ICI YAC library

The ICI YAC library (Anand et al., 1990) was screened by Southern hybridisation (Southern 1975). The library consists of approximately 35,000 YAC recombinants gridded at a high density on 23 filters of 8cm × 12cm. The average insert size is approximately 350Kb and the theoretical complexity of the library is therefore >3.5 genome equivalents. The presence of chimaerism in the library is estimated at approximately 15%. The ICI YAC library was obtained through the Human Genome Mapping Project (HGMP) Resource Centre,London. The YACs were received in the form of an agar plug which was then restreaked on an AHC agar plate and grown overnight. A single colony was then picked and grown overnight in AHC complete medium and the DNA extracted the following day. The method used to extract the DNA was very crude hence it was more accurate to estimate the DNA concentration on a 0.8% gel against a marker of known concentration than spectrophotometrically.

2.9.3 Pre-hybridisation protocols

Pre-hybridisation of filters was carried out in oligo hybridisation mix at 5°C below the T_m of the oligonucleotide and in hybridisation mix at 68°C. Filters were placed between two gauze sheets in a hybridisation bottle (Hybaid). 10ml of hybridisation mix was used for smaller bottles and 20ml for larger bottles. Denatured sonicated salmon sperm was added to the hybridisation mix at a concentration of 100μg/ml. Bottles were placed in a Hybaid hybridisation oven for 1-4 hours (a longer prehybridisation is required on the first use of a filter). The ICI YAC library was incubated in a perspex box (50ml of hybridisation buffer) with a sealed lid, shaking vigorously in a 68°C oven for at least 4 hours.

2.9.4 Hybridisation and washing protocols

End labeled oligonucleotide hybridisations were carried out for 4-16 hours at 5°C below the T_m of the oligonucleotide. Filters were then removed from the bottles and washed three times for 5 minutes each at room temperature in 4× SSC, 0.1% SDS, heated to the same temperature as the hybridisation reaction.

Random primed hybridisations were incubated for 16 hours at 68°C. Filters were then washed at increasing stringency with 2×SSC, 0.1% SDS; 1×SSC, 0.1% SDS; or 0.1×SSC, 0.1% SDS at 68°C.

2.9.5 Removal of hybridisation signal

Two different methods were used to remove the radioactive probe from a membrane enabling it to be rehybridised.

Gentle stripping

- Place membranes DNA side down in 0.4 M NaOH at 50°C
- Incubate, shaking, at 68°C, for 30 min
- Transfer membranes to 0.1×SSC; 0.1% SDS; 0.2M Tris HCl pH7.4 at 45°C
- Incubate, shaking, at 68°C for 30 min

Faster, more effective stripping

- Add filter DNA side down to a metal tray of boiling 0.1% SDS
- Incubate 30 secs, then remove tray from hotplate
- Allow to cool to room temperature
- In both cases check that all the radioactive signal has been effectively removed by exposing the filters to film overnight.

2.9.6 Hybridisation signal detection

2.9.6.1 Autoradiography

Filters were exposed to Kodak X-OMAT film in cassettes with intensifying screens. Films were exposed for a period of 30 minutes to two weeks (depending on the signal strength) before being developed, using an automatic X-ray film processor RGII (Fuji).

2.9.6.2 Phosphorimaging

A Molecular Dynamics Phosphorimager was also used for detecting and quantifying hybridisation. Filters were exposed on a phosphor screen for 30 minutes to 48 hours. They were then scanned on the Phosphorimager, where the radioactive signal is converted into a digital image, with variations in the pixel values proportional to the amount of radioactivity present. The image was displayed on a grey scale. The upper and lower limits of the grey scale were adjusted to give good image visualisation. The image was then printed out on a grey scale laser printer.

2.10 Radiolabelling of DNA

2.10.1 Random primed DNA labelling

The DNA fragment to be labelled was usually in the form of a block of LMP agarose, although occasionally purified DNA was labelled. The labelling technique was performed using a kit produced by Boehringer Mannheim and is described below. The kit is based on the hybridisation of a mixture of all possible hexanucleotides to the DNA to be labelled. Complementary strands are then synthesised from the 3'OH termini of the annealled random hexanucleotide primers, using Klenow enzyme. Modified deoxynucleotide triphosphates (in this case ^{32}P) are incorporated into the newly synthesised complementary DNA strand.

- Add 1-2 volumes dH_2O to the agarose block
- Heat to 68°C until melted, mix
- Remove 15µl (50-100ng) to a fresh tube, boil 5 min
- Add 5µl Random prime reaction mix + GAT (2µl hexanucleotide reaction mix in 10x reaction buffer, and 1µl of each of dG/T/ATP all at 0.5mM)
- Add 1µl Klenow enzyme
- Add 3µl [α-P^{35}] CTP (10µCi/µl)
- Centrifuge briefly, and incubate 37°C 45-60 min
- To test the incorporation of radioactivity: touch the labelled probe lightly with a Gilson P200 tip, then touch the tip onto a piece of filter paper. Place a monitor at a position relative to the filter such that it reads 100 counts. The filter is then washed twice under suction with 5% TCA, and the number of counts re-measured at the same distance from the monitor. This second reading therefore gives a rough measure of incorporation. If the incorporation is more than 50% the probe is boiled for 5 minutes, then added to the prehybridisation mix. If however the

incorporation is poor the probe is boiled for another 5 minutes, then another 1µl of Klenow enzyme is added and the mixture reincubated for another hour. The incorporation is again measured and a decision made whether or not to go ahead with the probe. If the incorporation is less than 20% the probe is abandoned. If however the incorporation is 20-50% a Nick™ column is used to remove the unincorporated radioactivity as described below

- Pour the top layer off the column
- Equilibrate with 2ml of TE
- Dilute the probe so in 50-100µl and apply to the column
- Apply 350-300µl of TE to the column
- Discard flow through
- Apply 400µl TE to the column
- Collect hot probe
- Boil 5 min then add to the prehybridisation mix

On hybridisation of a probe to the ICI YAC library it was necessary to measure the level of incorporation after removal of unincorporated label using a scintillation counter. 1×10^5 -1×10^6 counts per ml of hybridisation buffer were required which meant setting up 4-6 random prime reactions.

2.10.2 Non-random primed DNA labelling of small probes

Probes which are less than 200bp long do not label well using the standard Random prime kit, an adaptation of the kit where primers specific to the fragment to be labelled are used , is however very efficient and is described below.

- Denature 100ng of each primer for 3 min at 95°C
- Add 3µl GAT (1µl of each of dG/T/ATP all at 0.5mM)
- Add 1µl Klenow enzyme
- Add 3µl [α-P35] CTP (10µCi/µl)
- Centrifuge briefly, and incubate 37°C 30 min
- Add 1µl Klenow enzyme
- Centrifuge briefly, and incubate 37°C 30 min
- Repeat the last two steps
- Boil 5 min then add to the prehybridisation mix

2.10.3 End labelling of DNA oligonucleotides

Polynucleotide kinase catalyses the transfer of the terminal phosphate group of ATP to the 5' hydroxylated terminus of DNA or RNA.

- Prehybridise filters in oligonucleotide hybridisation buffer containing denatured salmon sperm at 10mg/ml for 5-30min at 68°C
- To 30 ng of oligonucleotide DNA
- Add 2µl PNK buffer (BM)
- Add dH_2O to 16µl
- Add 1µl Polynucleotide Kinase (BM)
- Add 3µl [γ-^{32}P] ATP (10µCi/µl)
- Incubate 40min 37°C
- Add Kinase reaction to prehybridisation solution
- Incubate overnight at 48°C
- Wash filters in 4×SSC 0.1%SDS 1%PPi for 4 sets of 5min

2.10.4 Preannealing of repetitive sequences

Two methods were tried to block non-specific hybridisation. On direct comparison, using sonicated human DNA proved to be more effective, and hence was the method used. In both cases the probe was labelled using a standard random prime reaction, but once a sufficient degree of incorporation had been achieved (70-90%) the probe was not put down a nick column, but repetitive sequences were suppressed using one of the methods described below.

2.10.4.1 Cot-1 suppressed probes

- Precipitate the probe by adding 50µl Cot-I DNA (1mg/ml), 10.5µl 2M NaOAc, and 140µl cold absolute EtOH
- Incubate 15min -70°C
- Centrifuge 15min in the cold room
- Resuspend the vacuum dried pellet in 100µl TE + Cot-I DNA at 10mg/ml
- Add 50µl 20×SSC, 25µl dH_2O, 20µl 0.1M Tris pH7.4; 0.1M EDTA, 1% SDS
- Vortex
- Incubate 100°C 5 min
- Incubate 65°C at least 20 min

- Add probe to prehybridisation mix

2.10.4.2 Sonicated human DNA suppressed probes

- Add 500µg of sonicated human DNA for every 25ng of probe
- Add 500µg of sonicated salmon sperm DNA for every 25ng of probe
- Incubate at 100°C for 5 minutes to denature the DNA
- Incubate at 68°C for 30-45 minutes to allow the DNA to reanneal
- Add to prehybridisation

2.11 Mutational analysis of the OMP gene

2.11.1 Usher type IB patients

Patient DNAs were provided by groups from Newcastle, Manchester and USA. Most of the work was done on the latter two since these came with parental and affected/unaffected sib DNAs, making it possible to assess whether or not a mutation/ polymorphism was inherited. Each group sent DNA stocks of a known concentration, which were then used to make dilutions to 100ng/µl for PCR. Laboratory numbers were given to these samples for simplicity during mutation analysis.

Families A, B, & C were supplied by William Kimberling (Kimberling et al., 1992), hence these had been linked to D11S527. The remaining families were British and these were typed as USHIB on the basis of clinical features and ethnic origin. The latter were therefore given to us in order to test this supposition at the candidate gene level.

Table 2.7 Usher 1B families

Family	Laboratory No.	Sample No.	Relationship
A	1	D3677	unaffected father
	2	D3680	unaffected mother
	3	D3678	affected sib
	4	D3681	affected sib
	5	D3679	unaffected sib
B	6	D2734	unaffected father
	7	D2735	unaffected mother
	8	D2377	unaffected sib
	9	D2378	affected sib
	10	D2376	affected sib
C	11	D2127	unaffected father
	12	D2139	unaffected mother
	13	D2137	unaffected sib
	14	D2134	unaffected sib
	15	D2138	unaffected sib
	16	D2129	affected sib
D	17	920922	unaffected father
	18	920921	affected sib
	19	920923	affected sib
E	20	920920	unaffected mother
	21	920919	affected child
F	23	92/959	unaffected mother
	24	92/957	unaffected father
	25	92/958	affected child
G	31	No.5	affected sib
	32	No.6	unaffected sib
H	34	No.17	unaffected sib
	35	No.18	affected sib

2.11.2 Chemical mismatch detection

The mutation detection method chosen was the HOT technique (so-named for the hydroxylamine and osmium tetroxide used in the method). This method was originally developed by Cotton and colleagues (1988), and is based on chemical modification of mismatched bases (mismatched Cs with hydroxylamine, Ts with osmium tetroxide) in heteroduplex molecules, followed by piperidine cleavage of the DNA molecule at the site of modification. The theory of the technique relies on the principles of Maxam and Gilbert sequencing (1980).

This technique is very reliable, and in cases where both wild type and mutant DNA fragments are labelled has been shown to identify 100% of mutations. Mutations still need to be confirmed by sequencing, although the nature of the mismatch (depending on the particular chemical it was identified by) and its location (a specific distance from either of the primers) are inherently indicated in the process of the technique. Another advantage of this technique is that it can be used to assess whether a mutation is homozygous or heterozygous, simply by labelling mutant DNA and heteroduplexing (see below) it back to cold self DNA. If the sample is heterozygous then a shortened fragment characteristic of the mutation is seen. The HOT technique can be routinely used to identify mutations in fragments of 1Kb or more (Cotton et al., 1988). The only drawback of this technique is that it uses hazardous chemicals and is relatively labour intensive, but these are far out weighed by its accuracy and the fact that it can be used to screen fragments 4-5 times the size of those used in the other major techniques.

The protocol described below is based on that used by Montandon et al (1990) , but modified for the osmium tetroxide reaction along the lines of Cotton *et al* and is published in Condie *et al.*

Both target and wild type DNAs are amplified by PCR and the products excised from a low melting agarose gel. The bands are then purified using the Wizard DNA prep kit and the recovery assessed by running an aliquot on a mini gel. The actual heteroduplex reactions are set up in two ways: labelling wild type and heteroduplexing to cold mutant DNA; and the converse, labelling mutant and hteroduplexing to cold wild type DNA. About 10 to 100ng of the purified DNA is kinase labelled

Kinase end labelling

- 10-100ng DNA
- 2µl 10xPolynucleotide Kinase Buffer
- 2µl [γ-^{32}P] ATP, specific activity ~5000 Ci/mmole
- 1µl (2.5 units) Polynucleotide Kinase
- distilled water to 20µl

- Incubate at 37°C for 30 minutes then transfer to ice
- Precipitate by adding:-
- 22.5µl dH_2O
- 7.5µl 2M Sodium Acetate
- 1.5µl mussel glycogen (1mg/ml)
- 750µl cold absolute ETOH
- Incubate 15-30 min -70°C
- Centrifuge 15 min
- Wash 2×70% ETOH
- Resuspend in 10µl dH_2O

Formation of heteroduplex

- 100ng cold DNA
- 10ng labelled DNA
- 1µl 10× salts (3M NaCl, 1M Tris pH8)
- dH_2O to 10µl
- Overlay with paraffin oil
- Incubate 100°C 5 min
- Incubate 65°C overnight

To clean and precipitate the heteroduplex

Pipette off the majority of the oil, then to remove the remainder of the oil add

- 50µl 0.3M Sodium Acetate, 0.1M EDTA
- 50µl chloroform
- Vortex, centrifuge and remove the top layer to a fresh eppendorf. The chloroform can be re-extracted with 150µl of 0.3M Sodium Acetate, 0.1M EDTA, if the majority of radioactive counts have not been removed at the first attempt.

Precipitate by adding

- 1.5µl mussel glycogen (10mg/ml)
- 750µl absolute ETOH
- Incubate 15 min in dry ice/methanol bath
- Centrifuge 15 min
- Wash 2×70% ETOH

- Dry 1 min under vacuum
- Resuspend in 13µl TE

  Divide into two tubes, 7µl for hydroxylamine modification, leaving 6µl for osmium tetroxide modification in the original tube.

Hydroxylamine modification

Make a fresh solution each time, combine

- 0.94g Hydroxylamine Hydrochloride
- 3.36ml dH_2O
- Vortex and add
- 1.02ml Diethylamine
- Vortex and add 20µl to each 7µl sample
- Mix, incubate at 37°C for 1 hour, then transfer to ice

Osmium Tetroxide modification

To the 6µl aliquot of heteroduplex add

- 2.5µl of 10× Osmium Tetroxide buffer (100mM Tris pH7.7, 10mM EDTA, 15% pyridine). This buffer should be stored frozen, and each aliquot used only once.
- 15µl freshly diluted (4% to 0.8% in dH_2O) Osmium Tetroxide
- Mix, do not centrifuge
- Incubate 5 min at 37°C, then transfer to ice

After each chemical modification precipitate add

- 200µl 0.3M Sodium Acetate, 0.1M EDTA
- 5µl Dextran (10mg/ml)
- 750µl absolute ETOH
- Incubate 15 min in dry ice/methanol bath
- Centrifuge 15 min
- Wash 2×70% ETOH
- Dry 1 min under vacuum

Piperidine cleavage

To dried pellet add

- 50µl freshly diluted piperidine (1/10 dilution to make 1M stock)
- Vortex 1 min
- Centrifuge briefly
- Incubate 90°C 30 min, then transfer to ice
- Precipitate by adding
- 200µl 0.3M Sodium Acetate, 0.1M EDTA
- 750µl absolute ETOH
- Incubate 15 min in dry ice/methanol bath
- Centrifuge 15 min
- Wash 2×70% ETOH
- Dry 1 min under vacuum
- Resuspend the dried pellet in 5µl TE
- Add 2µl denaturing dye
- Store at -20°C
- Before loading on denaturing polyacrylamide gel, heat samples to 100°C for 2 min

2.12 Sequencing of DNA

2.12.1 Conventional sequencing with a radioactive isotope

The method used here was that used by Stillman's lab at Cold Spring Harbour (pers. comm.) and is outlined below. Many of the constituents came from the Sequenase® Version 2.0 DNA sequencing Kit (United States Biochemical Corporation).

- 7µl of DNA, either 7µl from an alkaline lysis preparation which has been resuspended in 10µl, or 3-5µg of alternatively prepared DNA
- Add 1µl of RNase A (10mg/ml)
- Incubate 15 min room temperature

Annealing reaction

- Add 1µl of the appropriate primer (10ng/µl)
- Add 1µl of 1M NaOH
- Incubate 10 min 68°C

Sequencing reaction

- Add 4µl TDMN (0.28M TES (Sigma T-1375), 0.5ml chloroform, 50mM DTT (USB), 80mM $MgCl_2$, 0.2M NaCl, dH_2O to 50ml pH to ≈1.6)

- Incubate 10 min room temperature
- Place on ice until ready to use
- Add 4µl labelling mix
- 1µl 0.1M DTT
- 0.4µl 5×Sequenase labelling mix (USB - 7.5µM dCTP, 7.5µM dGTP, 7.5µM dTTP)
- 2.1µl LoTE (3mM Tris pH7.5, 0.2mM EDTA)
- 0.5µl [α-^{35}S]-dATP
- Add 2µl T7 DNA polymerase (USB) diluted 1:10 in enzyme dilution buffer (USB 10mM Tris.HCl pH7.5, 55mM DTT, 0.5Mg/ml BSA)
- Incubate 45 secs room temperature

Termination reaction

- Add the labelling reaction in 4µl aliquots to 2.5µl of each termination reaction (USB)
- Incubate 10 min 37°C
- Add 5µl stop buffer (USB)
- Store at -20°C immediately or denature at 100°C for 2 min and run on a polyacrylamide gel

In this method the first 50 bases of sequence is obscured because prior to this the salt forms a front which decreases the resolution severely. It is possible to get rid of the salt by precipitating and washing the DNA prior to the addition of stop buffer as described below

- To the sequencing reaction
- Add 0.4µl 0.5M EDTA
- Add 1.5µl 2M Sodium Acetate pH5.5
- Add 1µl tRNA (1µg/µl)
- Add 0.6µl dH_2O
- Add 2 volumes of cold absolute ETOH
- Incubate -70 C 15 min
- Centrifuge cold room 15 min
- Wash 70% ETOH
- Dry 1 min under vacuum
- Resuspend 6µl dH_2O
- Add 4µl stop buffer

2.12.2 Double stranded PCR sequencing with DMSO

This method also uses many constituents from the USB Sequenase® Version 2.0 DNA sequencing Kit, but has been specially adapted to allow sequencing of PCR products

Normally a 100µl PCR reaction was gel purified, and the DNA isolated by Gene Clean, then combine:-

Annealing reaction

- 6µl DNA solution (~100ng)
- 1µl DMSO
- 1µl sequencing primer (~1µg)
- 2µl sequencing buffer (USB - 200mM Tris.HCl pH7.5, 100mM MgCl2, 250mM NaCl)
- Incubate at 100°C for 2 min then place immediately in a dry ice/methanol bath

Sequencing reaction

Add to annealing reaction:-

- 1µl 0.1M DTT
- 0.5µl of deoxyadenosine-5'-triphosphate ($\alpha^{35}S$)
- 1.5µl Sequenase enzyme (USB) diluted 1:6 in enzyme dilution buffer (USB)
- 2µl labelling mix (USB) diluted 1:15 in dH2O
- Incubate 45 seconds at room temperature

Termination reaction

Add 3.2µl of sequencing reaction to:-

- 2µl of each termination reaction (USB) diluted 9:1 with DMSO
- Incubate 37°C 10 min
- Add 4µl stop solution

2.12.3 ALF sequencing

Two methods were used for fluorescent sequencing of templates on the Pharmacia ALF machine.

2.12.3.1 ALF cycle sequencing

In an attempt to improve the quality of sequence obtained PCR products for cycle sequencing were also purified using Mung Bean Nuclease or Isopropanol, since excess PCR primers interfere with subsequent PCR-like sequencing reactions..

Mung Bean Nuclease (Dowton and Austin 1993)

Mung bean nuclease is a single-strand-specific nuclease purified from germinated mung bean sprouts. It degrades DNA and RNA to 5'-phosphoryl mononucleotides.

To 6µl of the PCR reaction add

- 4.4µl of 25mM $MgCl_2$
- 1.2µl of 1mM EDTA
- 0.4µl of distilled water
- 2U Mung Bean Nuclease (NEB)
- Overlay with oil and incubate 25°C for 30 min
- Incubate 100°C for 5 min

Isopropanol

Isopropanol precipitation allows the purification of PCR-amplified DNAs away from excess nucleotides and amplification primers.

To 6µl of the PCR reaction add

- 20µl of 4M Ammonium Acetate, vortex
- 40µl of Isopropanol
- Incubate room temperature 10 min
- Centrifuge 20 min
- Wash pellet with 70% ETOH
- Resuspend pellet in 12µl dH_2O

Cycle sequencing

This was carried out using the BRL dsDNA Cycle sequencing system

Prereaction mix

- 5µl fluorescently labelled primer

- 4.5µl Taq sequencing buffer
- 26µl template DNA (~50fmol)
- 0.5µl Taq DNA polymerase (5U/µl)

Reaction mix

- Add 8µl from prereaction mix to each of four tubes containing 2µl termination mixes -A, -C, -G, and -T
- Overlay with oil, then amplify as described below

20 cycles: 30s 95°C →30s 55°C →60s 70°C then 10 cycles: 30s 95°C →60s 70°C

2.12.3.2 ALF single stranded sequencing

Primarily single stranded DNA must be prepared. This was done using Dynabeads as described below

Washing the Dynabeads and extracting the DNA

- Take 20µl well suspended Dynabeads per reaction
- Wash twice with 100µl 2M Binding and Washing Solution (2M BWS - 10mM Tris HCl pH7.5, 1mM EDTA, 2M NaCl) removing the liquid with a long tipped pastette
- Resuspend the beads in 20µl BWS
- Add up to 40µl PCR product (template amplified using one biotinylated and one nonbiotinylated template)
- Incubate at least 30 min at room temperature on an IKA Vibrax VRX at 1400rpm
- Wash twice with 100µl TE to remove unbound material

Preparation of the biotinylated strand

- Add 10µl 0.1M NaOH (Analar BDH 1M stock stored in aliquots at -20°C)
- Incubate 5 min room temperature
- Remove the NaOH, which contains the non-biotinylated strand
- Wash twice with 100µl TE, which neutralises the solution and removes the high salt
- Resuspend the Dynabeads in 13µl dH_2O to elute the biotinylated strand
- Store the biotinylated strand, ready to sequence, at 4°C

Preparation of the non-biotinylated strand

- Add 1µl of 1M HCl (Analar BDH 1M stock stored in aliquots at -20°C) to the NaOH supernatant, this allows the non-biotinylated strand to return to its native conformation.
- Mix well

- Add 2µl dH_2O
- Store the non-biotinylated strand, ready to sequence, at 4°C

Pharmacia Autoread sequencing protocol

- Add 2µl annealing buffer (a buffer solution containing $MgCl_2$) to each tube
- Add 2µl fluorescent primer (1 pmol)
- Mix by tapping
- Incubate in a 65°C water bath 10 min
- Allow to cool slowly for at least 10 min
- prewarm tubes containing the 2.5µl of dNTP mix in 42°C water bath
- Add 1µl extension buffer to each sample
- Add 2µl of T7 DNA Polymerase (diluted 1:2 in enzyme dilution buffer)to each sample
- Mix, and add 4.8µl to each dNTP tube
- Incubate 5 min at 48°C
- Add 5µl stop solution
- Store at -20°C immediately or denature at 100°C for 2 min and run on a polyacrylamide gel

2.12.4 ABI cycle sequencing

This involves a linear amplification of the sequencing reaction using 25 cycles of denaturation, annealing, and extension in the presence of Taq polymerase and labelled di-deoxy nucleotides (the primer is unlabelled). There are several advantages of this method over conventional sequencing. The dideoxy terminators are each labelled with a different fluorescent dye, which means that the entire reaction can be carried out in a single tube and run out in a single gel lane. In addition false stops go undetected since no fluorescent dye will have been incorporated. The reaction is very simple and easy to set up as it is just like setting up a PCR reaction. The only time consuming step is cleaning up the reactions after the PCR to remove any excess dye terminators, since these will obscure the signal at the start of the gel. A few recommendations were also made by the kit manufacturers as to the design of the primer and the preparation of the template.

Double stranded template was prepared in bulk by caesium chloride centrifugation, then any residual caesium removed by drop dialysis. Smaller amounts of plasmid template were prepared by PEG precipitation. PCR template was purified using the Promega Wizard PCR Kit.

Primers were recommended to be longer then 18 base pairs for good hybridisation and to avoid false priming. Long runs of single bases were avoided as far as possible, as were primers able to form secondary structure or able to dimerise. The GC content was recommended to be over 50%. If however this was not possible primer lengths were extended to longer than 18 bases ensuring that the Tm of the primer was greater then 45°C. These conditions were on the whole not that difficult to meet and every single primer used worked. The primers were designed using the Primer programme and purified by rapid butanol purification.

Reaction set-up

The reaction was set up using the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit as described in the booklet provided with the kit, a basic description of which is shown below.

For both double stranded sequencing and sequencing of PCR fragments (also double stranded) the primer was used at 3.2pmol per reaction. Less template is required for PCR sequencing than double-stranded plasmid sequencing i.e. 100-500ng compared to 1-1.5µg.

Firstly a reaction premix is prepared. This can be stored for at least a month at 4°C hence enough for several reactions was usually prepared. Amounts for the 1× mix are shown.

Reaction premix

- 4µl 5×TACS buffer (400mM Tris.HCl, 10mMMgCl_2, 100mM $(NH_4)_2SO_4$ pH9)
- 1µl dNTP mix (750µM dITP,150µM dATP, 150µM dTTP, 150µM dCTP)
- 1µl DyeDeoxy™ A Terminator
- 1µl DyeDeoxy™ T Terminator
- 1µl DyeDeoxy™ G Terminator
- 1µl DyeDeoxy™ C Terminator
- 0.5µl (4 units) AmpliTaq® DNA Polymerase

Sequencing reaction

- 9.5µl reaction premix
- 9.5µl template (1µg ds DNA, 100ng purified PCR products)
- 1µl primer (3.2pmol)
- Overlay the reaction with a drop of mineral oil

Cycling the reaction

The tubes are then placed on a thermal cycler preheated to 96°C. Ramping time for cycle sequencing is optimal at 1°C /sec hence Perkin Elmer Cetus Model 480 was used and the programme is shown below.

- 96°C for 30 secs
- 50°C for 15 secs
- 60°C for 4 minutes
- 25 cycles total
- Rapid thermal ramp to 4°C and hold

Removal of excess Taq dye terminator dyes using phenol/chloroform extraction

- Transfer the reaction products to a fresh tube, taking as little oil as possible
- Add 90µl dH_2O
- Add 100µl of ABI phenol/chloroform/water (68:14:18) reagent, mix first to ensure homogeneous
- Vortex 5 secs, and centrifuge 2 min
- Remove 105µl of the top aqueous phase to a fresh tube
- Add 100µl of phenol/chloroform/water reagent
- Vortex 5 secs, and centrifuge 2 min
- Remove 93µl of top aqueous phase to a fresh tube
- Add 15µl of 3M NaOAc pH4.5 followed by 300µl cold EtOH and mix thoroughly by vortexing
- Incubate on ice 15 min
- Centrifuge 15-20 min
- Remove as much EtOH as possible with a pastette
- Centrifuge briefly, and remove remainder of EtOH using a Gilson P2 micropipettor
- Dry pellet 85°C 1min
- Pellets can now be stored at -20°C or resuspended 4µl of formamide/EDTA
- Incubate 90°C 2min, store on ice for up to 1 hour before loading on gel

3. Cloning a repeat free resource

3.1 Optimizing the product complexity of Alu PCR

Inter Alu-PCR was used to amplify human specific fragments of DNA from human chromosome 11 reduced somatic cell hybrids (E67.4, WJX11.2, WJX3.4 see section 2.1.3.1). The primer chosen was designed to match a consensus sequence near the 3' end of the Alu repeat. Amplification only occurs if two sufficiently close Alu repeats are inversely orientated toward each other. Alu PCR produces a complex mixture of inter-Alu sequences which may be cloned in order to isolate specific products. Factors influencing the efficiency with which inter-Alu sequences are amplified include i) degree of match to the consensus sequence to which the primer sequence is directed, from one or both adjacent inversely orientated Alus; ii) the distance separating inversely orientated Alus; iii) the nature of the intervening sequence. It was desirable to minimize the effects of these factors and identify the conditions which gave the most uniform distribution of products and maximize their number, thus ensuring a comprehensive representation of products from all regions of the human component of the hybrid.

3.1.1 Comparison of different buffer and enzyme combinations

These experiments predate the licensing restrictions of the PCR technique, allowing the optimal enzyme and buffer combination for this particular experiment to be established. Initial experiments were carried out using Taq (isolated from *Thermus aquaticus* - Promega) and Tth (*Thermus thermophilus* - Hybaid) thermostable DNA polymerases. Each enzyme was tested with another two

buffers in addition to those supplied by the manufacturers. Tricine buffer (Ponce and Micol 1992) unlike the others does not contain KCl and has been shown to allow the preferential amplification of larger (3-6 kb) fragments. A fourth buffer was designed for Alu-PCR "in house" to give optimised reproducibility with a variety of DNA templates, and is referred to as Edinburgh buffer.

Initial experiments were carried out to find the most efficient annealing temperature. The templates used for these experiments were the somatic cell hybrids WJX11.2 and E67.4, both of which contain subfragments of chromosome 11 as their sole human component (~50 Mbp complexity in each case). Annealing at 65°C was found to give more distinguishable products than 60°C in a 35 cycle PCR reaction. Taq polymerase failed to give any products with Tricine or Edinburgh buffer, few products with Tth buffer and only gave a satisfactory result with Taq buffer. In contrast, Tth polymerase gave many distinguishable products, and more than Taq polymerase itself with Taq buffer, under all conditions tested. Fig. 3.1 shows Taq and Tth enzyme amplification with each buffer : Fig. 3.2 shows the range of products achieved with Tth enzyme.

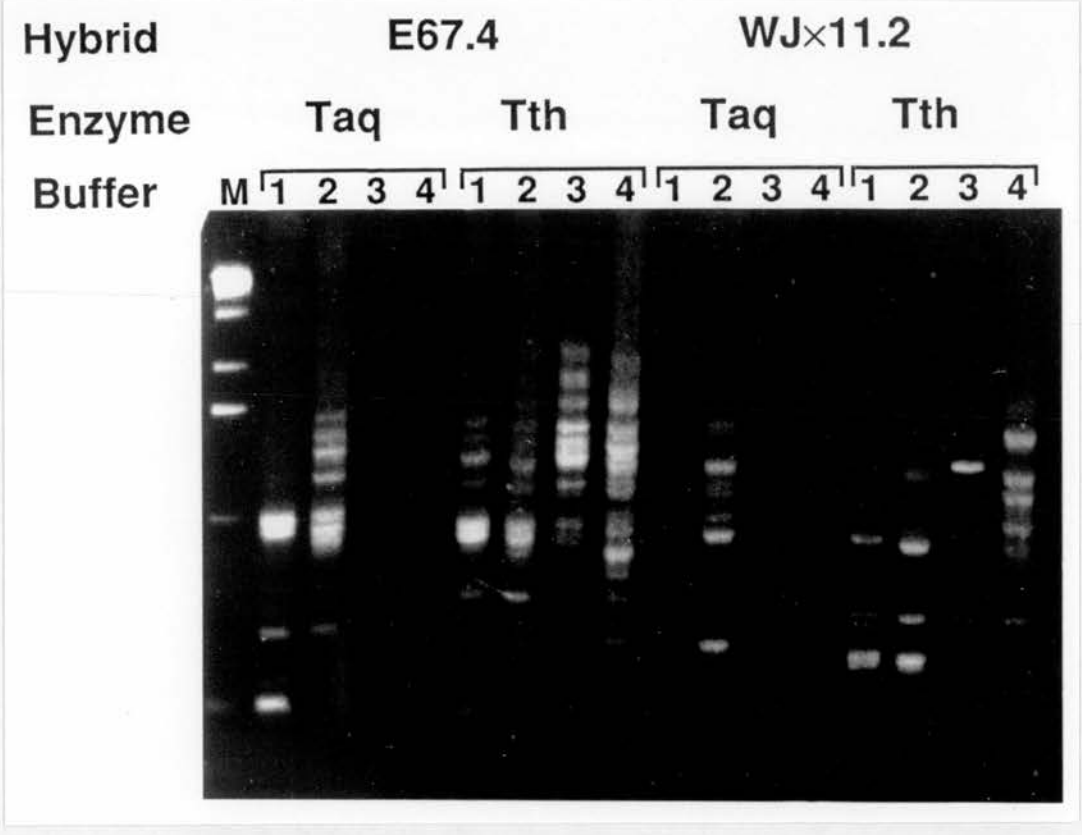


Figure 3.1 Comparison of different buffer and enzyme combinations

The range of products obtained by Alu-PCR amplification of the somatic cell hybrids E67.4 and WJX11.2, using the Alu repeat primer 614. Each hybrid was amplified using different enzyme and buffer combinations as indicated. For both hybrids, Tth enzyme gave a greater range of products, and was able to function with all of the buffers tested.

Key: M = 1kb ladder marker (Gibco BRL); 1 = Tth buffer; 2 = Taq buffer; 3 = Tricine buffer; 4 = Edinburgh buffer

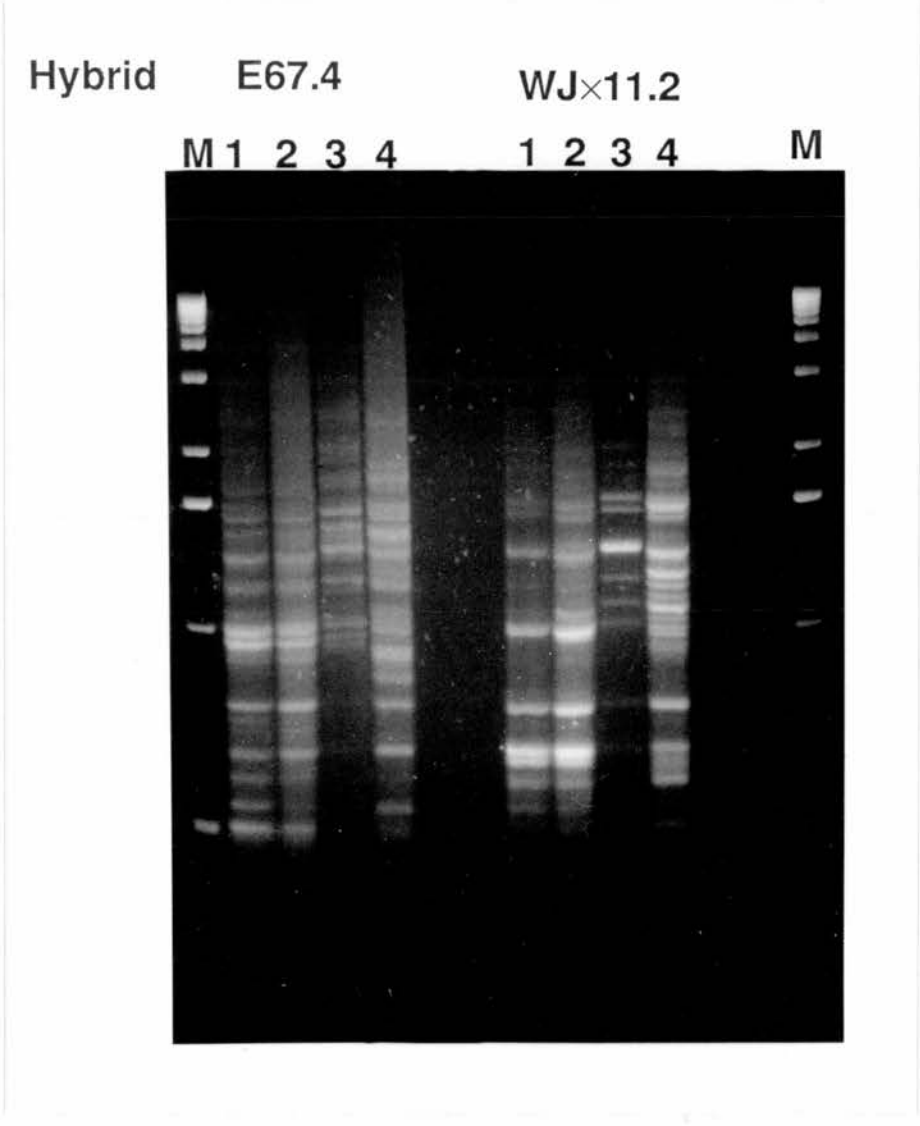


Figure 3.2 The range of products achieved with Tth enzyme

The range of products obtained by Alu-PCR amplification of the somatic cell hybrids E67.4 and WJX11.2, using the Alu repeat primer 614. Each hybrid was amplified using Tth enzyme and different buffer combinations as indicated. Human, hamster, and mouse controls were all blank but have been left off this gel for clarity.

Key: M = 1kb ladder marker (Gibco); 1 = Tth buffer; 2 = Taq buffer; 3 = Tricine buffer; 4 = Edinburgh buffer

3.1.2 Obtaining the point prior to saturation

The maximum product complexity is most likely to be obtained during the exponential phase of Alu-PCR. Therefore, using Tth enzyme with all four buffers, I attempted to establish the point at which each amplification reaction became non-exponential. Generally after 25 cycles of amplification (using 100ng of human genomic in a 50µl reaction) the amount of product is still doubling with each cycle, hence this was chosen as the start point. Samples were removed at cycles 25 to 35 with a final single long extension cycle at 72°C. In order to keep the total concentration of DNA constant, the amount of sample loaded on a gel was decreased by half for each additional cycle beyond 25. This was because the number of products should approximately double with every amplification cycle when the reaction is in exponential phase. If, however, the reaction has become saturated then some products will be amplified in preference to others, resulting in a stronger signal for specific products. Some of the reactions gave very faint products, so the complexity was assessed by Southern transfer of the DNA to a nylon membrane and hybridisation with the Alu primer probe 614. The greatest product complexity was gained after 27 cycles using Tth polymerase in combination with Taq buffer (Fig. 3.3). After 27 cycles the amount of product falls off indicating that the reaction is no longer exponential. A further comparison of the rodent background amplification by Tth showed that the enzyme produced a high background in combination with mouse and hamster with Tth buffer (Fig. 3.4) but none with Taq buffer (Fig. 3.4).

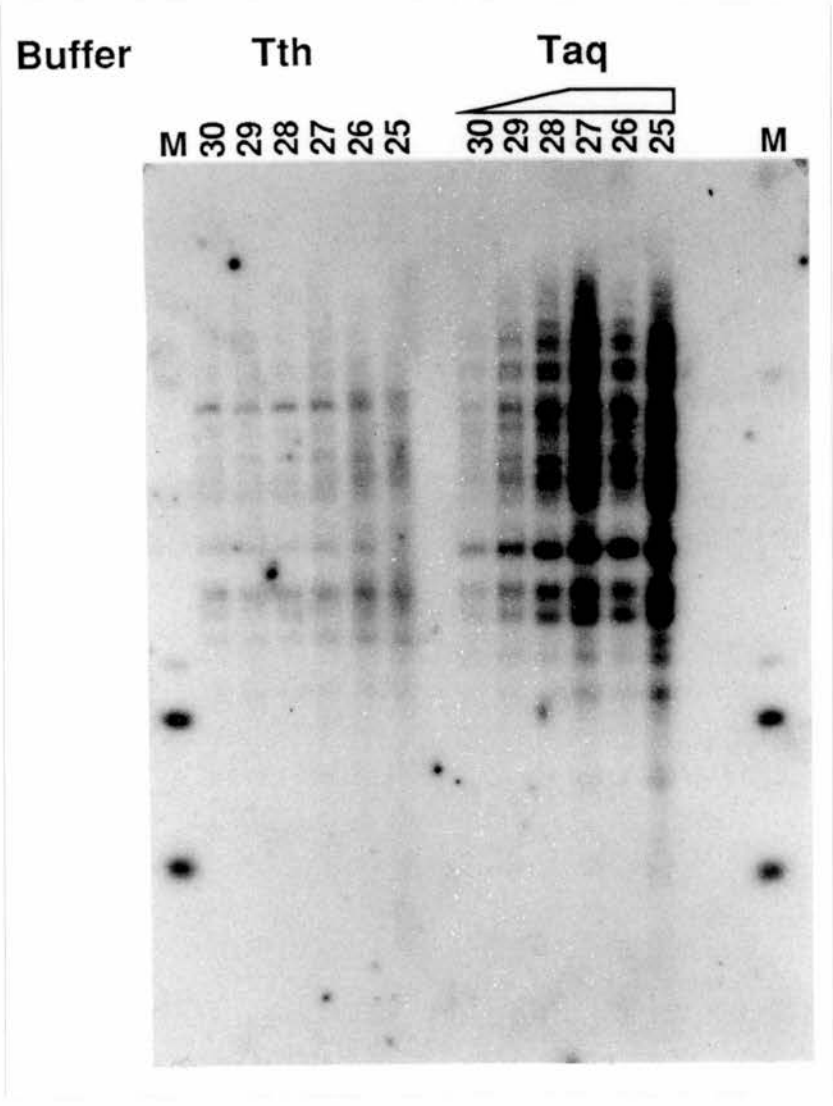


Figure 3.3 Assessment of product complexity

Alu PCR reactions using Tth enzyme in combination with Tth or Taq buffer were carried out for increasing numbers of cycles, using the somatic cell hybrid E67.4 as a template. With each extra cycle half as much sample was loaded on a gel, which was run until the 500bp products reached the bottom, then Southern blotted. The membrane was then hybridised with ^{32}P labelled Alu primer 614. The greatest product complexity, within the exponential phase of the reaction, was achieved after 27 cycles with Taq buffer. M = 1kb ladder. The Taq polymerase 26 cycle sample was underloaded.

