GENETIC AND PHYSICAL MAPPING IN THE TERMINAL REGION OF HUMAN CHROMOSOME 1p

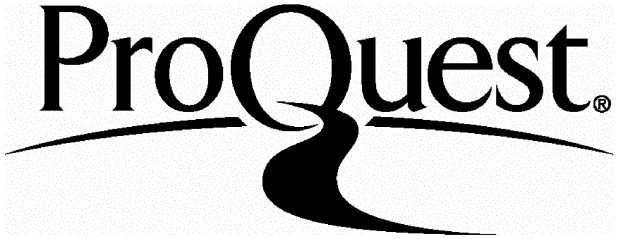
by

STAVROULLA LOUCA XENOPHONTOS

A thesis submitted for the degree of
Doctor of Philosophy in the
University of London

August 1993

The Galton Laboratory
Department of Genetics and Biometry
University College London



ABSTRACT.

Markers which lie in 1p36 have been ordered by different physical mapping techniques. Somatic cell hybrids with chromosome 1 translocations were analysed by the polymerase chain reaction and Southern hybridization resulting in the assignment of markers with respect to the translocation breakpoints. Dual colour fluorescence *in situ* hybridization (FISH) and single colour FISH allowed the accurate determination of the order of markers from 1pter to the centromere.

Pulsed field gel electrophoresis (PFGE) was also applied to study the physical relationship between loci in 1p36. Long range maps constructed around markers indicated that none of the loci analysed were physically linked. Variable methylation patterns were observed around the *PGD* locus in cell lines and peripheral white blood cell DNA. A Yeast artificial chromosome was isolated for one of these markers (p58) and will be an important nucleation point in YAC contig construction in the region. In addition this has been mapped by FISH experiments in a region commonly lost in tumours of neural crest origin and is a candidate gene whose inactivity contributes to these anomalies.

Partial exon-intron organization was determined at the *PGD* locus accompanied by a detailed characterization of the 5' CpG island. Additional 5' PGD coding sequence has been isolated by PCR screening of a cDNA library, followed by direct sequencing of the PCR product.

Genetic mapping studies facilitated by the CEPH pedigrees and data base have allowed a tentative location for the *PGD* locus in the map of chromosome 1. Segregation analysis of the D1F15S1 polymorphism in these pedigrees has led to its assignment at a location proximal to *PND*. The genetic mapping information from pedigree analysis is confirmed by physical data.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Ben Carritt for his continued advice and encouragement over the last few years. I would also like to thank Dr. D.A. Hopkinson for his valuable advice during the production of this thesis. Thanks are extended to Dr. M. Fox for technical advice and assistance in carrying out the Fluorescence *in-situ* hybridization experiments. I also thank John Attwood for all his help in carrying out the linkage analysis. I acknowledge the MRC Human Biochemical Genetics Unit for providing me with the opportunity to do this project. Finally, I would like to express my gratitude to my family and UCL friends for their interest in my work and for continued support.

To my parents and my sister.

TABLE OF CONTENTS

LIST OF FIGURES.

LIST OF TABLES.

LIST OF ABBREVIATIONS.

BSA	Bovine Serum Albumin
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
EDTA	Ethylene diaminetetra acetic acid
ESP	EDTA/SDS/Proteinase K
FISH	Fluorescence *In-Situ* Hybridization.
FITC	Fluorescein isothiocyanate
IPTG	Isopropyl-beta-D-thiogalactoside
kb	Kilobase
LMP	Low melting point
µl	microlitre
ml	millilitre
MTT	3-[4,5-Dimethylthiazol-2-yl] - 2,5 - diphenyltetrazolium bromide.
PBS	Phosphate buffered saline
PFGE	Pulsed field gel electrophoresis
PMS	Phenazine methosulphate.
PMSF	Phenylmethylsulphonylfluoride
PCR	Polymerase chain reaction
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
SDS	Sodiumdodecyl (Lauryl) sulphate
Tris	Tris (hydroxymethyl) methylamine
TRITC	Tetramethyl rhodamine isothiocyanate
Xgal	5-Bromo-4-chloro-3-indolyl-beta-D-galactoside
YAC	Yeast artificial chromosome

CHAPTER 1 INTRODUCTION

INTRODUCTION.

This thesis is concerned with the genetic and physical mapping of the distal part of human chromosome 1p. It therefore begins with a review of aspects of genetic and physical mapping. The genetic and physical map of chromosome 1 will then be presented followed by the comparative map in man and mouse. A brief reference will then be made to the significance of the comparative map in human disease associated with 1p aberrations. A more detailed section on diseases associated with 1p loci will follow. Finally, the function, isolation and polymorphism at the locus around which mapping is focused, *6-phosphogluconate dehydrogenase* will be introduced.

The international effort to map and sequence the human genome will result in an understanding of genome organization, gene function and regulation and ultimately the rapid localization of disease genes with the eventual opportunity for cure and prevention through gene therapy.

Localization of genes involved in genetic disease relies on the existence of, and the continued search for highly polymorphic markers to saturate the existing genetic map. This, if of a high enough resolution will allow rapid assignment of the gene to an interval in the map. The existence of such a map preferably at a resolution of one to five cM, will also facilitate prenatal and presymptomatic diagnosis.

Technological advances have been driven by the magnitude of the human genome mapping project; physical techniques for cloning larger inserts have evolved from cosmids to YACs, PFGE has been applied with great success to their manipulation and assays for their expression have been devised. The renaissance of irradiation reduced hybrids and the use of Alu PCR as a vehicle in the rapid generation of region specific probes, coupled with conventional

Southern analysis has allowed high resolution mapping at the 100-500 kb range of resolution. In addition computerized data analysis, accumulation and organization as well as access to international data bases has greatly advanced and assisted in this international project. Model organisms such as the prokaryote *E. coli*, the lower eukaryotes *S. cerevisiae, C. elegans, D. melanogaster* and the mouse have been thoroughly exploited for comparative mapping and to test the efficiency of developing techniques in gene targeting and gene therapy.

1.1 GENE MAPPING.

Mapping requires the products of both physical and genetic mapping technology. The two resulting maps provide different information; a most likely genetic order is generated by the frequency of recombination between markers with the interlocus distance defined in centiMorgans. One centiMorgan approximates to one recombination event in 100 opportunities. Physical maps on the other hand provide an absolute measurement of interlocus distance in nucleotides. Integration of the two maps is difficult because although in theory the order obtained from all methods should be the same, relative interlocus distances may not. There is no direct correlation between genetic and physical distance other than the approximation of one cM being equivalent to one Mb, a result of dividing total physical length by genetic length (Renwick, 1971).

A physical map in theory can never be inaccurate as to order, although the interlocus distances can vary from the true value by a small margin as a result of variation in DNA concentrations in PFGE plugs, mobility artefacts of the electrophoresis system used etc. Genetic order, for closely spaced markers at least, is sensitive to mistyping errors. With even the most rigorous internal checking regimes some of these may escape detection where alleles show Mendelian and legitimate segregation in the pedigrees being analyzed. This form of error will of course be undetected by any linkage programme used. Physical mapping may be used to reveal such errors in the genetic map and for the targeted isolation of markers that lie within predetermined intervals. We see here the dependence of the evolution of the genetic map on the physical but in general both can provide guidance for the development of the other. Physical maps produced by one method can be confirmed as well as refined by the use of alternative physical mapping strategies. Also physical mapping methods frequently generate physical resources which lead to expansion of the map.

To achieve the upper limits of resolution through meiotic mapping a great deal of effort and time is essential. Even with the most informative of markers, analysed with the most efficient techniques such as multiplex PCR, the existing small genetic distances require a large number of informative meioses for estimation of recombination frequency. This, and the undetectable error element produced by mistyping and mutation make genetic mapping a more complex approach to high resolution mapping. The large number of pedigrees necessary to which the Centre d'Etude du Polymorphismes Humain (CEPH) organization contributes 60 large three generation pedigrees, can only be used for the detection of a minimum recombination frequency of approximately 0.16%. This calculation is based on the assumption that there are approximately ten progeny per pedigree. If we then assume that this corresponds to about 160 kb, which is only a rough estimate as we know that this relationship is highly variable and dependent on sex and chromosomal location, we see a physical distance well within the limits of physical mapping strategies requiring the minimum effort and time for further studies.

The advantages of physical mapping however, do not overide the efficacy of genetic mapping since this is the only way in which any gene may be simultaneously localized to a chromosome and with respect to a large number of other loci, the only way in which any linkage to a disease may be detected and indeed the starting point for eventual isolation of disease genes. By HGM11(AUGUST 1991), almost 10,000 loci had been assigned in the human genome (Williamson *et al.*, 1991), 1416 have been assigned to one of 23 linkage groups by the efforts of the NIH/CEPH collaborative mapping group (1992). So we see the importance of present day and future advancement in these mapping technologies, and of course the success of collaboration in the eventual completion of the human genome mapping project.

1.1.1 GENETIC MAPPING.

1.1.1.1 Linkage Analysis.

Genetic mapping consists of the derivation of the order of markers and the distance between them. The frequency with which parental haplotypes remain intact after gametogenic meiosis can be used as a measure of genetic distance within them. In complex haplotypes, the patterns of meiotic rearrangement can be used to derive the order of markers making up the haplotype with or without distance measurements. The ability to define such haplotypes depends critically upon the differences between parental genotypes, and the more common these differences are, at the population level, the larger the potential data set for mapping studies.

Independent assortment of loci during meiosis can result from them being on different chromosomes or from the occurrence of cross overs between them. When the segregation of loci departs from independent assortment the loci are said to be genetically linked.

If the parental phase of markers is known, those progeny with the parental haplotypes are known as the nonrecombinants and those with a different combination of alleles relative to the parental phases are the recombinants. The fraction of these relative to the whole number of meioses analysed is the recombination fraction for that pair of loci in that family.

Where phases are unknown, recombination fractions cannot be directly estimated but are conventionally derived using a probability function, from the ratio (Z1) of the different combinatorial types in the progeny.

To determine whether two loci are linked it is necessary to calculate the ratio of two probabilities (the likelihood or odds) namely that the cosegregation of two loci would have arisen if they were linked, to the probability that the data would have arisen if the loci were unlinked. Numerically, lack of linkage is expressed as a recombination fraction of 0.5. Logarithms to the $base_{10}$ of the ratio of these likelihoods are referred to as LOD (Z) scores (logarithm of odds). By convention, in most situations a lodscore of three (odds of 1000:1) signifies linkage, that is that linkage at the recombination fraction for which a maximum lodscore is obtained, is 1000 times more likely than no linkage and further that this result would have arisen by chance. However, because of the prior probability of linkage of two markers, a lod of three actually represents a 20 fold greater likelihood for linkage versus non linkage (Morton 1955). This is expressed as follows:

Z (theta) = log10 (L (theta)/L (0.5)) where L = the probability.

Once such linkage has been established between a pair of loci it is then desirable to convert the recombination fraction to genetic distance. Mapping functions have been devised which take into consideration the fact that consecutive recombination fractions are not cumulative. This is due to the occurrence of crossovers in adjacent intervals that may cancel each other out. Also the existence of genetic interference i.e the prevention of recombination in the vicinity of an existing chiasma is considered in some mapping functions. The Haldane mapping function for example does not consider such interference. The map distance x is calculated by the following formula: x = -0.5 ln(1-2 theta) where theta is the recombination fraction (Haldane, 1919). The Kosambi mapping function;

$x = 0.5 \tanh^{-1}(2 \text{ theta})$ takes into consideration the interference (Kosambi, 1944).

Genetic interference was first noted in studies on *D. melanogaster*. Here the ratio of double recombinants to the product of single recombinants was lower than one, suggesting that the occurrence of one crossover reduced the probability of another cross over in the vicinity. The above ratio is referred to as the coefficient of coincidence (S). When S>1, negative interference is said to occur, when S<1 positive interference is signified and a value of 1 suggests no interference (reviewed by White and Lalouel, 1987).

1.1.1.2.1 Genetic Mapping by Sperm Typing.

An alternative approach to the conventional analysis of large pedigrees for genetic mapping is the analysis of single haploid sperm cells from one individual. Typing of single cells is only possible by the polymerase chain reaction since the small amount of DNA present (1.6×10^{-9}fmol) in a single cell has to be amplified. This means that only PCR based markers may be typed though the polymorphism does not necessarily have to be revealed by the PCR, as allele specific oligonucleotides can subsequently be used. The latter was demonstrated in the amplification of beta globin alleles in diploid cells homozygous for each allele and for two other loci in sperm cells from a doubly heterozygous individual (Li *et al*, 1988). Using a rigorous PCR based procedure in sperm typing of a triply heterozygous individual, three loci were ordered with confidence in a region of chromosome 3p (Goradia *et al*, 1991).

The advantages of genetic mapping via sperm typing are numerous. A very large number of meiotic events can be studied. This will produce more statistically significant information, since the maps derived will be more accurate as a result of there being such a large sample size. The ability to study such a large number of gametes means that even very short genetic distances may be measured because it is more likely that recombinants will be detected. In addition rare polymorphisms can be used provided that the donor is

informative. Last, the use of a technique to amplify most of the genome in individual sperm by the use of random primers, prior to specific locus amplification will allow the typing of numerous loci from one cell so that multipoint linkage analysis may be carried out from information derived from one cell. Another feature of this is the opportunity to confirm typings of all loci. The technique is called primer extension preamplification (PEP) and was shown to be efficient in a test on 12 sorted sperm cells and 12 loci (Zhang *et al*, 1992). The disadvantage of this is the fact that only male specific maps are derived and in humans the use of oocytes is simply not feasible.

1.1.1.2.2. Genetic Mapping with Ovarian Teratomas.

Originally ovarian teratomas (dermoid cysts), were said to be uniformly homozygous at centromeres and homozygous or heterozygous elsewere. Therefore it was suggested that they could be used for "centromere based" gene mapping. That is the further from the centromere, the greater the likelihood of being heterozygous. This is based on the supposed origin of ovarian teratomas i.e failure of meiosis II (Ott *et al*, 1976). The finding of ovarian teratomas with heterozygous centromeres raised doubts over the supposed origin and thus the suitability of ovarian teratomas for mapping. Therefore Carritt and others in 1982 investigated the origin of teratomas using somatic cell hybrids. This study illustrated the varied origins of teratomas, namely that heterozygous ovarian teratomas arose by suppression of meiosis I and homozygous ovarian teratomas arose by suppression of meiosis II. In addition, complex interchromatid exchanges must have occurred and unless the origin of ovarian teratomas is known they are not suitable for genetic mapping. But because the sites of interchromatid exchange had been mapped with respect to the then available markers, the teratomas and hybrids from them can be used to place new markers with respect to these points of exchange (Carritt *et al*, 1982b).

1.1.1.3. Genetic Variation.

The original polymorphisms used in linkage analysis were polymorphisms at the protein level. Although useful, their low heterozygosity and confinement to coding regions of the genome limits their potential in high resolution genetic mapping. The advent of DNA based polymorphism has allowed genetic mapping to proceed at a much faster rate. Restriction fragment length polymorphisms (RFLPs) resulting from point mutations, deletions or insertions have been instrumental in the construction of genetic maps and detection of linkage to disease. The usefulness of these was demonstrated by many labs in the construction of linkage maps (Botstein *et al*, 1980; White *et al*, 1985; Donis-Keller *et al*, 1987). A problem associated with RFLP based mapping is the fact that most are only dimorphic which sets an upper limit of 50% on heterozygosity.