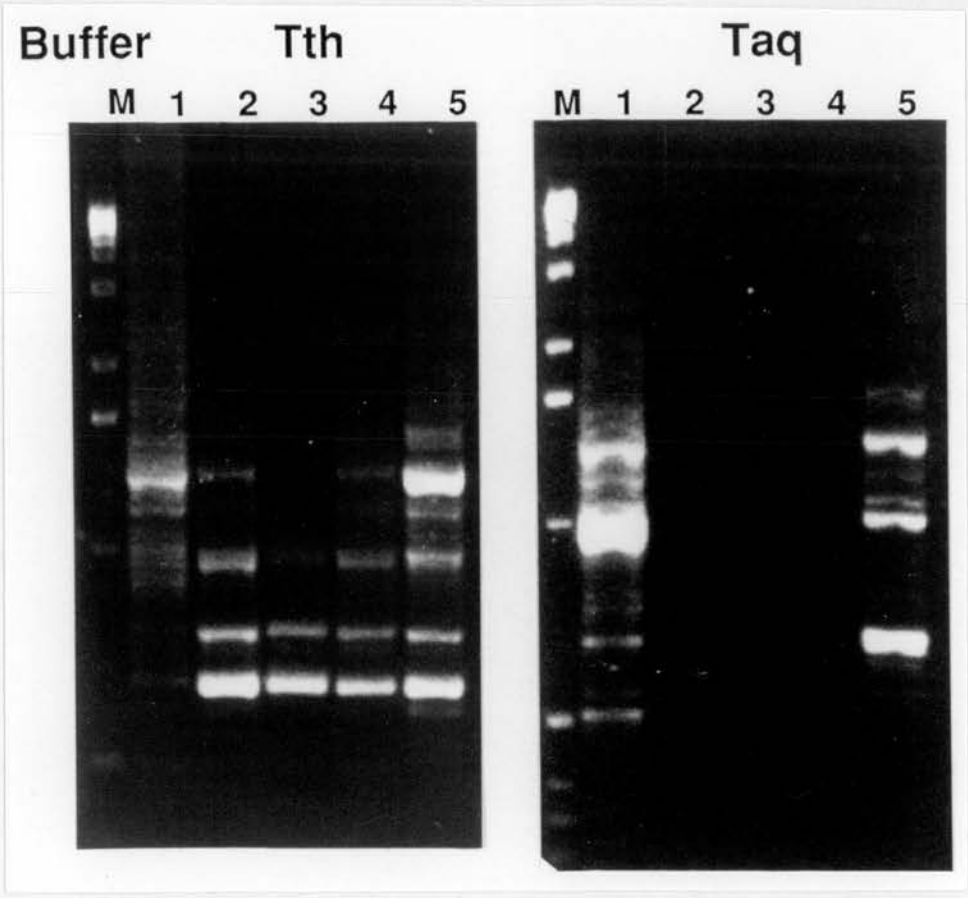


Figure 3.4 Assessment of buffer specificity

Here 35 cycle PCR was carried out using the Alu repeat primer 614, in combination with Tth enzyme and Tth or Taq buffer. The somatic cell hybrids E67.4 and WJX11.2 and their rodent cell backgrounds were used as template. Tth buffer appeared to be contaminated at source, and gave bands with rodent DNA and even in the total absence of template (3).

Key: M = 1kb ladder marker; 1 = E67.4; 2 = mouse genomic DNA; 3 = blank;4 = hamster genomic DNA; and 5 = WJX11.2.

3.2 Cloning Alu PCR products

Initially my aims were to devise a quick and simple ligation independent cloning (LIC) method for selectively cloning Alu PCR products. However, ultimately a new strategy for cloning Alu-PCR products was devised and proved to be more conveniently carried out. Thus, the aim of cloning Alu PCR products from a specific chromosomal region was achieved via a method called "Turbo cloning" which gave many more cloned products than LIC, which were also easier to map.

3.2.1 The principle of Ligation independent cloning

The pBluescribe vector was specially adapted to contain a 30 base pair GC rich linker with a central SmaI site (by Catherine Scahill, a summer student in or lab). Linearisation with SmaI results in a vector with GC rich ends which can be made single stranded using T4 DNA polymerase which in the absence of dGTP and dCTP and the presence of excess dATP and dTTP acts as a 3' - 5' exonuclease removing the G and C nucleotides. This activity terminates after 12bp where the enzyme reaches an A or T nucleotide. Alu-PCR fragments with a complementary 12bp tail are produced using a primer (C324 or C325) which has a 5' tail complementary to the vector. These primers comprise a slightly shortened version of the Alu primer 614, in addition to the GC rich tail, and were always used to reamplify fragments produced using 614 primer since a two stage PCR gave a greater complexity of products than using just C324/5 alone, probably due the fact that 614 is better matched to the consensus sequence. These products too could

be made single stranded with T4 DNA polymerase. Since the single stranded tails of vector and PCR product are complementary to each other, annealing in the absence of DNA ligase should result in cyclisation. Although the cyclised product contains four nicks (non-covalently linked ends), it may be used directly to transform competent *E.coli*. Once inside the bacterial cell the two nicks are repaired by *E.coli* DNA ligase and the recombinant plasmid is replicated as normal (Fig. 3.5).

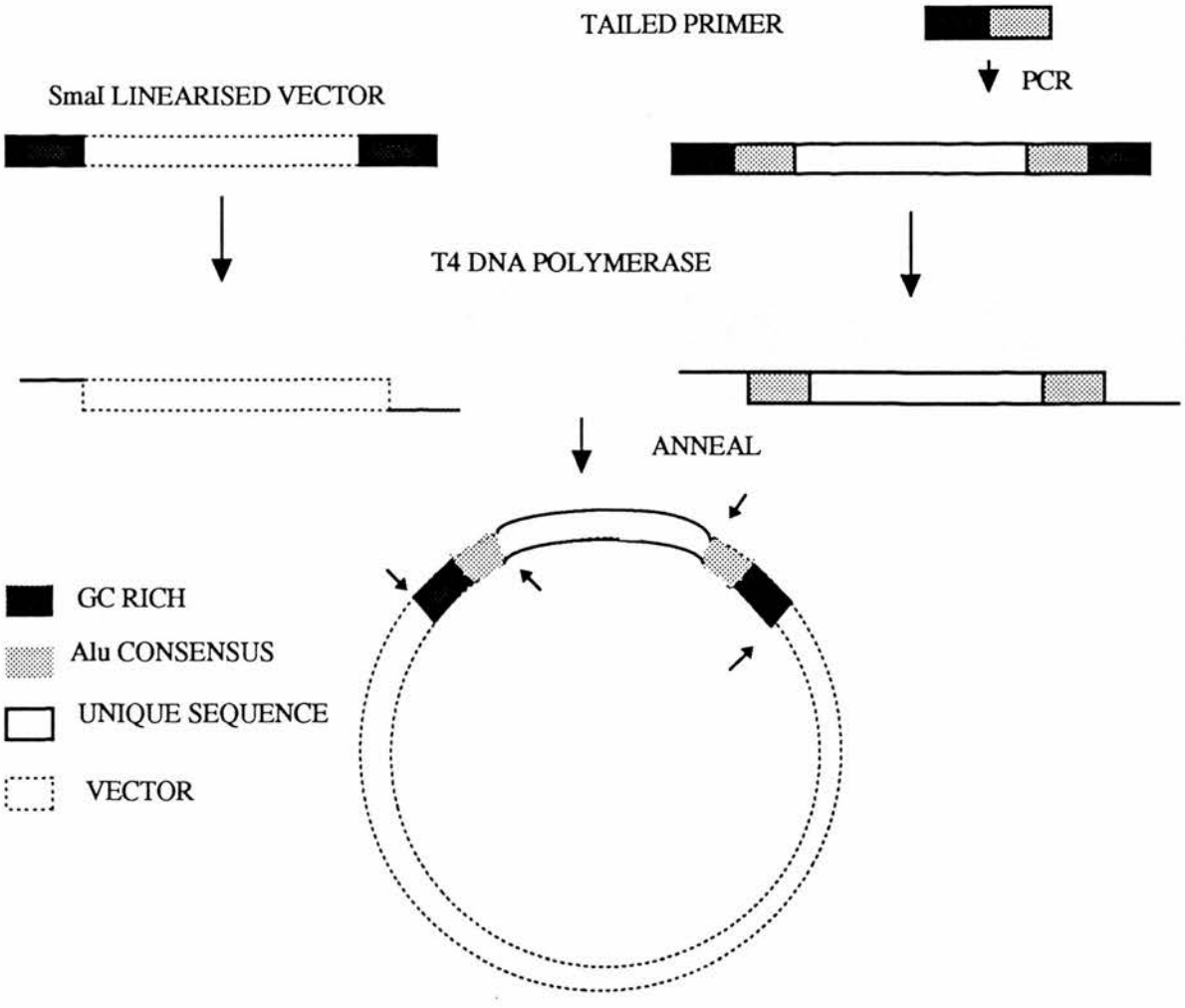


Figure 3.5 Ligation Independent Cloning

The overall scheme of ligation independent cloning is shown. Inter-Alu PCR products and LIC vectors are treated with T4 DNA polymerase to produce single stranded ends, which are then annealed in the absence of DNA ligase. The mixture can then be transformed as normal. Once inside the bacterial cell E.coli DNA ligase will repair the nicks indicated by arrows (↓).

3.2.2 Production of tailed products for Ligation Independent Cloning

3.2.2.1 Maximising the complexity

The hybrids E67.4 and WJX11.2 were chosen because mapping of markers to these showed that they retained shared as well as distinct regions of chromosome 11 (Fletcher et al., 1993), and as such were good substrates for testing the coincident sequence cloning of Alu-PCR products. PCR amplification reactions primed by 614 were each reamplified with a different tailed primer: E67.4 with C325 ; and WJX11.2 with C324 (these primers contain a shorter Alu consensus than 614 which makes them less specific, but here they are being used as nested primers on a resource of reduced complexity so it was thought unnecessary to alter the annealing temperature). In order to ensure the majority of products were tailed in this reaction it was necessary to remove any excess 614 primer from the first round of amplification. This was done by passing the whole 50µl 614 PCR reaction down a Nick™ column. The sample was now in 400µl. An eight fold range of volumes was reamplified (0.5, 1, 2,4µL) for 35 cycles at 65°C. All four volumes gave roughly the same amount of product so it was decided to proceed using 1µl template. In each case, however, the reaction was obviously non-exponential, so decreasing cycle PCR was carried out as before. Amplification was shown to be exponential between 29 and 34 cycles. Thus conditions had now been identified which produced the greatest complexity of Alu-PCR products.

3.2.3 Vector production

The vectors required for LIC had already been constructed in this laboratory by Catherine Scahill, and their ability to clone Alu PCR fragments tested. These early tests indicated that the technique had promise, but the cloning efficiency was not as high as had been expected.

The three vectors were constructed by insertion of 30bp linkers into the unique SmaI site of the pBluescribe vector (see Fig. 3.6). In each case the SmaI site was destroyed and a new SmaI site created in the middle of the linker flanked by a 12bp G/C sequence and terminated with a T/A only triplet. Each 30bp linker was produced by annealing together two complementary 30 base oligonucleotides, which were then blunt end cloned by ligating to SmaI linearised pBluescribe. Primers are readily available for the amplification and sequencing of inserts cloned into the Bluescribe polylinker (M13-20 and M13 reverse primers).

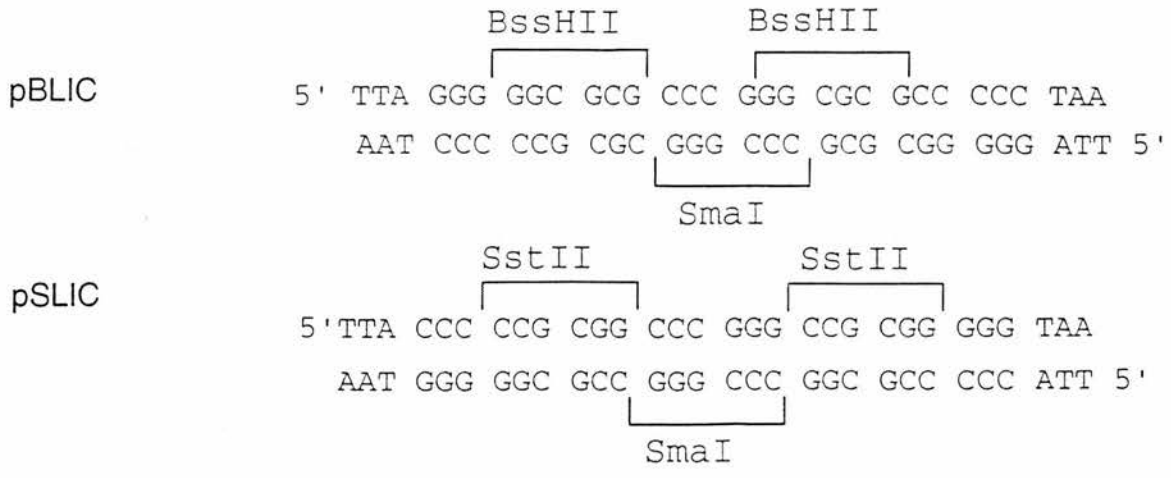


Figure 3.6 The oligonucleotides used to convert pBS into LIC vectors

The oligonucleotides used to convert pBS to LIC vectors are illustrated above, and Fig. 3.5 illustrates the way in which complementary single stranded tails were produced at the ends of the PCR products and the vector. The PCR primers are described in the previous table (2.4).

Both strands are shown in each case. The names of the vectors (pBLIC and pSLIC), and the positions of restriction sites are also indicated. are Ligation Independent Cloning vectors.

The obvious way of verifying the identity of such a short sequence would be to sequence it. However the sequence of the linkers meant that this was not as simple a process as it at first seemed. The polylinkers in pSLIC and pBLIC are entirely self complementary, in addition to which all of the linkers were by design GC rich. Sequencing was normally done using the Sequenase 2.0™ kit (USB). Sequences of dyad symmetry containing G and C residues can form hairpin loops and are often not fully denatured during electrophoresis, and this causes the regular pattern of migration of DNA fragments on the gel to be compressed to such an extent that the bands become difficult to read, and sequence information for such regions is lost. The ability of regions which are rich in G and C nucleotides to form secondary structure can be avoided by using dITPs (analogues which form weaker secondary structure) instead of dGTPs in the sequencing reaction. In cases of exceptional secondary structure, however, the enzyme is more likely to pause with dITPs than dGTPs, resulting in stops in the sequence. Presumably as a result of these factors the sequence of these vectors was only readable up to the SmaI site where the linker was inserted, after which point the sequence was compressed and there were many stops. It was thought that the problems with dyad symmetry might be avoided by first separating them by linearising the vector at the linker SmaI site and then sequencing out towards either end of the vector.

After several attempts, the linker at both ends of linearised pSLIC was sequenced and the deletion of a single base from one of the SstII sites was observed. pBLIC

proved less amenable to sequencing and only one end of the linearised vector was sequenced, but the sequence of this was correct.

The integrity of the vectors was further investigated by PCR amplification of the polylinker followed by restriction enzyme digestion to look for the presence of expected sites.

In each case 10pg of vector was amplified in a 50µl PCR reaction which was then divided into six 8µl aliquots. These were digested individually with each of the three enzymes and double digested with each of the three possible pairs. The expected and observed results are shown below.

Table 3.1 Expected and observed results on digest of vector polylinker.

The results of restriction enzyme digestion of the three vectors in comparison to the results expected from each if the required sites are present. The fragment sizes are shown in base pairs. The smaller fragments (<80bp) would not be observed on a 2% agarose gel.

Enzyme / Vector	BssHII	SmaI	SstII	BssHII + SmaI	BssHII + SstII	SmaI+ SstII	
pSLIC	**199**	**114/85**	**109/10/80**	**114/85**	**109/10/80**	**109/10/80**	**Expect**
	199	**114/85**	**109/90**	**114/85**	**109/90**	**109/10/80**	**Observe**
pBLIC	**111/6/82**	**114/85**	**199**	**111/3/3/82**	**111/6/82**	**114/85**	**Expect**
	111/82	**114/85**	**199**	**111/82**	**111/82**	**114/85**	**Observe**

These results showed that pBLIC did contain all the expected sites. pSLIC also gave the result predicted by sequencing : one of the SstII sites had been lost without a noticeable decrease in size, consistent with a single base deletion. The absence of digestion with SstII need not necessarily indicate the absence of a site since SstII is known to show preferential cleavage of some sites in the same substrate, but the fact that this conclusion agreed with the sequence made it highly likely that the site was missing. Thus pSLIC had to be remade to its original specification.

pSLIC was made by linker insertion. The efficiency of the blunt end ligation was very low, so these were repeated in the presence of PEG6000. This creates a

phenomena called macromolecular crowding and causes DNA molecules to be ligated together into concatemers (Upcroft and Healey 1989). This technique was a forerunner variant of "Turbo cloning" (Boyd 1993) and can be done using any vector. The vector was phosphatased and the oligonucleotides kinased to ensure that concatamers consisted predominantly of trimeric vector-insert-vector units. It was then possible to cut the vector at a unique site using an enzyme which produces cohesive ends (EcoRI), to give single vector-insert units which could then be ligated efficiently via the cohesive ends.

Initially an efficiency of 1×10^4 colonies/μg of vector was achieved, but sequencing of clones which were of the appropriate size, judged by PCR, showed that they contained two SstII sites although only one would be functional as the other had a single base deleted from the site just like the earlier version of pSLIC. After several failed attempts the ligation was repeated again, but this time with a new batch of oligonucleotides. Ten fold more colonies were produced. Colonies were screened for positives using ^{32}P labelled oligonucleotides on colony lifts of these plates, and 6 of the positive colonies analysed by restriction enzyme digestion of the PCR product. 3 gave the appropriate pattern, whereas 3 gave the pattern previously observed indicating that either the oligonucleotide stock was contaminated or some peculiarity of the oligonucleotide sequence caused it to rearrange.

The inserts were also characterised by sequencing. pSLIC was as expected. Both ends of pBLIC were sequenced this time. While still compressed, the end of the

linker previously non-amenable to sequencing was readable. One of the Ts from the AT triplet had been lost, but as this should have no affect on the usefulness of the vector it was ignored.

3.2.4 Testing the system

The vector structures were now correct so it was time to test their cloning ability. Conditions were developed which gave the highest cloning efficiency. The Alu-PCR products used are described in section 3.2.3.2. The conditions were developed using just the pSLIC vector and the appropriate C324 amplified insert. These optimal conditions were then compared to conventional cohesive end cloning and turbo cloning of the same Alu PCR products.

3.2.5 Optimisation of ligation independent cloning conditions

Initial attempts to clone purified insert into vector using the suggested conditions gave very few colonies on electroporation into XL-1 blue cells indicating that at least one of the stages was not optimal. Each stage, therefore, was altered in turn in an attempt to increase the yield of recombinants. In brief, Alu-PCR fragments reamplified with the appropriate primer were separated by gel electrophoresis and the DNA purified on a Spin-X column, to remove excess dGTP and dCTP which would prevent the proper formation of overhangs, then precipitated and resuspended in KGB buffer. These fragments and SmaI linearised plasmid were each incubated for 20 minutes at 37°C with 1 unit of T4 DNA polymerase, allowing the 3'-5' exonuclease activity of the enzyme to produce the single stranded

tails, in a buffer comprised of 1.5×KGB and 0.5mM dATP and dTTP. The enzyme was then heat inactivated by incubation at 68°C for 10 minutes. 40ng of T4 DNA polymerase treated Alu-PCR product was then combined with 10ng of vector and incubated at room temperature for at least 30 minutes to allow annealing prior to transformation.

Aslanidis and deJong (1990) used chemical transformation, not electroporation, to transform their cells. Something in the electroporation process could have been causing the insert to be lost, even if it had annealed properly to the vector in the first place. Hence the normal lab method of preparing competent DH5α cells was compared to the Hanahan method and the Calcium/Manganese-based method (CCMB) (Hanahan 1983), but here DH10B cells were used. *E.coli* strain DH10B has a similar genotype to DH5α but is capable of being made more transformation competent. The efficiencies of these three types of chemical transformations were compared to electroporation of XL1 blue cells. Both chemical transformation methods and electroporation gave the same low numbers of colonies when transformed with LIC reactions (CCMB gave none at all). Transformation of the cells with a uncut pSLIC plasmid, however, indicated that the cells prepared for electroporation were at least 10 fold more competent (100 fold for CCMB) than those made chemically competent. This indicated that something was occurring during electroporation of LIC products to cause the loss of insert annealed to plasmid, leading to a reduction in cloning efficiency. Perhaps the unligated tails were somehow separating under the high current applied. Whatever the reason,

the net effect was that the method of transformation used made no difference to the LIC efficiency.

It was thought that KGB may not have been optimal for the T4 DNA polymerase stage, so tests were carried out using the buffer recommended by Aslanidis and deJong (1990), but this failed to increase the efficiency. It was also possible that the enzyme was either degrading the DNA or was not sufficiently inactivated by incubation at 68°C. The occurrence of any degradation was tested by comparing, by transformation and on a gel, the nature of an insert after incubation for 5,10,15 and 20 minutes in the presence and absence of T4 DNA polymerase. On a gel the DNA appeared unchanged. On transformation the same number of colonies were produced by incubation at 5 as 15 minutes, but fewer at 20 minutes. In contrast, the number of background blue colonies decreased with increasing in time. Thus 15 minutes was chosen as the optimal length of time for incubation. The use of phenol/chloroform and ethanol precipitation to ensure the removal of T4 DNA polymerase did not increase the efficiency of the cloning process indicating that heat inactivation was sufficient.

The single stranded vector and insert ends all have a degree of self complementarity, are GC rich, and are at a relatively low concentration in the cloning reaction. All of these factors favour self annealing. In order to decrease the rate of self annealing, vector and insert were incubated at 60°C for 2 minutes then allowed to cool slowly to room temperature. This produced a further increase in cloning efficiency, and later work showed that incubation at 60°C for 2 minutes

could be carried out on a thermal cycler and cooled rapidly to 4°C without any decrease in efficiency.

Increasing the amount of insert to 120ng (12:1 weight ratio instead of 4:1) also produced a slight increase in efficiency. Thus the optimal conditions were: incubation for 15 minutes with T4 DNA polymerase, heat inactivation of enzyme for 10 minutes, combining vector and insert and annealing by brief incubation at 60°C and cooling rapidly to 4°C. All positive results were obtained with the pSLIC vector; pBLIC was never observed to contain any insert. This was thought to be due to the greater self complementarity of the vector ends, so pBLIC was abandoned at this stage.

The best way to get an idea of the efficiency of a novel cloning methodology is to compare it to existing methodologies. The methods chosen for comparison here were standard cohesive end cloning using the pSLIC vector, and Turbo cloning (Boyd 1993). It was decided to use Alu PCR to produce a library from the somatic cell hybrid WJX3.4. This hybrid was chosen because it covers the relatively poorly mapped region 11q12-13. Hence any products could be used to increase the density of markers in the region.

3.3 Comparison of three different cloning methods

3.3.1 Preparation of material for cloning

The inter-Alu PCR products were prepared by amplification of the hybrid WJX3.4 using the 614 primer using the conditions which gave the optimal number of

products (section 3.2.2.1). The products were electrophoresed on a 1.5% Seakem GTG agarose gel. A control track containing a small amount of sample was run between the preparative track and a track containing the 1Kb size marker (Gibco BRL). The control and size marker tracks could then be excised, stained with EtBr and realigned allowing size estimation in the preparative track. When the products had run the full length of the gel the bottom part was cut off to exclude everything under 350bp including any excess primers and small products. The gel was then electrophoresed in the opposite direction to concentrate the products. The concentrated products were cut out in three separate fractions (see Fig. 3.7). The DNA was purified from the agarose using a Spin-X column and resuspended in 10µl. These products were then cloned by three different methods as described.

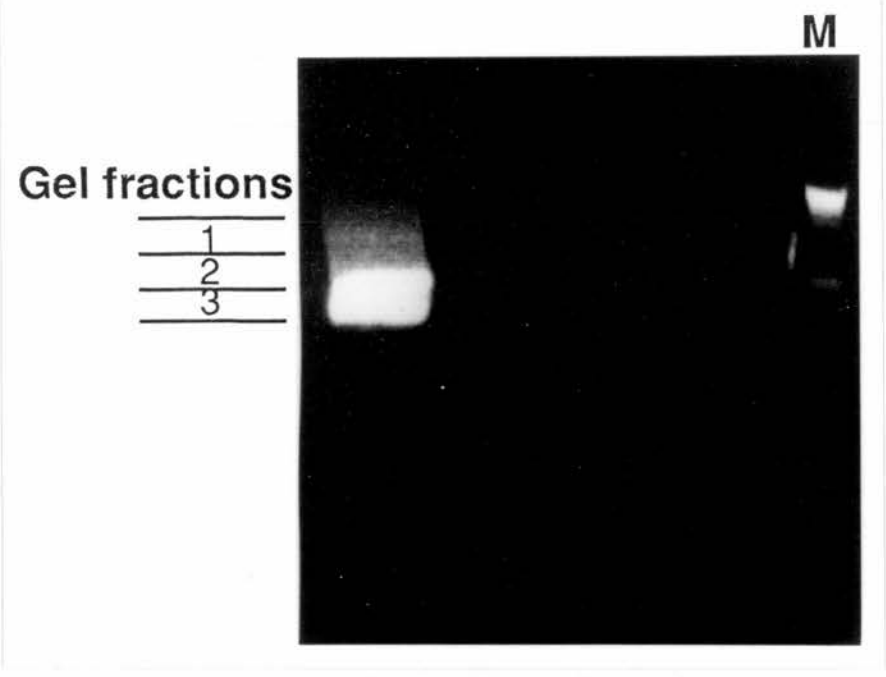


Figure 3.7 Concentration of WJX3.4 Alu-PCR products

Products from Alu-PCR amplification of WJX3.4 were run on a gel alongside the 1kb ladder marker in the absence of ethidium (see text). The three different size fractions are indicated

3.3.2 Ligation independent cloning

Either 1µl or 1µl of a 1/10 dilution of each fraction was reamplified using the primer C324. The products were electrophoresed on a 1.5% agarose gel to assess the range of fragments amplified (Fig. 3.8). The 1:10 dilutions of fractions 1 and 3 were chosen for cloning because unlike the others they did not contain over-amplified products. The products were purified using Wizard PCR columns (Promega). SmaI digested pSLIC vector was prepared by cutting out the linearised band from a 1.5% Seakem GTG gel. This vector DNA was then purified using a Spin-X column and resuspended in 10µl.

1µl aliquots of purified vector and C234 PCR products were electrophoresed adjacent to each other on a 1.5% agarose gel in order to assess their relative concentrations. It was estimated that there was four times as much PCR product as vector. Both the PCR product and linearised vector were treated with T4 DNA polymerase and combined in several different ratios (Table 3.2), and the ends annealed by incubating at 60°C for 10 minutes then immediately transferring to 4°C. The annealed mix was then electroporated into DH10B cells. The same proportion of the annealed mix was transformed in each case by altering the volume and precipitating where necessary. The amount of insert transformed ranged from 10-200ng.

3.3.3 Cohesive end cloning of PCR products

In cohesive end cloning a sequence containing an appropriate restriction site is added to the 5' end of the PCR primers. Vector and insert restriction digestion with the enzyme, then produces identical overhanging ends which can be ligated in the conventional manner. A C324 (a primer which contains an added SstII site) PCR of WJX3.4 was carried out on fractions as described above and again the product from the 1:10 dilutions of fractions 1 and 3 were chosen for cloning. Some products were present at higher concentrations, so prevent preferential cloning of these the volume was estimated which contained 400ng of the lowest concentration products and correspondingly 400ng of the highest concentration products, and these were cloned separately. The products were digested with SstII as was 100ng of vector. Products and vector digests were then combined, extracted with phenol:chloroform, ethanol precipitated and resuspended in ligation mix (1µl Ligase, 1µl 10 × Ligase buffer: Boehringer Mannheim) and incubated overnight at 16°C. 1µl of this ligation mix was then electroporated into DH10B cells. This corresponds to 40ng of insert.

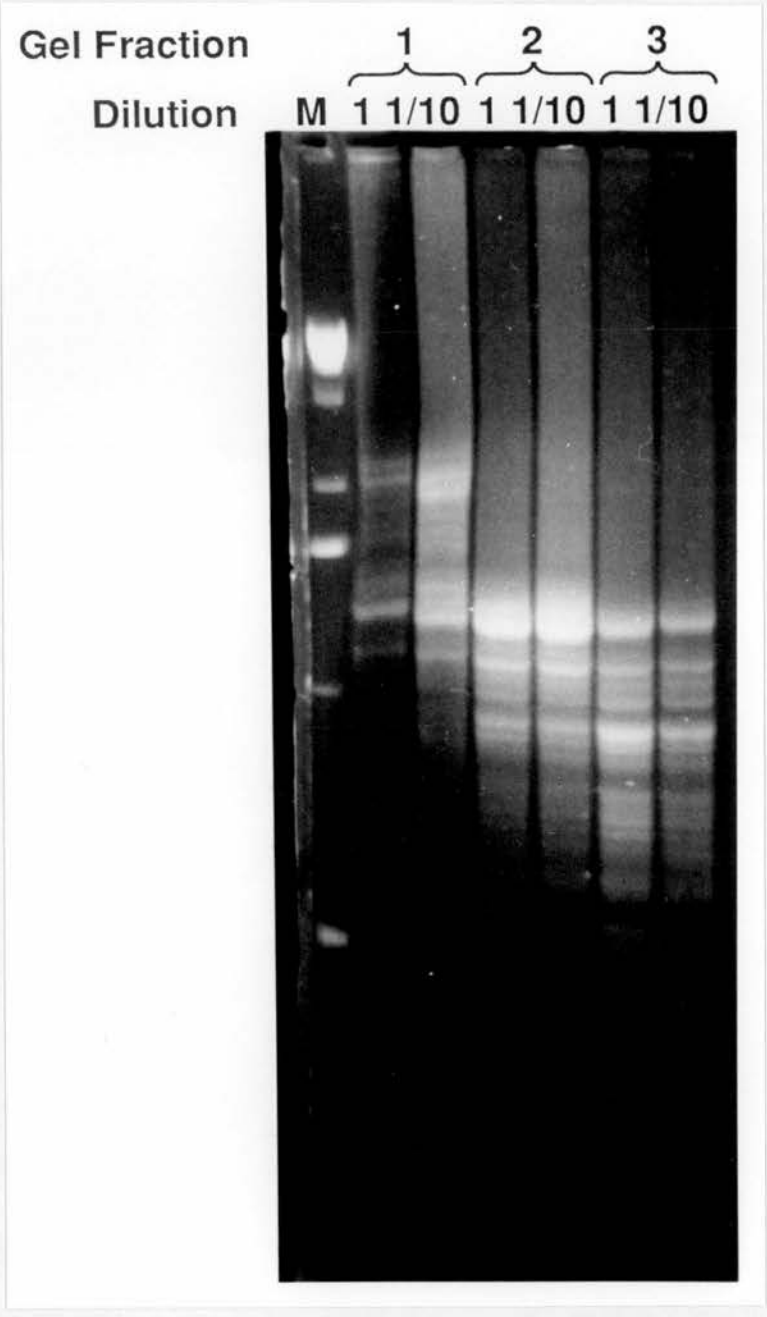


Figure 3.8 Tailed amplification products from gel fractions

Either 1µl or 1µl of a 1:10 dilution of each gel fraction from the concentrated WJX3.4 Alu-PCR products (Fig. 3.7) was reamplified using the tailed primer C324. Comparison here on a 1.5% agarose gel indicates that the 1:10 dilutions of fractions 1 and 3 together contain the full range of equally amplified products hence these were chosen for cloning. Marker (M) = 1Kb ladder.

3.3.4 Turbo cloning

Turbo cloning is a fast, efficient method for cloning PCR products and other blunt-ended DNA fragments into plasmids (Boyd 1993). Fragments are cloned into a vector containing the 34bp site-specific recombination sequence *lox* in the presence of the cognate Cre recombinase system of bacteriophage P1. The same three fractions of the WJX3.4/614 PCR were used as in the previous two methods, but here only the 1:10 solutions were amplified. The primer used was Alu_{18}, which first had to be kinased. This primer is homologous to the polyA end of the Alu consensus sequence, hence products contain negligible Alu sequence within them, and are therefore more useful as probes. Fraction one was selected to clone as this produced even intensity products across the size range (Fig. 3.9). The ends of the products were made blunt ended using T4 polymerase, and purified using a Wizard PCR column, before combining with dephosphorylated vector.

The vector needed to be dephosphorylated to reduce the background of regenerated vector, and prevent production of long concatemers with multiple *lox* sites. Several micrograms of vector (pBSlox : bluescribe containing a 34bp lox site) were prepared as a stock. Typically the vector was cleaved by SmaI and dephosphorylated with Calf intestinal phosphatase (Promega). To reduce background of uncut vector the linearised DNA was purified by excision from an agarose gel, and extracted on a Spin-X column. In this case two aliquots of purified SmaI - linearised dephosphorylated vector were kindly provided by Chris Boyd. These were called batch 1 and 2 in Table 3.2.

The concentrations of purified vector and PCR products were assessed on an agarose gel before combining in a 10:1 molar ratio (100ng vector:30ng insert) in a ligation mix containing 0.4 units of T4 DNA ligase (Boehringer), and a final concentration of 15% PEG 6000 (BDH). This was incubated for 15 minutes at room temperature, then the ligase was inactivated by incubation at 75°C for 5 minutes. The reaction was then diluted to below the level of PEG concentration where macromolecular crowding takes place by adding four volumes of M buffer (Boehringer medium salt restriction enzyme buffer) containing 0.1-0.3µg of Cre protein (NEN) which causes circularisation of vector insert trimers (vector-insert-vector). After incubation at 30°C for 30 minutes, Cre activity was destroyed by incubation at 75°C for 10 minutes. 5µl of the cloning reaction was desalted by drop dialysis (section 2.7.4), and 2µl of this corresponding to an estimated 750pg of insert was electroporated into DH10B cells.

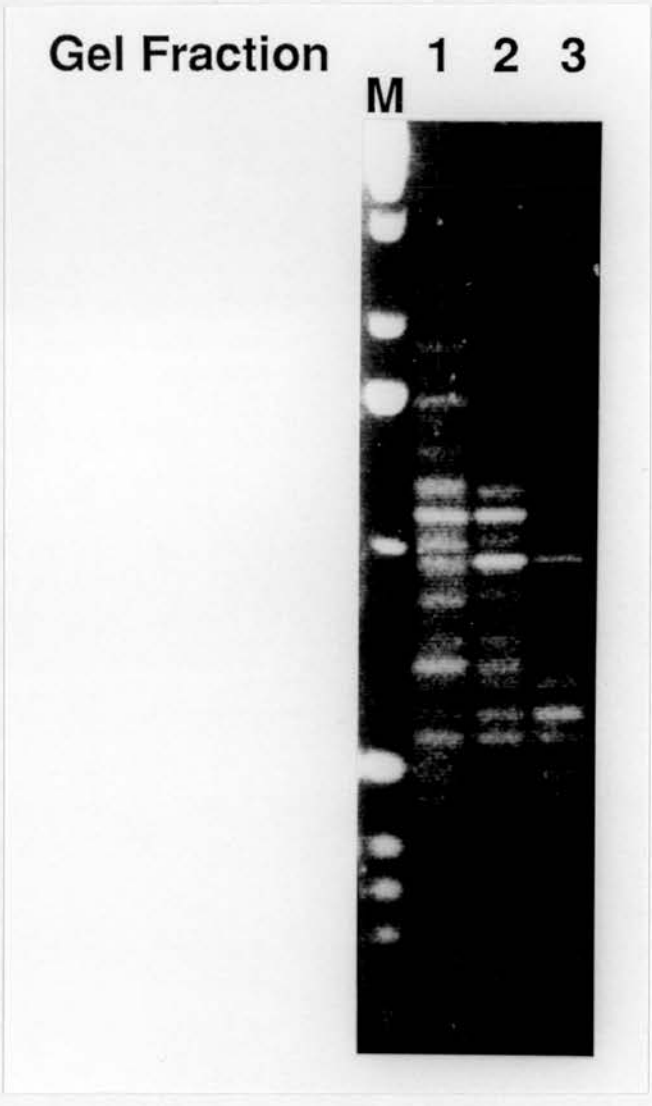


Figure 3.9 Reamplification of gel fractions with primer Alu_{18}

Shown here are the products from the reamplification of the 1:10 dilutions of each gel fraction from the concentrated WJX3.4 Alu-PCR products (Fig. 3.7). The Alu repeat end primer Alu_{18} was used in the reamplification. Fraction one gave the greatest range of products and hence was chosen for cloning. Marker (M) = 1kb ladder.

3.3.5 Tabulation and discussion of results

Table 3.2 shows the number of recombinants achieved by each method. Cohesive end cloning, although it produced the largest number of overall recombinants, gave the highest background of non-recombinant colonies (12-26%). Turbo cloning on the other hand also gave a large number of recombinants but a low background of non-recombinant colonies (2-5%). LIC gave a fairly low background of non-recombinant colonies (5-18%), but the actual number of recombinants was also fairly low due to the low efficiency of the cloning methodology (see Table 3.2).

Table 3.2 .1 and 3.2.2 The number of colonies achieved per ml of cells plated

The numbers of recombinants and background blue colonies achieved with each of the different cloning methodologies.

Cloning method	LIC 1:1[a]	LIC 2:1	LIC 3:1	LIC 5:1	LIC 10:1	LIC 20:1
No. of whites[c]	0	$1.8x10^3$	$1.3x10^4$	$5.8x10^3$	880	640
No. of blues[c]	0	200	$1.4x10^3$	318	120	60
% whites	-	90	90	95	88	91

Cloning method	Turbo cloning batch I	Turbo cloning batch II	Cohesive min[b]	Cohesive max[b]
No. of whites[c]	$2x10^6$	$3x10^6$	$2.8x10^5$	$4.6x10^5$
No. of blues[c]	$3.5x10^4$	$1.6x10^5$	$1x10^5$	$6x10^4$
% whites	98	95	74	88

[a] ratio of insert to vector; [b] here approximately 400ng of the lowest/highest concentration products were cloned; [c] the number of white(recombinant)/blue(non-recombinant) colonies per µg of insert DNA.

The next stage was to analyse these colonies in order to determine how many were real recombinants (i.e. contained an insert), and the range of fragments cloned. 96 colonies were picked from each ligation experiment and stored as glycerol stocks in microtitre plates, which could then be recovered by streaking out on an agar plate as required. The size of insert was assessed by carrying out PCR amplification using universal primers. Fig. 3.10 shows the results from the universal PCRs, and these results have been tabulated (Table 3.3) as has the overall efficiency of each cloning method.

In this case, turbo cloning is more efficient than cohesive cloning. Previous studies have shown the efficiency of the two to be equivalent (Boyd 1993). However, the ten fold decrease in the efficiency of cohesive end cloning observed here is likely to be due to the nature of the vector and insert ends.

Colonies obtained by LIC amplified poorly and very few of the clones appeared to contain any insert. PCRs from amplification of colonies obtained by cohesive end cloning contained a lot of primer dimers. Many of the latter clones appeared to be the same size indicating selection for a specific product may have occurred. However no size selection was observed in C324 PCR products cloned in pSLIC. The ability to reamplify inserts with Alu_{18} is important if they are to be used as markers since this removes the need for competition and stringent washing conditions in hybridisation. Neither cohesive clones nor LIC clones amplified well with this primer, even though Alu_{18} is exactly matched to the consensus. In general C324 PCR products cloned into pSLIC were difficult to amplify. The problem most likely results from the flanking GC rich sequences preventing the primer from

annealing, possibly by forming some sort of hairpin structure.

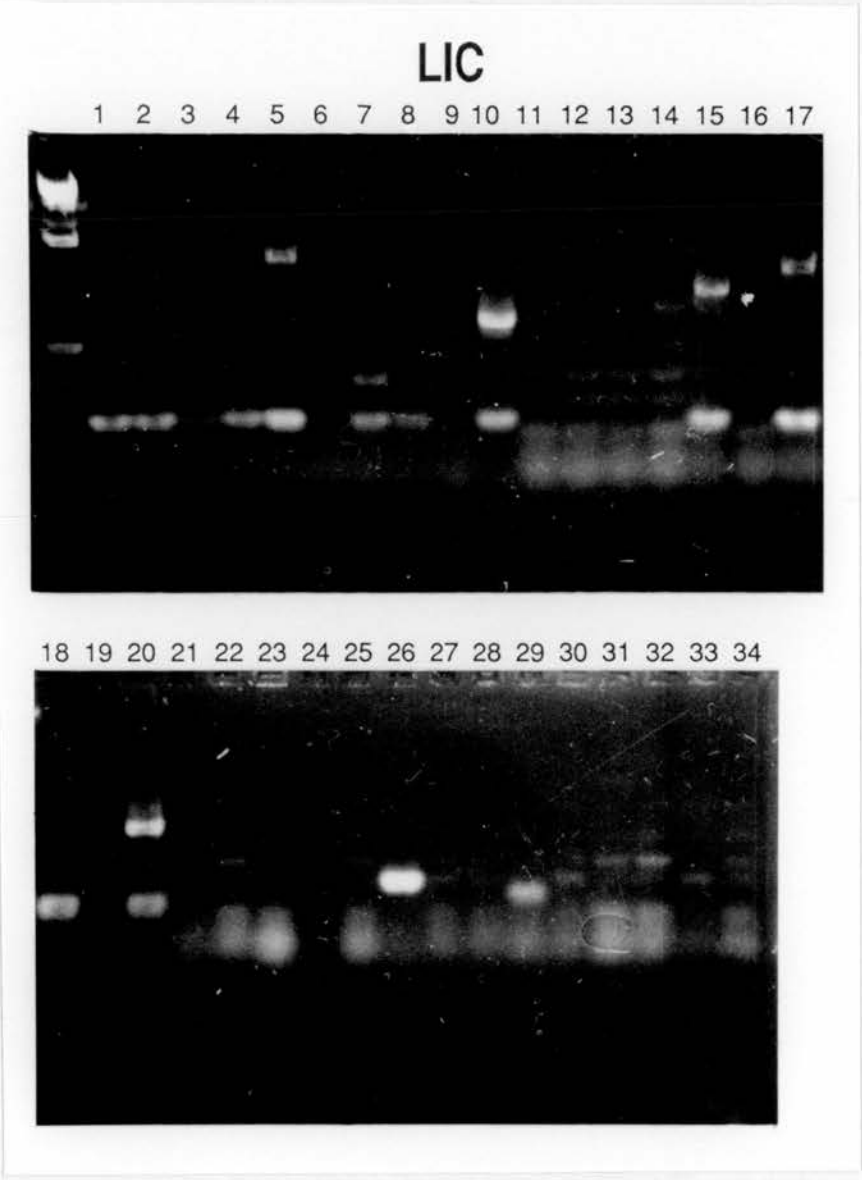
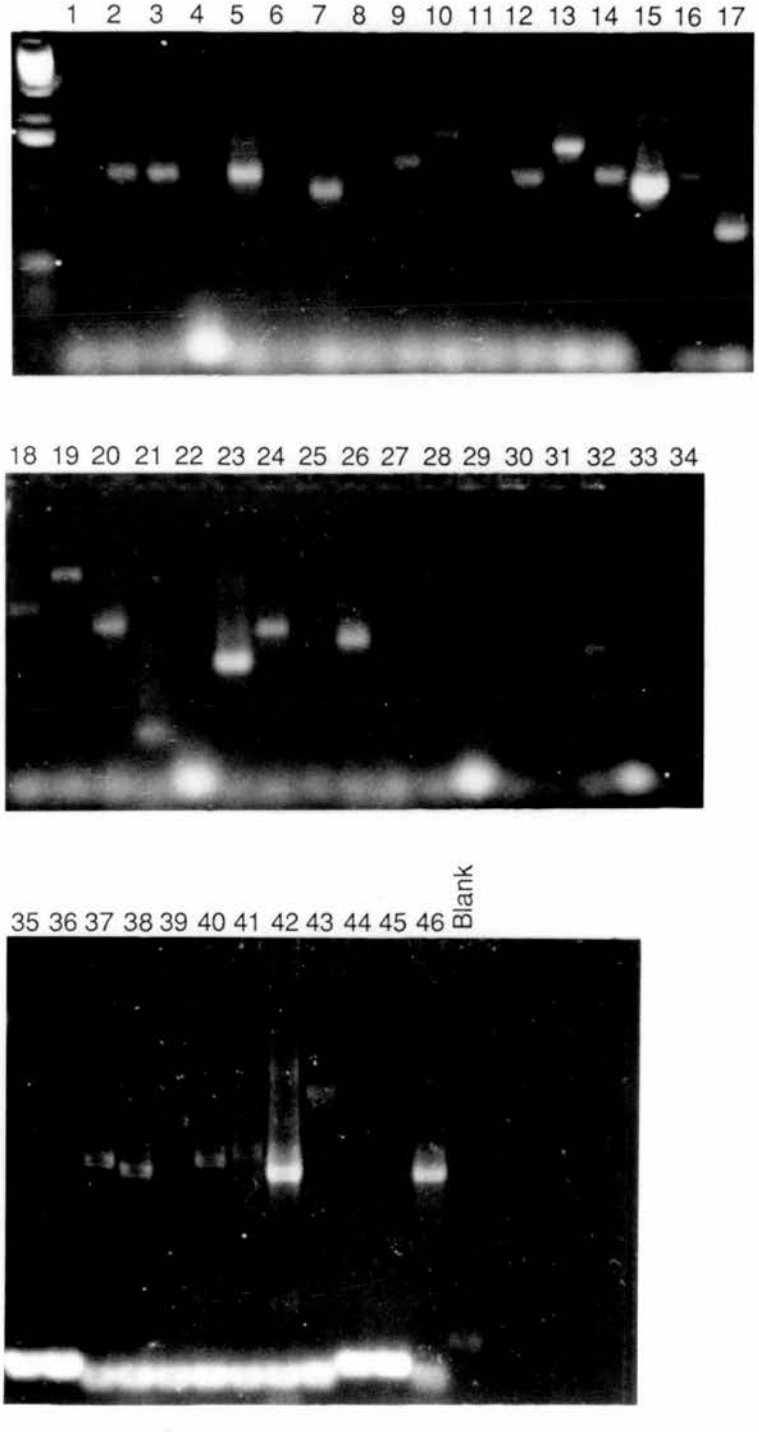


Figure 3.10 Universal PCRs of Inter-Alu PCR products

Clones were amplified using the M13 universal primers and run on a 1.5% agarose gel alongside the 1kb ladder marker (M). The flanking GC rich sequences appear to be hindering amplification by Taq polymerase of inserts cloned into the pSLIC vector (cohesive cloning and LIC), since theoretically the same range of products had been cloned into pBSlox by turbo cloning and these amplified well.

Cohesive


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34
35 36 37 38 39 40 41 42 43 44 45 46 Blank

Turbocloning

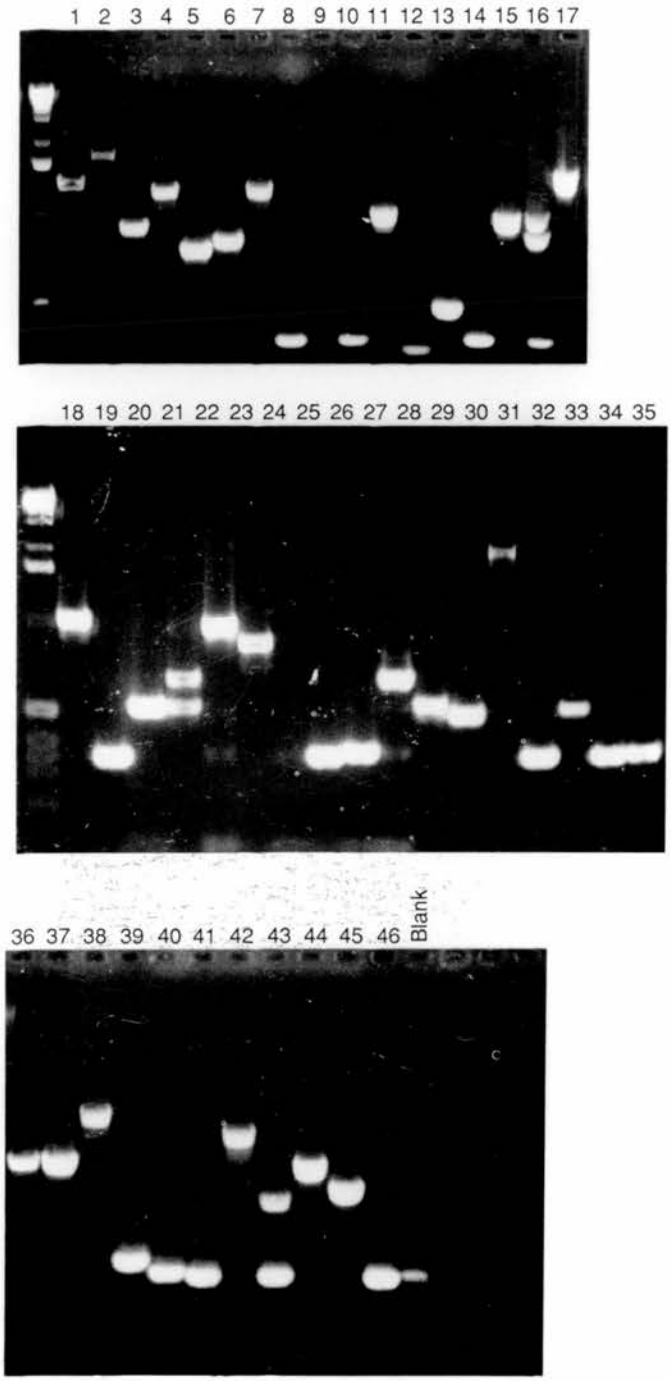


Table 3.3 Efficiency and range of insert size

Comparison of the cloning efficiency and the range of insert size for the three different cloning methodologies.

Cloning method	Efficiency colonies/μg insert	Size range base pairs
LIC	$660\text{-}1.2\times10^4$	300-1600
Cohesive	2.8×10^5-4.6×10^5	800-1600[a]
Turbo cloning	2×10^6-3.1×10^6	300-2000

[a] many of the clones obtained by cohesive end cloning appeared to be of similar size

The next step was to assess the repeated sequence content of the clones. Alu repeats are not the only repeat present in the human genome and others can be equally problematical to mapping. Human repeat content was assessed by hybridising ^{32}P labeled Cot1 DNA to blots of the Alu_{18} PCRs from all three cloning methods. The results of this are shown in table 3.4.

Table 3.4 Human repeated sequence content

The results of PCR amplification and CotI hybridisation are shown. A total of 46 clones were analysed in each case. The proportion of clones which contain repetitive DNA appears to be about the same in each case.

Cloning method	No. of clones which failed to amplify	Intensity of signal on CotI hybridisation				NO. suitable as probes[a]
		None	Low	Med[m]	High	
LIC	13	2	2	6	23	4
Cohesive	3	8	12	5	18	16[b]
Turbo cloning	0	24	4	5	12	16

[a] PCR products which were greater than 300bp and had either no or low repeat content

[b] only 6 of these were reamplifiable with Alu_{1e} primer, so could be used as probes;

3.3.6 CA repeats

Microsatellite markers, particularly CA repeat elements, have allowed the construction of genetic maps for a number of genomes as well as human (Gyapay et al., 1994). Such repeats appear to be homogeneously distributed along the genome (Weber et al., 1990), exhibit length polymorphisms, and can be typed by PCR (Weber et al., 1989). It was decided to screen the chromosome 11q12-13 libraries for CA repeats containing clones which would then be useful in refining linkage data for disorders associated with the region.

CA repeat containing clones were identified by hybridisation of an oligonucleotide probe consisting of 12 CA dinucleotides. The probe was labelled by kinase end labelling and hybridised to Southern blots of universal PCRs of 96 clones from each cloning method. In this way 7 CA repeats were identified: 3 produced by cohesive end cloning and 4 by turbo cloning. Plasmid template from each clone was prepared by PEG precipitation and sequenced by ABI cycle sequencing. The sequence obtained was then used to design primers for PCR amplification across the repeat region. Two clones contained a repeat of less than 12 nucleotides, which has been shown to be the threshold for repeats to be polymorphic (Weber et al., 1990) hence primers were not designed for these. In three cases the repeat lay very close to the end of the Alu repeat sequence, here primers were designed to be used in conjunction with the Alu_{18} primer. The optimised PCR conditions and primer sequences are shown in section 2.4.4.5.

3.3.7 Conclusions

Three different methods were used to clone Alu-PCR products which had been produced from the somatic cell hybrid WJX3.4. All three involved a two stage PCR and various vector and insert modifications prior to the cloning step. The simplest and most rapidly completed procedure was ligation independent cloning. Here the steps between the secondary PCR and transformation took under 2 hours to complete. The next fastest was turbo cloning which, when kinasing of the primer was taken into account, took approximately 5 hours to complete (although the actual ligation process takes under 2 hours). Cohesive end cloning took the

longest, requiring an overnight ligation step. Overall the most efficient method was turbo cloning, which gave the largest range of readily reamplifiable products. The small numbers of colonies produced by cloning in pSLIC indicated that in a coincident sequence cloning experiment where the amount of clonable DNA would be reduced this method of LIC would not be efficient enough. Attempts could be made to further optimise the conditions for LIC, but as these conditions become more complicated the method loses its advantage in being simple and rapid. It was therefore decided not to pursue LIC further and instead to assess the use of turbocloned Inter-Alu PCR DNA as probes.

Five CA repeat containing clones were isolated from the 11q12-13 libraries, primers and conditions allowing their PCR amplification were designed, but their heterozygosity for the CEPH families remains to be determined (Weissenbach et al., 1992).

4. Mapping a repeat free resource

4.1 Mapping repeat free clones to a somatic cell hybrid panel

In order to map the inter-Alu PCR products relative to each other, a resource was required which allowed the region of origin to be subdivided. The resource used to produce the library of inter-Alu DNA clones was the somatic cell hybrid WJX3.4, which previous studies (Fletcher et al., 1993) had shown to contain a single contiguous region from human chromosome 11p11.2-11q13.5. A set of hybrid cell lines, which retained different subfragments of chromosome 11q was also available. These hybrid cell lines had previously been used to successfully map a number of ESTs (Slorach et al., 1995) and to refine the localisation of established markers (Evans et al., 1995). This somatic cell hybrid mapping panel, then could be used to orient these new inter-Alu PCR markers relative to previously mapped markers, and possibly identify new mapping intervals..

Basically the somatic cell mapping panel consisted of six fragmentation hybrids. Four were X-irradiation hybrids, WJX3.4 (source), WJX5.4, WJX7.4 and WJX11.2 which contain fragments of chromosome 11 in the background of a Chinese hamster ovary cell line, WG3H (Fletcher et al., 1993). The other two were independent HRAS-1 selected chromosome mediated gene transformants (CMGT), E67.1 and E67.4 which carry fragments of chromosome 11, in the background of a murine cell line C127(Porteous et al., 1986). In addition, two

translocation cell lines were included : MAR1 which carries the derived 11 translocation chromosome (Fletcher et al., 1993); and CF52 which has the well characterised t(11;16)(q13;p11) translocation chromosome as its sole human component in the background of a murine cell line, A9 (Koeffler et al., 1981)). Cell lines WG3H (hamster background) and RAG (murine background) were used as negative controls. J1Cl4, a hybrid cell line which contains an intact chromosome 11 as its sole human component (Jones et al., 1984) and was the chromosome donor for the WJX series, and F4 (human female placental DNA) were used as positive controls.

Turboclones which were either repeat free or had shown a very low level of repeat when screened with CotI human DNA and gave a product of larger than 300bp when amplified with Alu 18 primer (section 3.3.5), were selected for mapping on the somatic cell hybrid panel. A probe was obtained by amplifying each clone with Alu18 and labelling the gel purified product in a Random Prime reaction. Those clones which had shown a low level of repeat in the CotI human DNA pre-screen were suppressed with sonicated salmon sperm and sonicated human DNA prior to hybridisation. Control and hybrid cell line DNAs were digested with EcoRI, and 5μg of each electrophoresed on a 0.8% agarose gel, which was then Southern blotted and hybridised with these labelled markers. The blots were washed to various levels of stringency depending on the probe, ranging from 0.1-1×SSC.

Initially, 18 turboclones were isolated which were suitable for mapping on the somatic cell hybrid panel. At a later date another 46 turboclones were amplified

with Alu18 and screened with Cot1 human DNA. Thus a further 26 clones which were suitable for mapping were obtained. In total an attempt was made to map 30 of a possible 44 clones. Of the first set of clones 5 of 12 tried were successfully mapped. Mapping the next set of 18 clones gave a similar success rate: 11 were successfully mapped. The mapping results are summarised in Table 4.1.

Table 4.1 Mapping turbo clones on the somatic cell hybrid panel

Summary of the results from mapping Alu_{18} PCR products from the WJX3.4 library on a panel of chromosome 11q somatic cell hybrids. Suitable markers were those which gave a low or negative result when probed with Cotl DNA, and gave a product larger than 300bp on amplification with Alu_{18} primer.

	Set 1	Set 2	Total	Percentage of probes mapped
No.of suitable markers	18	26	44	NA *
No. used as probes	12	18	30	NA*
No. mapped	5	11	16	53
No. giving ambiguous mapping result	1	1	2	7
No. hybridising to repeats	5	6	11	37
No. which were plasmid contaminants	1	0	1	3

* NA = not applicable

Even after prescreening with Cot1, 40% of clones still gave a hybridisation smear on the somatic cell hybrid panel. It was realised in retrospect that a proportion of these may have been due to poor transfer of DNA during the Southern blotting step, as many of those which were mapped gave positive results on the same blot, indicating that lack of signal may have been due to poor quality of the actual blot.

In some cases, however, a blot which had already given a good result with a previous probe produced a smear on re-hybridisation. This had to be due to the individual clone and indicated that some sort of repeat sequence was present in the probe which is not effectively stripped by Cot1 DNA. This may have been due to MERs or some other type of repeat (Kaplan et al., 1991). MERs belong to a family of medium reiteration repeats, and a number of members have been shown to lie within 500bp of Alu repeats.

Markers which gave an unambiguous result were ordered on the panel relative to markers previously positioned. This was done in such a way as to minimise the number of independent blocks of human DNA in each hybrid. The results are shown in Table 4.2. In all, 21 distinct patterns of hybridisation identifying 21 different mapping intervals were discerned, giving an average resolution less than 1.5Mbp.

Table4. 2 Location of turbo clones on the somatic cell hybrid panel

Turboclones (shaded) have been positioned relative to known genes, Expressed Sequence Tags (ESTs), and polymorphic markers (D11S-). The location of ET58 is also shown. This is an exon trap product of the myosin type VII related protein, mutations in which have recently been shown to cause Usher Syndrome Type IB (see Discussion). Some clones gave more than one band and these have been given the suffix A-D. Brackets denote the individual mapping intervals. The marker order in these intervals is non contiguous.

+ = band observed in this hybrid

- = no band observed in this hybrid

ND = this hybrid was not included in hybridisations with this particular marker.

LOCUS	CF 52	WJX 3.4	WJX 7.4	WJX 5.4	WJX 11.2	E67.1	E67.4	MAR1
TF11	-	+	+	-	-	+	+	+
TA6	-	+	+	-	-	+	+	+
TC9	-	+	+	-	-	+	+	+
EST 220	-	+	+	-	-	+	-	+
PGA	-	+	+	-	-	+	-	ND
TA12B	-	+	+	-	-	+	-	+
TA12A	-	+	+	-	-	-	-	+
TE5	-	+	+	-	-	-	-	+
TA8A	-	+	+	-	-	-	-	+
TA8B	-	+	+	-	-	-	-	+
TB11	-	+	+	-	-	-	-	+
TE7	-	+	+	-	-	-	-	+
TH8	-	+	+	-	-	-	-	+
TC10	-	+	+	-	-	-	-	+
ROM1	-	+	+	-	-	-	-	ND
MDU1	-	+	+	-	-	-	-	ND
TD1	-	+	+	-	-	-	+	+
TG1A	-	+	+	-	-	-	+	+
TF9B	-	+	+	-	-	-	+	+
TB10	-	+	-	-	-	-	+	+
EST 218	-	+	-	-	-	-	+	+
EST 1878	-	+	-	-	-	-	+	+
EST 2002	-	+	-	-	-	-	+	+
GSTP1	-	+	-	-	-	+	+	ND
TF7	-	+	-	-	-	+	+	+
FGF4	-	+	-	-	-	-	-	ND
D11S527	+	+	-	-	-	-	-	ND
TF9A	+	+	+	-	-	-	-	+
EST 294	+	+	+	-	-	-	-	+
TH1	+	+	+	-	-	+	-	+
D11S533	+	+	+	-	-	-	+	ND
OMP	+	+	+	-	-	-	+	ND
ET58	+	+	+	-	-	-	+	+
TG1B	+	+	+	-	-	-	+	+
EST 981	+	+	+	-	-	+	+	+
EST 2145	+	-	+	-	-	-	+	+
D11S901	+	-	+	-	-	-	+	ND
TYR	+	-	+	-	+	-	+	ND
D11S873	+	-	+	-	+	-	+	ND
D11S388	+	-	-	-	+	-	-	ND
EST 16	+	-	-	-	+	-	+	-
CLG	+	-	-	+	+	-	+	ND
STMY	+	-	-	+	+	-	+	ND
D11S385	+	-	-	+	+	-	+	ND
TF9C	+	-	-	+	+	-	+	ND
NCAM	+	-	+	-	+	-	-	ND
DRD2	+	-	+	-	+	-	+	ND
D11S351	+	-	-	+	+	-	-	ND
THY1	+	-	-	+	+	-	-	ND
TF9D	+	-	-	+	+	-	-	-
EST 116	+	-	-	+	+	-	-	-
EST 652	+	-	-	+	+	-	-	-
MIC9	+	-	-	+	+	-	-	ND

A number of clones mapped to previously defined intervals: TC10, TA8, TE5, TB11, TE7, and TH8 mapped to the same interval as ROM1 and MDU1; TB10 mapped to the same interval as the ESTs 218, 1878, 2002; TF7 mapped to the same interval as GSTP1 (figure 4.1); and TG1B mapped to the same interval as OMP. This provides a substantial increase in the density of markers in these intervals.

Four clones gave more than one band when hybridised to the panel. In two cases this appeared to be due to the insert containing an EcoRI site, as the bands were always present in the same hybrid. TA8 (figure 4.2) showed this type of pattern as did TA12. TA12, however, gave only one band in E67.1, indicating that it may in fact bridge the breakpoint in this hybrid. The presence of an EcoRI site in each of these clones was confirmed by restriction enzyme digestion. The two bands from TG1 and four from TF9 (figure 4.3) all mapped to different intervals, several of which were not present in WJX3.4 suggesting that the hybridisation was to a related sequence. TG1 gave a strong background smear on human genomic DNA indicating that it contained a repetitive sequence. TF9, on the other hand, gave about twelve distinct bands on human genomic DNA. A subset of four of these was present in J1Cl4, indicating that the clone contained a low copy number repeat.

TH1 defined a new E67.1 sub-fragment between EST294 and D11S533. Three clones (TF11, TA6, and TC9) gave the same result on the hybrid panel and suggested the identification of a new E67.4 sub-fragment above EST200.

Two clones (TD7 and TE6) gave ambiguous results. In order to position them on the hybrid panel more than one novel sub-fragment would have to be introduced. The status of these results awaits confirmation by other markers.

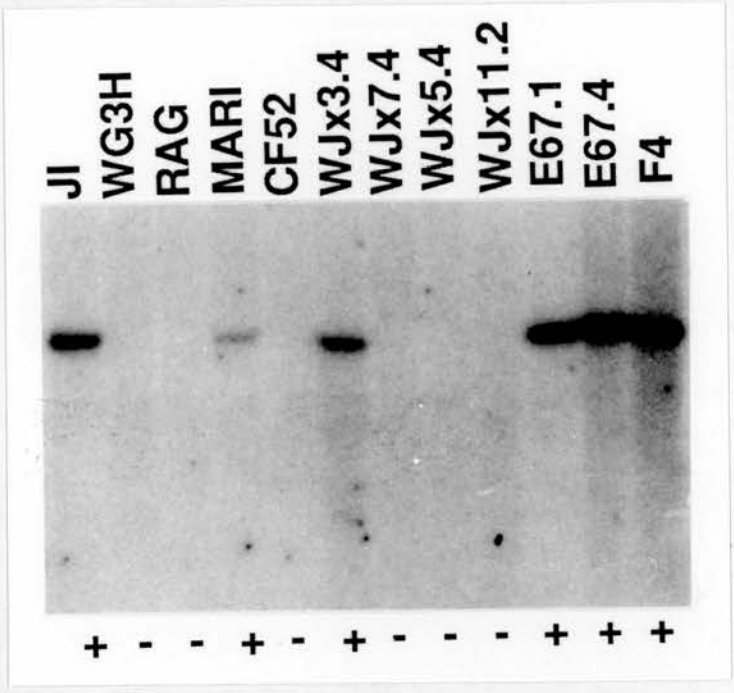


Figure 4.1 Mapping TF7 on the somatic cell hybrid panel

Clone TF7 was amplified with the Alu_{18} primer and the gel purified product labelled by random priming. This was then hybridised to a Southern blot of the chromosome 11 somatic cell hybrid mapping panel. This allowed the clone to be mapped to the same interval as GSTPI (Table 3.2). The hybrids are described in more detail in the text: F4 = total human genomic DNA; J1 = chromosome 11 only hybrid; WG3H and RAG = hamster and mouse background respectively; E67.1 and E67.4 = CMGT hybrids; WJX- = X irradiation hybrids.

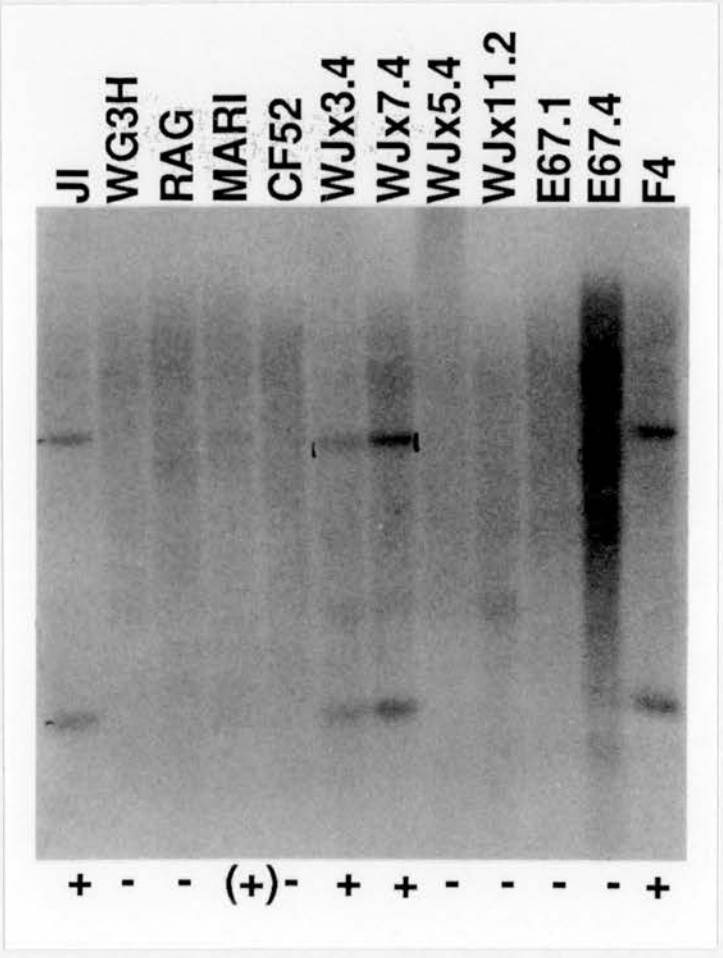


Figure 4.2 Mapping TA8 on the somatic cell hybrid panel

Clone TA8 was mapped in the same way as TF7 (Fig. 4.1). This clone gave two different bands, both of which were positive for the same hybrids (Table 4.2). This was later shown to be because the clone encompassed an EcoRI site. The hybrids are described in more detail in the text: F4 = total human genomic DNA; J1 = chromosome 11 only hybrid; WG3H and RAG = hamster and mouse background respectively; E67.1 and E67.4 = CMGT hybrids; WJX- = X irradiation hybrids.

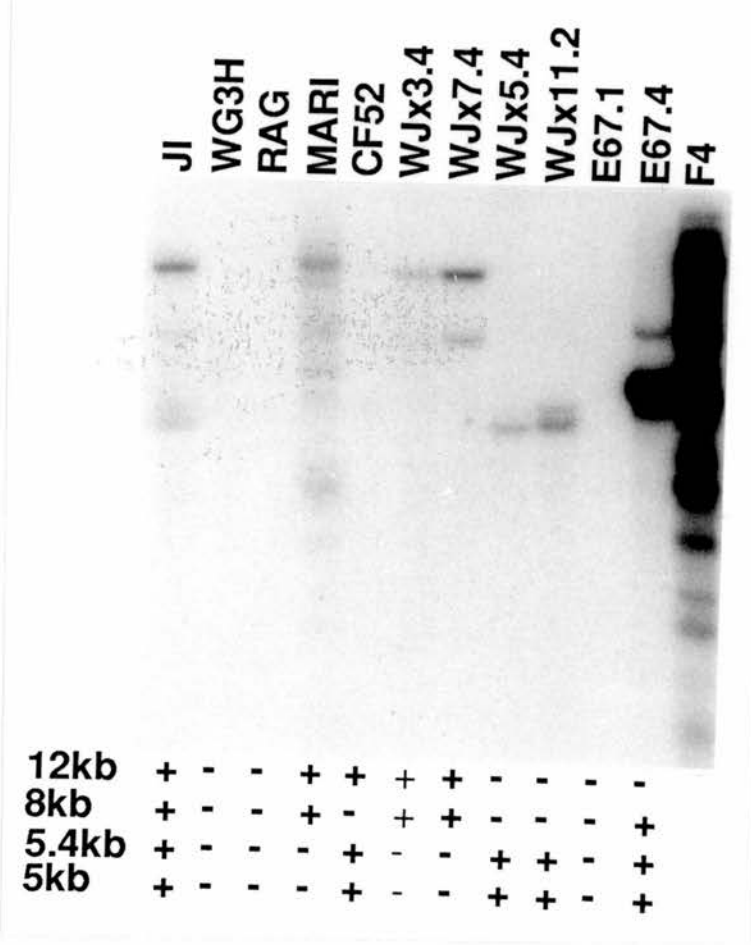
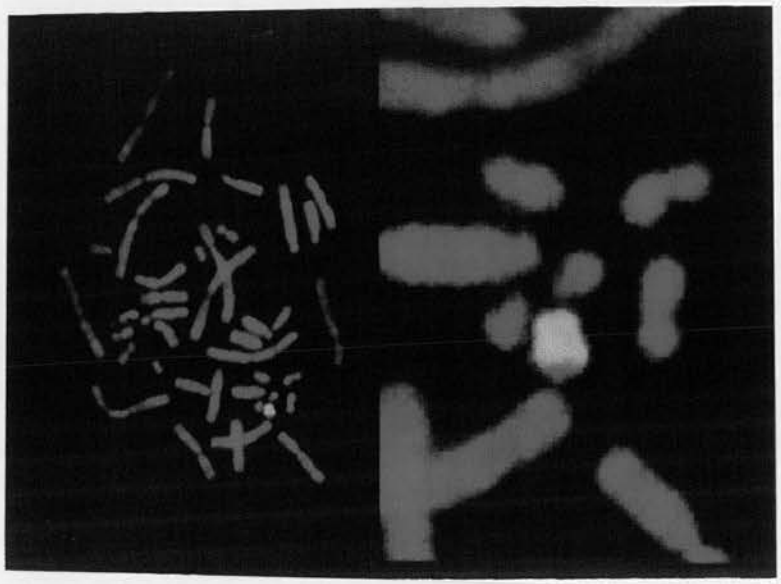


Figure 4.3 Mapping TF9 on the somatic cell hybrid panel

Clone TF9 was mapped in the same way as TF7 (Fig. 4.1). This clone gave four different bands each of which showed a different hybridisation pattern (Table 4.2). Three were negative for WJX3.4 indicating that the signal was actually due to hybridisation to a related sequence. The result showed that TF9 represented a low copy number repeat, four copies of which were located on chromosome 11 (J1 result). The hybrids are described in more detail in the text: F4 = total human genomic DNA; J1 = chromosome 11 only hybrid; WG3H and RAG = hamster and mouse background respectively; E67.1 and E67.4 = CMGT hybrids; WJX- = X irradiation hybrids.

4.2 FISH analysis

In parallel with mapping individual products of the WJX3.4 Alu-PCR library on the somatic cell hybrid panel, the cytogenetic location of pooled clones was confirmed by FISH. The results of this experiment, which was done by Dr Yoshiro Shibasaki, are shown in Fig.4.4c. Comparison to previous mapping results (previously published by others in our group - Fletcher et al., 1993) with complete Alu-PCR products from the hybrid, shows that the turboclones clearly map to the same chromosomal location. The FISH methods used here were as previously described (Fletcher et al., 1993).



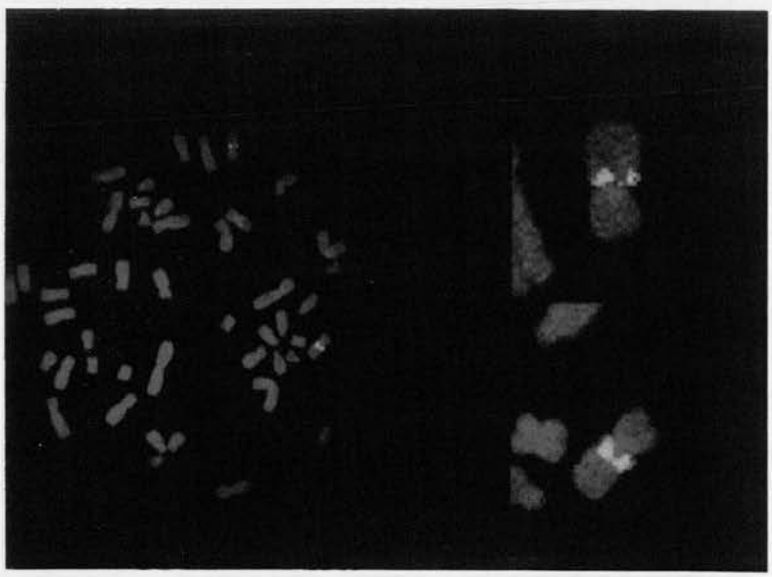


Figure 4.4 Mapping WJX3.4 on normal mitotic spreads

a) The human content of WJX3.4 was assessed by *in situ* hybridisation of the Cotl fraction of human DNA onto the hybrid. A representative mitotic spread is shown on the right, with an enlarged view of the human fragment-bearing chromosome on the left.

b) This shows human DNA from WJX3.4, isolated by Alu-PCR, painted back onto a normal human mitotic spread. A representative mitotic spread is shown on the right, with an enlarged view of the human fragment-bearing chromosome on the left.

Figures a and b show the apparent single contiguous region of chromosome 11q in WJX3.4.

c) This shows hybridisation of human DNA from a selection of turboclones, isolated by Alu_{18} PCR and combined in equimolar quantities, to normal mitotic spreads. Chromosome 11 was identified using a cosmid clone for IGF2 (green signal) which identifies 11p15-ter (de Patger-Hothuizen et al., 1987).

4.3 An integrated map of chromosome 11q12-13

Having shown that the turboclones mapped to the expected region by FISH and sublocalised them within the region on the somatic cell hybrid panel, it was now appropriate to investigate the region in order to determine the usefulness of the markers. Several of the genes which have been mapped to chromosome 11q12-13 have been shown to be directly involved in disorders (Royle et al., 1987; Rochelle et al., 1992; Smith et al., 1993), for example PC (pyruvate carboxylase) and F2

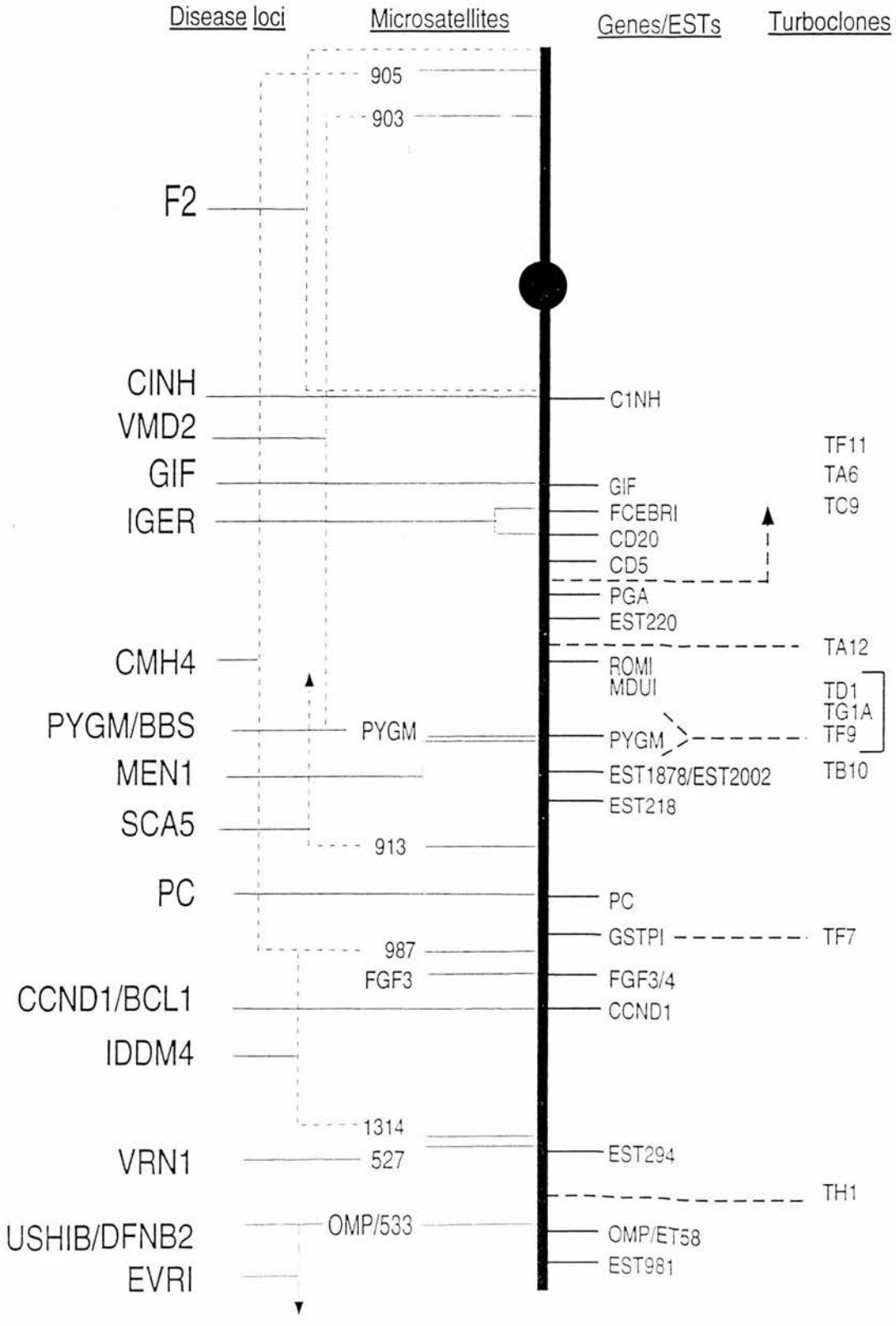
(human prothrombin). A number of other disorders have been linked to polymorphic markers, and in some cases candidate genes have been identified. Disorders mapped to the region include several types of cancer, eye retinopathies, diabetes and of course Usher syndrome type IB. These disorders have been described in the Introduction (section 1.1.3).

The recent Chromosome 11 Mapping Workshop (van Heyningen and Little 1995) was able to successfully consolidate much of the chromosome 11 mapping data. Using this, the latest genetic linkage map (Gyapay et al., 1994), other information in disorder mapping papers (described in section 1.1.3), and cloned DNA maps derived from disease associated regions (Smith et al., 1993; Carrier et al., 1993; Stafford et al., 1994;Byström et al., 1990; and Bascom et al., 1992a,b), an integrated map of 11q12-13 was compiled (Fig. 4.5).

Figure 4.5 An integrated map of chromosome 11q12-13

The construction of this map has been previously described in more detail (fig. 1.2). Here the relative positions of the turbo clones is also given. The orientation of turboclones TD1, TB10, TG1A and TF9 could not be determined relative to PYGM. As before the D11S- prefix has been omitted from microsatelitte marker designations.