A category of DNA sequence polymorphism also exhibiting a length polymorphism is the minisatellite content of the human genome. (Jeffreys *et al*, 1985). These are composed of unique tandemly repeated sequences (VNTRs) whose core sequences under low hybridization stringency have the ability to detect other similar minisatellites in genomic libraries which also show multiallelic variation (Nakamura *et al*, 1987). The heterozygosity of these VNTRs approaches 90% and so are of paramount importance in genetic maps, however, they tend to be preferentially located at the proterminal regions of chromosomes (Royle *et al*, 1987; Royle *et al*, 1988) and so do not allow unbiased coverage of the genome.

Variation is not just generated by a variation in the number of repeats. It was found that the actual repeat unit itself is variable. The prototype in this discovery is the VNTR D1S8, an extremely polymorphic VNTR with a heterozygosity of 97%. All other loci so far studied show this high degree of variation. The variation in D1S8 is in the form of an A to G transition giving

rise to a HaeIII site in some of the repeat units already known to possess a HinfI site each. Internal mapping of the alleles illustrates that the degree of this transition in the large number of repeats varies such that a binary code of presence or absence of this new site along the repeats is used to derive an MVR map (minisatellite variant map). Variation of this type was seen even in alleles of the same repeat unit number and a high order repetition was seen in the MVR maps. Haplotypes could be established from the similar regions of the binary codes of different alleles. (Jeffreys *et al*, 1990; Monckton and Jeffreys, 1991). The construction of the MVR map was made simpler by the use of the two types of variant repeat specific primers terminating at the polymorphic sites. One of these is used in conjunction with a minisatellite DNA flanking primer. The result of this at optimum PCR conditions is a set of PCR products whose common start is the flanking region and end is a different position which is the location of the polymorphic site (Jeffreys *et al*, 1991). The exploitation of this ultravariable system is primarily in forensic science and studies in evolution since genetic mapping would require far greater analyses for the full application.

The biased distribution of these VNTRs has been overcome by the discovery of yet another tandemly repeated variable sequence category in the human genome which is found to be distributed throughout the genome at a high frequency. These are known as microsatellites and the first to be discovered were dinucleotide repeats of the sequence TG with an estimated frequency of 10^5 per genome (Hamada and Kakunaga, 1982). These are easily analyzed and sized by amplification with flanking primers and polyacrylamide gel electrophoresis. Once found to be polymorphic and to show codominant Mendelian inheritance they can be used in linkage analyses in pedigrees. Their heterozygosity is found to be between 34% and 74%, this being lower than the maximum seen in VNTRs. Their copy number was between 10-60. (Weber and May,1988; Litt and Luty, 1989). The (CA)n dinucleotide repeat was found to exist in three different forms, the perfect repeat with no interruption by other sequences,

imperfect and compound repeats. The longest perfect repeats were found to be the most informative type (Weber, 1990).

Not long after the discovery and demonstration of the great potential of microsatellites in the field of mapping, many groups located many more of these in the genome and constructed high resolution maps composed mainly of these markers in regions of the genome devoid of conventional polymorphic markers. A tetranucleotide repeat for example was found in an intron of the *beta I GABA receptor* gene. This had seven alleles and was used to type the CEPH pedigrees resulting in its incorporation into the map of chromosome 4 (Dean *et al*, 1991). A flow sorted chromosome 20 library was screened for CA repeats and of those isolated, 16 were ordered in a 160cM map. Non random distribution of these was however noted , a lack of these in the terminal region of the chromosome. (Hazan *et al*, 1992). Similarly, a chromosome 9 specific library was constructed from a somatic cell hybrid with chromosome 9 only, and used to isolate GT repeat sequence containing cosmids. 13 were found to be highly polymorphic and were typed on the Venezuelan pedigrees giving rise to a 90 cM map of part of 9q (Kwiatkowski *et al*, 1992).

The latest and most striking endeavour to create a pure microsatellite map was infact for the whole genome (Weissenbach *et al*, 1992). In this unique attempt, rather like the earlier RFLP maps of the whole genome, a large number of markers were first isolated. Their informativeness was assessed and chromosomal assignment via somatic cell hybrids was achieved. Typing in eight of the largest CEPH pedigrees (amongst them 91 meioses) followed and led to the division of 813 markers into 23 linkage groups. The average physical spacing was one marker for every two Mb and average genetic distance was 5%. The total genetic distance covered was 90% of genome length. 605 markers had a heterozygosity of >0.7 and 553 were ordered with odds ratios >1000:1. A framework map was constructed for each chromosome composed of

5-18 markers with a mean spacing of 10-20cM. Further markers were then added to these framework maps.

Other locations of DNA sequence polymorphisms include the 3' region of Alu sequences. First, the polydeoxyadenylated tract has been shown to be variable in length and to show Mendelian inheritance. This was observed in flanking sequences of the *adenosine deaminase* gene and the *beta-globin* pseudogene. The polymorphism is referred to as an AluVpA polymorphism (Economou *et al*, 1990). Second, microsatellites have been found to be associated with the 3' end of Alu sequences. A trinucleotide repeat (TTA)n was found at the *HMG-CoA reductase* locus which showed codominant inheritance, had seven alleles and a heterozygosity of 62% . Further, a search in the GenBank Depository illustrated the existence of this repeat in the same relative location of a number of other human genes (Zuliani and Hobbs, 1990).

The existence of such a diversity of DNA sequence variation is crucial to gene mapping. The location of ultravariable VNTRs in terminal regions is a problem in that it prevents complete analysis of all regions. However, the ubiquitous distribution and rapid typing of microsatellites compensates for this kind of natural misfortune. In addition the detection of inherited silent mutations which give rise to variation was at one time limited to the fortuitous generation of an RFLP. Now, however, such changes, especially point mutations, can be detected in any region of the genome by a variety of new techniques. These allow the detection of polymorphisms which segregate in a Mendelian fashion and can be applied to regions of interest where other more conventional markers are lacking.

One of these techniques makes use of single strand conformational polymorphism (SSCP). This is based on the relation between the electrophoretic mobility of a single stranded DNA molecule and its folded conformation. This

in turn reflects its sequence and any change in sequence brings about a shift in mobility when compared to the wild type strand.

Originally DNA was digested, denatured, separated on a neutral polyacrylamide gel, blotted onto a hybridization membrane and the SSCP detected upon hybridization to the probe (Orita *et al*, 1989a). An alternative approach is PCR-SSCP. Here the region in which the search for polymorphism is carried out is amplified by PCR, either with or without simultaneous labelling, denatured and separated on neutral gels, (Orita *et al*, 1989b).

This technique was used in the search for polymorphism in Alu sequences through the use of flanking unique primers. SSCP was found that segregated in a Mendelian fashion around genes investigated (Orita *et al*, 1990). The ubiquitous distribution of these Alu sequences, their abundance in the human genome and diversity makes them ideal candidates for detection of polymorphism for use in construction of linkage maps.

Another procedure for the identification of point mutations which may be useful in genetic mapping is denaturing gradient gel electrophoresis. This technique involves migration of DNA through a gradient of denaturant (typically urea and formamide). Any difference in base composition between a wild type molecule and a mutant will be detected as a result of a difference in mobility caused by a different melting temperature (Burmeister *et al*, 1991a).

1.1.1.4. Sex Difference in Recombination Frequency.

A consistent observation when constructing genetic maps is a sex specific difference in genetic length. Generally the female autosomal maps exhibit a 90% excess in comparison to the male map in humans. An excess in female recombination was first observed in linkage studies between the *ABO* blood

group and nail patella syndrome (Renwick and Schulze, 1965). A similar difference has been observed in all the maps for chromosome 1 to date (Donis-Keller *et al*, 1987; Dracopoli *et al*, 1988; O'Connell *et al*, 1989; Rouleau *et al*, 1990; Dracopoli *et al*, 1991). However, the ratio of male to female genetic length is highly variable and infact in some cases in the male some intervals are longer than in the female. Examples of this include the intervals between the *LMYC* and *GLUT1* loci, and the marker D1S45 and *RH* on chromosome 1p (Dracopoli *et al*, 1988; Donis-Keller *et al*, 1987).

This suggests that the sex difference is not the result of an increase in trans-acting recombinational machinery in human females but that sex specific chromosomal sites for recombination may exist throughout the genome. (Thomas and Rothstein, 1991).

It is also suggested that actively transcribed regions are more likely to undergo a recombination event than non-coding sequences. This assumption is made on the basis of evidence from yeast where It is shown that recombination is stimulated by the transcription process (reviewed by Thomas and Rothstein, 1991). The application of this information to a possible explanation for sex based difference in human genetic maps was undertaken. Here it is postulated that the observed sex difference in transcriptional activation of genes during gametogenesis is reflected in the sex specific maps (Thomas and Rothstein, 1991).

1.1.1.5. The Molecular Basis of Recombination.

Investigations of the role of minisatellite sequences in recombination have demonstrated that the consensus sequence of hypervariable minisatellites was able to stimulate homologous recombination up to 13.5 fold. The stimulation was shown to occur at a distance and to be bidirectional (Wahls *et al*, 1990).

This sequence shows similarity to the octamer motif in lambda and *E coli* which was found to induce recombination of lambda and promote growth of the phage in *E. coli.* This sequence is known as a chi sequence (cross over hotspot instigator), (Smith, 1983). This minisatellite consensus sequence shows a great deal of variability, which is thought to result from unequal homologous recombination. Indeed their preferential location in terminal regions of chromosomes may explain the increase in the number of chiasmata reported for these regions (Laurie, 1985). In the mouse MHC region, where there is a great deal of recombination, sequence determination has shown sequences with an 80% similarity to the human minisatellite core sequence. The location of these is downstream to the recombination breakpoint, the same being true for chi sequences (Kobori *et al*, 1986; Uematsu *et al*, 1986).

In humans, hotspots for recombination have been reported for many sites. A classic example is the pseudoautosomal region of the Y chromosome, a region which undergoes an obligatory recombination event with the X chromosome during meiosis (Rouyer *et al*, 1986; Petit *et al,* 1988) .It is suggested that there are targets for recombinases and that variation within these targets also leads to variation in recombination. It is further suggested that the repeat unit of human recombinational hotspots converts to the Z-DNA conformation, this being the left handed helix of the classical right handed B-DNA. This conformation has been shown to be the substrate for rec1, the recombinase isolated from the fungus *Ustilgo* (Kmiec and Holloman, 1986).

1.1.1.6. The Relationship Between Genetic and Physical Distance.

The relationship between genetic and physical map distance is not linear because the recombination frequency and degree of interference throughout the genome are not constant (reviewed by White and Lalouel, 1985). The existence of

recombinational hotspots in certain regions results in the inflation of genetic length. Where pre-existing genetic maps are used to guide physical mapping studies it is always wise to bear in mind the presence and possible effects of recombination hotspots and coldspots. Many examples exist which show a great deal of variation in the exact relationship of these two forms of measurement, which is often accounted for by their location along the chromosome. In studies of loci near centromeric regions, suppression of recombination has been noted, so called cold spots for recombination, which was noted for example, in chromosome 9 studies (Fountain *et al,* 1992). In distal regions where excess chiasma formation has been reported (Laurie *et al,* 1981; Laurie and Hulten, 1985) inflation of genetic distance is seen. In chromosome 21 for example, 50-100 kb were found to be equivalent to one cM between pairs of markers near the telomere (Burmeister *et al,* 1991b). Other reports illustrate a similar correlation, 1cM=50kb; 1cM=70kb, (Steinlein *et al,* 1992; Sefton *et al,* 1990). On the other hand an approximate one cM to one Mb relationship has been found in the cystic fibrosis maps (Drumm *et al,* 1988; Poutska *et al,* 1988). Reports of a disparity in order derived using physical and genetic mapping are very infrequent in the literature. Two examples however which show this are the studies on chromosome 20 (Steinlein *et al,* 1992) and on chromosome 13q14 (Higgins *et al,* 1991). In the latter study for example, three markers were in the order: cen - D13S22 - D13S21 - D13S10 - tel, according to physical mapping studies. This order, however, was estimated to be 35000 times less likely in genetic mapping studies than the alternative order which inverted D13S22 and D13S21 (Higgins *et al,* 1990).

1.1. 2. PHYSICAL MAPPING.

Physical mapping at any resolution is made possible by the availability of a large number of techniques. These allow the localization of genes and DNA

segments to any part of the genome, establishment of order and interlocus distances. These will be reviewed in the sections that follow.

1.1.2.1. Somatic Cell Hybrids.

The establishment of interspecies hybrids which retain a few human chromosomes in a rodent background has been an important instrument in physical mapping. The cosegregation of loci and specific chromosomes or subchromosomal regions leads to their assignment to the latter. The first successfully propagated human-rodent somatic cell hybrids were those between a diploid fibroblast and an established mouse fibroblast cell line (Weiss and Green, 1967). At the time, characterization of hybrids was limited to karyotyping and the use of human antigen detection on the hybrid cell membrane. Later the establishment of hybrids from patients with cytogenetically defined translocations or deletions, coupled with other qualitative assays, led to the regionalization of many genes. Now it is possible to detect gene products in hybrids by enzyme assays and protein gel electrophoresis, and genes or anonymous DNA sequences by the techniques of Southern transfer and hybridization (Southern, 1975) and/or by PCR using species specific primers (Abbott and Povey,1991).

The large number of genes assigned to chromosomes over the last 20 years using hybrids illustrates the power and simplicity of the technique. However, instability and de novo rearrangements occur which, along with parental rearrangements, are only detectable at the level of the light microscope. Where marker density is high, invisible deletions or cytogenetically indistinguishable translocations can be characterized by co-mapping markers and breakpoints. An example demonstrating this is in the analysis of neuroblastomas. Here, somatic cell hybrids were constructed to separate the two homologues of chromosome 1. Subsequent analysis demonstrates that a locus usually lost from a particular

region of chromosome 1 has been retained in the mutant cell and translocated to another chromosome (Ritke *et al*, 1989).

A series of hybrids constructed from patients with deletions or translocations, coupled with screening for the presence of a number of loci, gives rise to a deletion map of the chromosomal region which is essentially an order of loci with respect to ordered breakpoints. This was achieved for example for chromosome 18 using PCR and genomic probes generated from a chromosome 18 specific hybrid (Kline *et al*, 1992). The use of multiplex PCR analysis for the simultaneous amplification is a more rapid route to deletion mapping of chromosomes when employing a large number of hybrids and markers. The reduction in time and labour over hybridization experiments will no doubt make multiplex PCR deletion analysis a popular screening technique. This has been accomplished for chromosome 16 employing 30 different breakpoints in 29 separate hybrids (Richards *et al*, 1991).

Radiation reduced hybrids are now widely used in physical mapping even though they are associated with some problems, such as the presence of multiple noncontiguous fragments in a single cell, instability and *de novo* rearrangements. Such hybrids are amenable to screening by all the standard somatic cell hybrid techniques and may also be used as a source for probe generation. They allow contiguous high resolution maps to be constructed which may also be complemented and corroborated by PFGE.