Disease locus	**Disorder**
F2	Hypoprothrombinemia/dysprothrombinemia
CINH	Hereditary angioedema
VMD2	vitelliform macular dystrophy/Best's disease
MENI	Multiple endocrine neoplasia
CMH4	Familial hypertrophic cardiomyopathy
PYGM	McArdle disease
BBS	Bardet-Biedl syndrome
EVRI	Familial exudative vitreoretinopathy
SCA5	Spinocerebellar ataxia type 5
PC	Pyruvate carboxylase deficiency
CCND1	Parathyroid adenomatosis I
BCL1	Chronic lymphatic B-cell leukemia
IDDM4	Insulin independent diabetes mellitus type 4
IGER	Atopy
VRN1	Neovascular inflammatory vitreoretinopathy
USHIB	Usher syndrome type IB
DFNB2	Neurosensory, non-syndromic deafness type 2
GIF	Congenital pernicious anemia


Disease loci
Microsatellites
Genes/ESTs
Turboclones
905
903
F2
CINH
VMD2
GIF
IGER
CMH4
PYGM/BBS
MEN1
SCA5
PC
CCND1/BCL1
IDDM4
VRN1
USHIB/DFNB2
EVRI
C1NH
GIF
FCEBRI
CD20
CD5
PGA
EST220
ROMI
MDUI
PYGM
EST1878/EST2002
EST218
PC
GSTPI
FGF3/4
CCND1
EST294
OMP/ET58
EST981
TF11
TA6
TC9
TA12
TD1
TG1A
TF9
TB10
TF7
TH1
913
987
FGF3
1314
527
OMP/533

Using distance estimates based on the number of breakages observed between two markers when exposed to a specific dose of radiation (Cox et al., 1990; van Heyningen and Little 1995) it was also possible to estimate the size of 11q12-13. The region spans 570.5 cR (606-35.5) and if 1cR is roughly equal to 50.2Kb (James et al., 1994) then the region covers 28.6Mb. This estimate is in rough agreement with the estimated size of 11q at 81.5Mb (James et al., 1994).

Figure 4.5 shows that a number of turboclones mapped to regions which had been linked to disorders. A significant number of the turboclones appeared to be concentrated in the region which had been linked to Best Disease (VMD2), Bardet Biedl Syndrome (BBS), and Spinal Cerebellar Ataxia (SCA5). It was decided to use a selection of these to build up a more universally useful cloned DNA map of the region, in the form of a YAC contig.

4.4 Isolation of YACs using repeat free clones

Clones which gave extremely low or no background on the somatic cell hybrid panel were selected and used to screen the ICI YAC library (Anand et al., 1990). Thus TD1, TB10, and TF7 were used individually and TA8, TA12, and TH8 as a pool. A number of ICI YAC spot filter sets were available which had previously been screened in the course of another mapping project. Initial experiments indicated that some of the filter sets were of poor quality, i.e. probes which had previously given good results on these now gave faint hybridisation. The use of these sets was discontinued. Some typical results obtained when screening the

YAC panel are shown in Figs. 4.6 and 4.7. These results are summarised in Table 4.3. 18 YACs were positive in the primary screen although 20 YACs were actually acquired from the HGMP resource centre, because one of the signals from TF7 and TD1 was difficult to position precisely on the filter due to low background, which meant that the layout of the gridded colonies could not be seen. Clearly at the stage of screening the YAC filter there were no obvious differences between correctly identified clones and those which later turned out to be negative. Thus it was necessary to acquire all of the primary positive YACs and carry out a secondary screen.

Table 4.3 Results of primary screen

ICI YAC spot filters were screened using pooled probes TA12, TA8, and TH8 and individual probes, TB10, TD1, and TF7, and the results are shown below.

+ = the YAC was positive for this probe

- = the YAC was negative for this probe

Shaded results represent those which were difficult to orientate, only one of the two YACs was actually positive.

YACs / Probes	2CF2	4FF6	4GB11	6If3	7AF3	11BB8	13FE2	14IA7	17IE4	22EF5	22GE5	27HD6	27IA8	35GF5	35HB5	36BA5	36BD5	38GE6	39BE6	40IH3
TA12	+	+	-	-	-	-	+	-	-	+	+	+	+	+	-	-	-	-	-	-
TA8	+	+	-	-	-	-	+	-	-	+	+	+	+	+	-	-	-	-	-	-
TH8	+	+	-	-	-	-	+	-	-	+	+	+	+	+	-	-	-	-	-	-
TD1	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-
TB10	-	-	-	-	-	+	-	+	+	-	-	-	-	-	+	-	+	-	-	+
TF7	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-

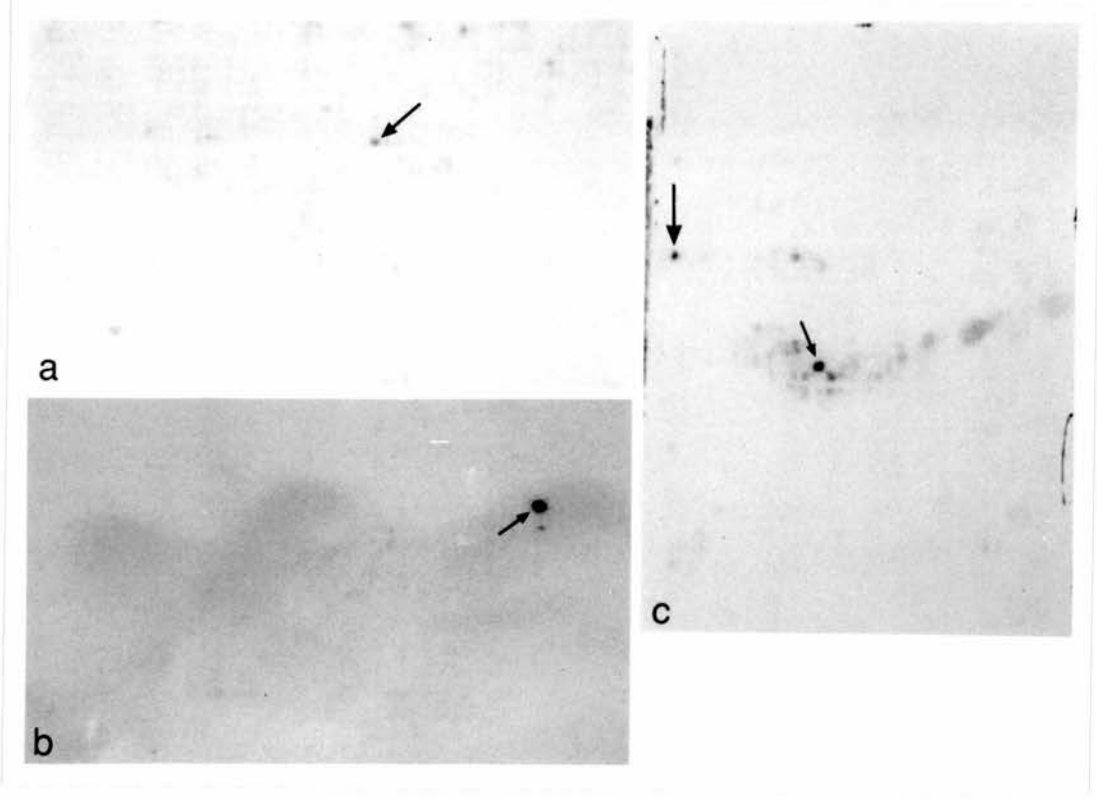


Figure 4.6 Screening YAC filters with pooled turboclones

Alu_{18} PCR products from turboclones TA8, TA12, and TH8 were labelled by random priming, pooled and used to screen the ICI YAC filters. Large arrows identify true positives, and small arrows show those later shown to be false. Four of the primary positives are shown here; a - 4FF6; b - 2CF2; and c - 27HD6 and 27IA8. Rescreening of isolated YACs, later showed 4FF6 to be a true positive for TA12, and 27IA8 to be a true positive for TH8.

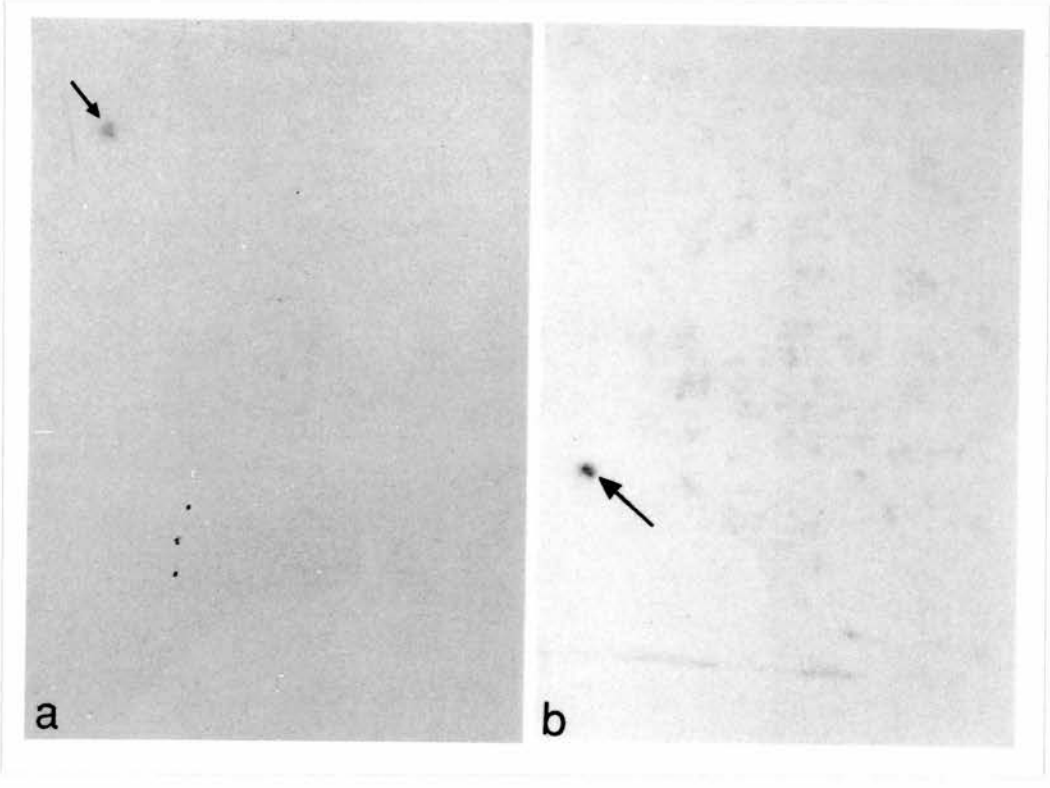


Figure 4.7 Screening the YAC filters with the turboclone TF7

TF7 identified three YACs in the primary screen, one of which, 2CF2, was also identified by the TA8, TA12 and TH8 pool, in neither case however was the primary result confirmed by the secondary screen suggesting that nonspecific hybridisation was occurring. Shown here are the primary positives a- 4GB11 and b - 36BA5, of which only 36BA5 was positive in the secondary screen.

Approximately 0.5µg YAC DNA was then digested with EcoRI and after re-estimating the relative concentration of each YAC on a gel, equal concentrations were electrophoresed on a preparative gel which was Southern blotted.

Initial positives were then confirmed by hybridising turboclones TF7, TB10, TD1, TH8, TA8, and TA12 to these blots. Blots were also screened with the remainder of the turboclones. In addition ESTs, known gene probes, and polymorphic markers which had been mapped to the somatic cell hybrids were hybridised to the blots. Primers were available for all of the latter set of markers so a probe was prepared by amplification of J1Cl4 DNA, and the gel purified product labelled by Random Priming. In some cases the product was less than 200bp, in which case the conditions were altered so that fragment specific primers were used in the random prime reaction.

Both the YAC and somatic cell hybrid Southern blots, which were screened with the turboclones, were produced from EcoRI digested DNA. In addition, the YAC library had also been prepared from EcoRI digested DNA. Thus each individual probe would be expected to produce a band of the same size on each positive YAC and hybrid. Moreover the same amount of each YAC had been loaded in each lane hence the intensity of signal would be expected to be the same for each true positive YAC. These provisos were used to confirm or disregard positive bands during the secondary YAC screen leading to the results illustrated in Table 4.4. Examples of some of the results and the justification behind their inclusion or exclusion are shown in Fig. 4.8. The size and intensity of any bands obtained on

the YACs was compared to genomic DNA which was run on the same blot. In the end only 5 of the 18 YACs from the primary screen were confirmed to be positives which in each case was only for the clone used in the initial selection process.

Table 4.4 Results of rescreening primary positives

Southern blots of EcoRI digested YACs which were positive in the primary screen were rescreened using turboclones, probes for known genes, and polymorphic markers from the region. Those probes used in the primary screen are shown in bold. Many of the probes failed to give a positive signal on any of the YACs : EST220; PGA; TE7; ROMI; MDUI; TGI; TF9; EST218; EST1878; EST2002; GSTPI; D11S527; EST294; and THI.

YACs	2	4	4	6	7	11	13	14	17	22	22	27	27	35	35	36	36	38	39	40
	C	F	G	I	A	B	F	I	I	E	G	H	I	G	H	B	B	G	B	I
	F	F	B	f	F	B	E	A	E	F	E	D	A	F	B	A	D	E	E	H
	2	6	11	3	3	8	2	7	4	5	5	6	8	5	5	5	5	6	6	3
Probes																				
TA12	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TE5	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	X	-	-	-	-
TA8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X
TB11	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
TH8	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
TD1	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-
TB10	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TF7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-

+ confirmed positive in the secondary screen

X established as false positive by the secondary screen

- negative in the secondary screen

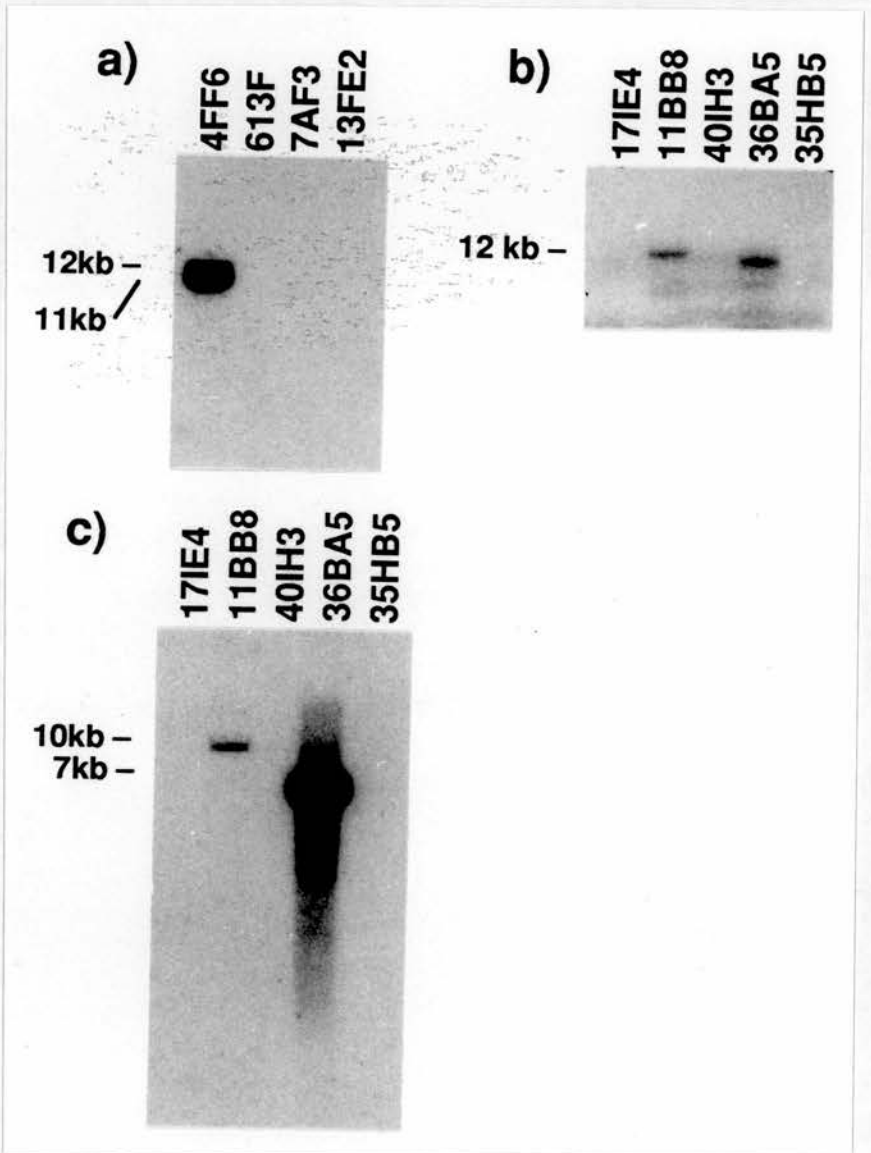

a)
4FF6
613F
7AF3
13FE2
12kb –
11kb
b)
17IE4
11BB8
40IH3
36BA5
35HB5
12 kb –
c)
17IE4
11BB8
40IH3
36BA5
35HB5
10kb –
7kb –

Figure 4.8 Confirmation of positive YACs

a) TA12 was one of the three markers used in a pooled probe. Secondary screening confirmed it to be responsible for selection of the 4FF6 YAC. The presence of two bands was also observed in screening of the somatic cell hybrid panel and was due to the fact that the turboclone insert contained an EcoRI site.

b) TE5 was not used in the primary screen, since it contained a small amount of repeated DNA sequence. In the secondary screen it appeared to be positive for the 11BB8 and 36BA5 YACs. The position of this clone on the somatic cell hybrid panel, separated by several markers from the other turboclones (TF7 and TB10) which were definitely positive for these YACs, and the fact that the band obtained on YACs was 12Kb compared to the 7Kb band obtained on the somatic cell hybrids indicated that these were false positives due to the repetitive DNA in TE5.

c) TF7 was used to select three YACs in the primary screen - 2CF2, 4GB11, and 36BA5. Secondary screening confirmed 36BA5 to be positive, as a band of 7Kb was obtained on somatic cell hybrids and YACs. 11BB8 was clearly negative in the primary screen, but appeared to be weakly positive in the secondary screen, although hybridising to a band of 10kb. This result was put down to weak hybridisation and therefore not scored as a positive result.

The 5 confirmed positive YACs were sized by preparing agarose plugs which were then run on a pulsed field gel. Not all of the YACs were clearly distinguishable from the yeast chromosomal background, so the gel was Southern blotted and probed with ^{32} P labelled CotI DNA to detect human specific DNA (see Figure 4.9).

- 4FF6 was identified using TA12 and on secondary screening two bands were seen, as had been observed when this clone was used to screen the somatic cell hybrid panel. Restriction enzyme digestion confirmed that this clone contained an EcoRI site, and the presence of only one band in the hybrid E67.1 indicated that it overlapped the end of this particular subfragment of E67.1.

- 27IA8 (280kb) and 7AF3 (350kb) both lay within the ROMI-MDUI subregion, but were identified with separate clones. In addition none of the other markers from this subregion came up positive on these YACs indicating that the region must be relatively large. The distance between ROMI and MDUI is estimated at 10.9cR (van Heyningen and Little 1995) which is equivalent to 500kb. The combined size of the YACs and that fact that they do no contain any other markers for the region makes it likely that the YACs themselves bridge the endpoints of the region.

- 36BA5 maps to the same region as GSTPI, and yet is negative for this marker indicating that this region is more than 150kb.

It was possible that the YACs, in particular 27IA8 and 7AF3, might overlap in regions not covered by markers. This was tested by comparing the Alu-PCR

profiles of the YACs. After amplification using the Alu_{614} primer. The annealing temperature found to give the highest complexity of products was 50°C. The Alu-PCR fingerprint, obtained at this annealing temperature, was then compared by electrophoresing the products in parallel on a 1.5% agarose gel (Fig.4.10). No bands of the same size were observed in the YACs indicating that there were no regions of overlap. FISH was then carried out by Dr Yoshiro Shibasaki, essentially as described by Breen et al., (1993) using the pooled Alu-PCR products to ensure that the YACs did in fact map to the right region (figure 4.11).

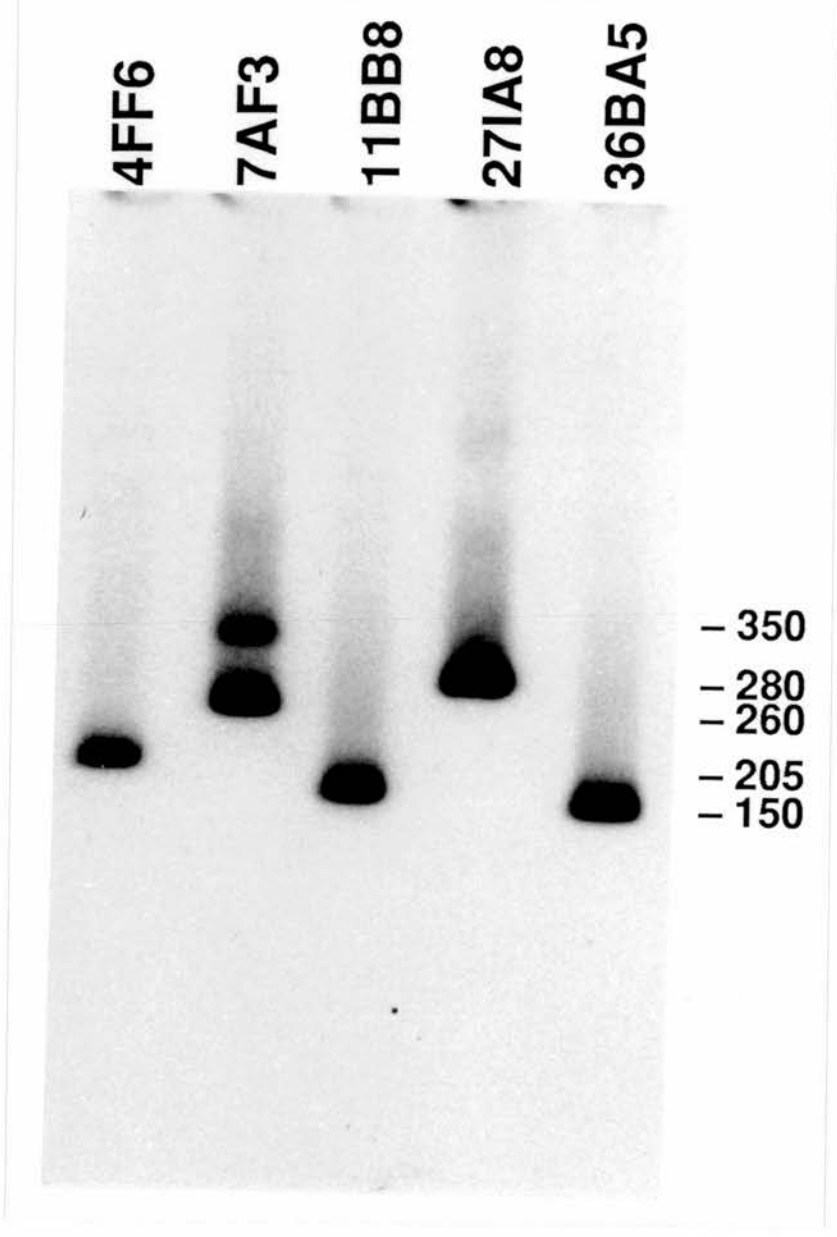


Figure 4.9 Sizing positive YACs

Agarose plugs were prepared and run on a CHEF pulsed field gel as described in Material and Methods. Some of the YACs were of the same size as the yeast chromosomes, so the gel was Southern blotted and hybridised with CotI DNA to identify the YACs. Each could be roughly sized relative to the markers: 4FF6 = 205kb; 7AF3 = 265kb, 350kb; 11BB8 = 150kb; 27IA8 = 280kb; and 36BA5 = 150kb. YAC 7AF3 is unstable and the deleted form is also present.

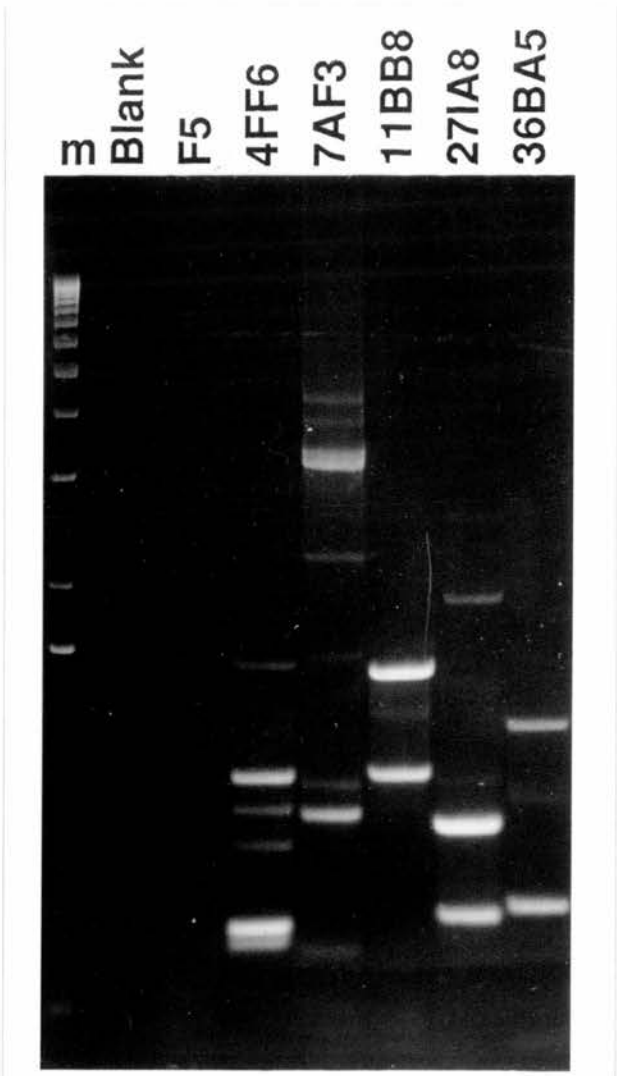


Figure 4.10 Alu-PCR profile of positive YACs

YAC s were amplified using Alu614 at Ta = 50°C. Profiles can be directly compared here on a full length 1.5% agarose gel. None of the bands are common between the YACs. M = 1kb ladder, F5 = human genomic DNA.

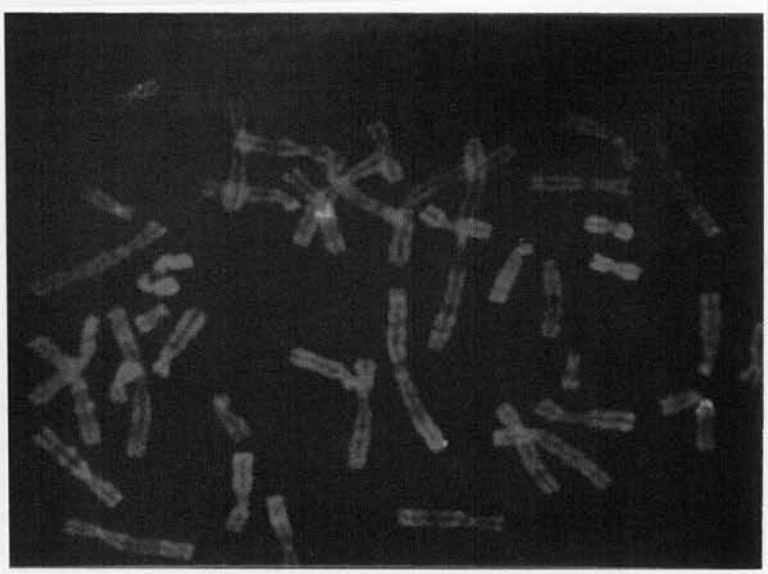


Figure 4.11 FISH with positive YACs

The Alu-PCR products shown in Fig. 4.10 were combined and used as a probe in *in situ* hybridisation to normal mitotic chromosome spreads. This was done by Dr Yoshiro Shibasaki as described in Breen et al (1992).

4.5 Discussion

A library of repeat free clones have been mapped to chromosome 11q12-13 defining a total of 21 mapping intervals in this region. One interval in particular (ROMI-MDUI) contained several clones. Clones mapping to interesting regions were used to screen the ICI YAC library, and 5 YACs were isolated. Screening these YACs with additional markers, and comparing Alu PCR fingerprints failed to identify any overlap between them. The results obtained were however consistent in that all the turboclones and YACs mapped to the expected region. The mapping process was simple and did not require complicated preannealing of the probe to remove Alu ends. Turbo cloning also allowed the cloning of a broad range of product sizes.

The results presented here are encouraging and demonstrate the feasibility of this method as a means of isolating new markers at a high density from a specific region. In order to build up a contig of 11q12-13 it would be necessary to either map additional markers; isolate larger insert containing YACs; or use end clones of the primary YACs to isolate more YACs.

5. Human Olfactory Marker Protein gene

In mice the *shaker-1* mutation was shown to lie on chromosome 7 within 200kb of the Omp gene. The fact that the human OMP gene lay within a homologous region, 11q13.5, to which Usher syndrome type I had been linked, and the observed ciliary defects, suggested the OMP gene as a candidate for this syndrome. Hence the gene was subcloned and mutation analysis was carried out as described in the following sections.

5.1 Subcloning

The human olfactory marker protein gene had already been isolated from a cosmid library using an exonic fragment as a probe, which in turn had been obtained by cross species PCR using primers directed against the published sequence of the rat OMP gene (Evans et al., 1993). To obtain the sequence of the human gene it was necessary to subclone the gene from the cosmid. This was done by digesting the cosmid, identifying OMP containing fragments by hybridisation analysis and cloning these into appropriately prepared plasmid vectors.

At the stage when I took over the project part of the OMP gene had already been subcloned as a PstI fragment into the Bluescribe plasmid cloning vector pBS (Stratagene™)(see Fig.5.1). The appropriate cosmid fragment had been identified by hybridisation of a rat PCR product to a cosmid PstI digest, and the identification of the cloned product confirmed by restriction enzyme digestion and sequencing. From the sequence it was clear that only part of the coding region and the 3' untranslated region had been cloned. Comparison with the rat sequence suggested

that about 80bp were missing from the human coding region. In order to clone the rest of the coding region and some of the 5' untranslated region the available human OMP sequence was examined for restriction sites compatible with the pBSK polylinker. TaqI recognises the tetranucleotide target sequence TCGA, which lies 173bp downstream of the PstI site, and produces ends compatible with the hexanucleotide target sequence ATCGAT recognised by ClaI in the pBSK polylinker. The 5' end of the gene was identified in a TaqI digest of the cosmid using a rat primer that covered the start codon (primer C) as a hybridisation probe (Fig.5.1).

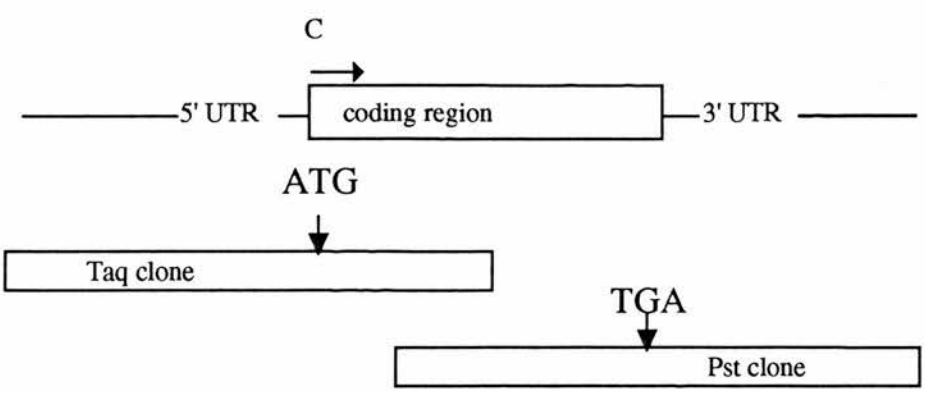


Figure 5.1 Subcloning the human OMP gene

The human gene was subcloned in two overlapping fragments into the Stratagene vectors pBS and pBSK. The overlapping fragments and their positions relative to the rat gene are shown. Primer C was derived from the rat coding sequence and used to identify the Taq fragment in the cosmid digest.

ClaI digested phosphatased pBSK vector was prepared. 1µg of cosmid was digested with TaqI, run out for the full length of a 1% gel and Southern blotted. Primer C was kinase end labelled and used as a probe on the blot and a hybridising

band of approximately 800bp was identified. The digest was then repeated, and the 800bp band cut out after the gel had been run its full length. DNA was purified from the agarose using a SpinX column. To increase the efficiency of DNA recovery 20µg of glycogen were added to the precipitation. A ligation was then set up containing 35ng insert:25ng vector, and 2µl of the ligation mix transformed into electrocompetent XL1 Blue cells. Colonies could have been analysed by colony lifts and hybridisation with primer C. Due to the low number of white colonies obtained, however, it was decided to use small scale plasmid preparations and colony PCR. Minipreps of six of the ten resulting white colonies were prepared. TaqI digestion of these showed that none of them contained any insert. A further 4 colonies were analysed by colony PCR, and only a fragment the size of amplified polylinker resulted. The insert preparation and cloning reactions were repeated several times, using different methods to purify the digested insert. This was an attempt to achieve higher returns and thus increase the number of different vector:insert ratios used in the cloning reaction. In addition, the ligation mix was precipitated and resuspended in 2µl allowing the whole reaction to be transformed. Again no positives containing an insert of the appropriate size were obtained. To increase the concentration of insert yet further, three 1µg aliquots of the cosmid were combined, precipitated (to decrease the volume) and run out in a single well. Success came when equimolar quantities (18ng insert:77ng vector; vector size ~3000bp, insert size ~700bp, recovered just 18ng of insert so 3000/700×18=77ng) of vector and insert were combined in the ligation mix, which was again

precipitated, resuspended in 4µl, and 2µl electroporated. PCR showed 8 of the 13 white colonies contained a fragment of the appropriate size and were positive for primer C by hybridisation. Sequence analysis of two clones confirmed the presence of the appropriate fragment, although in a different orientation in each case.

5.2 Sequence analysis of OMP

The foremost requirement of mutation analysis is an absolute certainty in the wild type sequence of the gene of interest. To confidently produce a consensus sequence for the entire human OMP gene I had to sequence both strands in both directions, which meant the insert had to be cloned in both orientations. The simplest way to clone the PstI fragment in the opposite orientation was to release the insert by a PstI digest, religate and use another restriction digest to determine the orientation of the insert. (Alternatively more colonies could have been screened from the original ligation mix, but unfortunately this was unavailable).

2µg of the PstI clone were digested with PstI, and run out on a gel to check the reaction had gone to completion. The reaction was then cleaned up using phenol/chloroform, ethanol precipitated, and the end concentration estimated on a gel. Ligations of 15ng/µl and 1.5ng/µl were set up. 1µl of the reaction was electroporated, and 200µl of the reaction plated, resulting in 20 and 5 white colonies respectively. 12 colonies were grown up and small scale plasmid preparations carried out. The orientation of the insert was determined using a HindIII/StuI double digest (Fig.5.2). Early sequencing of the PstI clone had

indicated that the HindIII site at the M13 reverse primer end of the pBS polylinker was unique. There were at least two sites for StuI: one in the 3' untranslated region (307bp from HindIII) and one in the coding region (280bp from the other StuI site). Initial single digests with each enzyme were used to prove that the enzymes cut at the expected sites; HindIII cut at one site and linearised the clone, whereas StuI released a 280bp fragment. The exact size of the PstI clone was not known, but it was estimated to be around 950bp by PCR. Thus if the orientation of the insert was reversed, then the size of the HindIII to StuI fragment would increase to approximately 360bp (Fig. 5.2).

The digests of the 12 small scale plasmid preparations of colonies were run out on a 2% gel to separate the two smaller products, and 3 had the insert in the correct orientation. In these the HindIII to StuI fragment had increased in size to around 360bp. Large scale plasmid preparations were prepared of two of these and the orientation confirmed by sequencing.

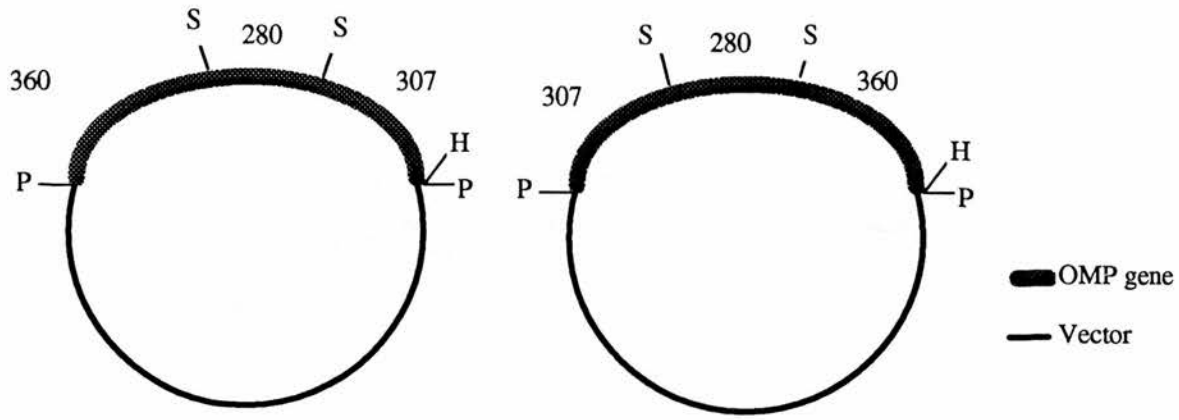


Figure 5.2 Cloning the PstI insert in the opposite orientation

The left-hand figure shows the insert as it was originally cloned. The right-hand figure shows the insert after it had been recloned in the opposite orientation. Fragment sizes are shown in base pairs. Restriction enzyme sites are indicated: H=HindIII; P=PstI; S=StuI.

The two parts of the OMP gene had been cloned into Stratagene™ vectors which contain sequences complementary to the M13 forward and reverse primers flanking the polylinker. These primers were initially used in PCR amplification and sequence analysis of inserted DNA. The sequence obtained from these reactions could then be used to design primers internal to the insert, allowing further sequence to be obtained. Conventional sequencing with a radioactive isotope was carried out first.

At first, only the PstI clone in one orientation was available for sequencing. Using the M13 primers almost 200bp of sequence was obtained from either end of the insert. Primer Y was designed to lie internal to the M13 reverse primer and thus another 260bp of sequence was obtained. 2μg of the PstI clone was double digested using StuI and HindIII and the 307 and 280 base pair fragments (see Fig. 5.2) were removed from the clone by gel purification of the large fragment. This removed part of the polylinker and previously sequenced OMP gene between the M13 reverse primer and the unknown region. The fragment was then made blunt ended using T4 DNA polymerase, precipitated, resuspended in 10μl and 1μl run out on a gel to assess the concentration. 50ng of the fragment was ligated overnight in 10μl, and 1μl electroporated. Plating 200μl produced 23 white colonies. Amplification of 14 of these showed 8 of them to contain insert of the appropriate size. This new clone was called the StuI clone and the M13 reverse primer was then used with it to sequence into the unknown region of a large scale preparation of a positive colony. Thus the sequence of the PstI clone was obtained, by walking out from the M13 reverse primer, although sequencing of both strands at each point was needed to produce a consensus sequence. This sequence information would however allow further primers to be designed.

Sequencing of the two TaqI clones proved less successful. It was possible to complete the sequencing of the OMP coding region, but due to the high proportion of G and C nucleotides the 5' untranslated region proved very difficult to sequence. Even using dITPs and deaza-dGTP, very little interpretable sequence information

was obtained. The use of dITPs in sequencing has been described in a previous section. The substitution of 7-deaza-dGTP for dGTP in the nucleotide mix can reduce compressions. When incorporated in the sequencing product, 7-deaza-dGTP reduces potential hydrogen bonding, and hence decreases band compressions during gel electrophoresis. Using the universal primers and primer X the clones were sequenced to 50bp beyond the start codon, but only a further 200bp were obtained of the 5' untranslated region from the other end of the clones. This left almost 300bp of sequence internal to the clones unsequenced.

At this point the reverse orientation PstI clone became available as did an automatic sequencer for fluorescent sequencing. Efficient completion of large DNA sequencing projects has been greatly facilitated by the development of fluorescence based dideoxynucleotide sequencing chemistries and instruments for real-time detection of fluorescence labelled DNA fragments during gel electrophoresis. In addition to eliminating the use of radioactive isotopes, these systems automate the task of reading sequences and provide computer readable data that may be directly analysed or entered into a sequence assembly project. The sequencer was the Automated Laser Fluorescent A.L.F.™ DNA Sequencer from Pharmacia LKB. Reactions are carried out using fluorescently labelled primer and run on a gel which is part of the sequencing apparatus. The sequence is obtained by a fixed point laser beam which reads and analyses the bases as they pass by. Cycle sequencing was used instead of conventional single primer annealing and extension sequencing. This involves the linear amplification of the

sequencing reaction using repetitive cycles of denaturation, annealing, and extension in the presence of Taq polymerase. In cycle sequencing, less DNA is required than in a normal non-cycling reaction, and the resulting sequencing data have less background, few (if any) stops, and can extend to over 500 bases from the primer. An initial attempt to sequence the original PstI clone, using the BRL cycle sequencing kit and fluorescently labelled M13 forward primer, achieved 380 bases of readable sequence. Comparison to rat and previously sequenced human showed the sequence to contain only a few ambiguities towards the end of the sequence. It was decided that this would be the quickest and most accurate way to complete the sequencing of the human gene and resolve any previous ambiguities. All of the clones were then sequenced using fluorescently labelled universal primers. Some of the caesium chloride plasmid preparations proved difficult to sequence and it was necessary to ensure the removal of any residual impurities by phenol/chloroform and ethanol precipitation. By comparison to the estimated size of the TaqI clone, the amount of sequence obtained from both ends of the TaqI clones should have revealed an overlap. Using the GCG fragment assembly program, all of the sequences obtained so far were put into contigs. Using Gel Merge, even under very non-stringent conditions, it was still not possible to produce one contig and the 5' untranslated region remained in a separate contig from the rest of the gene. It was hoped that the fact that the kit contained 7-deaza-dGTP and the reactions were extended at 70°C would be sufficient to reduce false stops induced by template secondary structure, but this was obviously not the case. The sequence information obtained so far, however, did make it possible to design

primers for sequencing the gene in genomic DNA (Fig. 5.3). Two sets of primers were required: an outer set to amplify the gene by PCR; and an inner set (nested) to specifically sequence these PCR products.