The production of these hybrids by the technique of Goss and Harris (1975) begins with the irradiation of the parent somatic cell hybrid with a lethal dose of X rays followed by rescue of the chromosomal fragments by fusion with a recipient rodent cell line. The fragments retained are healed and propagated by insertion into the rodent genome. The hybrids are then screened for the presence

of markers known to be in the parent hybrid and statistically analysed to infer the most likely order and distance of markers (Cox *et al*, 1990).

In the study by Cox and others (1990) 103 radiation hybrids were screened for the retention of 14 markers from the proximal region of 21q. The frequency (theta) of separation due to breakage by X rays between each pair of markers is calculated using the following formula:

Theta = $[(A^+B^-) + (A^-B^+)] / T (R_A+R_B - 2R_AR_B)]$.

Where (A+B-) and (A-B+) are number of clones retaining one marker (+) but not the other (-), T is the total number of clones analyzedR_A is the fraction of all hybrids analyzed for marker A that retained marker A and R_B is the fraction of all hybrids analyzed for marker B that retained marker B. The theta value varies from 0-1.0 where a value of 0 suggests that two markers are never separated and a value of 1.0 suggests the opposite. The theta value is then converted to a physical distance (D), the unit centiRay (cR) using the following function :

D = -ln(1-theta). A lodscore for each pairwise combination is calculated and considered statistically significant if greater than three A more sophisticated computer programme calculates the likelihood of each of 12 possible orders for a group of four markers.

PFGE was used in conjuction with this statistical analysis to confirm the order of these 14 markers which formed a contiguous group of loci over 4500kb. The correlation of these units was uniform over different regions across the map and it was found that ~53kb corresponded to $1cR_{8000}$. A later analogous study on the distal region of 21q with 28 markers found a similar correlation between $centiRays_{8000}$ and kilobases (Burmeister *et al*, 1991b).

An inherent problem for irradiation reduced hybrids is the presence of multiple fragments within one clone which may lead to the inaccurate inference of order.

Such a problem will of course be more marked if relatively few clones are screened. Another way to detect such anomalies is through the use of FISH, using the parent hybrid as probe to visualize the multiple fragments. This technique is known as chromosome painting, (reviewed by Trask, 1991). It is also possible to use the irradiation reduced hybrid itself as a probe on the parent hybrid to investigate this, the method here is known as reverse painting.

Hybrids are not only used to assign genes and markers to specific bands but are also a valuable resource for generation of probes, vital to high resolution mapping. A chromosome-specific hybrid can be used to construct a genomic library which has its human clones identified by hybridization to human DNA and then these clones are analysed for single copy markers which are in turn physically mapped and potential candidates for detection of polymorphisms. An example of this is a chromosome 21 specific hybrid for a study on Alzheimer's Disease. The clones from this hybrid were then physically localized using a somatic cell hybrid panel with different portions of human chromosome 21 (Van Camp *et al*, 1990). Another application of chromosome specific hybrids is the generation of region specific probes using Alu element mediated PCR. This has been applied to the X chromosome (Nelson *et al*, 1989), chromosome 10 (Brooks-Wilson *et al*, 1990) and chromosome 17 (Guzzetta *et al*, 1991).

Another productive way to saturate a chromosomal region with markers is the construction of a genomic library from a hybrid which has, for example, retained the closest flanking genetic markers to a locus. This was done in studies to clone the myotonic dystrophy gene. Here a somatic cell hybrid with 30Mb of human DNA previously shown to be involved in the disease was used to generate a series of radiation hybrids which were then screened for the presence of the nearest flanking markers in 9q13.3. One such hybrid which also lost a marker outside the interval of interest was used to make a genomic library from which clones could be used as region specific markers and for isolation of

cDNA clones likely to be derived from the region and therefore potential candidates for the gene (Brook *et al,* 1992).

Alu PCR coupled with generation of irradiation reduced hybrids is an even more productive route to high resolution mapping of the human genome. PCR products from parent hybrids and radiation hybrids are used to detect overlaps between the human component in different hybrids and to assess the existence and quantity of human material in the hybrid (Zoghbi *et al*, 1991). In the latter example 65 such hybrids were constructed and, through the exploitation of these interspersed repeat sequences, 61 single copy markers were derived. These not only increased the density of markers in the region but were subsequently used to divide this large hybrid panel into ten mapping intervals. Nine of the markers were also found to detect polymorphisms and will be of interest in positional cloning studies of the HLA-linked spinocerebellar ataxia gene on chromosome 6p. Another example of this kind of approach is in detection of cognate clones from genomic libraries. X chromosome-specific irradiation reduced hybrids were characterized using their Alu PCR products on X-chromosome-specific cosmid and YAC libraries (Monaco *et al*, 1991).

1.1.2.2. *In Situ* Hybridization.

In situ hybridization is a technique widely used in physical mapping and other areas of genetic research. It provides information on the location and order of DNA sequences and is fundamental to gene mapping and for the detection and characterization of structural abnormalities of the chromosome.

The original technique used radioactively labelled DNA probes (Harper and Saunders,1981), and, although informative and reliable, required a great deal of data in order to obtain a statistically reliable result and a great deal of time for autoradiography. It has now been replaced by a more rapid, precise, convenient

and versatile procedure known as fluorescence *in-situ* hybridization (Pinkel *et al*, 1986, Pinkel *et al,* 1988). This allows the simultaneous observation of probes on banded chromosomes. The probes can be single copy sequences, genomic clones in cosmid, phage or YAC vectors or Alu PCR products. Alu PCR products or total human DNA may be hybridized to hybrid cell metaphases to identify the human component and vice versa, a technique known as chromosome painting. The technique allows simultaneous visualization of probe and chromosomes because of the use of different fluorescent dyes. The probe is labelled with biotin-11-dUTP which after *in situ* hybridization is detected by avidin. which has a high affinity for biotin. Avidin is used as a fluorescein isothiocyanate-avidin conjugate which allows visualization of the immunofluorescent probe. Amplification of this fluorescent signal can be achieved by binding of an anti-avidin antibody conjugated to biotin. The use of any probe which contains repeat sequences within it has to undergo a preannealing step to human repetitive DNA to suppress the random hybridization of these sequences. Chromosomes are stained with fluorescent counterstains (for example propidium iodide and DAPI) which give R and Q banded patterns respectively.

The order of clones from a specific region can also be obtained by measurement of the position of the signal from the telomere or centromere. This distance is then divided by the total length of the chromosome and expressed as the fractional length (Lichter *et al*, 1990). Greater resolution is possible when using extended prometaphase chromosomes and greater still using interphase nuclei. Studies using genomic phage clones from the myotonic dystrophy locus (19q13.3) showed that clones could be identified as separate signals even when their actual physical distance was 125kb. To distinguish two separate signals on metaphase chromosomes they had to be separated by at least one Mb so we see here the tenfold increase in resolution using interphase nuclei as opposed to metaphase spreads (Lawrence *et al*, 1990).

The mapping of YACs to metaphase chromosomes by FISH has become a very important procedure for confirmation of locus specificity of YACs. The standard way to do this is to hybridize the whole YAC or cosmid subclones from these (Reithman *et al*, 1989). A more convenient way to map the YAC, and test for any chimeric or cotransformed sequences is hybridization of their Alu PCR products to metaphase chromosomes. Of course the ability to do this depends entirely on the existence of enough appropriately orientated Alu sequences in the individual clones to give rise to enough template DNA for hybridization. Although sequences as small as one kb have been visualized by immunofluorescence, the signal is always much stronger when the probe is larger. For example, pooled Alu PCR products from single YAC colonies were successfully located to specific metaphase bands where the total cumulative length of product from various YACs was 7-15kb (Breen *et al*, 1992; Lengauer *et al*, 1992). This rather recent exploitation of Alu PCR products in indirect FISH of YACs was extended to dual and triple colour immunofluorescence, thus ordering the clones on metaphase spreads (Baldini *et al*, 1992). We can see from even a limited survey of the applications of FISH that it is an extremely effective procedure for the assignment of any DNA sequence to its chromosome of origin and for the ordering of clones through interphase mapping. The availability of three different immmunofluorescent signal detection systems permits the positioning of three clones relative to each other in a single experiment.

1.1.2.3. Long Range Mapping by PFGE, CpG Islands and Methylation.

The technique of pulsed field gel electrophoresis (PFGE) was pioneered by Schwartz and Cantor in 1984. At the time, the maximum resolution offered was two Mb but now it has evolved into a technique which permits the separation of

DNA molecules in the 50 kb - 10 Mb size range. Numerous modifications were made to improve resolution range, to run straight tracks and improve band morphology upon hybridization (Carle and Olson, 1985; 1986; Chu *et al*, 1986; Gemmill *et al*, 1987).

Unlike conventional gel electrophoresis which is achieved by a unidirectional applied field, PFGE subjects DNA molecules to alternating electric fields, the time each field is on is known as the pulse time. The principle upon which separation is based is the differing abilities of molecules of different length to reorientate to the changed field direction. As the time required for reorientation is a function of molecular weight, then smaller molecules will reorientate faster and begin migration sooner and therefore further than the larger ones. The pulse time may be set to allow separation of a desired size range. As this is increased so is the size of fragment resolved.

In all systems, field pulses are applied at angles of greater than 90 ^{o} to each other and this results in the need to reorientate once the second field is switched on. The field inversion (FIGE) technique of Carle *et al* (1986) forces the reorientation of molecules by 180^{o}. This system applies a homogeneous/uniform electric field and the polarity is reversed at each switching cycle. Greater forward movement is achieved by having a greater pulse for forward movement. In contrast in all other systems pulse times are equal for both directions as this is essential for straight, forward migration.

Long range mapping by PFGE is assisted by the existence in the genome of restriction sites for enzymes which cleave DNA very infrequently. These so called 'rare cutters' produce large fragments of DNA which are >50kb in size and thus macrorestriction maps are produced. In this way genes are localized to subchromosomal regions, ordered with respect to others and abnormalities in patient DNA detected.

The applications of PFGE are indeed numerous. Electrophoretic karyotyping of many lower eukaryotes is made possible, for example in yeast (Carle and Olson , 1985). The resolution of chromosomes from *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* enables these chromosomes to be used as size markers in the analysis of the mammalian genome via PFGE. Long range mapping in *S.pombe* has allowed the alignment with its genetic map to take place. (Fan *et al,* 1991). In *S. cerevisiae* a map has been constructed at a resolution of 110kb thus allowing rapid localization of markers to defined regions (Link and Olson, 1991).

Detection of physical linkage between markers is made possible by sequential hybridization of probes to the same PFGE blot. Partial digests can be used which give larger fragments thus making it more likely to detect physical linkage. Many approaches allow partial digestion, amongst them is the use of the corresponding methylase for the restriction enzyme used to digest the DNA (Hanish *et al,* 1991). Regions of biological interest have also been investigated using PFGE. The relationship between alpha satellite arrays which are important in centromeric function was investigated around the centromeric region of chromosome 7p and were found to be linked on the same fragment and their maximum distance determined, (Wevrick *et al,* 1991). Preservation of large syntenic groups of man and mouse has been detected by PFGE, such as those in human chromosome 1 and mouse chromosome 3 (Kingsmore *et al,* 1990, Dracopoli *et al,* 1988). This information gives an idea of conservation of genomic organization and assistance in further comparative mapping approaches.

Central to the subject of long range mapping and concomitant detection of expressed sequences, is the CpG dinucleotide and its characteristic suppression in the genome as a result of methylation, and appearance at normal levels in

certain regions i.e CpG islands. Below, I will survey these features and their exploitation in long range mapping by PFGE. .

1.1.2.3.1. CpG Suppression and CpG Islands.

Nearest neighbour analysis led to the demonstration that the CpG dinucleotide in vertebrate DNA is markedly suppressed at about 30% of the expected frequency calculated from base composition (Swartz *et al,* 1962). This suppression is thought to stem from the methylation of cytosine to 5-methylcytosine followed by its deamination to give thymine. Thymine is not excised and replaced by a cytosine since it is a normal base. The result is that in replication a G-C pair is replaced by an A-T pair. Normally deamination (the replacement of an amino group by a keto group) of cytosine gives the base uracil which is recognised and replaced by the enzyme uracil glycosidase. The latter enzyme, however, does not recognise the intermediate in thymine production (i.e 5-methylcytosine) and so this results in the suppression of the CpG dinucleotide (Coulondre *et al,* 1978).

The significant role of methylation in CpG suppression is demonstrated by the analysis of the frequency of the CpG dinucleotide in homologous genes of methylated (humans, mouse and *Plasmodium falciparum*) and nonmethylated species (*Drosophila melanogaster, Caenorhabditis elegans* and *Saccharomyces cerevisiae)*, (Schorderet *et al*, 1992). In the nonmethylated species a greater frequency of the dinucleotide was observed in comparison to the methylated species. In the *phosphoglycerate kinase* gene for example, in yeast there are 39 CpGs whereas in the human gene there are only 15 at the same positions and in the human pseudogene only nine. Four CpGs were found in new positions in the pseudogene probably as a result of mutational balance. A comparable trend in incidence of CpG was observed for many other genes. Depression of this dinucleotide at a lower level than in methylated forms is however found in the

nonmethylated forms. This is thought to be the culmination of a methylation phase in their evolutionary history.

In spite of the foregoing, there is a fraction of DNA in the vertebrate genome which is enriched in the CpG dinucleotide, relative to other regions. Such CpG rich regions (CpG islands), were studied in the mouse and are also known as HpaII Tiny Fragment islands (HTF) because they are cleaved into tiny fragments by the enzyme HpaII (Bird *et al*, 1985). These regions are classically detected by digestion of genomic DNA with the methylation sensitive enzyme HpaII, end labelling with ^{32}P-dCTP and separation on a polyacrylamide gel to show the high frequency of the tetranucleotide sequence 5'-CCGG-3'. Characterization of this fraction by subcloning has revealed the following properties: The size range is 0.5-2kb, they are unmethylated single copy sequences, CpG:GpC ratio is >0.6, GC% is > 50% compared with 40% in the rest of the mouse genome and they show a clustering of sites for the methylation sensitive enzymes HpaII and HhaI. The average size of these islands is one kb and they represent 1% of the mouse genome ($3X10^6$kb). From this it is estimated that $3X10^4$kb of DNA is CpG rich i.e 30 thousand islands exist per haploid genome.

1.1.2.3.2. CpG Islands as Gene Markers.

Most of the expressed sequences in the mammalian genome have been shown to be associated with CpG islands at their 5' ends (reviewed by Bird ,1986; 1987). An analysis of the EMBL Database by Larsen and others (1992), confirms the already established characteristics of these islands and the types of genes with which they occur and the relative positions. All tRNA, rRNA, all constitutively or widely expressed genes are shown to be associated with a CpG island. Those genes in which CpG islands were not found were all tissue specific or of limited expression. 40% of all of the genes surveyed had an island at their 5' end. Some

pseudogenes whose expressed gene had an island were also found to possess such islands.

Long range maps derived using PFGE may be linked and complemented with restriction maps of cosmid and phage clones. However, individual clones of this size (20-40 kb) rarely suffice to bridge gaps between PFGE maps and even link themselves to the map. Contigs, of phage and cosmids have however, proved adequate and crucial to this linking procedure. Many ways have therefore been devised to achieve this. Cosmids, as a result of their larger size, have been thoroughly exploited in contig construction.

1.1.2.4. Cosmid Contig Construction.

Before the advent of YACs the only opportunity for chromosome walking experiments in positional cloning studies and large scale contig construction was provided by phage and cosmid clones. The latter were most commonly used due to their ability to hold larger inserts. Their main disadvantage is the fact that many regions of the genome appear to be unclonable or unstable once cloned and therefore underepresented in cosmid libraries. Although YACs are now at the frontier of long range mapping, cosmid analysis is the method of choice in positional cloning studies. The walking strategies devised resemble those described for YACs but will be mentioned briefly below.

Several vectors with SP6, T7 and T3 phage RNA polymerase promoters have been engineered, enabling the synthesis of single stranded end specific RNA probes (riboprobes) by transcription; these are then used to screen libraries for overlapping clones (Wahl *et al*, 1987). These promoter sites are also instrumental in restriction mapping of the clones. The individual clone to be analyzed is partially digested and then transcribed. The whole series of transcripts, that is the nested set, should provide a position for all the restriction

sites starting from the origin of transcription. The use of riboprobe synthesis coupled with cosmid multiplex analysis allows a more rapid way of detection of overlaps between clones. This was carried out for chromosome 11q specific cosmid clones (Evans *et al*, 1989).

The genome of the nematode *C. elegans* has been successfully organized into cosmid contigs using isolation of end probes and the detection of overlaps between clones through digestion of DNA, end labelling and digestion again with a different enzyme. Similarly sized fragments are detected by a computer programme and are used to declare overlaps and order cosmids (Coulson and Sulston,1986). Repetitive sequence fingerprinting is another approach used in human cosmid mapping . Here, cosmids are digested, separated on a gel and hybridized to a repeat sequence, again a similar pattern in restriction fragments detected by the probe allows the assignment of overlaps. Oligonucleotide probes such as (GT)n repeats may be used, an example is that on chromosome 16 the analysis here linked clones into 400 contigs (Stallings *et al*, 1990). Such repeat oligonucleotides are also employed in the screening of high density cosmid grids from flow sorted chromosome specific libraries (Nizetic *et al*, 1991)

1.1.2.5. Jumping and Linking Libraries.

Another way of bridging the gap between genetic and physical maps is through the use of jumping and linking libraries (Collins and Weissman, 1984; Poutska and Lehrach, 1986) These are able to cope with the large physical distance separating many loci in mammalian genomes. They are also able to overcome the problems of unclonable and repetitive sequences. Whilst chromosome walking is successful in application to cosmid, phage and YAC clones, jumping and linking clones provide a useful mapping resource when using infrequently cutting enzymes and PFGE.

The method of Collins and Weissman for construction of jumping clones essentially involves the ligation of ends of large fragments cut with infrequently cutting enzymes such as Not1 or partially with a frequently cutting enzyme such as Sau3A to a suppressor sequence containing plasmid and then cleaving outside the plasmid with a second enzyme to dispose of the intermediate DNA sequence. This small recombinant fragment is then ligated to a lambda vector which carries an amber mutation and so only recombinants containing the marker plasmid between the two end sequences of the large parent fragment are propagated in the bacterial host lacking a suppressor gene. The use of infrequently cutting enzymes for construction of the jumping library gives rise to a specific library which requires fewer clones to represent all the sites in the genome because of the large distance between each site. The use of partial digests of frequent restriction sites leads to the requirement of many more clones to cover all the sites and these clones are referred to as generalized jumping clones.

Hybridization of the single copy sequences within these clones to PFGE filters containing DNA cleaved with the same enzyme used in construction of the jumping library should result in identification of one fragment, the size of which is indicative of the size of the jump made. Both sequences on either side of the plasmid must also detect the same fragment since they are the very ends of that fragment. Clones from a generalized jumping library will also hybridize to a set of overlapping partially cut fragments and so they can be easily ordered. Clones from a specific jumping library, however, rely partly on a complementary linking library for establishment of order.

A new scheme for the construction of jumping libraries was devised which improves representation of most jumping clones from the source of DNA used in library construction. This is attributable to the use of two different cloning vectors with a wider range of cloning capacities, and the ligation of insert DNA

into the vector after cleavage with NotI thus mostly genuine jumping clones are ligated to phage vector. In addition sequencing primer sites are adjacent to the NotI cloning sites so that sequence for STS generation may be derived (Zabarovsky *et al*, 1991).

Linking libraries are constructed by cleaving DNA partially with an enzyme and then ligating this to a plasmid vector (carrying a selectable marker) with complementary ends. Cleavage then follows with the 'rare- cutting' enzyme and ligation to a phage vector with an amber mutation. Linking clones are detected by plating on a suppressor free host. Linking clones of this nature represent most of the unmethylated CpG islands since they have been found to contain unique and conserved sequences (Brockdorff *et al*, 1990) and so may be diagnostic for genes as well as instrumental in restriction mapping. Each of these clones, should detect two adjacent restriction fragments. A novel strategy based on this technique has been designed whereby the overall distance from one linking clone to the next is increased and the corresponding number of clones for mapping is decreased. This way would allow more rapid restriction mapping of the genome. Essentially the methylase BsuE is used to methylate any so called NotI-BsuE sites and therefore reduce the number of cleavable NotI sites. This approach was applied to the human X chromosome (Arenstorf *et al*, 1991).

Numerous reports have emerged where jumping and linking libraries have been successfully used. Examples of jumping clones include those of chromosome four (Poutska *et al*, 1987); mapping of the human major histocompatibility complex (Shukla *et al*, 1991); in the identification of the *CF* gene (Rommens *et al*, 1989); mapping in the human Xq27 region (Nguyen *et al*, 1989); in chromosome specific linking library construction flow sorted chromosomes have been used for human chromosome 17 (Wallace *et al*, 1989); and hybrids

for human chromosome 21 (Saito *et al*, 1991) and for the mouse X chromosome (Brockdorff *et al*, 1990).

1.1.2.6. Yeast Artificial Chromosomes.

The introduction of yeast artificial chromosomes (YACs) as cloning vectors of exogenous DNA sources at a capacity 10 times greater than that possible in the conventional vectors was realized in 1987 (Burke *et al.*) The advantages of this cloning and propagation system are numerous. First it allows rapid long range mapping accompanied by high resolution, the properties hitherto of linkage, somatic cell and cytogenetic analyses coupled with conventional recombinant technology. The application of hybridization and PCR in STS generation (sequence tagged sites), (Olson *et al*, 1989), allow saturation mapping.

The inherent problems of the system such as coligation, instability, and cotransformation are easily detected. Chimeric YACs for example, which arise either from cocloning of unrelated sequences or as a result of *in vivo* recombination between Alu sequences (Green *et al*, 1991), can be identified by FISH (Green *et al*, 1990). The presence of noncontiguous inserts would be a major problem if encountered and infact one total genomic YAC library was found to have a chimera frequency of 60-65% (Bronson *et al,* 1991). A way to overcome this is to carry out the ligation with dilute concentrations of insert DNA. In this study the frequency of chimeric YACs as determined by FISH was between 1-10% (McCormick *et al*, 1993).

The original cloning system of Burke and others has been modified to increase the size of inserts and representation of the genome (Imai *et al*, 1990; Larin *et al*, 1991) and also in the ability to map the clones after isolation (Sheiro *et al*, 1991). In addition a fragmentation vector to facilitate restriction mapping and recombination studies in inserts exists. This takes advantage of the Alu repeat

sequence to target homologous recombination with insert Alu sequences (Pavan *et al*, 1990).

Many libraries have been constructed from specific and enriched DNA sources such as somatic cell hybrids with human Xq24-Xq28 and clones were isolated by colony hybridization with assigned probes (Abidi *et al*, 1990; Wada *et al*, 1990). Single copy human genes have also been isolated by hybridization to a total human YAC library (Brownstein *et al*, 1989; Travers *et al*, 1989) and by PCR screening of a library constructed from a cell bearing a translocation (Feingold *et al*, 1990). The use of PCR in isolation of cognate YACs from YAC pools is more sensitive and accurate than hybridization to colonies, and has been extensively used for isolation of many PCR based loci (Green and Olson, 1990). The disadvantage of this approach is that it requires a constant supply of sequence information and synthesis of primers for each step in contig construction . An alternative way to isolate clones is to combine YACs in agarose blocks on PFGE blots which are durable and can be screened several times to locate YACs, to size them instantly and to detect possible physical linkage between probes. Each PFG sample is composed of four microtitre plates (384) colonies. Identification of the individual clone is then carried out by colony hybridization (Mendez *et al*, 1991).

A YAC grid may be probed with cosmid contig end probes to detect YACs which link up separate contigs thereby increasing the size of the contig and reducing the number of uncloned intermediate DNA sequences. Such was the case with regions of the *C. elegans* genome (Coulson *et al*, 1988). The advanced state of the mapping project of the *C. elegans* genome allows the examination of other mapping techniques for efficiency and precision and their potential for larger genomes. The use of repeat sequence fingerprinting as a way to YAC contig construction by detection of overlaps was tested and although capable of contig construction it had its limitations (Cangiano *et al*, 1990).

The isolation of end probes has been the most popular way to walk to the next YAC clone. These clones are used either directly in hybridization experiments if YAC colony filters are available, or as templates for sequencing to design primers for what may be defined as an STS. This may then be used to detect the overlapping YAC. Ways to isolate end probes have rapidly accumulated. The technique of plasmid rescue involves the inclusion in the YAC arms of antibiotic resistance genes and the pBR322 origin of replication such that when the whole recombinant is cleaved with an enzyme and then religated it produces two kinds of molecules. Both of these are propagated in *E. coli* and the insert manipulated as a conventional plasmid (Sheiro *et al*, 1991).

The use of Alu and Alu-vector PCR was successfully employed in the assignment of 110 X chromosome specific YACs to subchromosomal regions by hybridization of PCR products and STSs generated by sequencing of Alu-vector products (Nelson *et al*, 1991). In the absence of an Alu primer facing the vector primer an end specific PCR product cannot be obtained for use as a probe or for sequencing. A way to overcome this is to use the Vectorette system. This essentially involves digestion of the YAC to give blunt ends followed by ligation to the vectorette cassette. This is a double stranded molecule with a mismatched internal region which acts as a complementary site for a specially designed primer. This construct is then used as template in a PCR with a vector primer and so the only amplification which occurs is that where there is a vector arm and insert DNA sequence ligated to the vectorette cassette. The product is then used for sequencing or as a probe (Riley *et al*, 1990).

The above method was used in a bidirectional walk to construct a 1.5Mb contig of chromosome 21q a region implicated in Familial Alzheimer's Disease (FAD) and Down's syndrome (Butler *et al*, 1992). Sequence-tagged sites were specifically generated to facilitate the isolation of chromosome 7 specific YACs (Green *et al*, 1991) and a 1.5Mb contig assembled using such STSs within and

around the *CF* locus. None of the YACs isolated contained the whole gene so meiotic recombination between two overlapping clones was achieved through fusion of yeast strains of opposite mating type to give a contiguous gene (Green and Olson, 1990). Simultaneous screening of a total human YAC library by hybridization and PCR by Anand and coworkers also resulted in a CF contig of similar size but in derivation of a 310kb YAC which contained the *CF* gene and all of its flanking sequences (Anand *et al*, 1991).

YACs have not only shown great potential in long range mapping, they have also been specifically adapted to clone telomeric DNA fragments. This is made possible by the structural and functional similarity between the terminal repeat sequences of humans and yeast. The system here essentially uses a one arm vector with all the necessary sequences and upon ligation to 'insert' DNA propagation only occurs when the vector arm ligates to a fragment containing a healed telomere (Reithman *et al*, 1989).

The stable integration of YAC inserts into mammalian cells and expression of genes shows that they are not rearranged and that the gene remains active, thus showing great potential for their use in gene therapy through homologous recombination with the endogenous mutant counterpart. This has taken place in the case of the *G6PD* gene in monkey cells (D'Urso *et al*, 1990; Eliceiri *et al*, 1991), also with a HPRT YAC. In the latter case transfer of the YAC into HPRT deficient mouse cells by cell fusion with the yeast host transforms them to the positive phenotype (Huxley *et al*, 1991).

1.2. The Genetic and Physical Map of Chromosome 1.

Chromosome 1 is the largest human autosome constituting 8.4% (252Mb) of the human genome. By HGM11 (1991) 254 genes and 161 D segments had been assigned to chromosome 1, giving a total of 415 loci. 162 of these are polymorphic and therefore capable of incorporation into a genetic map. The contribution of many labs to the construction of genetic linkage maps for chromosome 1 has resulted in a map with a mean genetic distance of 6.7cM (Dracopoli *et al*, 1991). Since 1987 a new map has been published each year with some shared loci and some new loci generated for the purpose of increasing the density of markers. The map of Dracopoli and others in 1988 had a total of 27 ordered loci from the *pronatriodilatin (PND)* locus to the *Duffy* blood group antigen locus *(FY)*, from 1p36 to 1q23. The distance in males was 102cM and in females 230cM. The next map contained 28 loci, four distal of *PND* and eight proximal to *FY* (O'Connell *et al*, 1989). The length of this was 320cM in males and 608cM in females as a result of the inclusion of more distal loci in a region of a chromosome shown to undergo greater recombination. The genetic map of chromosome 1 derived by counting chiasmata in males was shown to be 194.3cM in length with approximately 40% of crossing over concentrated in the terminal 20% of each arm (Hulten *et al*, 1982).

For the construction of these linkage maps, the CEPH kindreds (Centre d' etude du Polymorphismes Humaine) were used. These are a random collection of large, three generation, outbread European families (Dausset *et al*, 1990).In a third publication an alternative reference panel was used, the Venezuelan Pedigree, a large interelated S. American Indian collection with a marked founder effect (Gusella *et al*, 1983).This allowed a comparison of the order generated by two different linkage programmes, as well as comparative information on order from two different sets of pedigrees. This was encouraging in that it deduced the same order and interlocus distance with two linkage

programmes (MAPMAKER and LINKAGE), the same order as the CEPH data gave, with variation only in distance, a result of using different panels and random sampling. This map encompassed 22 loci, the most distal being *alpha-fucosidase* (*FUCA*) on the short arm and the most distal on the long arm was the *renin* (*REN*) locus at 1q32.3. The markers here formed two linkage groups with a gap between the makers D1S22 and D1S10. The genetic lengths were 163cM and 182cM in males and females respectively (Rouleau *et al*, 1990). The penultimate contribution to this mapping has been the CEPH consortium map which has more than doubled the number of uniquely placed loci on this chromosome with a sex averaged length of 308cM (Dracopoli *et al*, 1991). These maps are presented in figure 1.