It was recommended by Pharmacia that the most accurate sequence information would be achieved using a single stranded template. For this, one of the primers used in the genomic amplification reaction had to be biotinylated. The biotinylated and non biotinylated strands could then be separated using Dynabeads™ (streptavidin coated magnetic beads), and individually sequenced using nested fluorescently labelled primers. The primers were designed using the OLIGO4 program (Hybaid) to overlap in a similar way to the clones, and it was hoped that four reactions would be sufficient to sequence the gene (see Fig.5.3).

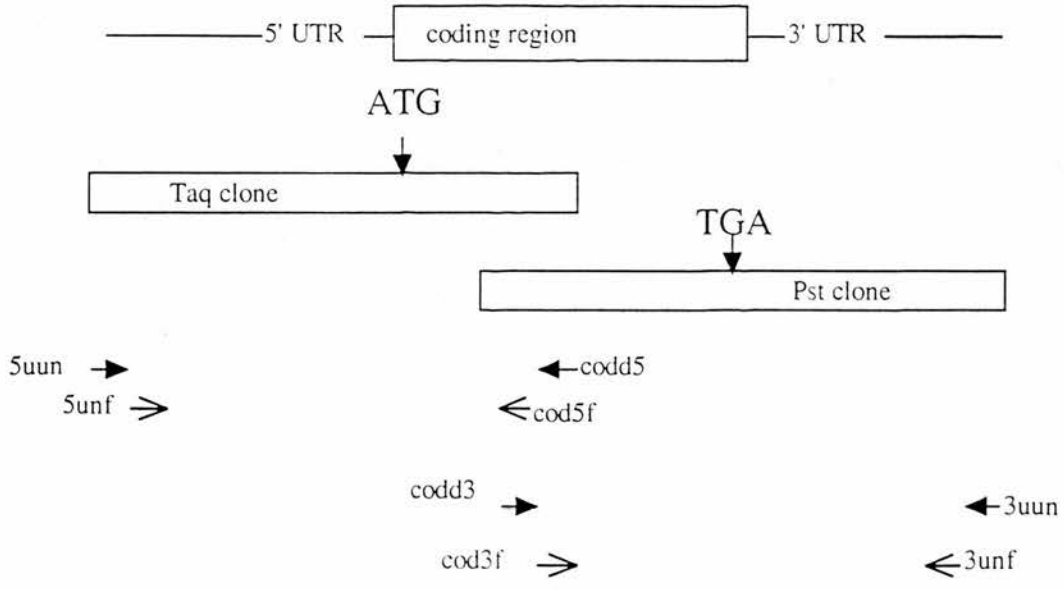


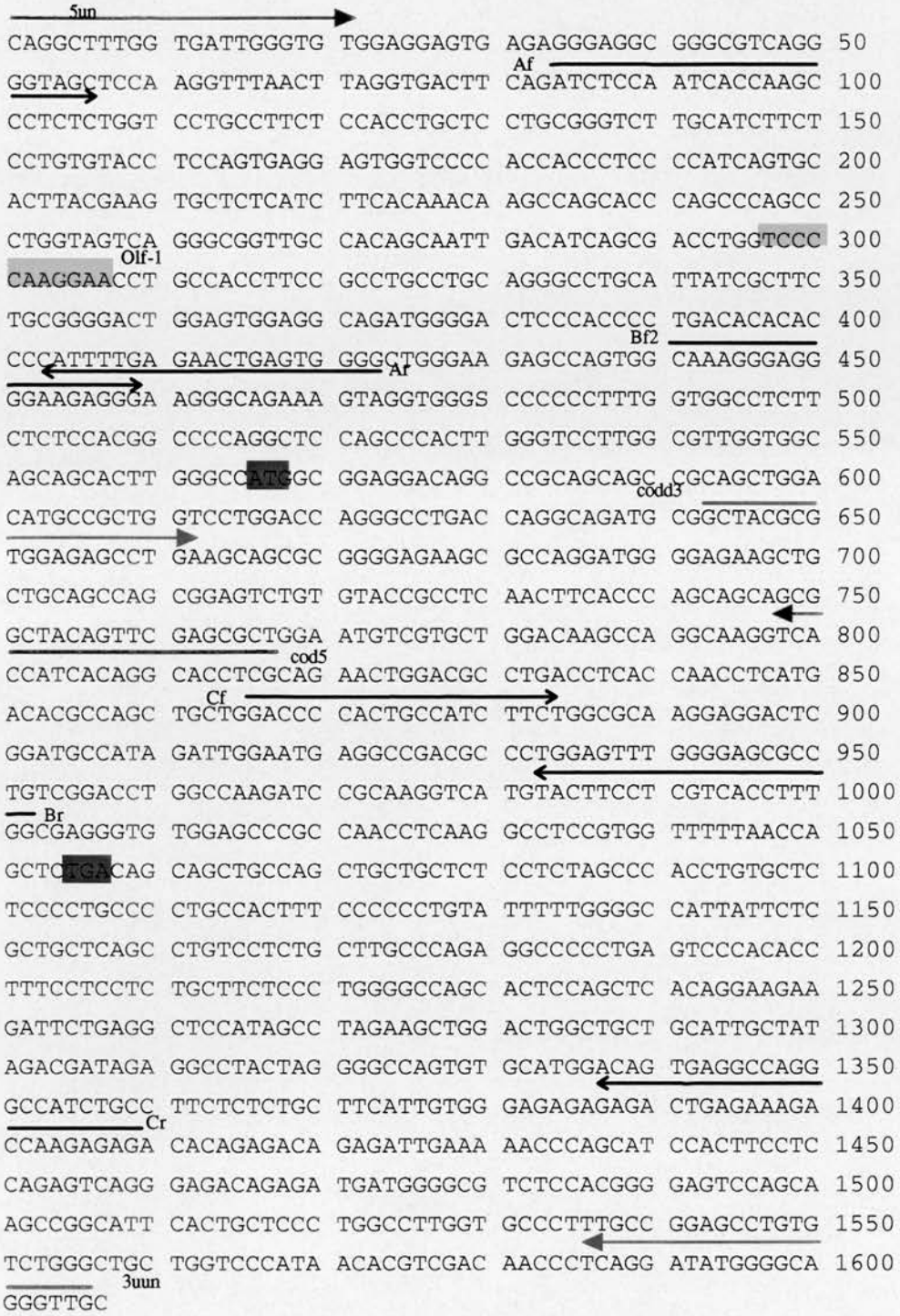
Figure 5.3 Primers designed for genomic amplification and sequencing

The location of the primers designed for genomic amplification and sequencing of OMP relative to the plasmid subclones and the gene itself. Start (ATG) and stop (TGA) codons are also shown, as are the untranslated regions (UTRs). The outer primers(→ 5uun,codd3,codd5 & 3uun) were ordered in biotinylated and non biotinylated forms so that either single strand could be pulled out using Dynabeads™. The inner primers were all fluorescently labelled, as indicated by the 'f' suffix.

The complete sequence of TaqI clones was finally achieved using the ABI sequencer in conjunction with the ABI cycle sequencing kit as described below.

The ABI sequencer is very similar to A.L.F., the only major difference being that the laser scans by actually moving across the glass plate. Special primers had to be

designed for use with this kit which contains fluorescently labelled dideoxy terminators, enabling all four termination reactions to be carried out in one tube. Basically the primers had to be longer than 18 bases, have a T_M of at least 45°C, and avoid long stretches of any single base. These were designed using the GCG primer design program Primer, and their locations on the clones are indicated below (Fig. 5.4). The optimum amount of accurate sequencing data expected was 300 bases, hence the primers were designed to produce even coverage of the gene. Sequencing of both the TaqI and PstI clones once using these primers was sufficient to close the gap in the 5' untranslated region and resolve all ambiguities (Fig. 5.4). The advantage of this kit over previously attempted sequencing methods was the combination of dITPs and extension at 60°C. dITPs are traditionally used to reduce secondary structure, but are incorporated less well than dGTPs at 72°C. Consequently lowering the temperature at which Taq polymerase extends allows efficient incorporation of dITPs and, as shown here, allows the enzyme to read through difficult regions.



5un
CAGGCTTTGG TGATTGGGTG TGGAGGAGTG AGAGGGAGGC GGGCGTCAGG 50
GGTAGCTCCA AGGTTTAACT TAGGTGACTT CAGATCTCCA ATCACCAAGC 100
CCTCTCTGGT CCTGCCTTCT CCACCTGCTC CTGCGGGTCT TGCATCTTCT 150
CCTGTGTACC TCCAGTGAGG AGTGGTCCCC ACCACCCTCC CCATCAGTGC 200
ACTTACGAAG TGCTCTCATC TTCACAAACA AGCCAGCACC CAGCCCAGCC 250
CTGGTAGTCA GGGCGGTTGC CACAGCAATT GACATCAGCG ACCTGGTCCC 300
CAAGGAACCT GCCACCTTCC GCCTGCCTGC AGGGCCTGCA TTATCGCTTC 350
TGCGGGGACT GGAGTGGAGG CAGATGGGGA CTCCCACCCC TGACACACAC 400
CCCATTTTGA GAACTGAGTG GGGCTGGGAA GAGCCAGTGG CAAAGGGAGG 450
GGAAGAGGGA AGGGCAGAAA GTAGGTGGGS CCCCCCTTTG GTGGCCTCTT 500
CTCTCCACGG CCCCAGGCTC CAGCCCACTT GGGTCCTTGG CGTTGGTGGC 550
AGCAGCACTT GGGCCATGGC GGAGGACAGG CCGCAGCAGC CGCAGCTGGA 600
CATGCCGCTG GTCCTGGACC AGGGCCTGAC CAGGCAGATG CGGCTACGCG 650
TGGAGAGCCT GAAGCAGCGC GGGGAGAAGC GCCAGGATGG GGAGAAGCTG 700
CTGCAGCCAG CGGAGTCTGT GTACCGCCTC AACTTCACCC AGCAGCAGCG 750
GCTACAGTTC GAGCGCTGGA ATGTCGTGCT GGACAAGCCA GGCAAGGTCA 800
CCATCACAGG CACCTCGCAG AACTGGACGC CTGACCTCAC CAACCTCATG 850
ACACGCCAGC TGCTGGACCC CACTGCCATC TTCTGGCGCA AGGAGGACTC 900
GGATGCCATA GATTGGAATG AGGCCGACGC CCTGGAGTTT GGGGAGCGCC 950
TGTCGGACCT GGCCAAGATC CGCAAGGTCA TGTACTTCCT CGTCACCTTT 1000
GGCGAGGGTG TGGAGCCCGC CAACCTCAAG GCCTCCGTGG TTTTTAACCA 1050
GCTCTGACAG CAGCTGCCAG CTGCTGCTCT CCTCTAGCCC ACCTGTGCTC 1100
TCCCCTGCCC CTGCCACTTT CCCCCCTGTA TTTTTGGGGC CATTATTCTC 1150
GCTGCTCAGC CTGTCCTCTG CTTGCCCAGA GGCCCCCTGA GTCCCACACC 1200
TTTCCTCCTC TGCTTCTCCC TGGGGCCAGC ACTCCAGCTC ACAGGAAGAA 1250
GATTCTGAGG CTCCATAGCC TAGAAGCTGG ACTGGCTGCT GCATTGCTAT 1300
AGACGATAGA GGCCTACTAG GGGCCAGTGT GCATGGACAG TGAGGCCAGG 1350
GCCATCTGCC TTCTCTCTGC TTCATTGTGG GAGAGAGAGA CTGAGAAAGA 1400
CCAAGAGAGA CACAGAGACA GAGATTGAAA AACCCAGCAT CCACTTCCTC 1450
CAGAGTCAGG GAGACAGAGA TGATGGGGCG TCTCCACGGG GAGTCCAGCA 1500
AGCCGGCATT CACTGCTCCC TGGCCTTGGT GCCCTTTGCC GGAGCCTGTG 1550
TCTGGGCTGC TGGTCCCATA ACACGTCGAC AACCCTCAGG ATATGGGGCA 1600
3uun
GGGTTGC

Figure 5.4 The consensus sequence of the human OMP gene

The sequence of the human OMP gene is shown. The Olf-1 binding site and the start and stop codons are shown boxed. The orientation of primers is indicated by the arrowhead. Those pointing right to left match the antisense strand sequence. Primers shown with a filled arrowhead were used for PCR amplification, others were used for sequencing.

5.3 Comparison to rat and mouse

Comparing the coding region of the rodent and human OMP genes could potentially allow conclusions to be drawn on the importance and putative function of specific regions of the gene. The mouse (Brown et al., 1994) and rat gene (Rogers et al., 1987) sequence have been reported earlier. Comparison at the nucleotide level revealed identity between human:rat, human:mouse, and mouse:rat of 90%, 88.6% and 98.2% respectively. Several of these differences did not result in changes at the amino acid level, consequently the identities here were 90.8%, 89.6% and 96.5%. These differences and the overall consensus sequence at the amino acid level are illustrated in the figure below.

```
           1                                                    50
Rat                                  k
Mouse                   e            q                       i
Human          r  qp           g     r                 r       q
Consensus  MAEDGPQKQQ LDMPLVLDQD LT-QMRLRVE SLKQRGEKKQ DGEKLLRPAE

           51                                                   100
Rat
Mouse
Human            n t    r   er
Consensus  SVYRLDFIQQ QKLQFDHWNV VLDKPGKVTI TGTSQNWTPD LTNLMTRQLL

           101                                                  150
Rat
Mouse
Human        t             i                            v
Consensus  DPAAIFWRKE DSDAMDWNEA DALEFGERLS DLAKIRKVMY FLITFGEGVE

           151         164
Rat                     *
Mouse                   *
Human                   *
Consensus  PANLKASVVF NQL-
```

Figure 5.5 Comparison of the deduced amino acid sequence of human, rat and mouse olfactory marker protein gene

This figure was compiled using the GCG pretty command on a 'file of file names' of the three sequences (i.e. pretty -con -dif -threshold=1.5 @fofn5.seq).The overall consensus sequence is shown below each line.

Since then the sequence of the human OMP gene has been confirmed by another group (Buiakova et al., 1994). They have, however, made several errors in their

alignment of the deduced amino acid sequence of OMP. Amino acids 31 and 49 appear as arginine (R) and glutamic acid (E) respectively, when in fact the sequence submitted by the same group to the data base indicates these amino acids were conserved between the species and are actually therefore serine (S) and alanine (A). There are also three typing errors: amino acids 58 and 103 are shown as isoleucine when they are in fact both threonine; and amino acid 62 is shown as glutamine when it is actually arginine. My sequencing of the gene confirmed that the sequence lodged in the database was correct and that the published discrepancies were errors. This paper was useful in confirming the location of an upstream Olf-I binding site, 260 bases upstream of the start codon (see Fig. 5.4).

6. OMP as a candidate for Usher Syndrome type IB

OMP is a relatively small single exon gene and it was necessary to decide exactly how much of the flanking regions to analyse. I had sequenced some although not all of the untranslated regions, but it was decided that given the fact that only 300bp of the untranslated region was required to give appropriate expression of OMP in transgenic mice (see the end of section 1.2.6.4), it would be sufficient to carry out mutational analysis on this extent of the 5' untranslated region. Very little is known in general about 3' untranslated regions, so it was decided to look for any major changes in the region encompassed by sequencing. A number of different methods were used to look for mutations in OMP, and the success of these is described in this section. All of the methods used were PCR based, hence conditions for amplification of the gene from genomic DNA had to first be optimised. This was a relatively simple task for the 3' end of the gene. Standard genomic PCR conditions, annealing at 53°C, were sufficient to produce a single 960bp PCR product, using the primers codd3 and 3uun (Fig. 5.4).

Sequencing had shown that the 5' end of the gene was GC rich, and this region had proved difficult to sequence. Similarly the region also proved troublesome to amplify. Primers had been designed such that the whole gene should be amplifiable with 5uun and 3uun, and the 5' end of the gene should be amplifiable with 5uun and codd5. Several attempts using annealing temperatures been 40 and 60°C produced either no product or, as was the case with 5uun-3uun at Ta=40°C, too

small a product (~500bp instead of 1600bp). Increasing the concentration of Mg^{2+} (from 1.5-4mM) in the reaction buffer also proved unsuccessful. Several sets of external primers were designed to replace 5uun and 3uun in the hope of increasing the specificity, none of which were successful, and any product was too small. 4 sets of external primers now existed and it was decided to try using them in combination in nested PCRs in the hope of increasing the specificity and sensitivity of the reaction.

Primary PCR A:Extf+Extr B:5uun+3uun C:Extf1+Extf2

Secondary PCR A_1:Extf1+Extr1 B_1:Extf2+Extr2 C_1:5uun+Extr1

A_2: Extf2+Extr2 B_2:Extf2+Extr1 C_2: Extf2+Extr1

The template used was cosmid DNA, and two genomic DNA samples from normal individuals. The annealing temperature for both PCR reaction stages was 45°C. After an initial 25 cycles of PCR, 1µl was transferred to the secondary PCR which was also amplified for 25 cycles. PCRs A_1, C_1, and C_2 gave a 500bp product; A_2 and B_2 gave a smear with faint bands of the wrong size; and B_1 gave a strong 1.6Kb band in the cosmid track and very faint bands of the same size in the genomic tracks (Fig. 6.1). In an attempt to increase the amount of product the number of cycles in either the primary or the secondary PCR were increased. Increasing the number of cycles in the primary PCR led to no bands even with cosmid template. Increasing the number of cycles in the secondary PCR failed to produce any more of the specific 1.6Kb product in the genomic samples, and

indeed appeared to increase the level of non specific product. Reducing the annealing temperature in the primary PCR to 40°C and increasing the annealing temperature to 60°C in the secondary PCR also failed to produce any more product from the genomic samples.

Dimethyl sulphoxide has been shown to prevent template reannealing and hence aid in the sequencing of double stranded templates (Winship 1989). A literature search showed that it had also been used to prevent the formation of complex secondary structures in some PCR templates (Shen and Hohn 1992). DMSO was shown by Shen et al to be the decisive factor for Taq mediated PCR amplification of DNA fragments containing complex secondary structures. In the test case, inclusion of DMSO at a final concentration of 5-10% proved sufficient to increase the product size from 200bp to the expected size of 1.5Kbp. Similarly it was hoped that inclusion of DMSO in amplification of the OMP gene would be capable of allowing specific amplification of the whole gene.

35 cycle 5uun-3uun PCRs were set up at T^a=45°C with DMSO at a final concentration of 2, 5 or 10%. At 10% a reasonably strong 1.6Kb product was visible in the genomic track. The genomic product was very weak at lower concentrations of DMSO although the cosmid amplified extremely well at all DMSO concentrations. The experiment was repeated with DMSO at a final concentration of 11, 12, and 13% and the level of genomic product increased with each increase in DMSO. Tests with 15 and 20% DMSO showed that it began to

inhibit the reaction above 13%. All further amplification of the whole gene included 13% DMSO in the reaction.

In the same way inclusion of DMSO in the amplification of the 5' end of the gene at 10% was sufficient to produce the expected 816bp genomic product, where combinations of all of the available 5' primers and nested PCRs on the whole gene product had shown little success.

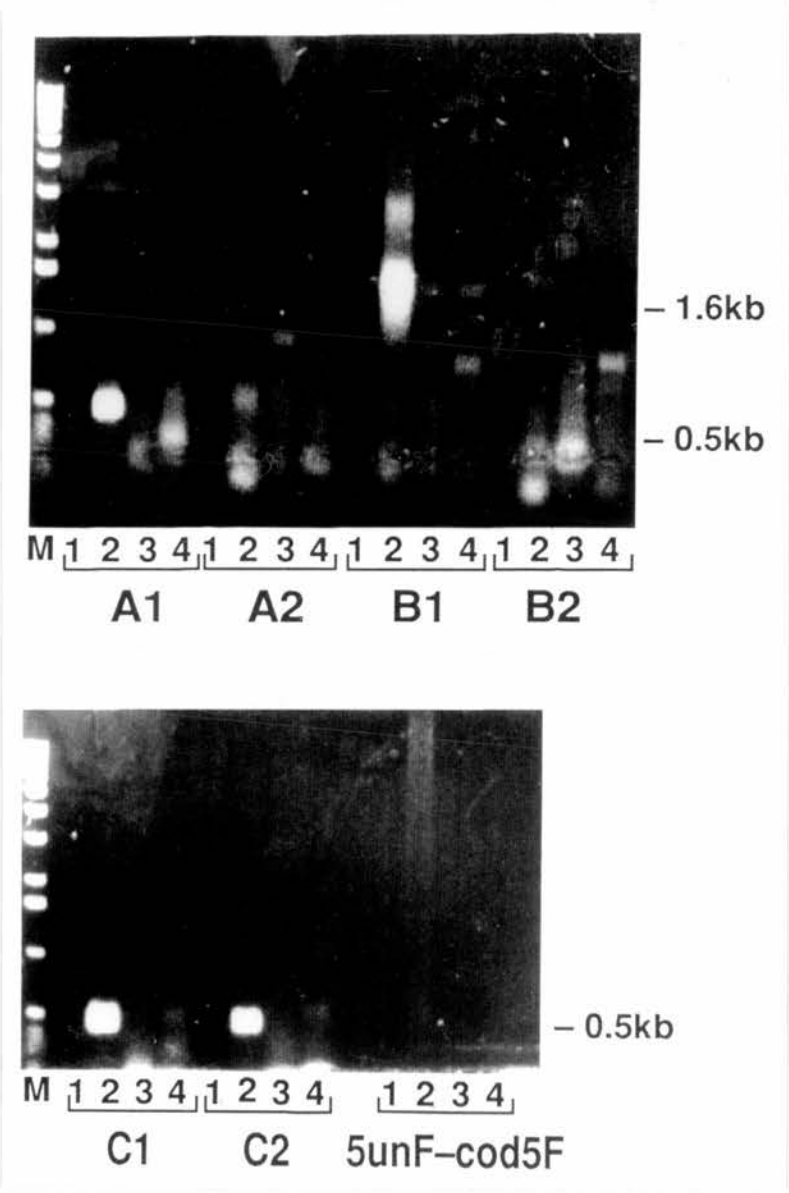


Figure 6.1 Attempt to amplify the 5' untranslated region of the OMP gene

The results of attempts to amplify the whole OMP gene are shown, A1-C2; see text for conditions used. Also shown is an attempt to amplify only the 5' end of the gene with primers 5unF and cod5F using 35 amplification cycles at 57°C annealing. Only B1 gave product of the correct size, although the amount of product obtained from genomic DNA was very poor.

Key: M = 1kb ladder marker (Gibco); 1 = blank; 2 = OMP cosmid DNA; 3-7 = human genomic DNA samples.

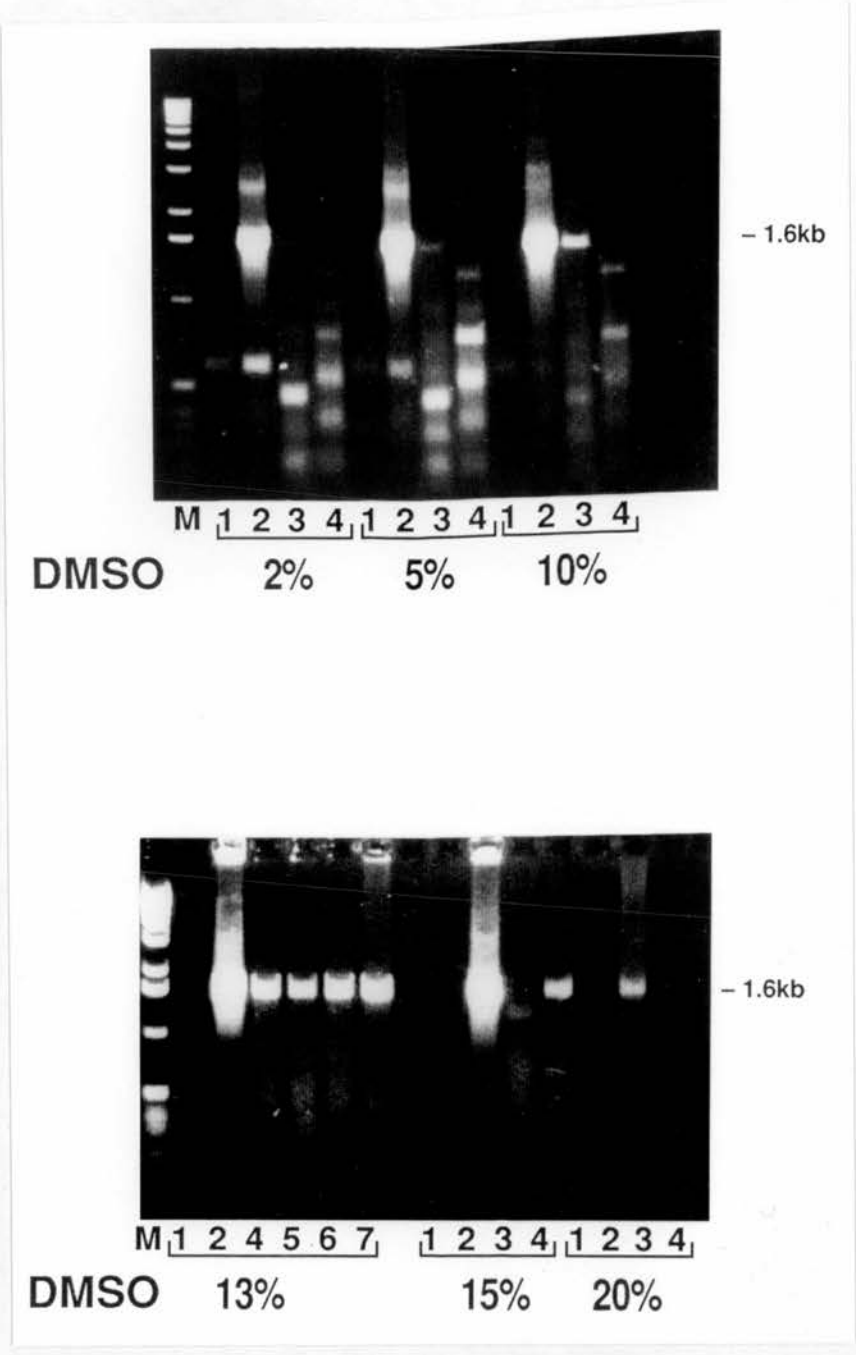


Figure 6.2 Successful amplification upon inclusion of DMSO

Reactions shown here were carried out in the presence of increasing amounts of DMSO, as indicated. Increasing the concentration of DMSO increased the amount of product. The reaction was found to be optimal in the presence of 13% DMSO, increasing the concentration beyond this point appeared to be inhibitory to the reaction.

Key: M = 1kb ladder marker (Gibco); 1 = blank; 2 = OMP cosmid DNA; 3-7 = human genomic DNA samples.

their locations are shown below (Fig. 6.3). Rodent primers were also tested in combination with human primers which lay within the coding region, and the relative positions of these are also shown.

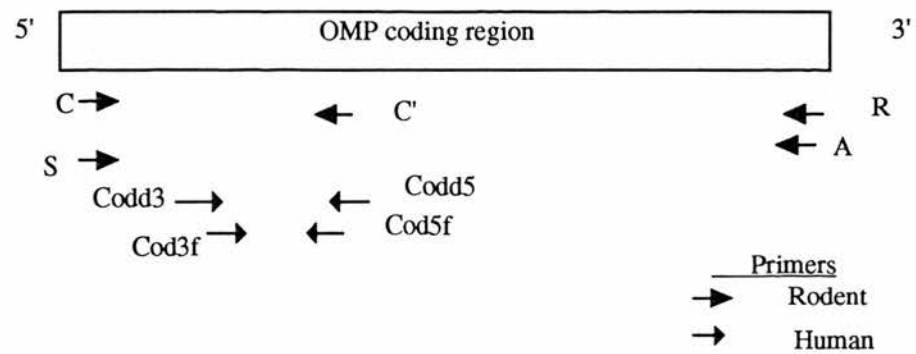


Figure 6.3 The location of primers used in interspecies comparison

Primers A and R match rat and mouse, primer C is designed to rat and primer S to mouse. Primers C and S begin at the first base after the start codon; primer C' is complementary to bases 239-259 in the coding region; primer R begins 4 bases downstream from the stop codon, and includes it in its sequence; primer A is nested within primer R and is complementary to bases 465-487. The relative location of the human primers is also shown.

All primer combinations were tested on genomic DNA from human (F4-human DNA obtained from a female placenta), mouse (RAG1- a mouse cell line), and rat (FA2A- a rat liver cell line from Dr Robert Hill). In the end, the primers C and C' were selected for amplification, since they gave a strong specific product. These primers can be used to amplify a section at the 5' end of the coding region (bases 3-259). C' was biotinylated and as such could be used to isolate single strands for sequencing. The non biotinylated strand could then be sequenced with the fluorescently labelled human primer cod5f, which was perfectly matched to the

rodent sequence. None of the primer combinations tested gave a specific human product hence it was decided to go ahead and use orphan peak analysis to compare rat and mouse C-C' PCR product, sequencing the non biotinylated strand with the human primer cod5f.

6.1.1.2 Orphan peak analysis

Here only two samples were being analysed so two different base reactions could be loaded adjacent to each other, as indicated below.

Table 6.1 Colours applied to base specific sequencing reactions

The A.L.F. automated DNA sequencer normally draws the four base profiles of a sample in different colours. Here rat and mouse base specific sequencing reactions were loaded in alternating lanes, such that those from A and C dideoxynucleotides (and similarly G and T) would be treated as one sample.

Lane	Associated Colour	Base loaded	Species
A	Green	A	Rat
C	Blue	A	Mouse
G	Yellow	C	Rat
T	Red	C	Mouse
A	Green	G	Rat
C	Blue	G	Mouse
G	Yellow	T	Rat
T	Red	T	Mouse

The region sequenced using the primer cod5f includes several differences between mouse and rat. It was possible to predict the nature of the expected orphan peak, and the time point at which it should occur. Similarly the time point and nature of any unknown mutations will be easily detected. Identifying the time point required

a standard sequencing reaction, using the cod5f primer, to be run on the same gel as the samples being analysed. The differences between rat and mouse and the expected result are tabulated below. All of the differences are on the antisense strand since this is the strand sequenced by cod5f. Normally, when loading as indicated above, each blue peak would be expected to be overlaid with a green peak and each yellow peak with a red peak, any unmatched peaks or overlaying of the wrong coloured peaks would indicate that the sequences differ at this point.

Table 6.2 Sequence differences and expected orphan peaks

The base differences, on the antisense strand, and their position in the coding region between rat and mouse are shown. The unmatched (orphan peaks) or overlaid peaks that these differences should produce are also indicated.

Base in coding region	Nature of difference	Expected peak
36	A rat - C mouse	Green (A) - Red (C)
66	A rat - G mouse	Green (A) orphan/ Blue (G) orphan
105	G rat - A mouse	Green (G) orphan/ Blue(A) orphan
138	A rat - G mouse	Green (A) orphan/ Blue (G) orphan
155	A rat - G mouse	Green (A) orphan/ Blue (G) orphan
167	G rat - T mouse	Green (G)-Red (T)

It became clear that it was imperative to have a good quality sequencing reaction in order to clearly identify orphan peaks. Poor quality sequencing was difficult to align and in such cases even matched peaks were difficult to clearly observe. Only the A/G mismatch at base 66 was clearly identified as an orphan peak (see Fig.6.4). The result was however so unequivocal that it was obviously worthwhile pursuing this method for mutational analyses of Usher IB patients.

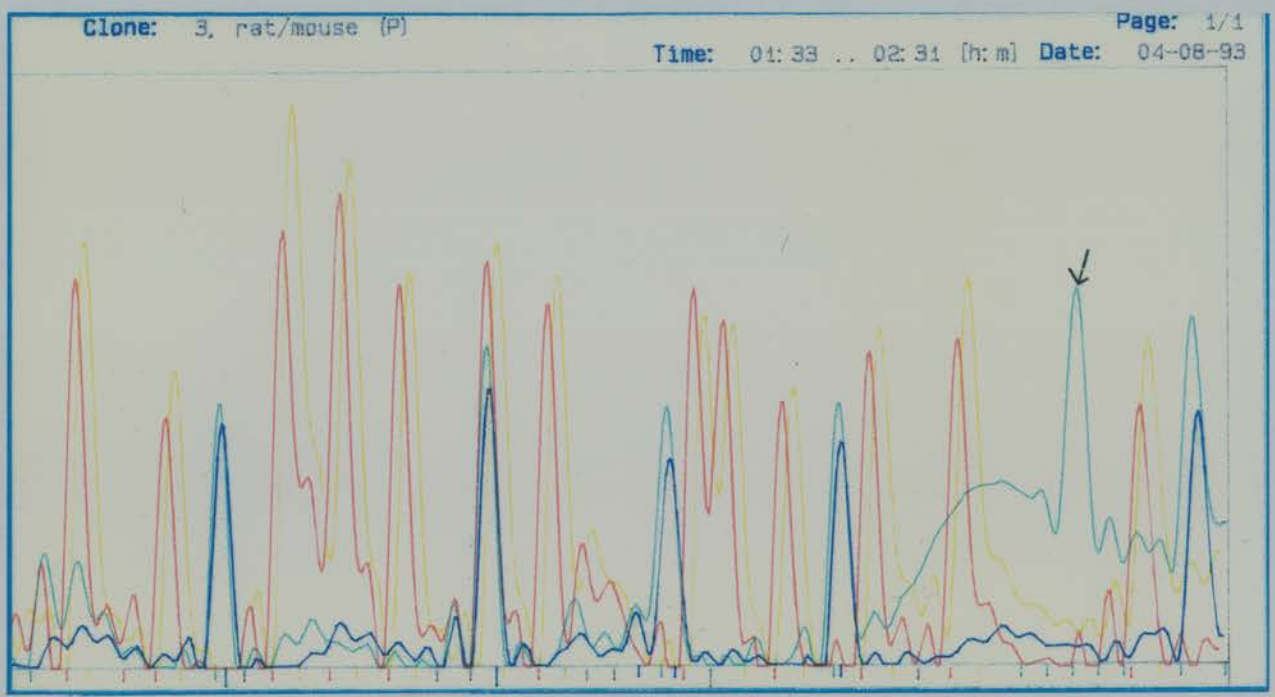


Figure 6.4 Orphan peak

Here the relative position of the A and C nucleotides in rat and mouse are shown. The arrow indicates the mismatched A nucleotide in the rat. The equivalent base in mouse is a G which had not been run here hence an orphan peak results.

Key: green = rat A; blue = mouse A; yellow = rat C; and red = mouse C.

6.1.2 Testing the technique by interhuman comparison

To test the efficacy of the method further and optimise the conditions, it was decided to compare four different normal genomic samples that were available in the laboratory. From each a single 960bp PCR product was amplified, using the primers codd3 and 3uun (see Fig. 5.3). Using the biotinylated version of either primer a single strand could then be isolated and sequenced using the appropriate internal fluorescently labelled primer. Multiple attempts were made at sequencing the biotinylated and non-biotinylated strands of several different genomic samples. Only the non-biotinylated strand was ever successfully sequenced and of these reactions all were fairly poor apart from one where 625bp of sequence was read. No reasonable explanation could be offered by Pharmacia or Dynabeads for the inconsistency of the method, as at all stages the conditions used appeared to be optimal. An experiment was undertaken to look at the efficiency of primer biotinylation. 300ng of primer 3uunB was run out on a gel adjacent to the supernatant from binding 300ng of primer to 20µl of Dynabeads. The primer lane showed a band of the expected size and intensity. The supernatant lane however contained no visible band indicating that all of the primer had been efficiently bound. In addition a sample of 300ng of 3uunB combined with 1µg of streptavidin immunoperoxidase, showed complete gel retardation of the primer on gel electrophoresis. Thus the problem lay not with the primer, but at some later stage. Since only the non-biotinylated strand was ever sequenced it was possible that the beads were somehow inhibiting the sequencing reaction.

An attempt was also made to directly sequence the double stranded PCR products. A PCR reaction was split into two aliquots. For one aliquot, single biotinylated and non biotinylated strands were isolated and sequenced as before. Samples of the remaining aliquot were purified using either mung bean nuclease (digests PCR primers and misprimed, single stranded amplification products) or isopropanol precipitation and cycle sequenced with each of the fluorescently labelled primers. These attempts were also unsuccessful. At this point the usage of the ALF sequencer became entirely devoted to the analysis of CA repeat polymorphisms and an ABI automated sequencer became available instead for sequencing. The success with this machine in completing the consensus sequence of the OMP gene (Fig. 5.4) indicated that more consistent results could be obtained and therefore it could also be used to look for differences in the OMP gene of patients.

6.2 ABI cycle sequencing

Orphan peak analysis, however, could not be carried out with the ABI sequencer since each dideoxy nucleotide is labelled with a different fluorescent dye allowing the entire sequencing reaction to be carried out in a single tube and run out in a single gel lane. Primers labelled with different fluorochromes could be used, but obtaining enough primers to cover the whole gene would prove prohibitively expensive. The ABI cycle sequencing kit is also very expensive, and all of the gene would have to be sequenced in each individual in order to identify potential mutations, hence the decision was made to use the HOT technique to identify

putative mutations and confirm these results by ABI cycle sequencing of the relevant part of the gene.

6.3 Chemical mismatch analysis / HOT

Equipment and expertise were available in the Unit for all three major mutational analysis techniques (SSCP, DGGE, & HOT) generally used in the detection of single base changes. It was decided to screen a small number of patients using HOT since this should detect 100% of all mismatches. The nature of the technique is such that the point at which a base change occurs is indicated by the size of the fragment on a gel, and that position can then be sequenced providing appropriate primers are available (see introduction for detailed description). Using labelled sense and antisense DNA meant that any cleavage fragments could be due to a base difference on either strand, thus for every fragment there were two possible points at which the mismatch could have occurred. Not only does the HOT technique give an accurate indication as to where the mismatch has occurred, it also indicates the nature of the mismatch. This gives less chance of ambiguities and compressions being mistaken for polymorphisms/mutations. HOT is a highly user dependent technique. The clarity of the results very much depend upon the level of expertise. The saving grace of the technique is that it is accurate, detects all types of mismatches and you often have two chances to identify a mutation. If both labelled wild type and patient DNA are used the mismatch is often expected to produce a fragment from each, so with a 50% percent success rate with respect to the chemical reaction you would still expect to detect 100% of the mismatches.

In these experiments the cosmid was used as the control DNA because this had been used to determine the consensus sequence.

It was decided to carry out HOT analysis on the two separate PCRs of the 5' and 3' end of the gene since these overlapped in the coding region, and were of a more manageable size for electrophoresis in a polyacrylamide gel. These were known as C5 and C3 PCRs and covered bases 2-817 and 644-1605 respectively (Fig. 6.5).

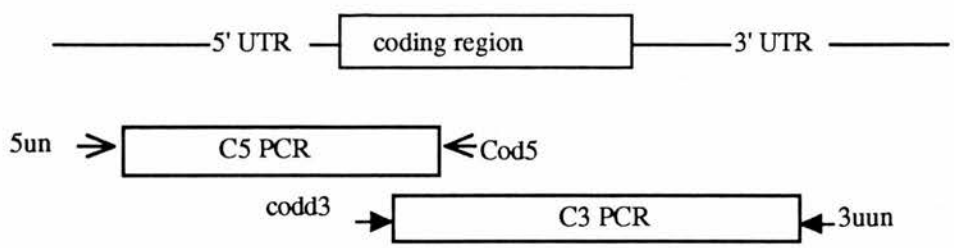


Figure 6.5 The PCR fragments used for HOT analysis

The two PCR fragments used in HOT analysis, and how these relate to the gene structure are shown. C5 PCR amplifies bases 2-817 and produces a fragment of 816bp, and C3 PCR amplifies bases 644-1605 and produces a fragment of 962 base pairs.

Nine patients were chosen for analysis. These were selected on the basis that either an affected sib or additional family member DNA was available, hence any potential mutations could quickly be followed through in the family and confirmed. Mismatches would be classified as mutations if they were seen in affected and not in normals, and altered to the coding sequence. Mismatches would be classified as polymorphisms if they were seen in both affected and normals. The presence of a

mismatch in only one of a sib pair could be an indication that it was due to a polymorphism or a new mutation.

The results are shown in two separate tables corresponding to the two different fragments analysed. These are then followed by the sequence of the gene annotated to show where the mismatches might lie and the primers designed to cover the region.

Table 6.3 Results of HOT analysis on the C5 PCR product

The size of products obtained by HOT analysis of the C5 815bp PCR fragment. PNo. indicates the laboratory number given to the samples which are described in more detail in table 2.7. Sizes are in base pairs. NC = no cleavage products observed.

	Patient labelled		Cosmid labelled	
	Hydroxylamine	Osmium tetroxide	Hydroxylamine	Osmium tetroxide
P3	550	172/550	NC	NC
P4	550	550	NC	NC
P9	NC	NC	NC	NC
P10	NC	172/550	NC	NC
P16	NC	NC	NC	NC
P17	73/550	73/172	NC	NC
P18	NC	NC	NC	NC
P21	NC	NC	NC	NC
P25	550	550	NC	NC

Table 6.4 Results of HOT analysis on the C3 PCR product

The size of products obtained by HOT analysis of the C3 962bp PCR fragment. Sizes are in base pairs. NC = no cleavage products observed.