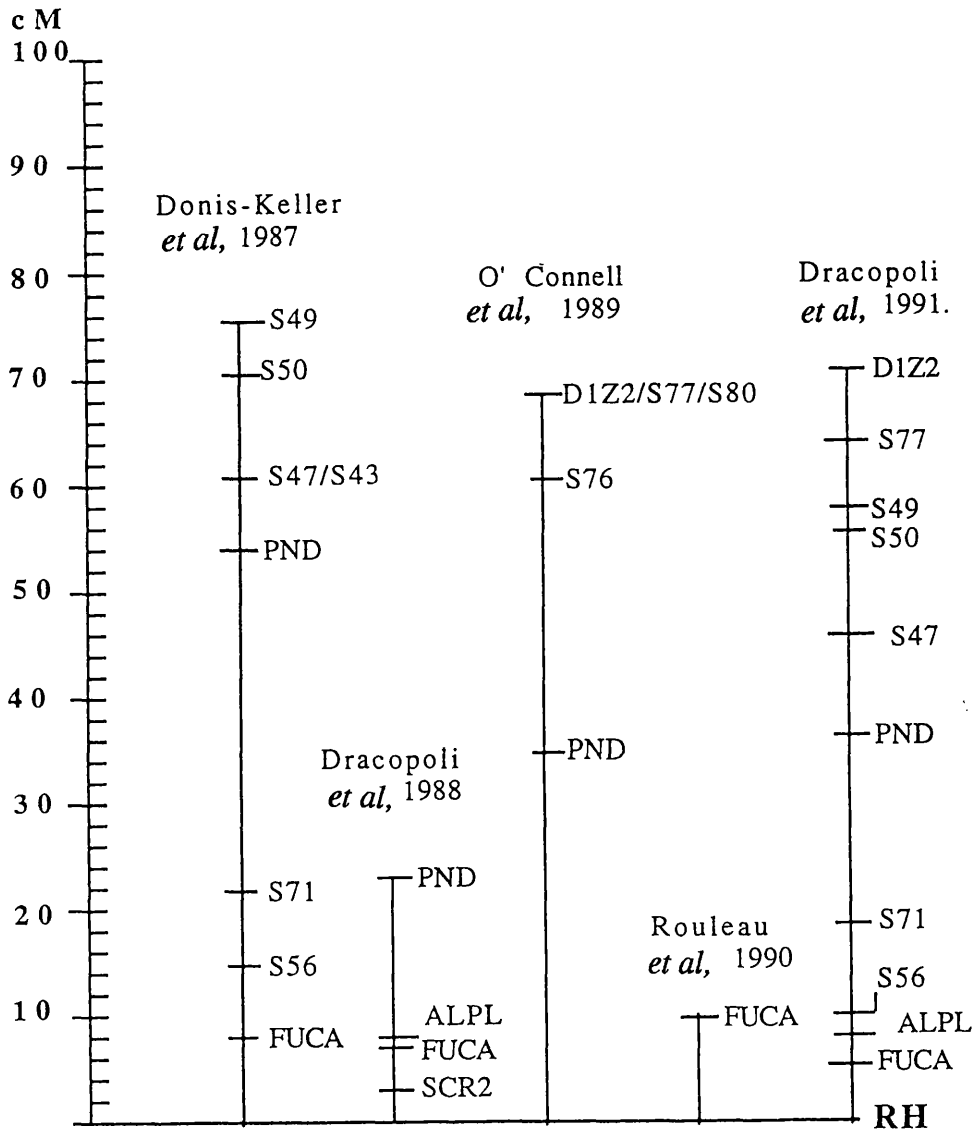


Figure 1. Comparison of Sex Averaged Genetic Maps of the Distal Region of Human Chromosome 1p.

The maps are presented from RH up to the most distal marker on 1p. The order of shared markers between the maps is the same but the distance between them varies.

Locus Key: FUCA- *alpha-fucasidase*, PND - *pronatriodilatin*, ALPL- *alkaline phosphatase*, SCR2 - *Gardner - Rasheed feline sarcoma viral oncogene homologue*, RH - *Rhesus* blood group, S43/47 - D1S43/47, S49 - D1S49, S50 - D1S50, S56 - D1S56, S71 - D1S71, S76 - D1S76, S77 - D1S77, S80 - D1S80.

The recent linkage map of the whole genome using CA repeats led to the construction of a chromosome 1 map with 69 loci, 48 of which were mapped with odds of at least 1000:1 (Weisssenbach *et al*, 1992). The average spacing of these was 4.28cM, giving a slightly more dense map than the previous. Since CEPH pedigrees were typed for these new markers, integration of these two maps is feasible. This will lead to the generation of a map with approximately twice as many loci, give greater coverage of the chromosome, and a mean distance below four cM would be obtained. This of course applies to the genome as a whole.

The distal region of chromosome 1 harbours three enzyme loci *glucose dehydrogenase (GDH) , enolase 1 (ENO1)* and *6-phosphogluconate dehydrogenase (PGD),* which still have not been ordered. Earlier studies using the existing enzyme polymorphisms placed *PGD* distal to *Rhesus (RH)* (Cook *et al,* 1974;1977). In MRC families *PGD* was linked to *GDH* and *RH* but the latter pair remained unlinked (King and Cook, 1981). On the other hand *ENO1* was found to be linked to *RH* and *FUCA* implying that this locus is proximal to *GDH* (Giblet *et al*, 1973). A possible order through physical studies was suggested, this being *GDH-ENO1-PGD* with *PGD* in a proximal position (Carritt *et al*, 1982a). Maximum likelihood analysis using two point data still failed to resolve with any confidence the order of *PGD* and *ENO1* whilst *GDH* was assigned a more distal location (Sherman *et al*, 1984). At the time of these analyses, the only polymorphisms available for these loci were relatively uninformative and detected at the protein level. The recent discovery of a BamHI RFLP at the *PGD* locus has opened the possibility for locating it within the present CEPH consortium map (Kleyn,1990). The development of more informative polymorphisms at the other two loci will of course be advantageous to their accurate inclusion in this map.

The homologous region in the mouse containing this group of loci is chromosome 4 (Lalley *et al*, 1978). *ENO1* is placed distal to *PGD* so this is valuable predictive information on the orientation of these two genes. The intergenic distance is seven cM (Davisson *et al*, 1991) which is in agreement with the result of two point analysis, that is a value of 8.3 cM in females (Sherman *et al*, 1984).

Physical mapping studies have given little additional information on physical order of large groups of loci and intergenic distance. The small number of genes which have been finely mapped is clustered on the long arm and just distal to the centromere on 1p. There does appear however to be a high degree of synteny and linkage conservation between human chromosome 1 and mouse chromosomes 1, 3 and 4 and mapping data obtained in the mouse can often be used to suggest gene order and location in man. Figure 2 presents the homology between distal human 1p and mouse chromosome 4.

Two genes in the human chromosomal region 1p13.3-1p13.1, *nerve growth factor, beta polypeptide (NGFB*) and *thyroid stimulating hormone, beta (TSHB)* were found to be syntenic on mouse chromosome 3 and to show by PFGE a similar maximum intergenic distance of 310kb in the human and 220kb in the mouse. Their syntenic conservation is thought to be the result of their mutual regulatory mechanism, being genes that are antithetically controlled by thyroid hormones (Dracopoli *et al*, 1988). These two loci and four others in the genetic map of human 1p22-1q21 are on mouse chromosome 3, and 15 genes in the human 1q21-1q32 region are on mouse chromosome 1 (Moseley and Seldin, 1989). Further physical mapping studies in the region of *TSHB* and *NGFB* led to the finding that other genes shown to be within the conserved linkage group were also in the same physical order within two Mb as determined by PFGE (Kingsmore *et al*, 1990). Similar long range mapping studies in the conserved groups of human chromosome 1q and mouse chromosome 1 illustrates the same

gene order and distance in both species. The similarities here are extended to the distribution of CpG dinucleotides which occur in the recognition sequence of infrequently cutting enzymes (Kingsmore *et al,* 1989).

In addition paralogous regions on human chromosomes 1p and 12p are seen on mouse chromosomes 4 and 6 respectively. There are gene families which have different members on different chromosomes. For example *glucose transporter 5 and glucose transporter 1* (*GLUT5/GLUT1*) and *guanine nucleotide binding protein, beta polypeptide 1* (*GNB1*) are on chromosome 1p, whilst other family members, *glucose transporter 3 (GLUT3*) and *guanine nucleotide binding protein, beta polypeptide 3 (GNB3*) appear on 12p together. In the mouse, the first pair appear as a syntenic pair on chromosome 4 and the pair on human 12p which are in similar families appear together on mouse chromosome 6

An important goal in gene mapping is to be able to integrate the genetic and physical maps. This would lead to expansion and refinement, in addition to the deduction of the relationship of the two different types of map in different regions of the genome. A recent attempt to achieve this aim for chromosome 1 (Collins *et al*, 1992), led to the production of a so called composite map incorporating 177 loci. 127 of these were ordered by multiple two point calculations using the present CEPH map as a framework map. A lodscore of one was used for linkage due to the greater error tolerance level of pairwise analysis and also error elimination by the programme used, (MAP90PROGRAM). Three loci were found not to be in the same order as the CEPH map and a further two pairs of loci were inverted when compared to the CEPH map. Integration of this new genetic map with 50 physically assigned loci gave a composite map of 177 continuous loci.

The discrepancies reported no doubt reflect the difficulty of this procedure although their frequency is quite low considering the number of loci

incorporated in this map. For example the genetic order of four loci disagrees with physical assignment and in addition contradictory orders are obtained for two loci by high resolution ISH and cytogenetic studies. Colinearity with the mouse map is not consistent and this is thought to be the consequence of error in both maps and also to the heterogeneous nature of mouse linkage data. Within this map there is still a large proportion of loci which are locally unordered (ranked-0) and without linkage information so they engender discontinuouty within this map. (Collins *et al*, 1992).

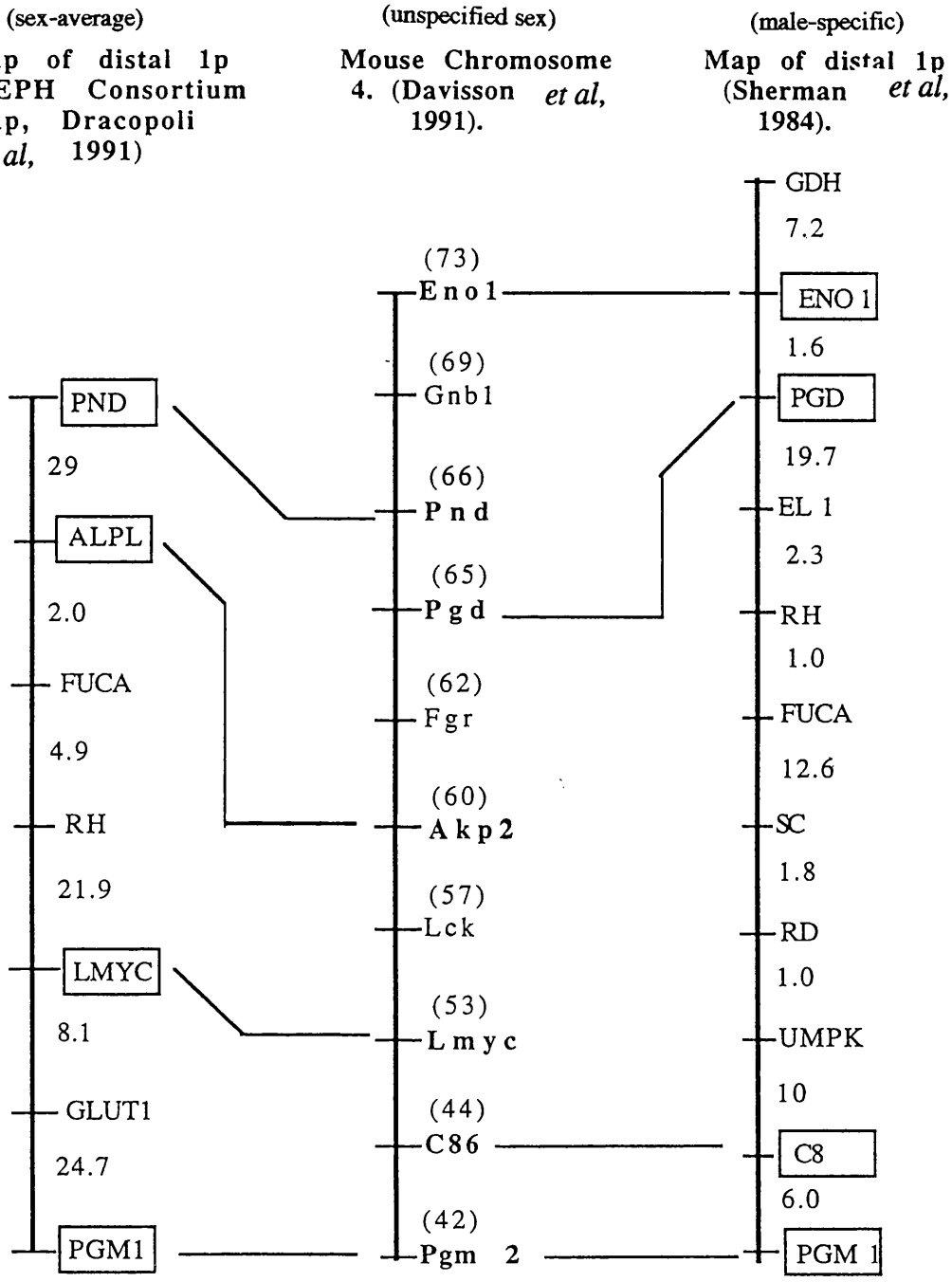


Figure 2. Man and Mouse Homologies, Human Chromosome 1p and Mouse Chromosome 4.

In the above figure two 1p maps are presented with the distance in cM between adjacent markers indicated. In the centre the map of mouse chromosome 4 is presented with the distance from the centromere for each marker indicated in brackets. Those loci showing linkage conservation are indicated in boxes in 1p and in bold in the mouse. Lines drawn across the 1p loci to the homologous loci in mouse connect those whose order is conserved.

1.3. The Importance of Comparative Mapping.

The most important role of comparative mapping information is in the prediction of gene location based on homologous maps. As already mentioned in section 1.2, a large syntenic group exists between distal 1p and mouse chromosome 4 and proximal 1p homology exists with mouse chromosome 3, figure 2 shows the homology between distal 1p and mouse chromosome 4.

Comparative mapping extends to disease homologies. Where a human disorder shows similarity to a murine phenotypic mutation then there is great potential for the localization of the defective gene(s) in the human through comparative mapping, and of course the exploitation of the mouse model for further experimentation. To date approximately forty five mouse mutants have been identified which might represent models for human single gene disorders (Darling and Abbott, 1992).

Chromosome 1 anomalies do not, however, enter into this group. The mouse phenotypic mutations recorded by Searle and others (1989) which map to that part of mouse chromosome 4 with homology to 1p are listed in table 1. These lie in the region which shows a high degree of syntenic and linkage conservation. It is possible therefore that some of the diseases already showing linkage to this region and/or bearing molecular aberrations in this region, may be phenotypically analogous to mouse mutations and so mouse models may be available for the isolation and analysis of the mutation. Also, diseases not yet found to be associated with chromosome 1p but which are homologous in the mouse and map to this conserved region, may be related to 1p mutations. Comparisons can for example be made with human diseases such as CMM, non insulin dependent diabetes melitus (NIDDM) and achondroplasia (Table 1).

Diabetes (db) in the mouse is mapped to chromosome 4, four cM away from the *Pgm-2* locus (Searle *et al*, 1989). In man the *glucose transporter 1 (GLUT1)* locus maps to 1p33, 23 cM from *PGM 1*. (Shows *et al*, 1987). Glucose transporter 1 is involved in the transportation of glucose into most cells. Insulin increases the uptake of glucose via GLUT1. A *glucose transporter 1* gene mutation is thought to be a cause for abnormality in response to insulin. An association has also been found between a GLUT 1 RFLP and 89 diabetics (Li *et al*, 1988). This information and the mapping of the db locus in the syntenic region in mouse suggest that this may be a disease homology between the two species.

A defect in the *cartilage matrix protein 2 (CRMP2)* gene may be involved in achondroplasia. This disease in the mouse maps to the region homologous to 1p, where *CRMP2* has been mapped. Its exact location is in 1p35, (Jenkins *et al*, 1990). Although 1p abnormalities have not been implicated in autosomal dominant achondroplasia, and *CRMP2* itself has not been linked to a disease, its presence in the extracellular matrix of cartilage, and the fact that the cn phenotype (achondroplasia) in mouse is linked to the homologous region leads to speculation that achondroplasia in man may be related to a *CRMP2* mutation.

In an interesting case, achondroplasia was found to coexist in a newborn with neuroblastoma (Verloes *et al*, 1991). The karyotype of the child was normal. However, as loss of heterozygosity observed in many neuroblastoma tumours and cell lines (Weith *et al*, 1989; Martinsson *et al*, 1989) involves markers residing in 1p36, it was proposed that a microdeletion may have occurred which resulted in the loss of both the neuroblastoma suppressor gene and a gene whose absence results in achondroplasia. Since the cn phenotype in mouse maps to this homologous region, there is good reason to search this part of the human genome for the gene involved in achondroplasia.

Other mouse phenotypes mapping to the conserved region are the following: meander tail, (mea); clasper, (cla); cribriform degeneration, (cri); head bleb, (heb) pupoid foetus, (pf); snubnose, (sn); polysyndactyly, (Ps) These phenotypes are manifestations of neuromuscular and developmental abnormalities. Any human disorders with similarities to these will have a mouse model to focus on.

Mouse Mutation Mapping to Conserved Region of Chromosome 4.	Possible Human Loci where mutation gives a homologous disease with the mouse
misty (m)	Cutaneous malignant melanoma (CMM)
diabetes (db)	Glucose transporter 1 GLUT1)
achondroplasia (cn)	Cartilage matrix protein 2 (CRMP2)

Table 1. Possible Disease Homologies in Man and Mouse.

1.4. Genetic Disease Associated with Distal 1p Aberrations and /or Linkage to Distal 1p Markers.

Chromosome 1 abnormalities and linkage to chromosome 1 loci has been established for several disease-related phenotypic traits. These include neuroblastoma (Gilbert *et al*, 1982) cutaneous malignant melanoma (Greene *et al*, 1983), medullary thyroid carcinoma (Mathew *et al,* 1987), breast carcinoma (Chen *et al*, 1989; Genuardi *et al*, 1989) and hypertension (Wilson *et al*, 1991). Various types of solid tumour are associated with chromosome 1 abnormalities. Although a relatively consistent and significant loss of chromosome 1 material is reported, solid tumours also show loss of at least one other chromosomal region. The region most frequently involved on chromosome 1 is 1pter-1p31. Of particular interest here is the similar conclusion of loss of heterozygosity studies (LOH) studies on four neoplasms which are embryologically derived from neural crest cells. These include neuroblastoma, cutaneous malignant melanoma, pheochromocytoma and medullary thyroid carcinoma. Whilst evidence has accumulated to implicate an alternative locus in the cause of the last three neoplasms, evidence still remains strong for the involvement of chromosome 1 loci in the progression of neuroblastoma.

1.4.1. Neuroblastoma.

Neuroblastoma is the second most common extracranial paediatric solid tumour with an annual incidence of eight per million individuals under 15 years. The incidence is greater in the first three years of life (Christiansen and Lampert, 1988). It is a disease of postganglionic sympathetic neurones, showing an autosomal dominant pattern of inheritance and said to fit Knudson's two mutation hypothesis (Knudson *et al*, 1971). Occurring in both familial and sporadic forms the non-hereditary form would result from two (postzygotic) somatic mutations in a single cell which develops into one tumour. The

hereditary form would arise when the first mutation is transmitted in the germline (prezygotic) and so it is present in all cells. A somatic mutation in the normal allele of the target tissue would result in tumourigenesis. Such individuals would have a greater predisposition to the disease which would occur earlier in life (Brodeur and Fong,1989).

Cytogenetic analyses of tumours and neuroblastoma cell lines have led to the detection of abnormal tumour-associated karyotypes whilst the constitutional karyotype of patients was normal. At the cytogenetic level, the most common observation was a deletion with the breakpoint in 1p32 whilst the most common LOH was from 1pter-1p34. 70% of tested material showed these cytogenetic defects and further, these were observed in all cells tested. Those tumours without apparent 1p abnormalities may have had deletions too small to be visible by karyotyping alone (Gilbert *et al*, 1982).

Loss of heterozygosity studies using RFLPs in tumours and tumour cell lines, as well as the construction of hybrids as a means to separate the normal from abnormal chromosomes, have shown that LOH is a frequent event in these cells. eight of nine tumours were shown to lose alleles in the region 1pter-1p35 with the consensus deletion in 1p36.12-1p35 (Martinsson *et al*, 1989; Weith *et al*, 1989). Another analysis found that 30% of tumours showed LOH in the region 1p36.1-1p36.3 and that this was strongly correlated with an advanced stage of the disease, as well as with amplification of the oncogene *N-myc* on chromosome 2. There was no LOH detected in the short arm of chromosome 1 in early stage tumours (Fong *et al,* 1989). Overall these studies narrow the region of interest to 1p36.3-1p36.1.

In addition two independent studies report the association of neuroblastoma with abnormal constitutional karyotypes involving 1p36 (Laureys *et al*,1990; Biegel *et al,* 1993). The first of these patients had a constitutional reciprocal

translocation, t(1;17)(p36;q12-q21). This abnormal karyotype was observed in skin fibroblasts as well as in peripheral white blood cells. FISH studies with the terminal marker on 1p, D1Z2 showed that this marker was translocated to 17q (Laureys *et al*, 1990). In the second patient, a constitutional interstitial deletion was found. Cytogenetic analysis showed that this involved 1p36.2-1p36.1, and molecular studies show that the region deleted is proximal to D1Z2 and distal to the *Na+/H+ antiporter (APNH.*) This deleted region includes the D1S47 locus which is shown by other studies to be lost in tumour tissue and tumour cell lines (Fong *et al,* 1989). It is found that the maternal allele is the one lost in the patient with the constitutional interstitial deletion.

These two reports on abnormal constitutional karyotypes involving 1p36 in association with neuroblastoma give strong evidence for the presence of a locus in the region whose loss or inactivation predisposes the children to the disease.

A second genetic lesion on chromosome 1 is the translocation of the *avian myelocytomatosis viral oncogene homologue 1(MYCL1)* to another part of the genome. This is thought to involve a complex rearrangement where a break occurs proximal to the gene and a deletion distal to it thus giving rise to what appears as an interstitial deletion. In this analysis the genes *PGD, PND* and*FGR (Gardner-Rasheed feline sarcoma viral oncogene homologue)* were also shown to be lost from one chromosome in all four lines tested (Ritke *et al,* 1989). This translocation which was detected in just one of four tumours in the above study was also seen in three out of five tumours in another study suggesting that it is also a common event. LOH studies with 20 probes still suggest that the loss of a gene present in the region 1p36-1p35 in some way contributes to neuroblastoma (Hunt and Tereba, 1990).

It is likely that the gene on 1p is a tumour suppressor, a function not consistent with the known properties of any of the genes used in the LOH studies. A

recently mapped gene in this region which perhaps fulfils this requirement is the protein kinase *p58* which maps to 1p36, the commonly deleted region in neuroblastoma, (Eipers *et al*, 1991). An analysis of the function of this gene has shown that it is involved in cell division; when overexpressed it stops division at late telophase, and when expression is lost cells replicate DNA at an accelerated rate (Bunnell *et al*, 1990). This gene is thus a candidate by function and location for the tumour suppressor lost in neuroblastoma. Its loss or inactivation would probably result in the expression of a dominantly acting neuroblastoma gene. An investigation to look for the loss of this *p58* gene in neuroblastoma cell lines seems at this stage to be a most plausible suggestion.

The amplification of the gene *N-myc* and regions around this gene is manifested as extrachromosomal elements known as double minutes (DM) or as homogeneously staining regions integrated into the genome (HSRs). The degree of amplification is found to range from 3-300 fold per haploid genome in 38% of tumours and 50% of neuroblastoma cell lines (reviewed by Brodeur and Fong, 1989).

A correlation has been established between the severity of the disease and both 1p loss detected and *N-myc* amplification. Essentially, those individuals showing partial 1p monosomy and *N-myc* amplification have only a 10% chance of survival, whereas those lacking these two abnormalities have a 90% chance of survival (Christiansen and Lampert, 1988).

1.4.2. Cutaneous Malignant Melanoma.

The suggestion of linkage between cutaneous malignant melanoma (CMM) and the associated precursor, dysplastic nevus syndrome (DNS), to a locus (*RH*, Lod=2.0 at a theta value of 0.3) on 1p was made assuming a dominant pattern of inheritance with incomplete penetrance (Greene *et al*, 1983). Further studies in

North American affected pedigrees show that CMM and DNS are linked (Lod=3.9 at theta=0.08), (Bale *et al*, 1986) and that the CMM gene is linked to markers in 1p36, 7.6 cM distal to the *PND* locus and proximal to the marker D1S47 (Bale *et al*, 1989). This localization is confirmed by other studies which place the gene 8cM distal to the *PND* locus (Dracopoli *et al*, 1989). In addition it was also found that in 43% of melanomas and 52% of melanoma cell lines there was LOH of 1p markers in the same region where linkage is shown. This LOH, as in neuroblastoma, was shown to occur in advanced stage tumours thus implicating the defect of a gene in this region as having an effect late in the disease (Dracopoli *et al*, 1989).

A number of reports have excluded the possible linkage of CMM/DNS to 1p markers already found to be linked by the above reports (Cannon- Albright *et al*, 1990: Kefford *et al*, 1991; Nancarrow *et al*, 1992). The last two groups had analyzed Australian pedigrees and the first, three Utah pedigrees. The loci D1S47 and *PND* were common to all three studies since these were closely linked to the disease in the earlier study (Bale *et al*, 1989). A total of eleven 1p markers were typed and analysed for linkage to CMM or CMM/DNS between the three groups. A common finding was the exclusion of linkage of this disease to any of the markers spanning a 135cM region of 1p. This therefore suggests that the disease is genetically heterogeneous, and also that the association with the precursor syndrome is variable.

We see here the agreement in an autosomal pattern of inheritance, of incomplete penetrance and in the existence of genetic heterogeneity in CMM. Another chromosome implicated in this disease is chromosome 9p (Fountain *et al*, 1992), this supports the suggestion of genetic heterogeneity in CMM. A 2-3Mb region on 9p just proximal to the *alpha-interferon* gene cluster is heterozygously or homozygously deleted in 85% of melanoma tumour and cell lines. It is

proposed that the *IFNA* genes are potential candidates due to their location in this commonly deleted region and because of their anti-proliferative role.

1.4.3 A Summary of Other Diseases Associated with Distal 1p Aberrations.

The above diseases and other disorders associated with distal 1p abnormalities are summarized in table 2. Biochemical or genetic lesions are mentioned for each. A review by Povey and Parrington (1986) mentions some of these diseases and others associated with more proximal chromosome 1 loci.

Disease	MIM no.	Biochemical or Genetic Lesion	References
Haemolytic aneamia, (Erythroblastosis faetalis; Rh-null disease)	111700	Rh incompatibilty during transfusion and pregnancy. Absence of red cell structural protein with Rh incompatibility.	Levine *et al,* 1941. Perez-Perez *et al,* 1992.
Elliptocytosis-1	130500	protein 4.1 deficiency,disorder of red cell membrane	Conboy *et al,* 1986
Erythrokeratodermia Variabilis	133200	Unknown	van der Schroeff *et al*, 1984
Galactose epimerase deficiency	230350	reduction in the interconversion of UDP-galactose and UDP-glucose, shortage of UDP-galactose a precursor for glycoproteins & glycolipids.	Mitchell *et al,* 1975
hypophosphatasia, Infantile	241500	Alkaline phosphatase deficiency of liver,bone and kidney, substrates elevated in plasma i.e phosphoethanolamine & inorganic pyrophosphate.	Chodirker *et al,* 1987.
fucosidosis	230000	alpha-fucosidase deficiency in liver, accumulation of fucose in all tissues.	Darby *et al,* 1988

porphyria cutanea tarda	176100	Uroporphyrinogen decarboxylase deficiency in liver & red blood cells, reduction in production of coproporphyrinogen for haem biosynthesis	Kushner *et al,* 1976
Neuroblastoma	256700	Loss of a putative tumour suppressor in 1p36	Martinsson *et al,* 1989; Weith *et al,* 1989.
Cutaneous malignant melanoma	155600	Loss of a putative tumour suppressor? Familial predisposing mutation linked to *PND*	Bale *et al,* 1986; Dracopoli *et al,* 1989.
Neuronal Ceroid Lipofuscinosis (CNL1)		Neural and extraneural accumulation of ceroid lipofuscin-like storage cytosomes due to an unknown biochemical defect. (Defective gene linked to D1S7, D1S79 & D1S57.)	Jarvela *et al,* 1991.

Table 2. A Summary of Diseases Associated with Distal 1p Gene Product Deficiencies or Showing Linkage to 1p Loci.

1.5. 6-Phosphogluconate Dehydrogenase: Function, Isolation and Polymorphism.

1.5. 1. Biochemical Role of 6 - Phosphogluconate dehydrogenase.

6-phosphogluconate dehydrogenase (E.C.1.1.1.44) catalyzes the oxidative decarboxylation of 6-phosphogluconate (6-PG) to ribulose-5-phosphate, and forms part of the pentose phosphate pathway (Horecker *et al,* 1951). This pathway is an alternative to the major glycolytic pathway and generates NADPH and ribose 5-phosphate. NADPH is used in reductive biosynthesis of fatty acids and steroids, and ribose 5-phosphate is used in nucleotide biosynthesis. This pathway is also important in red blood cells as it maintains their redox potential (Eaton and Brewer, 1974).

The oxidative part of the pentose phosphate pathway (figure 3) starts with the dehydrogenation of glucose 6-phosphate to form 6 phosphoglucono-lactone. This is then hydrolyzed to give 6-phosphogluconate, the substrate for 6-PGD, which is oxidatively decarboxylated to give ribulose 5-phosphate. The final step in the synthesis of ribose 5-phosphate is the isomerization of ribulose 5-phosphate by phosphopentose isomerase.

A reversible link exists between the pentose phosphate pathway and glycolysis.This involves the nonoxidative branch of the pathway which leads to the conversion of ribose 5- phosphate to glyceraldehyde 3- phosphate which is an intermediate compound of the glycolytic pathway. This reversible link allows the levels of NADPH, ATP, ribose 5-phosphate and pyruvate to be controlled to meet cellular demands.