	Patient labelled		Cosmid labelled	
	Hydroxylamine	Osmium tetroxide	Hydroxylamine	Osmium tetroxide
P3	409/800	NC	454	NC
P4	112/144/409 / 800	112/144	NC	NC
P9	800	112/144	454	NC
P10	800	NC	65/454	NC
P16	409/800	NC	NC	NC
P17	112/144/800	112/144	NC	NC
P18	800	NC	NC	NC
P21	112/144/409 / 800	112/144	NC	NC
P25	800	NC	NC	NC

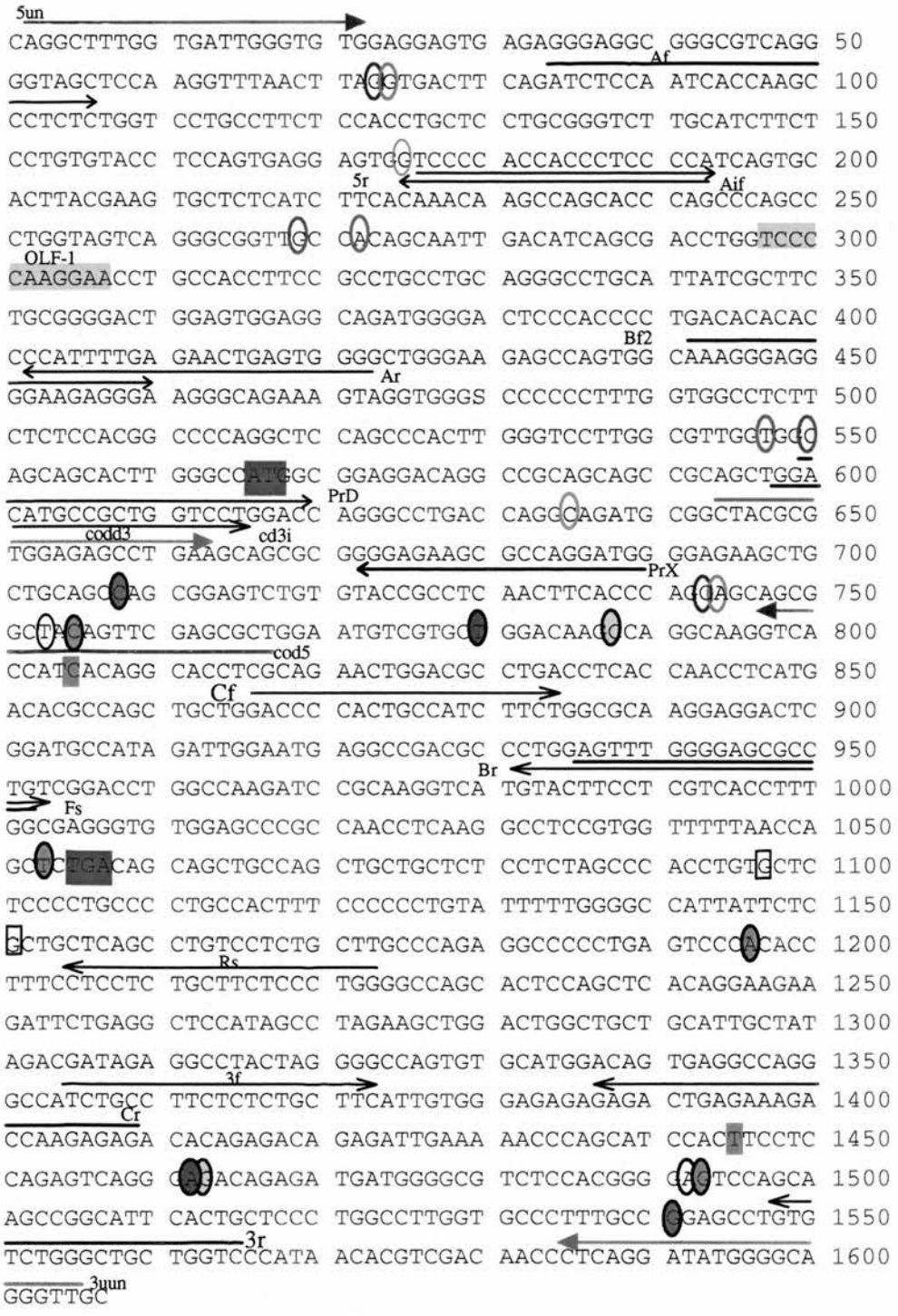


Figure 6.6 Compilation of potential mismatches in the OMP gene of Usher IB affected individuals

Boxed sequence indicates the binding side for the Olf-1 transcription factor, start (ATG)and stop (TGA) codons. Primers used in PCR amplification are represented as red (C3) and dark blue (C5) arrows. All other arrows represent primers used in sequencing. Ovals and squares indicate potential position of mismatches, each pair corresponding to the two possible sites.

Initially attempts were made to resolve the mismatches by ABI cycle sequencing. Each of the patients which appeared to have a particular mismatch by HOT were sequenced through the two possible regions where the mismatch may have occurred. The sequences were then overlaid using the GCG pile up command, to compare and highlight the differences between each of these and the consensus sequence. Problems were encountered at this stage which are inherent to sequencing with this particular kit. The conditions used are such that Taq polymerase does not incorporate all dideoxynucleotides with the same efficiency. The efficiency of incorporation is influenced by local base composition and the structure of the different dyes attached to the ddNTPs. This results in a number of characteristic base patterns: C's following G's are often weak; T's following G's are often weak; in a string of four or more G's the third G often shows a reduced signal, and the following G's are elevated; and A's after T's can show reduced signal. Point mutations leading to a nonfunctional gene may be identical in both alleles, but sometimes each allele will be mutated at a different point resulting in a compound heterozygote. In such a situation a mismatch would be visualised as a reduction in peak size, and it quickly became clear that a difference in just one allele would be almost impossible to resolve from inefficient incorporation of a particular base. Sequencing on the opposite strand reverses the orientation of the bases and therefore resolves the ambiguity, although at the same time creates more ambiguities particular to the base composition of this strand. This meant that for each patient at least four sequencing reactions were required to resolve each mismatch, and any mismatch had to be detected by visually scanning the

chromatograms obtained by sequencing through the region in both orientations and comparing these to control sample. It was decided therefore that it would be more accurate to carry out ^{35}S sequencing on the PCR products, in the presence of DMSO.

In each case the size of the HOT product was estimated from a sequencing reaction of the PstI clone which was run alongside (see Tables 6.3 & 6.4). Both potential sites of the mismatch were then sequenced and the presence of a mismatch confirmed if the base change would have resulted in the observed HOT result. Thus a number of mismatches were confirmed in the human OMP gene as described below.

C5 PCR fragment sequencing results

P17 patient DNA produced a very strong 73bp fragment when heteroduplexed to cosmid and treated with osmium tetroxide (Table 6.3). Sequencing with primer 5r showed the mismatch was not in the 5' untranslated region. Subsequent sequencing with primer cod5 revealed a C/A mismatch at base 743 (Fig. 6.10). This would result in the T nucleotide on the antisense strand being mismatched and would produce the observed 73bp fragment when treated with osmium tetroxide. The complementary 743bp product resulting from hydroxylamine treated labelled cosmid was also faintly visible. In addition, a band of 73bp was also faintly visible in the patient hydroxylamine track, which was probably due to a slight effect of the mismatched T on the adjacent C nucleotide. This mismatch should also be covered by the C3 PCR fragment (see Fig. 6.6), and would be expected to produce a band

of 862bp (when labelled patient heteroduplex is treated with osmium tetroxide). The failure to detect this band is most likely due to its proximity to the excess of full length end labelled fragments. The 99bp fragment expected from hydroxylamine treatment of the cosmid was also not observed. The observed nucleotide change from C to A would change amino acid 59 from a glutamine (CAG) to lysine (AAG).

A band of approximately 634bp was observed with hydroxylamine treated heteroduplex between labelled cosmid and P25 patient DNA (Table 6.3). Sequencing with primer X revealed no mismatches around base 634, which meant the change had to have occurred in the 5' untranslated region probably at base 175, since this is the closest G base 182 (816-634=182). Unfortunately primer Af gave very poor sequencing data, and the mismatch lay within primers Aif and 5r (see Fig. 6.6), therefore the status of this putative mismatch was unresolved.

Several fragments, in particular those at 172bp and 550bp (Table 6.3) were observed in reactions analysed from the C5 PCRs which could not be confirmed by sequencing. They appeared to be gel or reaction dependent, such that analysis carried out on the same PCR product on a different occasion did not show the same fragments. These may have resulted from secondary structure formation in the highly GC rich 5' untranslated region, particularly the 172bp fragment since sequencing with primer 5r revealed a compression in this region.

C3 PCR fragment sequencing results

The 65bp fragment produced by hydroxylamine treatment of labelled cosmid heteroduplexed to P10 patient DNA (Table 6.4) could have been either due to mismatched C at 708 or G at 1540 in the gene (Fig. 6.6). Sequencing with primer cod5 showed it was not at base 708 and hence not in the coding region. Subsequent sequencing with primer 3f revealed a G/A mismatch at 1540 (Fig. 6.10). The wild type sequence was G at this point hence on labelled wild type antisense strand a C/A mismatch would occur resulting in the 65bp cosmid-hydroxylamine product. Similarly a T/G mismatch would be expected to be picked up by osmium tetroxide treatment of labelled patient heteroduplex. This was not observed.

The 409bp band in patients 3,4,16, and 21 produced by hydroxylamine treatment of labelled patient heteroduplex DNA (Fig. 6.5 and Table 6.4) could have been due to a mismatch at base 1196 or 1053 (Fig. 6.6). Sequencing with primer Fs gave perfect sequence at 1053, whereas primer Cr revealed a G/A polymorphism at base 1196 in patient 3,4 and 21 and a change from A to G in both alleles from P16 patient DNA (Fig. 6.8). The wild type sequence at this point was an A, hence on labelled patient antisense strand a C/A mismatch would occur resulting in the 409bp patient hydroxylamine product. The complementary cosmid labelled osmium tetroxide 552bp band was not seen.

The 454bp hydroxylamine treated cosmid labelled band in patients 3,9, and 10 (Fig. 6.7 and Table 6.4) could have been due to a mismatch at base 1098 or 1151 (Fig.

6.6). Sequencing with primers Fs (Fig. 6.8) and Rs revealed the cause to be a G/A polymorphism at base 1151. The wild type sequence was G at this point hence on labelled wild type antisense strand a C/A mismatch would occur resulting in the 454bp cosmid-hydroxylamine product. The complementary patient labelled osmium tetroxide product was not however observed.

Potentially the most interesting product was the 800bp fragment observed for all the patients by hydroxylamine modification, which could have been due to a mismatch at base 805 or 1444. ^{35}S sequencing with primer Br was unable to resolve the region around 800bp and resulted in a large compression. ABI cycle sequencing, however, revealed a G in the patients at base 790 whereas the consensus sequence at this point was an A. Hence on labelled patient antisense strand a C/A mismatch would result giving the observed 800bp patient-hydroxylamine product. The complementary cosmid labelled, osmium tetroxide treated 790bp product, due to a T/G mismatch, was not seen. This mismatch lay within the coding region, but unfortunately changed the third base of the triplet encoding proline from CCC to CCA hence still coding for proline. Both patient alleles differed from the consensus at this point. Consequently, it was the only base change to be unequivocally detected by ABI sequencing. At a later date another group published the OMP gene sequence (Buiakova et al., 1994) and comparison of my sequence to this revealed the change was probably due to a mutation in the cosmid sequence and the patients in fact matched the wild type consensus sequence.

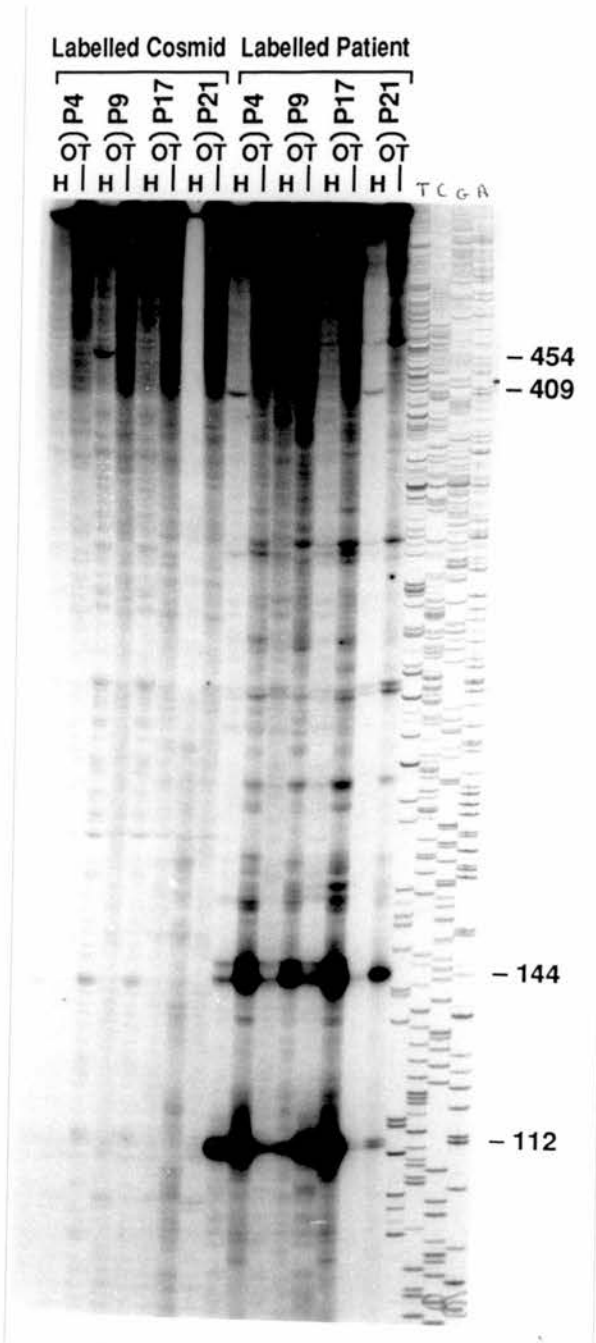


Figure 6.7 HOT analysis of the C3 PCR fragment

An aliquot of a sequencing reaction of the OMP PstI clone was run alongside the products of HOT analysis of the C3 fragment. This allowed bands produced by mismatched bases to be sized. Shown here is the 454bp band in patient 9 produced by hydroxylamine treated heteroduplex in which the cosmid DNA was labelled; the 409bp band in patients 4 and 21 produced by hydroxylamine treatment of patient labelled heteroduplex DNA; the 112 and 144bp bands in patients 4, 9, 17 and 21 detected by both hydroxlyamine and osmium tetroxide treatment of patient labelled heteroduplex DNA.

Key: H = hydroxylamine; OT = osmium tetroxide

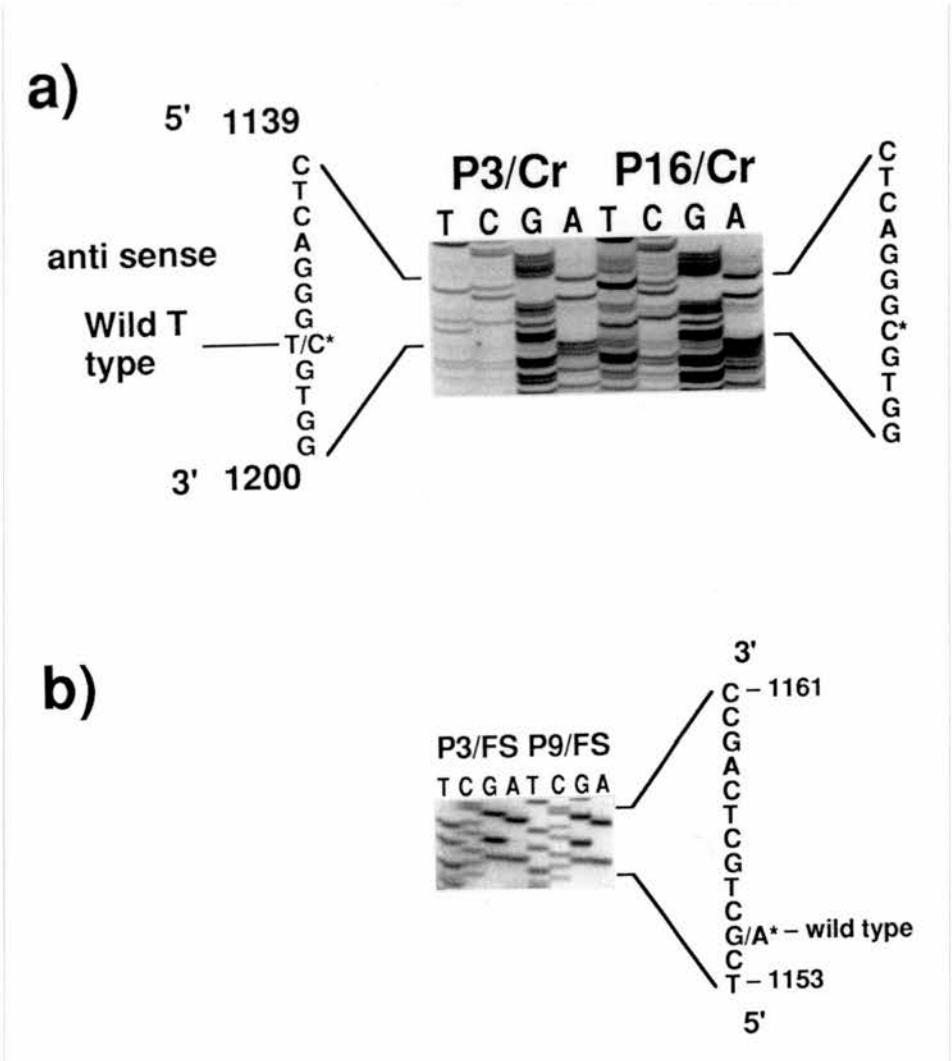


Figure 6.8 Sequence analysis of mismatched bases

a) The T normally found on the antisense strand at base 1196 in wild type DNA, has been mutated to a C in one allele from patient 3, whereas both alleles in patient 16 have been altered.

b) The G normally found in wild type DNA has been mutated to an A in one allele from patients 3 and 9.

In common with the C5 PCR fragment, several bands were observed in HOTs on the C3 PCR fragment which appeared to be gel or reaction dependent. In general these were ruled out due to their inconsistency. Two bands, however, were particularly strong. These were the 112 & 144bp products, detected in patients 4,9,17 and 21 by hydroxylamine and osmium tetroxide treatment of patient labelled heteroduplex (Fig. 6.7 and Table 6.4). These could have been due to mismatches at 852bp, 820bp or 1495bp, 1463bp respectively (Fig. 6.6). Sequencing with primer 3f indicated that there were no polymorphisms at 1463 or 1495. Similarly sequencing with primer Br indicated an exact match to the consensus at 852bp. The sequence around base 820 was extremely compressed and impossible to read. Thus it was impossible to rule out the fact that the 144bp band might have resulted from a mismatch in this region. The sequencing of the 3' untranslated region with 3r gave a very interesting result. Those patients which gave the two bands and patient 18 (sib of patient 17) all produced sequence which appeared to be overlaid until 1350bp where there was a strong compression, after which the sequence was clear and easy to read. In patient P3 , however, which did not give the two bands (112 & 144bp) in HOTs there was a compression at 1350, but prior to this point the sequence was still easy to read (Fig. 6.9). It appeared that in the other samples some sort of inversion or small deletion had occurred which put the sequence out of alignment downstream of base 1350. This could mean that primer 3f was only sequencing one allele and hence explain why no mismatch was detected. In order to test the theory that an inversion might have occurred, nested PCRs on a 1:100 dilution of the C3 PCR product were set up as indicated below.

The products obtained from patient P4 and the control patient P3 were all of the size expected in the normal. Thus the overlaying of the sequence had not resulted from an inversion, but probably from a small (since the PCR product size appeared to be unaltered) deletion then insertion or vice versa at base 112 and base 144. The exact nature of the change could be determined by cloning and sequencing each of the alleles independently

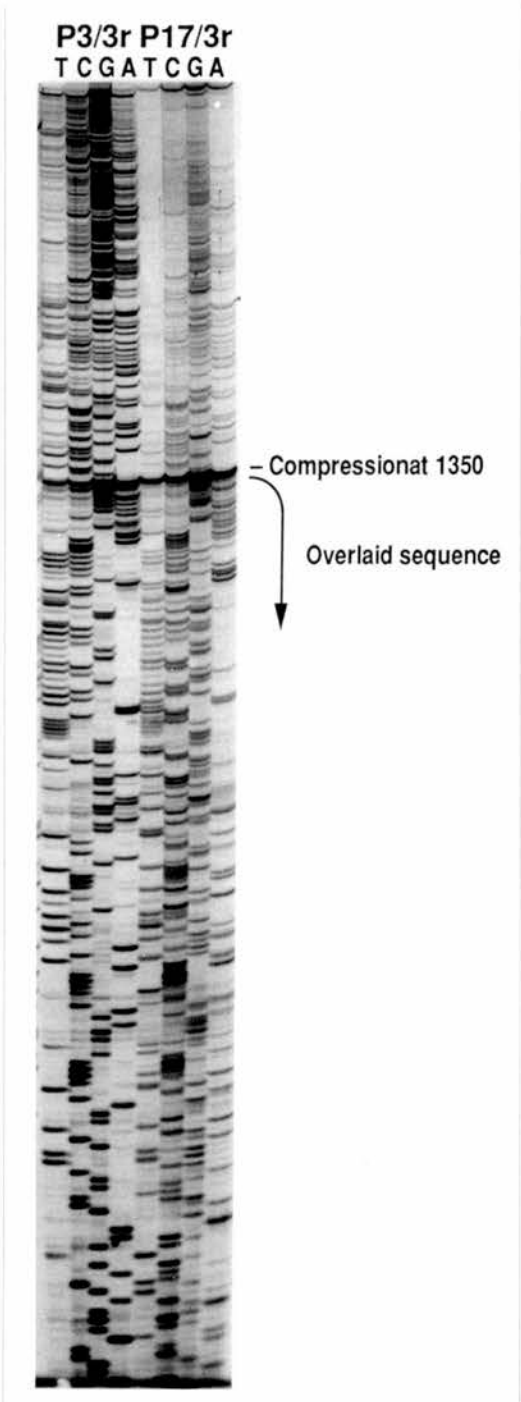
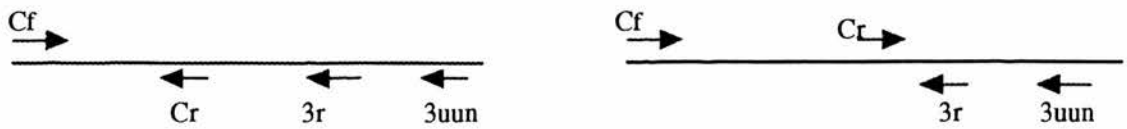


Figure 6.9 Sequence of the 3' untranslated region

Here sequencing reactions of P3 and P17 patient DNA with primer 3r have been run out adjacent to each other. Sequence in P17, while fainter then P3, was easy to read beyond 1350, where a compression has occurred. Prior to this point, however, the sequence in P17 contained many more bands than P3 and indeed appeared to be overlaid with a different sequence.



Primers	Normal primer order	Inverted primer order
	Expected products	Expected products
CF-3uun	⇒740bp	⇒740bp
Cf-Cr	⇒ 543bp	⇒same strand so no product
Cr 3uun	⇒same strand so no product	⇒197bp

Figure 6.10 Orientation of primers in the presence of an inversion

The line drawing on the left indicates the orientation of the primers as they were designed, whereas the right-hand drawing indicates the orientation if an inversion occurred between primers Cf and 3r. If such an inversion had occurred, primer Cr would be complementary to the opposite strand and a different set of products would result as indicated The product sizes expected using the indicated primers for each orientation are given.

Figure 6.11 Confirmed mismatches in the OMP gene of Usher IB affected individuals

The consensus sequence of the human OMP gene. Yellow and green boxed sequence indicates the binding site for the Olf-1 transcription factor, and start (ATG) and stop (TGA) codons respectively. Ovals and squares indicate confirmed mismatches, the normal base is shown above, the colour of these correspond to the potential mismatches shown in figure 6.6.

6.4 Comparative Mapping

Having excluded Omp as a candidate for *Shaker-1,* Steve Brown's group began to construct a YAC contig across the *sh-1* region using Omp as a startpoint. Using microsatellites they were able to delineate the extent of the non-recombinant region and from this several exon traps were isolated (Gibson et al., 1994). Two of these, ET17 and ET58, were shown to contain open reading frames, and be conserved in a number of species. Hence they were hybridised to a somatic cell hybrid panel containing sub fragments of 11q to see if they mapped to the region linked to USH1B.

The hybrid panel used, and the preparation of the Southern blots for mapping have been previously described in section 4.1. The exon traps had been cloned into a vector which had sequence complementary to M13 primers flanking the cloning site. The PCR amplified insert was isolated as a gel fragment which was labelled by random priming, then hybridised to a Southern blot of the somatic cell hybrid panel. In this manner ET58 was mapped to the same interval as OMP and hence to the region linked to USH1B (Table 4.2). ET17 did not produce any signal on the panel, which was not unexpected since on hybridisation to a zoo blot it failed to give any signal in the human track.

Benoît Arvéiler, while working in our laboratory, began isolating a human OMP YAC contig, and 5 of the YACs which he had isolated were available for screening. This was done by making a Southern blot of an EcoRI digest of these YACs and hybridising it with labelled ET58 insert as described above. Hence the

two YACs which were positive for OMP were also shown to be positive for ET58. These YACs were sized by preparing agarose plugs which were then run on a pulsed field gel. The YACs were not clearly distinguishable from the yeast chromosomal background, so the gel was Southern blotted and probed with ^{32}P labelled CotI DNA to detect human specific DNA (see Figure 6.12). Thus both YACs were shown to be 330kb and might in fact contain the same EcoRI fragment. The two genes were now known to lie within the same 330Kb EcoRI fragment.

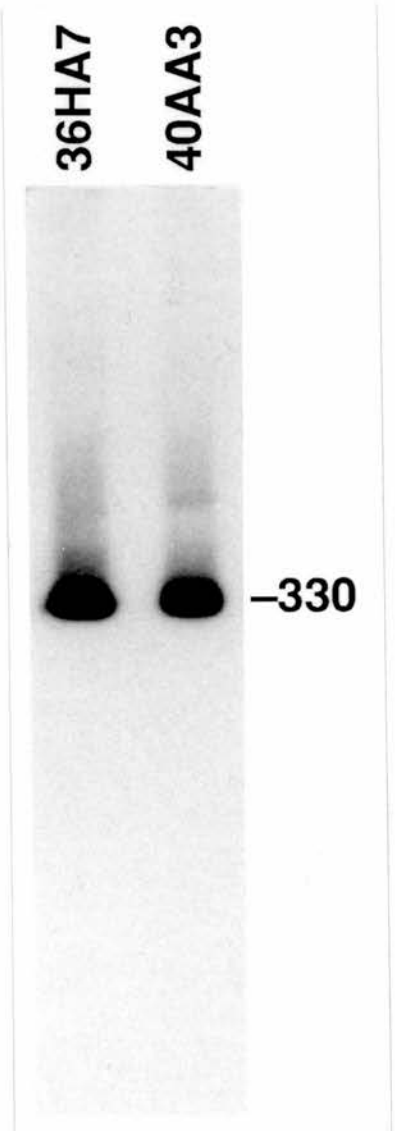


Figure 6.12 Sizing the OMP YACs

Agarose plugs were prepared and run on a CHEF pulsed field gel as described in material and methods. Some of the YACs were the same size as the yeast chromosomes hence the gel was Southern blotted and hybridised with CotI DNA to identify the YACs. Hence each could be roughly sized relative the markers, and both were in fact 330Kb.

6.5 Discussion

The overall results of mutational analysis on the human OMP gene are illustrated in Fig. 6.11. Only one of the HOT products resulted from a change at the amino acid level. Amino acid 59 lies within a region which is relatively poorly conserved between human and rodents. There are 5 amino acid differences between aa56 to aa67, and hence a change at this point may have very little effect on the protein product. This change was not shared by patient P18 who was the sib of patient P17, indicating that it did not co-segregate with the disorder in this family, given that both patient P17 and patient P18 are affected. This made OMP unlikely to be the cause of Usher syndrome in this family. Given the proximity of OMP to ET58, which was later shown to be part of the myosin type VII gene (see section 7), there is a possibility that this family is not linked to USHIB. Analysis of the haplotypes segregating in this family would confirm if they were unlinked, or if a recombination event had occurred between OMP and the disease locus. All of the other changes were located in the untranslated region and were therefore unlikely to have any affect on function of OMP. A single base alteration in the untranslated region is unlikely to affect the function or expression of OMP. Thus it was unlikely that OMP was the gene underlying Usher syndrome. In parallel to carrying out mutational analysis on OMP, I was carrying out comparative mapping using exon traps isolated from the shaker-1 region in mouse, to see if any other likely candidates could be identified. The results of this analysis are discussed in the next section.

6.5.1 The efficacy of the different mutation detection methods used

No mismatches were detected by sequencing which HOT had failed to identify. Several expected products were not seen in HOT analysis, although the mismatch was identified by the presence of the complementary product. Of the 5 mismatches which were not seen, 3 were T-G mismatches which are frequently missed (Jane Prosser, personal communication). The remaining 2 were a T-C and a C-T mismatch, while these are normally easy to detect they may have been missed due to effects of the sequence context around the mismatch. Since the release of the original paper on the HOT technique (Cotton et al., 1988) further analysis on the reactivity of mispaired bases has been carried out. Bhattacharyya et al (1989) observed that mismatching can result in chemical hyper-reactivity, which can depend on the flanking sequences. Thus, some at least of the mismatched base-pairs contain the potential to be recognised as chemically different, but this may be modulated by sequence context. This, however, should not deter use of the technique since as demonstrated here all of the mismatches are generally identified, and the absence of the complementary product is therefore insignificant.

Perhaps more worrying was the identification of strong false positives in the c5 product at 172 and 550bp. False positives are not generally observed in the HOT technique. Most studies however concentrate on analysis of the coding region, and it is feasible that sequences in the 5' untranslated region, which is highly GC rich, are looping out to form stem loops which contain mismatches and result in false positives.

7. Discussion

Summary and concluding remarks

In the final stage of confirming base differences in the OMP gene of USH1B affected sib pairs, mutations were identified in a novel myosin type VII gene in *shaker-1* mice and USHIB patients. The approach taken to identify the gene for USHIB was that of the positional candidate (Collins 1995, see later), where mapping to a chromosomal region was combined with a search for suitable candidate genes in that region. Selection of a candidate gene depends on a number of factors: expression in an appropriate tissue; location in the chromosomal region linked to the disorder; and where the function is known it must be reconcilable with the phenotype observed. Meeting these criteria does not necessarily distinguish the actual gene from other likely candidates. The only way to do this is to show (in affected individuals) there are no mutations in the false candidates; identify mutations in the real gene; or as was the case with OMP a combination of both. Candidate genes must be tested and ruling out a candidate gene is an important step to finding the real gene.

In USHII, phosducin and hCHML were both screened for mutations in affected individuals (section 1.2.4.2) with negative results, and hence ruled out as candidates. In developmental genetics, candidate genes must also be screened after applying a similar set of selection criteria. Hence ZFY was considered as the testis

determining gene (Page et al., 1987), but ruled out when its expression pattern (Koopman et al., 1989) and absence from XX males (Palmer et al.,1989) proved incompatible with this theory. The new candidate became SRY and its mouse equivalent Sry (Sinclair et al., 1990; Gubbay et al., 1990) which were localised to the smallest region of the Y chromosome known to be male determining and conserved in other eutherian mammals. The case was strengthened when the expression of Sry in the mouse was shown to be entirely consistent with a role in sex determination (Koopman et al., 1990). In humans too SRY was shown to be important for testis development when *de novo* mutations were found in the gene in two sex reversed XY females (Berta et al., 1990; Jäger et al 1990). Conclusive proof that Sry was indeed the testis determining factor came when introduction of Sry on a genomic fragment gave rise to normal testis development in female transgenic mice (Koopman et al., 1991).

While eliminating a gene as the cause of a disorder or as a key to a particular stage in development can be disappointing, information gained in the analysis may prove useful in other investigations. This was the case for the fibroblast growth factor receptor 3 gene (FGFR3) which was initially identified during attempts to identify the Huntington's chorea disease gene (Thompsom et al 1991) and later shown to be responsible for achondroplasia (Shiang et al., 1994).

The myosin type VII gene was isolated by Gibson et al (1995) from a mouse inner ear cDNA library using the exon trap probe ET58, which was isolated from a YAC mapping to the non recombinant region surrounding the *shaker -1* locus. ET58

had been shown to contain an open reading frame conserved across a number of species and during the course of this project was mapped to the same YAC as human OMP and hence to lie within the region linked to USHIB.

There are a number of features which made myosin type VII a good candidate for both syndromes. The gene is expressed in the mouse inner ear and indeed has been shown to be transcribed in the mouse cochlea (Weil et al., 1995) where neuroepithelial defects are observed in *shaker-1* mice. Previous studies have instigated a myosin type 1 molecule as playing a pivotal role in the adaptation motor. The adaptation motor is thought to allow hair cells of the inner ear to maintain their optimal sensitivity to minute displacements by adapting to sustained stimuli at the molecular level. The protein has been shown to be expressed near the tip of stereocilia where transduction and adaptation occur (Gillespie et al., 1993) and given the features of Usher syndrome and *shaker-1* it is thought that the myosin type VII gene is also involved. Defective cilia have been observed in affected individuals (section 1.2.5) and this could explain the involvement of 3 different sensory systems in Usher syndrome, and the combination of vestibular dysfunction and sensorineural deafness in *shaker-1* mice. The most conclusive evidence was of course the identification of mutations in the head region of the myosin type VII gene in both spontaneous and ENU induced mouse *sh-1* mutants (Gibson et al., 1995). Of the seven available *sh-1* mutants screened, three mutations were confirmed: a 3' splice site mutation which produced an in frame 10 codon deletion; and two non conservative arginine to proline changes. The small

soil nematode *Caenorhabitis elegans* has been intensively studied both at the molecular genetic and the developmental level for a number of years. The major myosin heavy chain gene, which was identified during analysis of muscle defective strains, has been particularly well studied and a number of mutants and their resultant phenotypes have been described (Dibb et al., 1985). Alteration of one of the arginines which was changed to a proline in a *sh-1* mutant in *C.elegans* produces a slow moving phenotype and some thick filament disorganization, confirming the importance of this amino acid for function. Analysis of USHIB families also revealed mutations in the motor domain of the myosin type VII gene in affected individuals (Weil et al., 1995): in two individuals two different heterozygous missense mutations were observed; in another two individuals heterozygous mutations which resulted in a premature stop codon, and hence a protein truncated before the ATP and actin binding sites were detected; and in one of the latter and another unrelated individual the same heterozygous 6bp deletion within two 5bp direct repeats (suggesting replication slippage or slipped strand mispairing) was seen. The presence of two defective alleles in one individual was strong evidence that the myosin type VII gene was responsible for USHIB. The mouse and human gene were therefore proposed to be orthologous.

Every known myosin molecule comprises a conserved head domain which includes the ATP (hydrolysis of which allows myosin to convert chemical energy to force) and actin binding domains and is responsible for generation of force, and a tail domain which includes regions responsible for regulation and for interaction with

proteins or lipids (Gillespie et al., 1995). Mutations in the head domain would therefore prevent adaptation to sustained stimuli resulting in the features observed in these syndromes. The absence of retinal defects in the mouse could be explained by either the nature of the mutations observed or compensation by another myosin molecule. Neurosensory nonsyndromic deafness type two (NSRD2) has also been linked to 11q13.5 (Guilford et al., 1994). This disorder may also involve mutations in myosin type VII of a similar nature to those observed in *shaker-1*. Similarly mutations in other myosin molecules may cause other forms of sensorineural deafness.

Screening of the remainder of USHIB patients must now be carried out in order to identify mutations, and the likely involvement of other myosin genes in the remaining Usher Syndrome subtypes investigated. It may be that myosin type VII is not always the gene responsible for USHIB, and as has been shown with Retinitis Pigmentosa a number of genes may result in the syndrome (section 1.2.1.2), in which case it is important to consider other proteins which myosin type VII might interact with and hence generate the same phenotype. Obviously anything which prevents ATP or actin binding will prevent the protein from functioning; binding of Ca^{2+} appears to modulate the rate of ADP release so this pathway, which involves calmodulin, is also important (Warwick and Spudich 1987). So to is the interaction of the protein with the stereocilium membrane and the microfilament core (Gillespie et al., 1993). Myosin type VII is the first

molecule to be identified that is known to be involved in auditory transduction and hence may prove to be the key to understanding neurosensory deafness.

In this work, analysis of the USHIB region involved mutational analysis of the human OMP gene, mapping candidate genes from the *sh-1* non recombinant region, and physical mapping of human chromosome 11q12-13. Physical mapping involved the development of an improved method for isolation and efficient mapping of inter-Alu PCR products. The modified method encompasses two important features : a secondary PCR to remove Alu end sequences and hence removes the need for a long preannealing step prior to hybridisation of such products; and a cloning methodology which is nonselective resulting in a broad range of markers. This method could be used to increase the density of markers in any chromosomal region, although obviously sites with an abundance of Alu repeats would result in more markers. Future application of this method would, however, benefit from a few minor alterations:

- in addition to CotI the turboclones should be screened for MERs and other medium frequency repeats
- pools of turboclones could be used to directly isolate YACs the end clones of which could then be mapped on a somatic cell hybrid panel by Southern hybridisation, or even the whole of an identified YAC could be mapped by FISH

- the ends of turboclones could be sequenced to see if the polymorphic nature of the Alu 3' ends could be used in linkage analysis and homozygosity mapping approaches

Rapid changes are at present occurring in the realm of physical mapping. YACs are seen as a weakness in any mapping project due their frequent deletions, tendency to be chimaeric, and the inability to separate them from the yeast host chromosome background by simple methods (Zabarovski et al., 1994). Consequently several alternative cloning systems based on bacterial host systems have been developed (Monaco and Larin 1994). These show a lower frequency of chimaerism, a higher transformation efficiency in generating libraries, and it is simpler to purify insert DNA away from the host genome. The two major *E.coli* cloning systems are BACs (bacterial artificial chromosomes) which use an F-factor based vector, and PACs (P1 artificial chromosomes) where the vector incorporates features of both the P1 and F-factor systems. Ultimately the desire is to construct MACs (mammalian artificial chromosomes) which could be used for transgenic complementation analysis in tissue culture, but perhaps mostly due to their potential use in somatic cell gene therapy (Huxley 1994). First, however, human centromeres and origins of replication must be defined.

Similarly technical developments are occurring which will greatly simplify the search for candidate genes particularly in polygenic and multifactorial disorders. Many human genetic disorders are complex in their aetiology, for example diabetes type I, and identification of candidate genes may be hampered by the fact that a

number genes and environmental factors may be involved. The direct identification of candidate genes without prior knowledge of their location in the genome, or function, would represent a significant advance and may in the near future be technically feasible (Jonsson et al., 1995). Such methods rely on the concept of identity by descent (IBD) which is outlined below

- genomic sequences from two unrelated individuals are different (i.e. polymorphic)
- given low mutation rates, identical by descent segments in the genomes of two unrelated individuals are identical in sequence
- through segregation and recombination of chromosomes, sequences IBD become progressively fewer and shorter with increasing numbers of meioses separating two relatives

Hence, sequences IBD should contain genes contributing to a phenotype which is shared by distant relatives.

These provisos are not novel (Sanda et al., 1986) and indeed form the basis of homozygosity mapping which is presently in vogue (Davies et al., 1994). In homozygosity mapping regions IBD are sought by looking for sharing of microsatellite alleles in affected individuals, and a physical region is thus identified. Theoretically, however, it should be possible to take the method one stage further and directly isolate sequences IBD by mixing, denaturing, and annealing genomic DNA from affected distantly related individuals and enriching for heteroduplexes which contain no mismatches while discarding homoduplexes and heteroduplexes

with internal mismatches. Early attempts were fraught with difficulties, but current technological advances in coincident sequence cloning methodologies have made the process more tractable (Nelson et al., 1993 : Lisitsyn et al., 1993). Both methods while at an early stage have shown a lot of potential and this approach will in time save a great deal of repetitive work involved in the analysis of multiple markers.

The human genome mapping project : progress and implications

Mapping

The progress of the human genome mapping project to date has depended on developments in four major areas: the construction of a high resolution genetic map; the construction of a high resolution physical map; rapid and facile identification of expressed sequence tags (ESTs); and the development of networked information systems to permit search and retrieval over the Internet.