The major role of the pentose phosphate pathway appears to be in the generation of NADPH for the reductive biosynthesis of fatty acids and steroids

(Rognstead and Katz, 1979), since there are greater levels of activity in adipose tissue, liver and adrenal glands (Glock and McLean, 1954).

The third role of the pentose phosphate pathway is in the protection of red blood cells from damaging superoxides and free radicals (Eaton and Brewer, 1974). Catalase and reduced glutathione (a tripeptide) protect the red blood cell from oxidant damage. Catalase reduces hydrogen peroxide and it is thought that its activity may be NADPH dependent (Eaton *et al*, 1972). The only source of NADPH in red blood cells is through the pentose phosphate pathway. Similarly, NADPH in red blood cells reduces the disulphide (oxidized) form of glutathione to the sulfhydryl form. This reaction is catalyzed by glutathione reductase. In red blood cells deficient in G6PD the level of NADPH is reduced, resulting in decreased life span and, if the level is very low, haemolytic anaemia. The oxidation of reduced glutathione protects haemoglobin from oxidation of its cysteine residues and keeps it in the ferrous state. For the maintenance of protection it is essential that the oxidized form of glutathione (GSSG) be converted back to the reduced form (GSH) by NADPH. These redox reactions constitute the glutathione cycle.

The pentose phosphate pathway in liver is controlled by diet and hormonal status (Glock and McLean, 1955; Micsicek and Towle, 1982; 1983). Starved rats have a reduced level of both G6PD and PGD activity and, once fed a high carbohydrate diet, G6PD and PGD activity is not only induced but at a level three-ten times greater than that of control animals (Glock and McLean, 1955). This phenomenon is known as "overshoot". The metabolism of polyunsaturated fatty acids suppresses this induction by inhibition of transcription of *G6PD* and *PGD* as well as that of other lipogenic enzymes (Tomlinson *et al,* 1988). Lipogenesis itself appears to influence the rate of the pentose phosphate pathway through the consumption of reducing power and the concomitant accumulation of $NADP^+$, (Tepperman and Tepperman, 1963). When the rate of lipogenesis is

low, other NADPH dependent reactions such as fatty acid chain elongation induce the pathway in the liver (Tepperman and Tepperman, 1965).

In the rat,hormonal control is exerted by the thyroid hormones triiodothyronine (T3) and thyroxine (T4). An increase in transcription occurs of both *G6PD* and *PGD* genes in response to thyroid hormones (Miksicek and Towle, 1982; 1983).

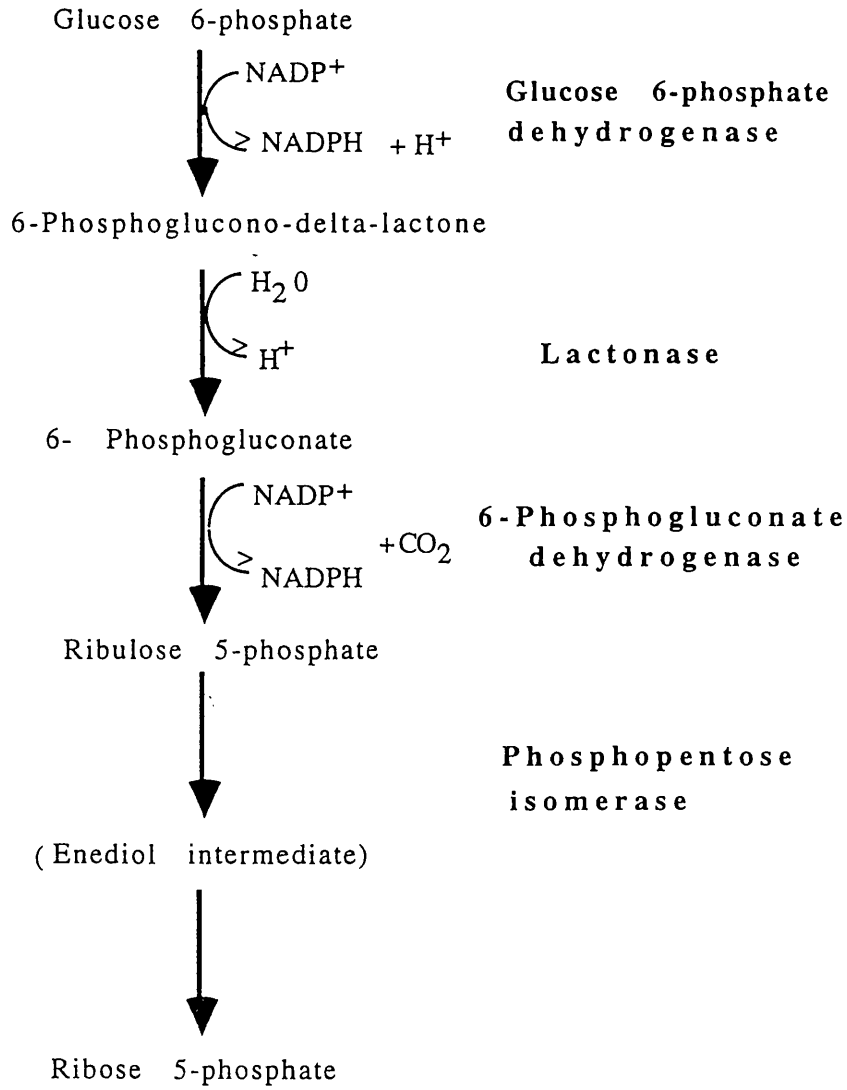


Figure 3. The Oxidative Branch of the Pentose Phosphate Pathway.

The pentose phosphate pathway generates NADPH and ribose 5-phosphate . NADPH is used in reductive biosyntheses and ribose 5-phosphate is used in nucleotide biosynthesis. The enzyme catalysing each reaction is shown in bold at the right.

1.5.2. Isolation of the Human *6-PGD* Gene.

Isolation of the *PGD* gene would allow the development of DNA based polymorphisms suitable for linkage analysis. This was considered desirable in the light of the poor informativeness of the protein polymorphism. This work and the characterization of the PGD RFLP is described in the Ph.D thesis of P. Kleyn, 1990. Below I give a brief summary of the details relevant to the continuation of this work.

A rat partial cDNA clone was used for the initial isolation of cDNA clones from human cDNA libraries. These were denoted clones pPGDH1-pPGDH4 and restriction mapping showed that clones one to three had a common 3' end. pPGDH2 and pPGDH3 were the same size and pPGDH1 had a truncated 5' end in comparison to pPGDH2 and pPGDH3. pPGDH4 was the most 5' clone, which was 800 bp. pPGDH3 was 900 bp. pPGDH3 and pPGDH4 share a natural EcoRI site, so that together they give a continuous cDNA sequence of 1.7 kb. The mRNA detected by the cDNA pPGDH4 is 2.4 kb in size. The human PGD protein subunits are 52 KD requiring approximately 1.4 kb of translated sequence in man. Figure 4 illustrates the relationship between these PGD cDNA clones

Partial sequencing of the clones and computer aided translation of the sequence showed that homology existed between the *E. coli* and sheep PGD polypeptide sequences. This provided proof that these clones were from the human *PGD* locus. This comparison allowed the calculation of how much more 5' coding sequence was still missing. This was approximately 23 amino acids.

Somatic cell hybrid analysis confirmed the chromosomal origin of these cDNAs but also showed that, as with other dehydrogenases, *PGD* had at least 2 pseudogenes. *In situ* hybridization with cDNA clones pPGDH4 and pPGDH1

confirmed that they were from chromosome 1, in the region to which PGD activity had been previously assigned, i.e 1pter-1p36.1, (Carritt, *et al*, 1982a). Evidence from these analyses pointed to chromosome 18q12-q21 as a probable location for one *PGD* pseudogene. The *PGD* locus was shown to span at least 17 kb (Kleyn, 1990).

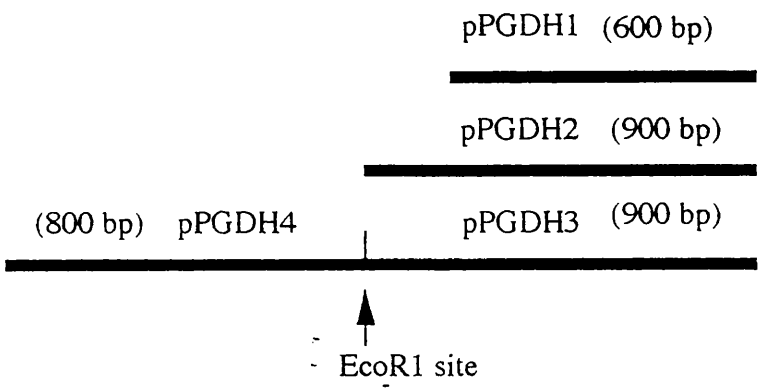


Figure 4: Diagram of PGD cDNA Clones.

The above figure is taken from Kleyn,1990. This illustrates the arrangement of the PGD cDNA clones isolated initially using the rat PGD cDNA as probe. pPGDH1 is the shortest 3' cDNA. pPGDH2 and pPGDH3 were isolated separately but were found to be the same according to their restriction maps. pPGDH3 and pPGDH4 were isolated together as a contiguous 1.7 kb clone and were then separated by EcoR1 digestion and subcloned. The EcoR1 site is indicated above. All clones were subcloned in to pAT153 (Kleyn,1990).

1.5.3. PGD Enzyme and Restriction Fragment Length Polymorphisms.

The original polymorphism at the *PGD* locus was a diallelic protein polymorphism showing Mendelian inheritance (Fildes and Parr, 1963; Parr 1966). This was detected in haemolysates electrophoresed in 0.01M phosphate buffered starch gels, pH 7.0, followed by the addition of the substrate 6-phosphogluconate in the following mix: NADP/PMS/MTT in agar at 37 ^{0}C. Three phenotypes were observed typical of a diallelic system in a dimeric protein. Alleles are referred to as A and B, the latter being the minor allele.

A heterozygosity of 4% was shown for Londoners (Parr, 1966) and a minor allele frequency in the range 0.014-0.039 for Caucasians giving an average polymorphism information content (PIC) of 0.09. In Negroid populations this range is within 0.000-0.139, in the Middle East, 0.028-0.092 and in Asiatics, 0.014-0.231. There is an apparent decrease in the frequency of the minor allele from East to West (Bowman *et al*, 1966; Tashian *et al*, 1967).

It is obvious from these frequencies that this polymorphism is not optimal for genetic mapping studies since in most Caucasian populations very little information would be derived for this locus. This led to the demand for a more informative polymorphism at this locus for genetic mapping studies to place *PGD* on the genetic map of chromosome 1.

A restriction fragment length polymorphism was detected following the isolation of a PGD cDNA clone. (Kleyn, 1990). This is a BamHI polymorphism detected by the 3'cDNA clone pPGDH1. It is a diallelic system with variant bands at 11 kb and 9.5 kb and constant bands at 11 kb and 4.2 kb. The 11 kb fragment is denoted the slow (S) allele, (allele 1)and has a frequency of 0.72, that at 9.5 kb is the fast (F) allele, (allele 2) and has a frequency of 0.28. Thus

the RFLP has a heterozygosity level of 40%, 10 fold greater than that at the protein level.

The protein polymorphism and the RFLP were shown to undergo no recombination in doubly heterozygous pedigrees of European and South African origin. A test of allele association using both phase known and phase unknown pedigrees yielded a chi squared value of 4.12 (1df) for the 2 X 2 table ($p<0.05$). In 17 unrelated phase known individuals the 11 kb (S/1) BamHI allele was in linkage disequilibrium with the minor (PGD*B) allele (kleyn,1990).

Aims of this Study.

The main aims of this study were to refine the genetic and physical maps of distal chromosome 1p. In so doing, future attempts to isolate the genes involved in human disease will be greatly facilitated. The knowledge on order and distance of markers mapped here and the resources generated in the form of YAC and cosmid clones will prove useful in this endeavour. In addition, the generation of both physical and genetic maps allowed such issues as the ratio of genetic to physical distance in the region, and the extent of synteny and linkage conservation between man and mouse to be investigated. Comparative data provided valuable predictive information on the possible location of mutations leading to disease. At a higher level of resolution a further contribution was made to the *PGD* gene sequence and organization.

CHAPTER 2 MATERIALS and METHODS

2.1 DNA Isolation.

2.1.1 Small-Scale Isolation of Plasmid DNA

Ten ml of L-broth containing 50µg/ml of ampicillin or 30µg/ml kanamycin was inoculated with a single colony and grown overnight at 37°C with vigorous shaking. 1.5ml of this culture was then centrifuged in a microfuge, the supernatant aspirated off and pellet suspended in 0.1ml Lysozyme buffer (see Appendix, section 5.2). 0.2ml alkaline SDS was added and left for 5 min. 0.150ml ice cold acid potassium acetate (3M) was added and mixed and after 5 min the tubes were centrifuged for 3 min. The supernatant was then extracted with 0.250ml buffer-saturated phenol (equilibrated with 0.1M Tris.Cl. pH 7.8).Two volumes of 95% ethanol were added to the deproteinized supernatant and left at 20-22°C to precipitate nucleic acids for 5 min. The pellet obtained after spinning for 10 min was dissolved in 50µl 1X TE .20µl of DNA were used in analytical digests in 40µl .Spermidine was included at 5mM in these digests and RNase A at 5µg per reaction. Digestion was allowed to proceed for 90 min only.

2.1.2 Large-Scale Isolation of Plasmid DNA.

Five ml of L-broth containing 50µg/ml of ampicillin or 30µg/ml of kanamycin was inoculated with 100µl of a glycerol stock and incubated in a 100ml conical flask at 37°C with vigorous shaking (200rpm) for 6-8 hrs. 2.5 ml of this daytime culture was then used to inoculate a 250ml aliquot of L-broth with the same antibiotic in a 1L conical flask and was allowed to grow overnight.

The cells were then harvested by centrifugation at 3500 x g for 10 min at 4°C, in an MSE Europa 24M Centrifuge. The supernatant was then poured off and the cell pellet was resuspended in 10ml of lysozyme buffer. The cells were lysed and the DNA simultaneously denatured by the addition of 20ml of alkaline

SDS. The mixture was left on ice for 5 min then mixed gently and left for another 5 min on ice. The DNA was then neutralized by the addition of 15ml of acid potassium acetate. This was left on ice for 30 min after moderate shaking and then centrifuged at 5500 x g at 4°C for 10 min. To the supernatant, 0.6 volumes of isopropanol was added, mixed and left at 20-22°C for 15 min. Centrifugation then followed to collect the DNA at 5500 x g for 15min at 20-22°C. The pellet was then washed with 70% ethanol, dried and resuspended in 11ml TE.(10mM Tris.Cl. pH 8.0; 1mM EDTA) Caesium chloride was added to 1.56g/ml (11.44g), 0.44ml 0.2M potassium phosphate buffer (pH 8.0) and 1.140ml Ethidium Bromide were also added. This was transferred to a Sorvall sealable centrifuge tube then filled with paraffin oil. Spinning then followed at 200 000 x g at 20-22°C for 17hrs in a Sorvall OTB75B ultracentrifuge. The plasmid/cosmid band was viewed with 360 nm UV light, and extracted with a needle and syringe. The ethidium bromide was extracted with caesium chloride saturated isoamyl alcohol. Dialysis then followed against 6L 10mM EDTA to reduce the level of caesium chloride. The DNA was precipitated with 0.1 volumes 4M sodium chloride and 2.5 volumes ethanol at -20°C/overnight. A second precipitation with 0.5 volumes 7.5M ammonium acetate and 2.5 volumes ethanol followed. Finally, the pellet was dissolved in 0.5ml TE and kept at -20°C. The concentration was determined by measurement of absorbance at 260nm.