Development of a genetic map has been greatly facilitated by the discovery of simple sequence repeats such as CA repeats and microsatellites. These repeats occur on average every 30,000bp throughout the human genome and are generally highly polymorphic (>80% informative). In addition analysis of the degree of polymorphism of these repeats in individuals is inexpensive, fast, and technically simple - requiring a straightforward PCR (Hoffman 1994). Their primary utility is to facilitate localisation of disease genes by family based genetic linkage studies. The major source of characterised CA repeats to date has been the French Institute

Généthon which was established and funded by the French muscular dystrophy association, using monies acquired through nonprofit, private fundraising (Gyapay et al., 1994). Polymorphic markers are the cornerstone of linkage analysis and while the results obtained from such analysis allow them to be orientated relative to each other, and provide estimates of the genetic distance between them, it is their physical location on the human chromosome, which provides the key to localisation of the disease gene. While estimates of the relationship between physical and genetic distance may allow rough localisation of genetic markers on the chromosome the real key lies in the integration of physical and genetic maps (Cox et al., 1994). This problem is being addressed and graphical user interfaces which allow the visualisation and retrieval of mapping data derived from different mapping approaches are being constructed (Boguski and Schuler 1995).

Through time a number of different approaches have been taken in the identification of disease genes, with varying degrees of success and potential levels of application (Collins 1995). Functional cloning is where the identification of a gene is based on the presence of a basic biochemical defect. The chromosomal localisation of the gene is not required, and in most cases where such information is available the gene has been cloned. The candidate gene approach also relies on the availability of some functional information, in combination with educated guesses as to possible candidate genes. In its purest form positional cloning assumes no functional information, and the responsible gene is located purely on the basis of its map position. The majority of positional cloning successes have

relied on the presence of cytogenetic rearrangements, large deletions or expanded trinucleotide repeats. The approach predicted to supersede all of the above is that of the positional candidate. Here mapping to a chromosomal subregion, generally by linkage analysis, is combined with a search for suitable candidates in that region. The expanding success of such an approach depends on the availabilty of a high resolution transcript map.

It is predicted that by the end of 1997 most human genes will be sequence-tagged and placed on various physical maps (Boguski and Schuler 1995). Single pass sequencing into the 3' untranslated regions (3'UTRs) of directionally cloned human cDNAs is being used to uniquely identify human genes and in parallel efficiently generate gene-based sequence-tagged-sites (Wilcox et al., 1991, and Olsen et al., 1989). There are a number of advantages in using 3'UTRs rather than other portions of an mRNA sequence : it is unusual for them to contain introns, therefore the PCR product size is the same from genomic DNA as cDNA templates; the sequence is less well conserved so it is easier to distinguish between individual genes and closely related paralogous family members; the presence of a poly A tail and polyadenylation signal identifies true human mRNAs; and the sequence of this region is more species specific. The total length of ESTs now sequenced amounts to 5 million base pairs which accounts for 0.15% of the total length of the human genome. More than 55,000 ESTs have to date been identified which correspond to authentic genes, although only 10,000 of these are presently logged in public databases (Maddox 1995). In order to eliminate redundancy in the Genbank

database a set of unique 3'UTRs were selected against which all new ESTs are screened before mapping (Boguski and Schuler 1995). Similary, at the mapping level redundancy is prevented by the establishment of another database (Rhalloc which stands for radiation hybrid allocation) into which groups deposit the sequences which they intend to map. Previously ESTs were assigned to a particular chromosome, efforts now are however concentrating on assignment to specific chromosomal regions using a combination of megabase insert yeast artificial chromosomes and somatic cell hybrids (Berry et al., 1995).

Sequencing

The fact that obtaining the human genome in its purest form (i.e. the complete sequence) would not immediately yield much practical information for application to human diseases, and the sheer expense involved in such a project has meant that to date this area has been of least concern to the Human Genome Project (Hoffman 1994). Indeed there is a fear that funding bodies will believe the identification of genes to be sufficient and shrink from obtaining the complete sequence of the genome (Maddox 1995). This would be unwise, since only 3% of the human genome encodes protein. The function of the remaining 97% while unclear may hold the key to explaining evolution and function of the human genome (Nowak 1994). Regulatory elements for genes are frequently located outside the protein coding sequence and without these it would be difficult to gain a real conception of how a gene functions. Repetitive elements are found throughout the genome of higher organisms and, unlike the structure of genes which are frequently highly

conserved between organisms, such repeats are often very simple in structure and may even vary at the cellular level within a single organism (Wooster et al., 1994). Their importance has recently come to the fore however, with the discovery that mutations in a minisatellite may contribute to as much as 10% of all cases of breast, colorectal, bladder cancer and acute leukaemia. It is thought that mutations in the minisatellite upregulate the Harvey *ras* gene, which is involved in a major growth regulatory pathway. Some repetitive elements also seem to have a function in maintaining the integrity of chromosomes such as satellites located at telomeres. Even introns are being investigated to see if they can encode regulatory RNAs: support for this approach of course comes from the XIST gene which shuts down one of the two X chromosomes in female cells via a putative regulatory RNA which it encodes. Sequencing the human genome will not supersede, but complement existing EST maps which will shed new light on global aspects of gene organisation, evolution and expression (Boguski and Schuler 1995).

The importance of sequencing the human genome is thus firmly established. How will this huge volume of work be approached? The production of the Human Sequence Map was discussed recently by John Sulston at the Annual HGMP / Gene Therapy Workshop (July 1995). Firstly he confirmed the importance of complete sequence rather than just cDNAs in order to place all of the accumulated information on a 'genomic ruler' and to formulate an archive for the future. Sequencing sample sections of the genome will not only miss potentially interesting regions, prevent the formation of a local map, but also fail to produce enough

information to function as a genomic ruler for physical and genetic marker placement. Curation of such a map would be an ongoing expense and in the end it would more cost effective to obtain the entire sequence. Instead of base by base sequencing the entire genome, however, the proposal is to shotgun sequence bacterial clones which can then be automatically assembled on a radiation hybrid map. This method is predicted to have an overall cost of 10p/ base which as a percentage cost of the actual overall cost of the HGP is actually rather small.

Model organisms

In addition to mapping the human genome, mapping of genomes of a number of other organisms, both vertebrate and invertebrate, has been undertaken. In the simpler invertebrates basic principles of genetics and biochemistry are easily studied, and the smaller genome size makes them more amenable to complete sequencing and identification of genes. At the simplest eukaryotic level, the yeast *Sacharomyces cerevisiae* has been an integral part of studies of basic eukaryotic functions such as the cell cycle for many years, and has played an important role in the identification of many homologous human genes. Sequencing of the yeast genome (14Mb) is progressing at a rapid rate and it is hoped that as more genes are identified then yeast mutants can be used to screen cDNA libraries and identify their human counterparts, by their ability to complement, as well as determining their function. Another important model invertebrate organism is the nematode *Caenorhabditis elegans.* The small genome size (100Mb) and short life cycle of this organism have lead to a number of studies into its development and behaviour,

and most importantly a number of the genes required for normal development and movement have extensive similarity to their mammalian counterparts. While nowhere near as complex as the human genome, sequencing of this simple animal genome was seen as a forerunner for evaluating technology and as a provider of information essential for interpreting human sequence (Sulston et al., 1992). After establishment of the *C.elegans* database (ACEDB) and publication of the first 2.2Mb of contiguous sequence (Wilson et al., 1994) this project is proceeding at a rate of 2Mb/month, with 18 Mb sequenced by July of 1995 (Sulston the Annual HGMP / Gene Therapy Workshop). While invertebrates have proven useful in identifying human genes, differences in morphology and development mean that they will never serve as comprehensive paradigms for human disease, nor will they encode the whole subset of genes required for vertebrate development.

From the perspective of gene identification the ideal model would be a vertebrate of minimum size and complexity, but maximum homology to the human genome. This role is filled by the Japanese pufferfish *Fugu rubripes* (Brenner et al., 1993). *Fugu* is a tetraodontoid fish and these share the same general body plan, and many of the specialised functions of higher vertebrates. The *Fugu* genome is just 400Mb which is 7.5 times smaller than the human genome, yet encodes a comparable number of genes. In general this is due to smaller intergenic and intronic sequences and less repetitive DNA (Baxendale et al., 1995). Evidence to date indicates conservation of syntenic regions as well as in the sequence of individual genes. Comparison of the Huntington gene of human and *Fugu* identified five

regions which are highly conserved (>80%) between these two distantly related species and hence may be functionally important. Another potential use of *Fugu* is in the identification of gene family members. The size, complexity and large number of repetitive elements in the human genome have made the identification of gene family members a very complicated process. In *Fugu* however four dopamine receptor like genes were identified by a straightforward degenerate PCR (Macrae and Brenner 1995). The fact that teleosts are the most distant vertebrate precursors of mammals means that all unconstrained sequences will have had the maximum amount of time to randomise through mutations and only those sequences required for functions common to all vertebrates will be conserved i.e. coding and regulatory. This view was supported by the results analysis of the function of non-coding regions conserved between mouse and *Fugu Hoxb-4* genes (Aparicio et al., 1995). Thus *Fugu* could be used to identify functionally important regions prior to transgenic analysis.

The ultimate experimental model for many human diseases has been the mouse, due to their physiological similarity and the conservation of syntenic regions (see section 1.2.6.5). Many naturally occurring mouse variants exist, several of which model human disorders. Such physiologically similar disorders may result from mutations in homologous genes, and the genes themselves may be located within homologous chromosomal segments. The identification of the myosin type VII gene, which maps to homologous segments of the genome in mouse and man, and results in disorders which are phenotypically very similar, is a clear example of this

(Gibson et al., 1995; Weil et al., 1995). In some cases, however, the same gene may have a slightly different function in different organisms and the resultant phenotype may not be the same, as was observed with PAX3 (Steel and Harvey 1992). Hence mapping of the mouse genome has been proceeding at a similar rate to human (Dietrich et al., 1994). Now, however, a genetic linkage map of the rat is being rapidly compiled (Jacob 1995). Rats have proved valuable models for several human diseases or traits, such as hypertension, diabetes and cancer (Gill et al., 1989) and hence the ability to genetically map these would be invaluable. While theoretically this is possible, the amount of space required to house enough rats for a high resolution positional cloning project is prohibitory (Frankel 1995). In addition techniques are well developed which allow the manipulation of the mouse genome, and genetically altered mouse strains can be produced which then allow the phenotype and function of genes to be studied (Fässler et al., 1995). Hence perhaps the best route would be to carry out initial mapping in the rat, followed by manipulation of candidate genes in the mouse.

Politics and ethics

It has been proposed that the identification of a gene that predisposes individuals to a human illness may represent the journey through the bottleneck in the understanding of that disease, allowing the development of diagnostic and therapeutic advances (Collins 1995). Once a gene has been identified, however, the way forward is not as simple as it may at first seem.

Identification of the affected gene improves diagnostic capabilities for a disorder, but to what end? Once the function of the gene is understood, probably by study at the biochemical and physiological level in model organisms, more appropriate drug therapy, means of preventative medicine, and possibly even gene therapy can be devised. The gap between the identification of the defective gene and the ability to treat the disorder has ethical implications. In addition, the choice of which disorders are funded at this level and who carries out the research is embedded in politics.

The scientific community was shaken a few years ago when J Craig Venter listed his first set of 600 ESTs and indicated his intention to patent them. Feelings that such basic information should remain in the public domain were high, and fortunately the application was refused. Renewed controversy has arisen, however, with the impending publication of the Genome Directory (Maddox 1995). The Genome Directory will not only contain Venter's latest compilation of ESTs, but also the latest version of the physical map produced by Généthon, and hence will be of immense benefit to groups trying to locate disease genes. The problem comes from the restrictions attached to the use of the directory. Venter is head of The Institute of Genome Research (TIGR) which although in itself non profit making has an agreement with the commercial organisation Human Genome Sciences (HGS) which in turn has an agreement with the pharmaceutical company SmithKline Beecham. The restrictions require any discoveries made on access to the directory leading to potentially valuable applications to be disclosed first to

HGS which would then have the first option to exploit them. This problem is being circumvented by the US pharmaceutical company Merck, which has provided a grant of $10 million to Washington University to replicate construction of the EST catalogue entirely in the public domain. Ultimately, however the choice of which database scientists select to search will probably depend on the reliability of the information contained in them rather than any restrictions.

At the outset of the HGP it was decided to direct at least three percent of its annual budget to a programme called Ethical, Legal and Social Implications (ELSI) (Nelkin 1992). The aim of the programme was to formulate a responsible social policy on the implications of research and potential misuses of information about genetic predisposition, but even this was not immune to political complications (Anderson 1992). In a number of cases, academics and even committee members were discovered to have commercial interest in companies to whom they were recommending the allocation of funds or areas of research. The end conclusion was that in all circumstances scientists should reveal all existing or potential conflicts of interest.

Perhaps the greatest ethical dilemma comes when one considers human dignity. Many people believe the HGP to be a precursor to an inevitable return to eugenics because it will provide the rigorous science and practical techniques necessary for large scale public health strategies (de Dinechin et al., 1993). The history of human eugenics is the source of some embarrassment, and Nazi atrocities produced a backlash against the view that the state had a legitimate interest in who

reproduced (Paul and Spencer 1995). Understandably this led to the belief and indeed hope that eugenics was dead. This may not in fact be the case, since forty years on recommendations with a eugenics feel are again being made - the International Huntington Association and the World Federation of Neurology in their 1990 guidelines deemed it acceptable to refuse to test women 'who do not give complete assurance that they will terminate a pregnancy when there is an increased risk' of Huntington's disease (Went 1990). Freedom of choice must be maintained at all times and the availability of genetic screens must not be allowed to buttress crude eugenic arguements or erode public support for health care (Butler 1994).

In general the recommendations are that screening should be limited to diseases which are treatable or preventable and should always be preceded by counselling (Knoppers and Chadwick 1994). This then highlights another problem - the lack of medically trained geneticists. Science regulates public debate to a much lesser extent than many scientists believe and the importance of good non-coercive communication cannot be emphasised enough (de Dinechin et al., 1993).

The final problem is that of autonomy and potential discrimination. Genetic constitution is by its very nature shared by members of the same family, hence in some cases a person's *right not to know* may clash with another's *right to know* and vice versa. While respect for privacy and confidentiality of genetic information is crucial, it is felt that it is important to communicate information to family members that are at high risk, but consent is essential before disclosure to any

other third party (Knoppers and Chadwick 1994). Third parties may take the form of insurance companies or employers, and it is important to protect individuals against genetic discrimination. At present health care is focussed on treatment rather than prevention of genetic disorders, and such information would be of little positive use to employers or insurance companies. For this reason the formulation of a 'Genetic Bill of Rights' may be envisaged i.e. 'an explicit statement of protections and duties to try to guarantee human self-determination in spite of an increasingly transparent genome' (de Dinechin 1993). It is important that governments do legislate on the outcome of the HGP, because as stated by Knoppers and Chadwick (1994) - 'The accountability of the HGP is at stake. So is our present obligations of stewardship to humankind and to future generations.'

It is not enough for the Human Genome Project to simply be a source of new knowledge, and those involved have a responsibility to society to help in interpretation and correct use of the information gained. Scientists must be aware of all potential outcomes of their results and be prepared to be involved in preventing the misuse. It is also important to ensure health care professionals are prepared for the impending ethical problems imposed by the HGP. Everyone needs some sort of basic understanding, and in the next few years perhaps the most important aspects of the HGP will be the education of the general public, and the provision of legislation to protect their rights.

8. References

Abe, T., Kikuchi, T. & Shinohara, T. 1994. The sequence of the human phosducin gene (PDC) and its 5'-flanking region. *Genomics*, **19**, 369-372.

Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. & Venter, J.C. 1991. Complementary-DNA sequencing - Expressed Sequence Tags and Human Genome Project. *Science*, **252**, 1651-1656.

Anand, R. 1992. Yeast artificial chromosomes (YACs) and the analysis of complex genomes. *Trends In Biotechnology*, **10**, 35-40.

Anand, R., Riley, J.H., Butler, R., Smith, J.C. & Markham, A.F. 1990. A 3.5 genome equivalent multiaccess YAC library - construction, characterization, screening and storage. *Nucleic Acids Research*, **18**, 1951-1956.

Anderson, C. 1992. Conflict concerns disrupt panels, cloud testimony. *Nature*, **355**, 753-754.

Aparicio, S., Morrison, A., Gould, A., Gilthorpe, J., Chaudhuri, C., Rigby, P., Krumlauf, R. & Brenner, S. 1995. Detecting conserved regulatory elements with the model genome of the Japanese puffer fish, *Fugu rubripes*. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **92**, 1684-1688.

Ara, F., Jacobson, S. & Inana, G. 1993. Phosducin as a candidate gene for Retinitis-Pigmentosa and Usher Syndrome typeII. *Investigative Ophthalmology & Visual Science*, **34**, 1458-1458.

Ara, F., Jacobson, S.G. & Inana, G. 1992. Analysis of Usher Syndrome type-II by the candidate gene approach. *Investigative Ophthalmology & Visual Science*, **33**, 791-791.

Arden, G.B. & Fox, B. 1979. Increased incidence of abnormal nasal cilia in patients with retinitis pigmentosa. *Nature*, **279**, 534-536.

Aslanidis, C. & deJong, P.J. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research*, **18**, 6069-6074.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. & Struhl, K. 1995. *Current protocols in molecular biology*. Wiley Interscience, New York.

Baker, H., Grillo, M. & Margolis, F.L. 1989. Biochemical and immunocytochemical characterization of olfactory marker protein in the rodent central nervous-system. *Journal Of Comparative Neurology*, **285**, 246-261.

Barrong, S.D., Chaitin, M.H., Fliesler, S.J., Possin, D.E., Jacobson, S.G. & Milam, A.H. 1992. Ultrastructure of connecting cilia in different forms of Retinitis- Pigmentosa. *Archives Of Ophthalmology*, **110**, 706-710.

Bascom, R.A., Garciaheras, J., Hsieh, C.L., Gerhard, D.S., Jones, C., Francke, U., Willard, H.F., Ledbetter, D.H. & McInnes, R.R. 1992. Localization of the photoreceptor gene ROMI to human chromosome-11 and mouse chromosome-19 - sublocalization to human 11q13 between PGA and PYGM. *American Journal Of Human Genetics*, **51**, 1028-1035.

Bascom, R.A., Manara, S., Collins, L., Molday, R.S., Kalnins, V.I. & McInnes, R.R. 1992. Cloning of the cDNA for a novel photoreceptor membrane-protein (ROMI) identifies a disk rim protein family implicated in human retinopathies. *Neuron*, **8**, 1171-1184.

Batzer, M.A. & Deininger, P.L. 1991. A human-specific subfamily of ALU-sequences. *Genomics*, **9**, 481-487.

Baxendale, S., Abdulla, S., Elgar, G., Buck, D., Berks, M., Micklem, G., Durbin, R., Bates, G., Brenner, S., Beck, S. & Lehrach, H. 1995. Comparative sequence-analysis of the human and pufferfish Huntingtons-disease genes. *Nature Genetics*, **10**, 67-76.

Berger, W., van de Pol, D., Warburg, M., Gal, A. & Bleeker-Wagemakers, L. 1992. Mutations in a candidate gene for Norrie disease. *Human Molecular Genetics*, **1**, 461-465.

Berry, R., Stevens, T.J., Walter, N., Wilcox, A.S., Rubano, T., Hopkins, J.A., Weber, J., Goold, R., Soares, M.B. & Sikela, J.M. 1995. Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nature Genetics*, **10**, 415-423.

Berson, E.L. & Adamian, M. 1992. Ultrastructural findings in an autopsy eye from a patient with Ushers Syndrome type-II. *American Journal Of Ophthalmology*, **114**, 748-757.

Berta, P., Hawkins, J.R., Sinclair, A.H., Taylor, A., Griffiths, B.L., Goodfellow, P.N. & Fellous, M. 1990. Genetic-evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448-450.

Bhattacharyya, A. & Lilley, D. 1989. Single base mismatches in DNA - long-range and short-range structure probed by analysis of axis trajectory and local chemical-reactivity. *Journal Of Molecular Biology*, **209**, 583-597.

Bittles, A.H. & Neel, J.V. 1994. The costs of human inbreeding and their implications for variations at the DNA level. *Nature Genetics*, **8**, 117-121.

Bock, S.C., Skriver, K., Nielsen, E., Thogersen, H.C., Wiman, B., Donaldson, V.H., Eddy, R.L., Marrinan, J., Radziejewska, E., Huber, R., Shows, T.B. & Magnusson, S. 1986. Human CI-inhibitor - primary structure, cDNA cloning, and chromosomal localization. *Biochemistry*, **25**, 4292-4301.

Boguski, M.S. 1995. The turning-point in genome research. *Trends In Biochemical Sciences*, **20**, 295-296.

Boguski, M.S. & Schuler, G.D. 1995. ESTablishing a human transcript map. *Nature Genetics*, **10**, 369-371.

Bonné-Tamir, B., Korostishevsky, M., Kalinsky, H., Seroussi, E., Beker, R., Weiss, S. & Godel, V. 1994. Genetic-mapping of the gene for Usher syndrome - linkage analysis in a large Samaritan kindred. *Genomics*, **20**, 36-42.

Bonneau, D., Raymond, F., Kremer, C., Klossek, J.M., Kaplan, J. & Patte, F. 1993. Usher syndrome type-I associated with bronchiectasis and immotile nasal cilia in 2 brothers. *Journal Of Medical Genetics*, **30**, 253-254.

Boughman, J.A., Vernon, M. & Shaver, K.A. 1983. Usher syndrome - definition and estimate of prevalence from 2 high- risk populations. *Journal Of Chronic Diseases*, **36**, 595-603.

Boyd, A.C. 1993. Turbo cloning - a fast, efficient method for cloning PCR products and other blunt-ended DNA fragments into plasmids. *Nucleic Acids Research*, **21**, 817-821.

Breen, M., Arveiler, B., Murray, I., Gosden, J.R. & Porteous, D.J. 1992. YAC mapping by FISH using Alu-PCR-generated probes. *Genomics*, **13**, 726-730.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B. & Aparicio, S. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature*, **366**, 265-268.

Brooks-Wilson, A.R., Goodfellow, P.N., Povey, S., Nevanlinna, H.A., deJong, P.J. & Goodfellow, P.J. 1990. Rapid cloning and characterization of new chromosome-10 DNA markers by Alu element-mediated PCR. *Genomics*, **7**, 614-620.

Brown, K.A., Sutcliffe, M.J., Steel, K.P. & Brown, S. 1994. Sequencing of the olfactory marker protein gene in normal and shaker- 1 mutant mice. *Mammalian Genome*, **5**, 11-14.

Brown, S.D.M., Brown, K.A., Sutcliffe, M.J., Cavanna, J.S., Greenfield, A.J. & Steel, K.P. 1991. Reverse genetic approaches to cloning deafness genes. *Annals Of The New York Academy Of Sciences*, **630**, 93-99.

Browne, D.L., Gault, J., Thompson, M.B., Hauge, X.Y., Evans, G.A. & Litt, M. 1991. Dinucleotide repeat polymorphism at the D11s527 locus. *Nucleic Acids Research*, **19**, 4790-4790.

Brueckner, M., Deustachio, P. & Horwich, A.L. 1989. Linkage mapping of a mouse gene, *iv*, that controls left right asymmetry of the heart and viscera. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **86**, 5035-5038.

Buch, N.H. & Jorgensen, M.B. 1963. Pathological studies of deaf mutes. *Archives of Otolarnygology*, **77**, 246-253.

Buiakova, O.I., Krishna, N.S.R., Getchell, T.V. & Margolis, F.L. 1994. Human and rodent OMP genes - conservation of structural and regulatory motifs and cellular-localization. *Genomics*, **20**, 452-462.

Bull, P.C., Barwell, J.A., Hannah, H.T.L., Pautler, S.E., Higgins, M.J., Lalande, M. & Cox, D.W. 1993. Isolation of new probes in the region of the Wilson disease locus, 13q14.2-q14.3. *Cytogenetics And Cell Genetics*, **64**, 12-17.

Butler, D. 1994. Ethics treaty to target genome implications. *Nature*, **371**, 369-369.

Bystrom, C., Larsson, C., Blomberg, C., Sandelin, K., Falkmer, U., Skogseid, B., Oberg, K., Werner, S. & Nordenskjold, M. 1990. Localization of the MENI gene to a small region within chromosome 11q13 by deletion mapping in tumors. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **87**, 1968-1972.

Carrier, L., Hengstenberg, C., Beckmann, J.S., Guicheney, P., Dufour, C., Bercovici, J., Dausse, E., Berebbibertrand, I., Wisnewsky, C., Pulvenis, D., Fetler, L., Vignal, A., Weissenbach, J., Hillaire, D., Feingold, J., Bouhour, J.B., Hagege, A., Desnos, M., Isnard, R., Dubourg, O., Komajda, M. & Schwartz, K. 1993. Mapping of a novel gene for Familial Hypertrophic Cardiomyopathy to chromosome-11. *Nature Genetics*, **4**, 311-313.

Chance, P.F., Bird, T.D., O'Connell, P., Lipe, H., Lalouel, J.M. & Leppert, M. 1990. Genetic-linkage and heterogeneity in type-I Charcot-Marie-Tooth disease (hereditary motor and sensory neuropathy type-I). *American Journal Of Human Genetics*, **47**, 915-925.

Charlieu, J.P., Laurent, A.M., Carter, D.A., Bellis, M. & Roizes, G. 1992. 3' Alu PCR - a simple and rapid method to isolate human polymorphic markers. *Nucleic Acids Research*, **20**, 1333-1337.

Chou, Q., Russell, M., Birch, D.E., Raymond, J. & Bloch, W. 1992. Prevention of pre-PCR mis-priming and primer dimerization improves low-copy-number amplifications. *Nucleic Acids Research*, **20**, 1717-1723.

Chuah, M.I. & Zheng, D.R. 1992. The human primary olfactory pathway - fine-structural and cytochemical aspects during development and in adults. *Microscopy Research And Technique*, **23**, 76-85.

Clark, J.M. 1988. Novel non-templated nucleotide addition-reactions catalyzed by procaryotic and eukaryotic DNA-polymerases. *Nucleic Acids Research*, **16**, 9677-9686.

Cobianchi, F. & Wilson, S.H. 1987. Enzymes for modifying and labeling DNA and RNA. *Methods In Enzymology*, 94-110.

Cole, C.G., Dunham, I., Coffey, A.J., Ross, M.T., Meierewert, S., Bobrow, M. & Bentley, D.R. 1992. A random STS strategy for construction of YAC contigs spanning defined chromosomal regions. *Genomics*, **14**, 256-262.

Cole, C.G., Patel, K., Shipley, J., Sheer, D., Bobrow, M., Bentley, D.R. & Dunham, I. 1992. Identification of region-specific yeast artificial chromosomes using pools of Alu element-mediated polymerase chain-reaction probes labeled via linear amplification. *Genomics*, **14**, 931-938.

Collins, F.S. 1995. Positional cloning moves from perditional to traditional. *Nature Genetics*, **9**, 347-350.

Condie, A., Eeles, R., Borresen, A.L., Coles, C., Cooper, C. & Prosser, J. 1993. Detection of point mutations in the p53 gene - comparison of Single- Strand Conformation Polymorphism, Constant Denaturant Gel- Electrophoresis, and Hydroxylamine and Osmium-Tetroxide techniques. *Human Mutation*, **2**, 58-66.

Cotton, R.G.H., Rodrigues, N.R. & Campbell, R.D. 1988. Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium-tetroxide and its application to the study of mutations. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **85**, 4397-4401.

Cowell, I.G., Dixon, K.H., Pemble, S.E., Ketterer, B. & Taylor, J.B. 1988. The structure of the human glutathione s-transferase pi-gene. *Biochemical Journal*, **255**, 79-83.

Cox, D.R., Burmeister, M., Price, E.R., Kim, S. & Myers, R.M. 1990. Radiation hybrid mapping - a somatic-cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science*, **250**, 245-250.

Cox, D.R., Green, E.D., Lander, E.S., Cohen, D. & Myers, R.M. 1994. Assessing mapping progress in the Human-Genome-Project. *Science*, **265**, 2031-2032.

Craig, J.M. & Bickmore, W.A. 1993. Chromosome bands - flavors to savour. *Bioessays*, **15**, 349-354.

Cremers, F.P., Molloy, C.M., van de Pol, D.J., van den Hurk, J.A., Bach, I., van Kessel, G.A.H. & Ropers, H.H. 1992. An autosomal homologue of the choroideremia gene colocalizes with the Usher syndrome type II locus on the distal part of chromosome 1q. *Human Molecular Genetics*, **1**, 71-77.

Cunningham, A.M. & Reed, R.R. 1992. A sense of smell. *Current Biology*, **2**, 116-118.

Danciger, E., Mettling, C., Vidal, M., Morris, R. & Margolis, F. 1989. Olfactory marker protein gene - its structure and olfactory neuron- specific expression in transgenic mice. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **86**, 8565-8569.

Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Reed, P.W., Gough, S., Jenkins, S.C., Palmer, S.M., Balfour, K.M., Rowe, B.R., Farrall, M., Barnett, A.H., Bain, S.C. & Todd, J.A. 1994. A genome-wide search for human type-1 diabetes susceptibility genes. *Nature*, **371**, 130-136.

Davisson, M.T., Roderick, T.H. & Doolittle, D.P. 1989. *Recombination percentages and chromosomal assignments*. Oxford University Press, Oxford.

de Dinechin, O., Harris, R., Kettner, M., Koch, L. & Zwierlein, E. 1993. Workshop of the Commission of the European Communities on Ethics of Human Genome analysis - Survey of the European discussion. *Journal Of Medical Genetics*, **30**, 257-260.

Dean, M. 1995. Resolving DNA mutations. *Nature Genetics*, **9**, 103-104.

Deininger, P.L., Batzer, M.A., Hutchison, C.A. & Edgell, M.H. 1992. Master genes in mammalian repetitive DNA amplification. *Trends In Genetics*, **8**, 307-311.

Dibb, N.J., Brown, D.M., Karn, J., Moerman, D.G., Bolten, S.L. & Waterston, R.H. 1985. Sequence-analysis of mutations that affect the synthesis, assembly and enzymatic-activity of the unc-54 myosin heavy-chain of caenorhabditis-elegans. *Journal Of Molecular Biology*, **183**, 543-551.

Dietrich, W.F., Miller, J.C., Steen, R.G., Merchant, M., Damron, D., Nahf, R., gross, A., Joyce, D.C., Wessel, M., Dredge, R.D., Marquis, A., Stein, L.D., Goodman, N., Page, D.C. & Lander, E.S. 1994. A genetic-map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genetics*, **7**, 220-245.

Ding, C.L., Li, X., Griffin, C.A., Jabs, E.W., Hawkins, A.L. & Levine, M.A. 1993. The gene for human phosducin (PDC), a soluble-protein that binds G- protein beta-gamma dimers, maps to 1q25-q31.1. *Genomics*, **18**, 457-459.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K. & Mattick, J.S. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, **19**, 4008-4008.

Dowton, M. & Austin, A.D. 1993. Direct sequencing of double-stranded PCR products without intermediate fragment purification - digestion with mung bean nuclease. *Nucleic Acids Research*, **21**, 3599-3600.

Dracopoli, N.C., Haines, J.L., Moir, D.T., Morton, C.C., Seidman, C.E., Seidman, J.G., Smith, D.R. & Boyle, E.D. 1995. *Current protocols in Human Genetics*. Wiley Interscience, New York.

Duyk, G., Gastier, J.M. & Mueller, R.F. 1992. Traces of her workings. *Nature Genetics*, **2**, 5-8.

Economou, E.P., Bergen, A.W., Warren, A.C. & Antonarakis, S.E. 1990. The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **87**, 2951-2954.

Eichenbaum, H. 1995. The hippocampus and the sense of smell. *Chemical Senses*, **20**, 105-105.

Evans, K.L., Fantes, J., Simpson, C., Arveiler, B., Muir, W., Fletcher, J., van Heyningen, V., Steel, K.P., Brown, K.A., Brown, S.D.M., St Clair, D. & Porteous, D.J. 1993. Human olfactory marker protein maps close to tyrosinase and is a candidate gene for usher syndrome type-I. *Human Molecular Genetics*, **2**, 115-118.

Evans, K.L., van Heyningen, V. & Porteous, D.J. 1995. Placement and refined mapping of established and new markers on human-chromosome 11q using a small panel of somatic-cell hybrids. *European Journal Of Human Genetics*, **3**, 42-48.

Evers, M., Zelle, B., Bebelman, J.P., Vanbeusechem, V., Kraakman, L., Hoffer, M., Pronk, J.C., Mager, W.H., Planta, R.J., Eriksson, A.W. & Frants, R.R. 1989. Nucleotide-sequence comparison of 5 human pepsinogen-a (PGA) genes - evolution of the PGA multigene family. *Genomics*, **4**, 232-239.

Farbman, A.I. & Margolis, F.L. 1980. Olfactory marker protein during ontogeny: immunohistochemical localization. *Developmental Biology*, **74**, 205-215.

Fassler, R., Martin, K., Forsberg, E., Litzenburger, T. & Iglesias, A. 1995. Knockout mice - how to make them and why - the immunological approach. *International Archives Of Allergy And Immunology*, **106**, 323-334.

Fletcher, J.M., Evans, K., Baillie, D., Byrd, P., Hanratty, D., Leach, S., Julier, C., Gosden, J.R., Muir, W., Porteous, D.J., St Clair, D. & van Heyningen, V. 1993. Schizophrenia-associated chromosome 11q21 translocation - identification of flanking markers and development of chromosome 11q fragment hybrids as cloning and mapping resources. *American Journal Of Human Genetics*, **52**, 478-490.

Fodde, R., van der Luijt, R., Wijnen, J., Tops, C., Vanderklift, H., van Leeuwencornelisse, I., Griffioen, G., Vasen, H. & Khan, P.M. 1992. 8 novel inactivating germ line mutations at the APC gene identified by Denaturing Gradient Gel-Electrophoresis. *Genomics*, **13**, 1162-1168.

Frankel, W.N. 1995. Of rats, mice and men. *Nature Genetics*, **9**, 3-4.

Freytag, S.O. & Collier, K.J. 1984. Molecular-cloning of a cDNA for human pyruvate-carboxylase - structural relationship to other biotin-containing carboxylases and regulation of messenger-RNA content in differentiating preadipocytes. *Journal Of Biological Chemistry*, **259**, 2831-2837.

Gaston, S.M., Tanzi, R.E., Stewart, G.D., Romano, D.M., Gardiner, K., Dang, T., Bowlin, R., Rider, S., McCormick, M.K., Kurnit, D.M., Gusella, J.F. & Patterson, D. 1991. Cloning of human chromosome-21 into yeast artificial chromosomes - progress of the chromosome-21 joint YAC effort. *Cytogenetics And Cell Genetics*, **58**, 2036-2036.

Gibson, F., Lehrach, H., Buckler, A.J., Brown, S. & North, M.A. 1994. Isolation of conserved sequences from yeast artificial chromosomes by exon amplification. *Biotechniques*, **16**, 453.

Gibson, F., Walsh, J., Mburu, P., Varela, A., Brown, K.A., Antonio, M., Beisel, K.W., Steel, K.P. & Brown, S. 1995. A type-VII myosin encoded by the mouse deafness gene *shaker-1*. *Nature*, **374**, 62-64.

Gill, T.J., Smith, G.J., Wissler, R.W. & Kunz, H.W. 1989. The rat as an experimental animal. *Science*, **245**, 269-276.

Gillespie, P.G. & Hudspeth, A.J. 1993. Adenine nucleoside diphosphates block adaptation of mechanoelectrical transduction in hair-cells. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **90**, 2710-2714.

Gillespie, P.G., Wagner, M.C. & Hudspeth, A.J. 1993. Identification of a 120 kd hair-bundle myosin located near stereociliary tips. *Neuron*, **11**, 581-594.

Goldberg, M.F. 1994. Molecular heterogeneity in Retinitis-Pigmentosa - more mutations. *Ophthalmic Genetics*, **15**, 47-50.

Gottesdiener, K.M., Karpinski, B.A., Lindsten, T., Strominger, J.L., Jones, N.H., Thompson, C.B. & Leiden, J.M. 1988. Isolation and structural characterization of the human 4F2 heavy-chain gene, an inducible gene involved in lymphocyte-T activation. *Molecular And Cellular Biology*, **8**, 3809-3819.

Graziadei, P.P.C. & Monti-Graziadei, G.A. 1978. *The olfactory system: a model for the study of neurogenesis and axon regeneration in mammals*. Raven Press, New York.

Grillo, M., Morris, R. & Akeson, R. 1992. Transgenic analysis of the OMP promoter using two reporter genes. *Sociology and Neuroscience Abstracts*, **18**, 239.

Grondahl, J. 1986. Tapeto-retinal degeneration in 4 Norwegian counties .1. diagnostic evaluation of 89 probands. *Clinical Genetics*, **29**, 1-16.

Grondahl, J. 1987. Estimation of prognosis and prevalence of Retinitis Pigmentosa and Usher Syndrome in Norway. *Clinical Genetics*, **31**, 255-264.

Grondahl, J. & Mjoen, S. 1986. Usher syndrome in 4 Norwegian counties. *Clinical Genetics*, **30**, 14-28.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P. & Lovellbadge, R. 1990. A gene-mapping to the sex-determining region of the mouse Y- chromosome is a member of a novel family of embryonically expressed genes. *Nature*, **346**, 245-250.

Guilford, P., Ayadi, H., Blanchard, S., Chaib, H., Lepaslier, D., Weissenbach, J., Drira, M. & Petit, C. 1994. A human gene responsible for neurosensory, nonsyndromic recessive deafness is a candidate homolog of the mouse *sh-1* gene. *Human Molecular Genetics*, **3**, 989-993.

Guilford, P., Benarab, S., Blanchard, S., Levilliers, J., Weissenbach, J., Belkahia, A. & Petit, C. 1994. A non-syndromic form of neurosensory, recessive deafness maps to the pericentromeric region of chromosome-13q. *Nature Genetics*, **6**, 24-28.

Guldberg, P. & Guttler, F. 1994. Broad-range DGGE for single-step mutation scanning of entire genes - application to human Phenylalanine-Hydroxylase gene. *Nucleic Acids Research*, **22**, 880-881.

Guzzetta, V., Deocaluna, R.M., Lupski, J.R. & Patel, P.I. 1991. Isolation of region-specific and polymorphic markers from chromosome- 17 by restricted Alu polymerase chain-reaction. *Genomics*, **9**, 31-36.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M. & Weissenbach, J. 1994. The 1993-94 Genethon human genetic-linkage map. *Nature Genetics*, **7**, 246-339.

Hallgren, B. 1959. Retinitis pigmentosa with congenital deafness, with vestibulo-cerebellar ataxia and mental abnormality in a proportion of cases: A clinical and genetico-statistical study. *Acta Psychiatry Scandinavian Supplement*, **138**, 5-101.

Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *Journal Of Molecular Biology*, **166**, 557-580.

Hanahan, D., Jesse, J. & Bloom, F.R. 1991. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods in Enzymology*, **204**, 63-103.

Hashimoto, L., Habita, C., Beressi, J.P., Delepine, M., Besse, C., Cambonthomsen, A., Deschamps, I., Rotter, J.I., Djoulah, S., James, M.R., Froguel, P., Weissenbach, J., Lathrop, G.M. & Julier, C. 1994. Genetic-mapping of a susceptibility locus for Insulin-Dependent Diabetes-Mellitus on chromosome 11q. *Nature*, **371**, 161-164.

Hattori, M., Shibata, A., Yoshioka, K. & Sakaki, Y. 1993. Orphan peak analysis - a novel method for detection of point mutations using an automated fluorescence DNA sequencer. *Genomics*, **15**, 415-417.

Heckenlively, J. 1988. *Retinitis Pigmentosa*. Philadelphia, PA.

Heckenlively, J.R., Chang, B., Roderick, T.H., Peng, C. & Erway, L.C. 1994. A mouse model for Usher Syndrome. *Investigative Ophthalmology & Visual Science*, **35**, 1833-1833.

Hewitt, J.E., Gordon, M.M., Taggart, R.T., Mohandas, T.K. & Alpers, D.H. 1991. Human Gastric Intrinsic-Factor - characterization of cDNA and genomic clones and localization to human chromosome-11. *Genomics*, **10**, 432-440.

Hinderink, J.B., Mens, L.H.M., Brokx, J.P.L. & van den Broek, P. 1994. Results from 4 cochlear implant patients with Ushers-Syndrome. *Annals Of Otology Rhinology And Laryngology*, **103**, 285-293.

Hinds, P.W., Dowdy, S.F., Eaton, E.N., Arnold, A. & Weinberg, R.A. 1994. Function of a human cyclin gene as an oncogene. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **91**, 709-713.

Hoffman, E.P. 1994. The evolving genome project - current and future-impact. *American Journal Of Human Genetics*, **54**, 129-136.

Holmquist, G.P. 1992. Chromosome bands, their chromatin flavors, and their functional features. *American Journal Of Human Genetics*, **51**, 17-37.

Hultman, T., Bergh, S., Moks, T. & Uhlen, M. 1991. Bidirectional solid-phase sequencing of *in vitro*-amplified plasmid DNA. *Biotechniques*, **10**, 84.

Humphries, P., Kenna, P. & Farrar, G.J. 1994. New dimensions in macular dystrophies. *Nature Genetics*, **8**, 315-317.

Hunter, D.G., Fishman, G.A., Mehta, R.S. & Kretzer, F.L. 1986. Abnormal sperm and photoreceptor axonemes in Ushers Syndrome. *Archives Of Ophthalmology*, **104**, 385-389.

Hutchinson, G.B., Andrew, S.E., McDonald, H., Goldberg, Y.P., Graham, R., Rommens, J.M. & Hayden, M.R. 1993. An Alu element retroposition in 2 families with Huntington disease defines a new active Alu subfamily. *Nucleic Acids Research*, **21**, 3379-3383.

Huxley, C. 1994. Mammalian artificial chromosomes - a new tool for gene-therapy. *Gene Therapy*, **1**, 7-12.

Inaba, T., Matsushime, H., Valentine, M., Roussel, M.F., Sherr, C.J. & Look, A.T. 1992. Genomic organization, chromosomal localization, and independent expression of human Cyclin-D genes. *Genomics*, **13**, 565-574.

Iwahana, H., Yoshimoto, K., Mizusawa, N., Kudo, E. & Itakura, M. 1994. Multiple fluorescence-based PCR SSCP analysis. *Biotechniques*, **16**, 296.

Jacob, H.J., Brown, D.M., Bunker, R.K., Daly, M.J., Dzau, V.J., Goodman, A., Koike, G., Kren, V., Kurtz, T., Lernmark, A., Levan, G., Mao, Y.P., Pettersson, A., Pravenec, M., Simon, J.S., Szpirer, C., Szpirer, J., Trolliet, M.R., Winer, E.S. & Lander, E.S. 1995. A genetic-linkage map of the laboratory rat, rattus-norvegicus. *Nature Genetics*, **9**, 63-69.

Jäger, R.J., Anvret, M., Hall, K. & Scherer, G. 1990. A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. *Nature*, **348**, 452-454.

James, M.R., Richard, C.W., Schott, J.J., Yousry, C., Clark, K., Bell, J., Terwilliger, J.D., Hazan, J., Dubay, C., Vignal, A., Agrapart, M., Imai, T., Nakamura, Y., Polymeropoulos, M., Weissenbach, J., Cox, D.R. & Lathrop, G.M. 1994. A radiation hybrid map of 506 STS markers spanning human-chromosome- 11. *Nature Genetics*, **8**, 70-76.

Jodice, C., Malaspina, P., Persichetti, F., Novelletto, A., Spadaro, M., Giunti, P., Morocutti, C., Terrenato, L., Harding, A.E. & Frontali, M. 1994. Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia-i. *American Journal Of Human Genetics*, **54**, 959-965.

Jones, C., Bill, J., Larizza, L., Pym, B., Goodfellow, P. & Tunnacliffe, A. 1984. Relationships between genes on human chromosome-11 encoding cell- surface antigens. *Somatic Cell And Molecular Genetics*, **10**, 423-428.

Jones, C.P., Janson, M. & Nordenskjold, M. 1989. Separation of yeast chromosomes in the megabase range suitable as size markers for pulsed field gel electrophoresis. *Technique*, **1**, 90-95.

Jonsson, J.J. & Weissman, S.M. 1995. From mutation mapping to phenotype cloning. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **92**, 83-85.

Kajiwara, K., Berson, E.L. & Dryja, T.P. 1994. Digenic Retinitis-Pigmentosa due to mutations at the unlinked peripherin/RDS and ROMI loci. *Science*, **264**, 1604-1608.

Kaplan, D.J., Jurka, J., Solus, J.F. & Duncan, C.H. 1991. Medium reiteration frequency repetitive sequences in the human genome. *Nucleic Acids Research*, **19**, 4731-4738.

Kaplan, J., Gerber, S., Bonneau, D., Rozet, J.M., Delrieu, O., Briard, M.L., Dollfus, H., Ghazi, I., Dufier, J.L., Frezal, J. & Munnich, A. 1992. A gene for Usher Syndrome type-I (USHIA) maps to chromosome-14q. *Genomics*, **14**, 979-987.

Kas, K., Weber, G., Merregaert, J., Michiels, L., Sandelin, K., Skogseid, B., Thompson, N., Nordenskjold, M., Larsson, C. & Friedman, E. 1993. Exclusion of FAU as the Multiple Endocrine Neoplasia type-1 (MENI) gene. *Human Molecular Genetics*, **2**, 349-353.

Keats, B., Todorov, A.A., Atwood, L.D., Pelias, M.Z., Hejtmancik, J.F., Kimberling, W.J., Leppert, M., Lewis, R.A. & Smith, R. 1992. Linkage studies of Usher Syndrome type-1 - exclusion results from the Usher Syndrome consortium. *Genomics*, **14**, 707-714.

Keele, C.A., Neil, E. & Joels, N. 1983. *Samson Wright's applied physiology*. Oxford University Press, Oxford.

Kimberling, W.J., Moller, C.G., Davenport, S., L.H., Lund, G., Grissom, T.J., Priluck, I., White, V., Weston, M.D., Bisconehalterman, K. & Brookhouser, P.E. 1989. Usher Syndrome - clinical findings and gene localization studies. *Laryngoscope*, **99**, 66-72.

Kimberling, W.J., Moller, C.G., Davenport, S., Priluck, I.A., Beighton, P.H., Greenberg, J., Reardon, W., Weston, M.D., Kenyon, J.B., Grunkemeyer, J.A., Dahl, S.P., Overbeck, L.D., Blackwood, D.J., Brower, A.M., Hoover, D.M., Rowland, P. & Smith, R. 1992. Linkage of Usher Syndrome type-I gene (USHIB) to the long arm of chromosome-11. *Genomics*, **14**, 988-994.

Kimberling, W.J., Weston, M.D., Moller, C., Davenport, S.L.H., Shugart, Y.Y., Priluck, I.A., Martini, A., Milani, M. & Smith, R.J. 1990. Localization of Usher Syndrome type-II to chromosome-1q. *Genomics*, **7**, 245-249.

Kimberling, W.J., Weston, M.D., Moller, C., Vanaarem, A., Cremers, C.W.R.G., Sumegi, J., Ing, P.S., Connolly, C., Martini, A., Milani, M., Tamayo, M.L., Bernal, J., Greenberg, J. & Ayuso, C. 1995. Gene-mapping of Usher Syndrome type IIA - localization of the gene to a 2.1-cm segment on chromosome 1q41. *American Journal Of Human Genetics*, **56**, 216-223.

Kimberling, W.J., Weston, M.D., Pieke Dahl, S., Kenyon, J.B., Shugart, Y.Y., Moller, C., Davenport, S.L.H., Martini, A., Milani, M. & Smith, R.J. 1991. Genetic-studies of Usher Syndrome. *Annals Of The New York Academy Of Sciences*, **630**, 167-175.

Klein, V., Piontek, K., Brass, N., Subke, F., Zang, K.D. & Meese, E. 1993. Identification of chromosome-specific Sequence-Tagged Sites by Alu- PCR. *Genetic Analysis Techniques And Applications*, **10**, 6-9.

Knoppers, B.M. & Chadwick, R. 1994. The Human-Genome-Project - under an international ethical microscope. *Science*, **265**, 2035-2036.

Koeffler, H.P., Sparkes, R.S., Stang, H. & Mohandas, T. 1981. Regional assignment of genes for human alpha-globin and phosphoglycollate phosphatase to the short arm of chromosome-16. *Proceedings Of The National Academy Of Sciences Of The United States Of America Biological Sciences*, **78**, 7015-7018.

Komatsu, H., Iida, S., Yamamoto, K., Mikuni, C., Nitta, M., Takahashi, T., Ueda, R. & Seto, M. 1994. A variant chromosome-translocation at 11q13 identifying PRADI cyclin DI as the BCL-1 gene. *Blood*, **84**, 1226-1231.

Koopman, P., Gubbay, J., Collignon, J. & Lovellbadge, R. 1989. Zfy gene-expression patterns are not compatible with a primary role in mouse sex determination. *Nature*, **342**, 940-942.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovellbadge, R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature*, **351**, 117-121.

Koopman, P., Munsterberg, A., Capel, B., Vivian, N. & Lovellbadge, R. 1990. Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature*, **348**, 450-452.

Korenberg, J.R. & Rykowski, M.C. 1988. Human genome organization - Alu, lines, and the molecular-structure of metaphase chromosome bands. *Cell*, **53**, 391-400.

Krishna, N., Getchell, T.V., Margolis, F.L. & Getchell, M.L. 1992. Amphibian olfactory receptor neurons express olfactory marker protein. *Brain Research*, **593**, 295-298.

Kudrycki, K., Steinizsak, C., Behn, C., Grillo, M., Akeson, R. & Margolis, F.L. 1993. Olf-1-binding site - characterization of an olfactory neuron-specific promoter motif. *Molecular And Cellular Biology*, **13**, 3002-3014.

Kwitekblack, A.E., Carmi, R., Duyk, G.M., Buetow, K.H., Elbedour, K., Parvari, R., Yandava, C.N., Stone, E.M. & Sheffield, V.C. 1993. Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic-heterogeneity. *Nature Genetics*, **5**, 392-396.

Lake, K. & Sharmo, O.P. 1973. Kartagener syndrome and deaf mutism: an unusual association. *Chest*, **64**, 661-663.

Lalley, P.A., Davisson, M.T., Graves, J.A.M., O'Brien, S.J., Womack, J.E., Roderick, T.H., Creaugoldberg, N., Hillyard, A.L., Doolittle, D.P. & Rogers, J.A. 1989. Report of the committee on comparative mapping. *Cytogenetics And Cell Genetics*, **51**, 503-532.

Lancet, D. 1988. *Molecular components of olfactory reception and transduction*. Plenum, New York.

Lanchantin, G.F., Hart, D.W., Friemann, J.A., Saavedra, N.V. & Mehl, J.W. 1968. Amino acid composition of human plasma prothrombin. *Journal of Biological Chemistry*, **243**, 5479-5485.

Largetpiet, D., Gerber, S., Bonneau, D., Rozet, J.M., Marc, S., Ghazi, I., Dufier, J.L., David, A., Bitoun, P., Weissenbach, J., Munnich, A. & Kaplan, J. 1994. Genetic-heterogeneity of Usher Syndrome type-1 in French families. *Genomics*, **21**, 138-143.

Larsson, C., Skogseid, B., Oberg, K., Nakamura, Y. & Nordenskjold, M. 1988. Multiple endocrine neoplasia type-1 gene maps to chromosome-11 and is lost in insulinoma. *Nature*, **332**, 85-87.

Lebo, R.V., Anderson, L.A., Dimauro, S., Lynch, E., Hwang, P. & Fletterick, R. 1990. Rare McArdle disease locus polymorphic site on 11q13 contains CpG sequence. *Human Genetics*, **86**, 17-24.

Lebo, R.V., Gorin, F., Fletterick, R.J., Kao, F.T., Cheung, M.C., Bruce, B.D. & Kan, Y.W. 1984. High-resolution chromosome sorting and DNA spot-blot analysis assign McArdles syndrome to chromosome-11. *Science*, **225**, 57-59.

Leigh-Brown, A. 1984. Molecular evolution - on the origin of the Alu family of repeated sequences. *Nature*, **312**, 106-106.

Leppert, M., Baird, L., Anderson, K.L., Otterud, B., Lupski, J.R. & Lewis, R.A. 1994. Bardet-Biedl syndrome is linked to DNA markers on chromosome 11q and is genetically heterogeneous. *Nature Genetics*, **7**, 108-112.

Levan, G., Klingalevan, K., Szpirer, J. & Szpirer, C. 1993. Rat gene map and comparative mapping with the mouse and other species. *Transplantation Proceedings*, **25**, 2775-2776.

Lewis, R.A., Otterud, B., Stauffer, D., Lalouel, J.M. & Leppert, M. 1990. Mapping recessive ophthalmic diseases - linkage of the locus for Usher Syndrome type-II to a DNA marker on chromosome-1q. *Genomics*, **7**, 250-256.

Li, Y., Muller, B., Fuhrmann, C., Vannouhuys, C.E., Laqua, H., Humphries, P., Schwinger, E. & Gal, A. 1992. The autosomal dominant Familial Exudative Vitreoretinopathy locus maps on 11q and is closely linked to D11s533. *American Journal Of Human Genetics*, **51**, 749-754.

Link, A.J. & Olson, M.V. 1991. Physical map of the Saccharomyces-cerevisiae genome at 110-kilobase resolution. *Genetics*, **127**, 681-698.

Lisitsyn, N., Lisitsyn, N. & Wigler, M. 1993. Cloning the differences between 2 complex genomes. *Science*, **259**, 946-951.

Little, P.F.R. 1992. Generating cloned DNA maps. *Trends In Biotechnology*, **10**, 33-35.

Liu, W.M., Maraia, R.J., Rubin, C.M. & Schmid, C.W. 1994. Alu transcripts - cytoplasmic localization and regulation by DNA methylation. *Nucleic Acids Research*, **22**, 1087-1095.

Lord, E.M. & Gates, W.H. 1929. *Shaker*, a new mutation of the house mouse (mus musculus). *American Naturalist*, **63**, 435-442.

Lyon, M.F. & Searle, A.G. *Genetic variants and strains of the laboratory mouse*. Oxford University Press, Oxford.

Macrae, A.D. & Brenner, S. 1995. Analysis of the dopamine-receptor family in the compact genome of the puffer fish *Fugu rubripes*. *Genomics*, **25**, 436-446.

Maddox, J. 1995. Directory to the human genome. *Nature*, **376**, 459-460.

Marchuk, D., Drumm, M., Saulino, A. & Collins, F.S. 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Research*, **19**, 1154-1154.

Margolis, F.L. 1988. *Molecular cloning of olfactory specific gene products*. Plenum, New York.

Mashal, R.D., Koontz, J. & Sklar, J. 1995. Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genetics*, **9**, 177-183.

Maule, J.C. 1994. Electrophoretic analysis- pulsed field gel electrophoresis.*In J.R.Gosden (Ed) Chromosome analysis protocols*. New Jersey: Humana Press Inc

Maxam, A.M. & Gilbert, W. 1980. Sequencing end-labelled DNA with base. *Methods in enzymology*, **65**, 499-560.

Meese, E.U., Meltzer, P.S., Ferguson, P.W. & Trent, J.M. 1992. Alu-PCR- characterization of a chromosome-6-specific hybrid mapping panel and cloning of chromosome-specific markers. *Genomics*, **12**, 549-554.

Monaco, A.P., Lam, V.M.S., Zehetner, G., Lennon, G.G., Douglas, C., Nizetic, D., Goodfellow, P.N. & Lehrach, H. 1991. Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of Alu-PCR products. *Nucleic Acids Research*, **19**, 3315-3318.

Monaco, A.P. & Larin, Z. 1994. YACs, BACs, PACs and MACs - artificial chromosomes as research tools. *Trends In Biotechnology*, **12**, 280-286.

Montandon, A.J., Green, P.M., Bentley, D.R., Ljung, R., Nilsson, I.M. & Giannelli, F. 1990. 2 factor-IX mutations in the family of an isolated hemophilia-B patient - direct carrier diagnosis by amplification mismatch detection (AMD). *Human Genetics*, **85**, 200-204.

Moyzis, R.K., Torney, D.C., Meyne, J., Buckingham, J.M., Wu, J.R., Burks, C., Sirotkin, K.M. & Goad, W.B. 1989. The distribution of interspersed repetitive DNA-sequences in the human genome. *Genomics*, **4**, 273-289.

Muller, B., Orth, U., Vannouhuys, C.E., Duvigneau, C., Fuhrmann, C., Schwinger, E., Laqua, H. & Gal, A. 1994. Mapping of the Autosomal-Dominant Exudative Vitreoretinopathy locus (EVRI) by multipoint linkage analysis in 4 families. *Genomics*, **20**, 317-319.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. 1986. Specific enzymatic amplification of DNA *in vitro* - the polymerase chain-reaction. *Cold Spring Harbor Symposia On Quantitative Biology*, **51**, 263-273.

Nadeau, J.H., Davisson, M.T., Doolittle, D.P., Grant, P., Hillyard, A.L., Kosowsky, M.R. & Roderick, T.H. 1992. Comparative map for mice and humans. *Mammalian Genome*, **3**, 480-536.

Nadeau, J.H., Kosowsky, M. & Steel, K.P. 1991. Comparative gene-mapping, genome duplication, and the genetics of hearing. *Annals Of The New York Academy Of Sciences*, **630**, 49-67.

Naylor, S.L., Chinn, R., Linn, R., Reus, B., Moore, S., Kebacher, K. & Leach, R.J. 1995. Regional mapping of 225 markers on chromosome-3. *Cytogenetics And Cell Genetics*, **68**, 140-140.

Nelkin, D. 1992. Gene-mapping - using law and ethics as guides - (Annas,G.J., Elias,S. *Nature*, **360**, 380-381.

Nelson, S.F., Mccusker, J.H., Sander, M.A., Kee, Y., Modrich, P. & Brown, P.O. 1993. Genomic mismatch scanning - a new approach to genetic-linkage mapping. *Nature Genetics*, **4**, 11-18.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C. & Markham, A.F. 1989. Analysis of any point mutation in DNA - the amplification refractory mutation system (ARMs). *Nucleic Acids Research*, **17**, 2503-2516.

Nouri, N., Litt, M., Pelias, M.Z., Deininger, P.L. & Keats, B.J. 1993. Tightly linked flanking microsatellite markers for the Usher Syndrome type-1 locus on the short arm of chromosome-11. *American Journal Of Human Genetics*, **53**, 1052-1052.

Nowak, R. 1994. Mining treasures from junk DNA. *Science*, **263**, 608-610.

Olson, M., Hood, L., Cantor, C. & Botstein, D. 1989. A common language for physical mapping of the human genome. *Science*, **245**, 1434-1435.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E., Mardon, G., Pollack, J., McGillivray, B., DeLachapelle, A. & Brown, L.G. 1987. The sex-determining region of the human Y-chromosome encodes a finger protein. *Cell*, **51**, 1091-1104.

Palmer, M.S., Sinclair, A.H., Berta, P., Ellis, N.A., Goodfellow, P.N., Abbas, N.E. & Fellous, M. 1989. Genetic-evidence that Zfy is not the testis-determining factor. *Nature*, **342**, 937-939.

Paul, D.B. & Spencer, H.G. 1995. The hidden science of eugenics. *Nature*, **374**, 302-304.

Piazza, L., Fishman, G.A., Kaplan, R.D., Horowitz, A.L., Hindo, W.A. & Mafee, M.F. 1987. Magnetic-resonance imaging of central nervous-system defects in Ushers Syndrome. *Retina The Journal Of Retinal And Vitreous Diseases*, **7**, 241-245.

Pieke Dahl, S., Kimberling, W.J., Gorin, M.B., Weston, M.D., Furman, J., Pikus, A. & Moller, C. 1993. Genetic-heterogeneity of Usher Syndrome type-II. *Journal Of Medical Genetics*, **30**, 843-848.

Pieke Dahl, S., Weston, M.D., Kimberling, W.J., Gorin, M.B., Shugart, Y.Y. & Kenyon, J.B. 1991. Possible genetic-heterogeneity of Usher Syndrome type-2 - family unlinked to chromosome 1q markers. *American Journal Of Human Genetics*, **49**, 200-200.

Pieke Dahl, S., Weston, M.D., Overbeck, L.D., Evans, K.L., Porteous, D.J. & Kimberling, W.J. 1993. Dinucleotide repeat polymorphism at the human olfactory marker protein (OMP) locus on chromosome-11q13.5 near tyrosinase (TYR). *Human Molecular Genetics*, **2**, 822-822.

Pinckers, A., Vanaarem, A. & Brink, H. 1994. The electrooculogram in heterozygote carriers of Usher Syndrome, Retinitis-Pigmentosa, Neuronal Ceroid-Lipofuscinosis, Senior Syndrome and Choroideremia. *Ophthalmic Genetics*, **15**, 25-30.

Ponce, M.R. & Micol, J.L. 1992. PCR amplification of long DNA fragments. *Nucleic Acids Research*, **20**, 623-623.

Porteous, D.J., Bickmore, W., Boyd, P.A., Gosden, J.R., van Heyningen, V. & Hastie, N.D. 1987. Hras1-selected chromosome transfer generates markers which co-localize Aniridia and genitourinary dysplasia associated translocation breakpoints and the Wilms Tumor gene within band 11p13. *Cytogenetics And Cell Genetics*, **46**, 676-676.

Porteous, D.J., Morten, J.E.N., Cranston, G., Fletcher, J.M., Mitchell, A., van Heyningen, V., Fantes, J.A., Boyd, P.A. & Hastie, N.D. 1986. Molecular and physical arrangements of human DNA in Hras1-selected, chromosome-mediated transfectants. *Molecular And Cellular Biology*, **6**, 2223-2232.

Prosser, J., Thompson, A.M., Cranston, G. & Evans, H.J. 1990. Evidence that p53 behaves as a tumor suppressor gene in sporadic breast-tumors. *Oncogene*, **5**, 1573-1579.

Ranum, L., Chung, M.Y., Banfi, S., Bryer, A., Schut, L.J., Ramesar, R., Duvick, L.A., McCall, A., Subramony, S.H., Goldfarb, L., Gomez, C., Sandkuijl, L.A., Orr, H.T. & Zoghbi, H.Y. 1994. Molecular and clinical correlations in Spinocerebellar Ataxia type-I - evidence for familial effects on the age at onset. *American Journal Of Human Genetics*, **55**, 244-252.

Reardon, W. 1992. Genetic deafness. *Journal Of Medical Genetics*, **29**, 521-526.

Rinchik, E.M., Carpenter, D.A. & Selby, P.B. 1990. A strategy for fine-structure functional-analysis of a 6-centimorgan to 11-centimorgan region of mouse chromosome-7 by high-efficiency mutagenesis. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **87**, 896-900.

Rinchik, E.M., Johnson, D.K., Margolis, F.L., Jackson, I.J., Russell, L.B. & Carpenter, D.A. 1991. Reverse genetics in the mouse and its application to the study of deafness. *Annals Of The New York Academy Of Sciences*, **630**, 80-92.

Rochelle, J.M., Watson, M.L., Oakey, R.J. & Seldin, M.F. 1992. A linkage map of mouse chromosome-19 - definition of comparative mapping relationships with human chromosome-10 and chromosome-11 including the MENI locus. *Genomics*, **14**, 26-31.

Rogers, K.E., Dasgupta, P., Gubler, U., Grillo, M., Khewgoodall, Y.S. & Margolis, F.L. 1987. Molecular-cloning and sequencing of a cDNA for olfactory marker protein. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **84**, 1704-1708.

Rosenfeld, P.J., Mckusick, V.A., Amberger, J.S. & Dryja, T.P. 1994. Recent advances in the gene map of inherited eye disorders - primary hereditary-diseases of the retina, choroid, and vitreous. *Journal Of Medical Genetics*, **31**, 903-915.

Royle, N.J., Irwin, D.M., Koschinsky, M.L., Mac Gillivray, R.T.A. & Hamerton, J.L. 1987. Human genes encoding prothrombin and ceruloplasmin map to 11p11-q12 and 3q21-24, respectively. *Somatic Cell And Molecular Genetics*, **13**, 285-292.

Ruano, G. & Kidd, K.K. 1991. Coupled amplification and sequencing of genomic DNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **88**, 2815-2819.

Rudiger, N.S., Gregersen, N. & Kiellandbrandt, M.C. 1995. One short well conserved region of Alu-sequences is involved in human gene rearrangements and has homology with prokaryotic-chi. *Nucleic Acids Research*, **23**, 256-260.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., mullis, K.B. & Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA-polymerase. *Science*, **239**, 487-491.

Saleeba, J.A., Ramus, S.J. & Cotton, R.G.H. 1992. Complete mutation detection using unlabeled chemical cleavage. *Human Mutation*, **1**, 63-69.

Sambrook, J., Fritsch, E.F. & Maniatis, T. 1989. *Molecular Cloning*. Cold Spring Harbour Press,

Samuelson, S. & Zahn, J. 1990. Ushers syndrome. *Ophthalmic Paediatrics And Genetics*, **11**, 71-76.

Sanda, A.I. & Ford, J.P. 1986. Genomic analysis .1. inheritance units and genetic selection in the rapid discovery of locus linked DNA markers. *Nucleic Acids Research*, **14**, 7265-7283.

Sandford, A.J., Shirakawa, T., Moffatt, M.F., Daniels, S.E., Ra, C., Faux, J.A., Young, R.P., Nakamura, Y., Lathrop, G.M., Cookson, W.C.M. & Hopkin, J.M. 1993. Localization of atopy and beta-subunit of high-affinity IgE receptor (fc-epsilon-rI) on chromosome-11q. *Lancet*, **341**, 332-334.

Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H. & Roe, B.A. 1980. *Journal of Molecular Biology*, **143**, 161-178.

Sankila, E.M., Pakarinen, L., Kaariainen, H., Aittomaki, K., Karjalainen, S., Sistonen, P. & de Lachapelle, A. 1995. Assignment of an Usher Syndrome type-III (USH3) gene to chromosome 3q. *Human Molecular Genetics*, **4**, 93-98.

Sankila, E.M., Pakarinen, L., Simola, K.O.J., Kaariainen, H. & Delachapelle, A. 1993. Genetic-linkage studies in Usher Syndrome type-III. *American Journal Of Human Genetics*, **53**, 1712-1712.

Sarkar, G., Yoon, H.S. & Sommer, S.S. 1992. Dideoxy fingerprinting (ddF) - a rapid and efficient screen for the presence of mutations. *Genomics*, **13**, 441-443.

Sawadogo, M. & Vandyke, M.W. 1991. A rapid method for the purification of deprotected oligodeoxynucleotides. *Nucleic Acids Research*, **19**, 674-674.

Schaefer, B., Kimberling, B., Thompson, J., Behm, G. & McConnell, J. 1991. Computerized image-analysis of brain morphometry in Usher Syndrome. *American Journal Of Human Genetics*, **49**, 162-162.

Servenius, B., Vernachio, J., Price, J., Andersson, L.C. & Peterson, P.A. 1994. Metastasizing neuroblastomas in mice transgenic for simian-virus-40 large-t (SV40T) under the olfactory marker protein gene promoter. *Cancer Research*, **54**, 5198-5205.

Service, R.F. 1994. Genome mapping - closing in on human and mouse maps. *Science*, **264**, 1404-1404.

Sheffield, V.C., Kimura, A.E., Fold, J.C., Bennet, S.R., Streb, L.M., Nichols, B.E. & Stone, E.M. 1992. The gene for autosomal dominant neovascular inflammatory vitreoretinopathy maps to 11q13. *American Journal of Human Genetics*, **s1 (suppl)**, A35 only.

Shen, W.H. & Hohn, B. 1992. DMSO improves PCR amplification of DNA with complex secondary structure. *Trends In Genetics*, **8**, 227-227.

Shiang, R., Thompson, L.M., Zhu, Y.Z., Church, D.M., Fielder, T.J., Bocian, M., Winokur, S.T. & Wasmuth, J.J. 1994. Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, Achondroplasia. *Cell*, **78**, 335-342.

Shinkawa, H. & Nadol, J.B. 1986. Histopathology of the inner-ear in Ushers Syndrome as observed by light and electron-microscopy. *Annals Of Otology Rhinology And Laryngology*, **95**, 313-318.

Shirakawa, T., Morimoto, K., Hashimoto, T., Furuyama, J., Yamamoto, M. & Takai, S. 1991. Linkage between IgE responses underlying asthma and rhinitis (atopy) and chromosome-11q in Japanese families. *Cytogenetics And Cell Genetics*, **58**, 1970-1971.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovellbadge, R. & Goodfellow, P.N. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240-244.

Slorach, E.M., Polymeropoulos, M.H., Evans, K.L., Seawright, A., Fletcher, J.M., Porteous, D.J. & Brookes, A.J. 1995. Regional localization of 19 brain Expressed Sequence Tags to human-chromosome-11 using PCR amplification of somatic-cell hybrid DNAs. *Cytogenetics And Cell Genetics*, **70**, 71-75.

Smith, M.W., Clark, S.P., Hutchinson, J.S., Wei, Y.H., Churukian, A.C., Daniels, L.B., Diggle, K.L., Gen, M.W., Romo, A.J., Lin, Y., Selleri, L., Mc Elligott, D.L. & Evans, G.A. 1993. A sequence-tagged site map of human chromosome-11. *Genomics*, **17**, 699-725.

Smith, R., Berlin, C.I., Hejtmancik, J.F., Keats, B.J.B., Kimberling, W.J., Lewis, R.A., Moller, C.G., Pelias, M.Z. & Tranebjaerg, L. 1994. Clinical-diagnosis of the Usher Syndromes. *American Journal Of Medical Genetics*, **50**, 32-38.

Smith, R.J.H., Lee, E.C., Kimberling, W.J., Daiger, S.P., Pelias, M.Z., Keats, B., Jay, M., Bird, A., Reardon, W., Guest, M., Ayyagari, R. & Hejtmancik, J.F. 1992. Localization of 2 genes for Usher Syndrome type-I to chromosome-11. *Genomics*, **14**, 995-1002.

Smith, R.J.H., Pelias, M.Z., Daiger, S.P., Keats, B., Kimberling, W. & Hejtmancik, J.F. 1992. Clinical variability and genetic-heterogeneity within the Acadian Usher population. *American Journal Of Medical Genetics*, **43**, 964-969.

Solomon, S.D., Geisterferlowrance, A.A.T., Vosberg, H.P., Hiller, G., Jarcho, J.A., Morton, C.C., Mc Bride, W.O., Mitchell, A.L., Bale, A.E., Mc Kenna, W.J., Seidman, J.G. & seidman, C.E. 1990. A locus for Familial Hypertrophic Cardiomyopathy is closely linked to the cardiac myosin heavy-chain genes, cri-l436, and cri-l329 on chromosome-14 at q11-q12. *American Journal Of Human Genetics*, **47**, 389-394.

Solomon, S.D., Jarcho, J.A., McKenna, W., Geisterferlowrance, A., Germain, R., Salerni, R., Seidman, J.G. & Seidman, C.E. 1990. Familial Hypertrophic Cardiomyopathy is a genetically heterogeneous disease. *Journal Of Clinical Investigation*, **86**, 993-999.

Southern,E.M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of molecular biology,* **98,** 503-517

Stafford, A.N., Rider, S.H., Hopkin, J.M., Cookson, W.O. & Monaco, A.P. 1994. A 2.8 mb YAC contig in 11q12-q13 localizes candidate genes for atopy - fc-epsilon-rI-beta and CD20. *Human Molecular Genetics*, **3**, 779-785.

Steel, K.P. 1991. Similarities between mice and humans with hereditary deafness. *Annals Of The New York Academy Of Sciences*, **630**, 68-79.

Steel, K.P. & Bock, G.R. 1983. Hereditary inner-ear abnormalities in animals - relationships with human abnormalities. *Archives Of Otolaryngology Head & Neck Surgery*, **109**, 22-29.

Steel, K.P. & Brown, S.D.M. 1994. Genes and deafness. *Trends In Genetics*, **10**, 428-435.

Steel, K.P. & Harvey, D. 1991. *Development of auditory function in mutant mice*. Elseveir Science Publishers BV, Amsterdam.

Steel, K.P. & Smith, R.J.H. 1992. Normal hearing in Splotch (sp/+), the mouse homolog of Waardenburg Syndrome type-1. *Nature Genetics*, **2**, 75-79.

Stohr, H. & Weber, B.H.F. 1995. A recombination event excludes the ROMI locus from the Bests Vitelliform Macular Dystrophy region. *Human Genetics*, **95**, 219-222.

Stone, E.M., Kimura, A.E., Folk, J.C., Bennet, S.R., Nichols, B.E., Streb, L.M. & Sheffield, V.C. 1992. Genetic linkage of autosomal dominant neovascular inflammatory vitreoretinopathy to chromsome 11q13. *Human Molecular Genetics*, **1**, 685-689.

Stone, E.M., Nichols, B.E., Streb, L.M., Kimura, A.E. & Sheffield, V.C. 1992. Genetic-linkage of vitelliform macular degeneration (Bests disease) to chromosome-11q13. *Nature Genetics*, **1**, 246-250.

Sudhof, T.C., Goldstein, J.L., Brown, M.S. & Russell, D.W. 1985. The LDL receptor gene - a mosaic of exons shared with different proteins. *Science*, **228**, 815-822.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierrymieg, J., Qiu, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R. & Waterston, R. 1992. The C-elegans genome sequencing project - a beginning. *Nature*, **356**, 37-41.

Tagle, D.S. & Collins, F.S. 1992. An optimized Alu-PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. *Human Molecular Genetics*, **1**, 121-122.

Tamayo, M.L., Bernal, J.E., Tamayo, G.E., Frias, J.L., Alvira, G., Vergara, O., Rodriguez, V., Uribe, J.I. & Silva, J.C. 1991. Usher Syndrome - results of a screening-program in Colombia. *Clinical Genetics*, **40**, 304-311.

Theriault, A., Whaley, K., McPhaden, A.R., Boyd, E. & Connor, J.M. 1990. Regional assignment of the human C1-inhibitor gene to 11q11-q13.1. *Human Genetics*, **84**, 477-479.

Thierfelder, L., Macrae, C., Watkins, H., Mc Kenna, W., Vosberg, H.P., Seidman, J.G. & Seidman, C.E. 1993. 2 new loci for hypertrophic cardiomyopathy map to chromosome-15q2 (CMH3) and chromosome-11q (CMH4). *Circulation*, 573-573.

Thierfelder, L., Macrae, C., Watkins, H., McKenna, W., Vosberg, H.P., Seidman, J.G. & Seidman, C.E. 1993. 2 new loci for hypertrophic cardiomyopathy map to chromosome-15q2 (CMH3) and chromosome-11q (CMH4). *Circulation*, **88**, 573-573.

Thierfelder, L., Macrae, C., Watkins, H., Tomfohrde, J., Williams, M., McKenna, W., Bohm, K., Noeske, G., Schlepper, M., Bowcock, A., Vosberg, H.P., Seidman, J.G. & Seidman, C. 1993. A familial hypertrophic cardiomyopathy locus maps to chromosome-15q2. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **90**, 6270-6274.

Thompson, L.M., Plummer, S., Altherr, M. & Wasmuth, J.J. 1991. A novel growth-factor receptor gene isolated from the Huntington disease gene region of human-chromosome. *American Journal Of Human Genetics*, **49**, 420-420.

Toda, T., Iida, A., Miwa, T., Nakamura, Y. & Imai, T. 1994. Isolation and characterization of a novel gene encoding nuclear- protein at a locus (D11s636) tightly linked to Multiple Endocrine Neoplasia type-1 (MENI). *Human Molecular Genetics*, **3**, 465-470.

Todd, S., Roche, J., Hahner, L., Bolin, R., Drabkin, H.A. & Gemmill, R.M. 1995. YAC contigs covering an 8-megabase region of 3p deleted in the small- cell lung-cancer cell-line u2020. *Genomics*, **25**, 19-28.

Tsujimoto, Y., Yunis, J., Onoratoshowe, L., Erikson, J., Nowell, P.C. & Croce, C.M. 1984. Molecular-cloning of the chromosomal breakpoint of B-cell lymphomas and leukemias with the t(11-14) chromosome-translocation. *Science*, **224**, 1403-1406.

Tsujino, S., Shanske, S. & Dimauro, S. 1993. Molecular-genetic heterogeneity of myophosphorylase deficiency (McArdles-disease). *New England Journal Of Medicine*, **329**, 241-245.

Upcroft, P. & Healey, A. 1987. Rapid and efficient method for cloning of blunt-ended DNA fragments. *Gene*, **51**, 69-75.

Usher, C.H. 1914. On the inheritance of retinitis pigmentosa with notes of cases. *Royal London Ophthalmology Hospital Report*, **19**, 130-236.

van Heyningen, V. & Little, P.F.R. 1995. Report of the 4th international workshop on human-chromosome-11 mapping 1994. *Cytogenetics And Cell Genetics*, **69**, 128-158.

Verpy, E., Biasotto, M., Meo, T. & Tosi, M. 1994. Efficient detection of point mutations on color-coded strands of target DNA. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **91**, 1873-1877.

Vidaud, D., Vidaud, M., Bannak, B.R., Siguret, V., Sanchez, S., Laurian, Y., Meyer, D., Goossens, M. & Lavergne, J.M. 1993. Mutations due to Alu element transcription. *European Journal of Human Genetics*, **1**, 30-36.

von Bokhoven, H., von Genderen, C., Molloy, C.M., van der Pol, D.J.R., Cremers, C.W.R.J., von Aarem, A., Schwartz, M., Rosenberg, T., von Kessel, A.H.G.M., Ropers, H.H. & Cremers, F.P.M. 1994. Mapping of the choroideremia-like (CHML) gene at 1q42-qter and mutation analysis in patients with Usher Syndrome type-II. *Genomics*, **19**, 385-387.

Walter, M.A. & Goodfellow, P.N. 1993. Radiation hybrids - irradiation and fusion gene-transfer. *Trends In Genetics*, **9**, 352-356.

Wang, M.M. & Reed, R.R. 1993. Molecular-cloning of the olfactory neuronal transcription factor OLF-1 by genetic selection in yeast. *Nature*, **364**, 121-126.

Warrick, H.M. & Spudich, J.A. 1987. Myosin structure and function in cell motility. *Annual Review Of Cell Biology*, 379-421.

Watkins, H. 1994. Multiple disease genes cause hypertrophic cardiomyopathy. *British Heart Journal*, **72**, S 4-S 9.

Watkins, H., Macrae, C., Thierfelder, L., Chou, Y.H., Frenneaux, M., McKenna, W., Seidman, J.G. & Seidman, C.E. 1993. A disease locus for Familial Hypertrophic Cardiomyopathy maps to chromosome-1q3. *Nature Genetics*, **3**, 333-337.

Watson,J.D., Hopkins,N.H. Roberts,J.W., Argetsinger Steitz,J., and Weiner,A.M., (Eds) 1987 *Molecular biology of the gene*. 4th Ed.

Weber, B.H.F., Vogt, G., Stohr, H., Sander, S., Walker, D. & Jones, C. 1994. High-resolution meiotic and physical mapping of the Best vitelliform macular dystrophy (VMD2) locus to pericentromeric chromosome-ii. *American Journal Of Human Genetics*, **55**, 1182-1187.

Weber, J.L. 1990. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics*, **7**, 524-530.

Weber, J.L. & 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.

Weil, D., Blanchard, S., Kaplan, J., Guilford, P., Gibson, F., Walsh, J., Mburu, P., Varela, A., Levillers, J., Weston, M.D., Kelley, P.M., Kimberling, W.J., Wagenaar, M., Leviacobas, F., Largetpiet, D., Munnich, A., Steel, K.P., Brown, S.D.M. & Petit, C. 1995. Defective myosin VIIA gene responsible for Usher Syndrome type IB. *Nature*, **374**, 60-61.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., vaysseix, G. & Lathrop, M. 1992. A 2nd-generation linkage map of the human genome. *Nature*, **359**, 794-801.

Went, L. 1990. Ethical issues policy statement on Huntingtons-disease molecular- genetics predictive test. *Journal Of Medical Genetics*, **27**, 34-38.

Weston, M.D., Kimberling, W.J., Moller, C.G., Dahl, S.P., Smith, R.J., Martini, A. & Milani, M. 1991. A progress report on the localization of Usher Syndrome type-II to chromosome-1q. *Annals Of The New York Academy Of Sciences*, **630**, 284-287.

Wilcox, A.S., Khan, A.S., Hopkins, J.A. & Sikela, J.M. 1991. Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs - implications for an expression map of the genome. *Nucleic Acids Research*, **19**, 1837-1843.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Burton, J., Connell, M., Bonfield, J., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonnhammer, E., Staden, R., Sulston, J., Thierrymieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinsonsproat, J. & Wohldman, P. 1994. 2.2 mb of contiguous nucleotide-sequence from chromosome-III of C-elegans. *Nature*, **368**, 32-38.

Winship, P.R. 1989. An improved method for directly sequencing PCR amplified material using dimethyl-sulfoxide. *Nucleic Acids Research*, **17**, 1266-1266.

Wooster, R., Cleton-Jansen, A-M., Collins, N., Mangion, J., Cornelis, R.S., Cooper,C.S., Gusterston, B.A., Ponder, B.A.J., von Deimling, A., Wiestler O.D., Cornelisse, C.J., Devilee,P., and Stratton,M.R. Instability of short tandem repeats (microsatellites in human cancers). *Nature Genetics,* 6, 152-156

Wright, A.F. 1992. New insights into genetic eye disease. *Trends In Genetics*, **8**, 85-91.

Wu, D.Y. & Wallace, R.B. 1989. Specificity of the nick-closing activity of bacteriophage-T4 DNA- ligase. *Gene*, **76**, 245-254.

Yoshida, T., Miyagawa, K., Odagiri, H., Sakamoto, H., Little, P.F.R., Terada, M. & Sugimura, T. 1987. Genomic sequence of HST, a transforming gene encoding a protein homologous to fibroblast growth-factors and the INT-2-encoded protein. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **84**, 7305-7309.

Youil, R., Kemper, B.W. & Cotton, R.G.H. 1995. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease-7. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, **92**, 87-91.

Young, R.P., Sharp, P.A., Lynch, J.R., Faux, J.A., Lathrop, G.M., Cookson, W.O.C.M. & Hopkin, J.M. 1992. Confirmation of genetic-linkage between atopic IgE responses and chromosome-11q13. *Journal Of Medical Genetics*, **29**, 236-238.

Zabarovskii, E.R., Domninskii, D.A. & Kiselev, L.L. 1994. Physical mapping of human genome - on the way to optimal strategy (a review). *Molecular Biology*, **28**, 765-773.