2.1.3 Isolation of Genomic DNA for Southern Analysis and PCR Amplification.

2.1.3.1 From Peripheral White Blood Cells.

This method was developed by John *et al* ,1990. 5ml of blood were mixed with 5ml of a solution containing 10mM Tris HCl pH 7.6; 10mM potassium chloride; 10mM magnesium chloride. To this, 120µl Nonidet P40 (BDH) were

added in order to lyze the cells. The nuclear pellet was collected by centrifugation at 500 x g for 10 min. The supernatant was poured off and the nuclear membranes lyzed by the addition of 0.8ml of 10mM Tris HCl pH 7.6; 10mM potassium chloride; 10mM magnesium chloride; 0.5M sodium chloride; 0.5% SDS; 2mM EDTA. This was extracted once with 0.4ml buffer-saturated phenol, once with 0.2ml phenol; 0.2ml chloroform:isoamylalcohol (24:1) and finally once with 0.7ml of this mix The DNA was precipitated by mixing the upper aqueous phase with two volumes of ice cold ethanol. The tip of a glass pasteur pipette was sealed by heating briefly and used to transfer the DNA fibres to 1ml 70% ethanol for washing. Spinning for 5 min in a microfuge followed to collect the DNA which was dried for 15 min in a freeze drier and resuspended in 200µl sterile water at 65°C overnight.

2.1.3.2 From Lymphoblastoid Lines.

Approximately 10^7cells were pelleted, washed twice in 0.9% saline and the pellet suspended in 10ml STE (0.1M sodium chloride; 10mM Tris.HCl pH 8.0; 1mM EDTA), 200µg/ml proteinase K and SDS added to 0.5% in the solution. This was incubated at 37°C overnight and then extracted twice with phenol and once with chloroform:isoamylalcohol (24:1). Precipitation followed with two volumes of ethanol, the precipitate was hooked out into 1ml TE and the DNA was left to dissolve at 4°C for 48 hrs. Then 0.5ml of 7.5M ammonium acetate and 3ml ethanol were added, mixed and the DNA recovered as before into 200µl TE.

2.1.3.3 From Hamster and Human-Hamster Somatic Cell Lines.

The medium was aspirated from flasks and the confluent cells detached from the surface by firstly washing with 10ml of versene (5 mM EDTA in HANK's balanced salt solution) to inhibit antitrypsin activity and then after aspiration of

versene, 2.5ml of TV (versene plus 0.025% trypsin) were added. The TV was spread over the cells by agitation and left over the cells at 37°C for 5 mins. 7.5ml medium (as appropriate to the cell type) was added and the cells were resuspended. They were then collected by centrifugation at 300 x g/3 min and washed in 0.9% saline as above. All other steps were as already described above for DNA isolation from lymphoblastoid lines.

2.1.4 Isolation of High Molecular Weight Genomic DNA for Pulsed Field Gel Electrophoresis.

2.1.4.1 From Peripheral White Blood Cells

Ten ml of blood were mixed with 35ml of red blood cell lysis solution (155mM ammonium chloride ; 10mM potassium carbonate; 1mM EDTA) and left on ice for 1hr. The white blood cells were collected by centrifugation at 300 x g for 10 min at 4°C and resuspended in 25ml lysis buffer for washing. Two more washes followed in 0.9% saline before resuspension in 10ml 0.9% saline. The cells were counted using a haemocytometer and after centrifugation were suspended in 1% low melting point agarose in 0.9% saline to give a final concentration of $3X10^7$ cells/ml. 100µl aliquots of this suspension were pipetted into moulds and placed on ice for 30 min to promote rapid setting and an equal distribution of cells within the blocks. The white blood cells were then lysed and proteins inactivated by incubation in ESP solution (0.5M EDTA; 1% sodium lauroyl sarcosinate; 2mg/ml proteinase K) at 50°C for 48 hrs. 1ml of ESP solution was used for each block. Washing in TE followed at 20-22°C (4X10 min) and then inactivation of proteinase K by incubation in 50ml TE containing 0.04mg/ml PMSF (phenylmethylsulphonylfluoride) at 20-22°C Lastly, the blocks were washed in 50ml TE for 30 min and then stored in 50ml 0.5M EDTA at 4°C.

The quality of the blocks was then assessed by running 3 X 50µl blocks from each preparation on a pulsed field gel. One was undigested, a second incubated with restriction enzyme buffer at 37°C for 3 hrs, a third as in the latter case but with an infrequently cutting restriction enzyme. Collectively these three tests show that there is no degradation during preparation, complete inactivation of nucleases and proteinase K.

2.1.4.2 From Cell Lines.

The Epstein Barr virus transformed cell line MGL8B2 was used as the source of human genomic DNA. These cells grew in suspension medium (RPMI 1640 plus 10% fetal calf serum), they were washed in 0.9% saline to remove the medium three times and collected by centrifugation at 300 x g for 3 min. All steps after this were as for peripheral white blood cells from counting the cells using a haemocytometer. Hamster cells and hamster-human somatic cell hybrid cells were harvested by trypsinization and subsequent steps were as above for DNA block production.

2.1.5 Isolation of Yeast DNA

2.1.5.1 Yeast DNA Isolation for PCR Amplification.

A single colony was inoculated into 50ml of S.D medium (synthetic dextrose minimal medium) in a 250ml flask and grown overnight at 30 °C with agitation (200rpm). The next day, the cells were harvested by centrifugation at 1000 x g for 5 mins. The cell pellet was resuspended in 2.5ml YRB (Yeast Resuspension Buffer) + 14mM beta-mercaptoethanol. 100 units of Lyticase (Sigma) were added to the cells which were microscopically monitored for spheroplast formation over 1hr at 37°C. These were then collected by spinning at 400 x g for 5 min and suspended in 5ml of 50mM Tris Cl pH 7.4; 20mM EDTA. The

membranes were then lysed by incubation with 0.5ml of 10% SDS at 65°C for 30 min. 1.5ml of 5M potassium acetate were then added and incubated on ice for 1hr then the precipitate collected by spinning at 15000 x g for 10 min. To the supernatant, two volumes of absolute ethanol was added, mixed, left at 20-22°C for 10 min and centrifuged at 4000 x g for 15 min. The pellet was dissolved in 3ml TE and then centrifuged at 15000 x g for 15 min and the supernatant transferred to another tube. A standard phenol chloroform extraction was performed once, followed by a chloroform extraction. An equal volume of isopropanol was added mixed gently and centrifuged at 4000 x g for 10 min. The pellet was dried and resuspended in 0.5ml TE. 150µg of RNAse were then added and incubated at 37°C for 30 min. 300µg of proteinase K were then added under the same conditions. Two phenol/chloroform steps were then carried out and one chloroform. 1/10th volume of 3M sodium acetate pH 6.0 was added along with two volumes of absolute ethanol, incubation at -20°C for 30 min and then spinning at 15000 x g for 5min. The DNA pellet was washed in 70% ethanol, dried and suspended in 200µl TE. The concentration was approximately 0.5mg/ml.

2. 1.5.2 Isolation of Yeast DNA for PFGE.

Yeast cells were spheroplasted as in section 2.1.5.1 and then for every original 20 ml culture 0.5 ml 1% LMP agarose in YRB was added and mixed gently. 100µl were then dispensed into moulds and left to set at 4°C for 30 min. The plugs were then transferred into 5ml of yeast lysis buffer (YLB) and left at 20-22°C for 1 hr. This solution was then replaced with a fresh 10 ml aliquot and the plugs were incubated at 50°C overnight. The following day the plugs were washed in 15 ml TE and then stored in 20 ml YLB at 20-22°C

2.2. DNA Digestion and Agarose Gel Electrophoresis.

All DNA digests were performed in BRL or NBL buffers recommended by the suppliers with added spermidine at 4mM and BSA at 1μg/10μl of buffer. Digestion was at 37°C for 3-6 hrs.

2.2.1 Analytical Gels.

500ng of cloned DNA was digested in 20μl volumes with 20 units of restriction enzyme. For genomic DNA 5μg were digested in a 40μl volume and for hybrid DNA 10μg were used per digest. 40 units of enzyme were used in each case.

2.2.2 Preparative Gels.

Inserts were prepared by digestion of the appropriate amount of recombinant clone to release 1μg of insert. The band was then excised from the gel weighed and 3X its weight was added in volume of water. Boiling of the probe followed for 7 min and 30μl were aliqouted into eppendorfs. Storage was at -20°C.

All agarose gels were 0.8-1.4% in 100-150ml 1X TBE buffer with 0.5μg/ml ethidium bromide. Preparative gels were made from Low Melting Point Agarose. The running buffer was also TBE. Gels were run overnight at 2v/cm.

2.2.3 Pulsed-Field Gel Electrophoresis.

PFGE digests were carried out in 150μl volumes of restriction enzyme buffers. DNA plugs were washed three times in sterile deionized water for 15 min to wash out the EDTA, and a 50μl segment was added to the buffer with 40 units of enzyme and left on ice for 30min to allow diffusion of enzyme into the plugs. Incubation at 37°C or 50°C then followed for 3-4 hrs. For double digests with

enzymes with different requirements the plug was washed briefly in TE after digestion with one enzyme and then added to the next desired buffer and enzyme for another 3-4hrs.

The PFGE apparatus used was an LKB-Pharmacia Pulsafor unit with the hexagonal electrode array. The buffer was cooled to 10°C before loading and maintained at this temperature throughout the run. Table 3 lists the various separation programmes used.

Gels were made with 0.5XTBE and were 150ml in volume. Percentage of agarose was between 0.9-1.2 and a specific running programme was used for separation of fragments in the desired size range. The plugs were loaded and sealed into the gel with 1% LMP agarose. Size markers used included Lambda 50kb ladders (Pharmacia) and *S.cerevisiae* chromosomes (Pharmacia). Staining with ethidium bromide (0.5μg/ml) in 0.5XTBE then followed for 30min. Destaining with the used running buffer proceeded for 1hr and then a photograph of the gel was taken.

Table 3. Parameters for Pulsaphor Separations.

A) Basic Programmes. B) Ramped Programmes.

	Size Range (kb)	Voltage (V)	Pulse Time (sec)	Run Time (hours)	Agarose %
A	50-1600	170	90	40	1.2
A	50- 800	170	50	40	1.2
B	50- 700	200	15 to 45	11	1.0
			45 " 15	11	
B	100-900	200	20 " 90	12	1.0
			90 " 20	12	
B	100-1600	120	20 " 120	18	1.0
			120	30	
			120 to 400	18	

2.3 Southern Blotting .

All gels were denatured in 1.5M sodium chloride/0.5Msodium hydroxide for 1hr and blotted in the same solution onto Genescreen Plus Membranes. The latter were neutralized in 0.2M Tris-HCl/2X SSC, pH8.0 for 10 min after transfer.

Cloned DNA was blotted onto three filters using a longer time for each one i.e 20min, 40min and 60min for the last one. Genomic DNA was transferred overnight and PFGE DNA was transferred for 48 hrs. The DNA was UV crosslinked by a 2 min exposure to 260 nm UV light.

2.4 Radiolabelling of Probe DNA.

DNA inserts to be labelled with ^{32}P-dCTP were firstly denatured by boiling for 1 min if in low melting point agarose (30µl volume) or 5 min if in solution. An Amersham Multi-Prime labelling Kit was used here as recommended. Labelling was allowed to proceed for 3-24hrs and then unincorporated labelled nucleotides were separated from the probe after addition of 50µl TE and spinning through a G 50 Sephadex column for 3 mins at 300 x g.

2.5 Filter Hybridization, Washing and Stripping.

2.5.1 Genescreen Plus Membranes.

Membranes were prehybridized in plastic bags containing 10ml of prehybridization solution for each membrane. This solution was 1M sodium chloride/2% SDS/10% Dextran Sulphate. Prehybridization was carried out at 65 ^{o}C for 2-3 hrs. The labelled DNA probe was denatured with 250µl (1.250 mg) sonicated salmon sperm DNA by boiling for 5 min, cooled on ice (to prevent

renaturation) and added to the prehybridization solution. Hybridization was allowed to proceed overnight in a 65°C waterbath with constant shaking.

Washing at low stringency at 20-22°C in 2X SSC was performed first for 30 min. Then washing at 65 °C followed in 2X SSC/1% SDS, (2X30min).

High stringency washes if necessary were at 65 °C in 0.1X SSC, (1X 15-30min) and finally 0.1XSSC/1%SDS, 65 °C, (1X15-30min). The filters were wrapped with Saran Wrap and placed in cassettes for autoradiography at -70 °C.

2.5.2 Nitrocellulose Membranes .

Nitrocellulose filters (Schleicher & Schuell) were prehybridized in 5X Denhart's/5XSSC for 6 hrs. The labelled DNA probe was added and hybridization allowed to proceed overnight. in a 65°C waterbath with constant shaking.

Washing at 20-22°C in 2XSSC, (2x 30min) followed. Then washes at 65 °C in 2XSSC/0.1% SDS, (2X30min) were carried out. The filters were air dried and placed in cassettes for autoradiography at -70 °C.

2.5.3 Hybond N+ Membranes.

Hybond N+ membranes were used for the transfer of YAC DNA at the Resource Centre, Northwick Park. These were high density gridded YAC filters in 3x3, nine microtitre plate arrays. The prehybridization solution contained the following: 75 ml of 20X SSC; 90 ml of distilled water; 5 ml of 0.5M EDTA; 12.5 ml of 100X Denhardts solution; 12.5 ml of 10% SDS; 5 ml of 10% sodium pyrophosphate. 5 ml of salmon sperm DNA (5mg/ml) was denatured and added to 160 ml of the above solution along with 35 ml of distilled water. This

was then added to the 40 YAC filters which were previously soaked in 20X SSC. Prehybridization followed for 3 hrs at 65 ^{0}C. The probe was denatured, quenched on ice then added to the filters. Hybridization was allowed to proceed overnight at 65^{o}C whilst shaking.

Washing followed in 2X SSC at 65 ^{o}C (2X 5 min). Then more stringent washes followed in 0.2X SSC/ 0.1% SDS at 65 ^{0}C (3X 15 min). The filters were then wrapped in Saran Wrap and placed in cassettes for autoradiography at -70 ^{o}C overnight. Overnight exposure allowed the detection of positive clones. Up to 10 days exposure was necessary to allow backround signal to develop as this was important in accurate alignment of the autoradiogram with the colonies on the filters.

2.5.4 Stripping of Probe from Genescreen Plus Membranes.

The stripping solution was 1X TE/1% SDS. This was heated to 85 ^{o}C poured over the filters and incubated at 65 ^{o}C for 30min. Then filters were rinsed in distilled water, wrapped with Saran Wrap and autoradiographed overnight to ensure complete removal of probe before a subsequent hybridization.

2.6 Cosmid Library Screening .

One total genomic cosmid library was screened for all the clones used in this study (vector was Lorist B, Kanamycin resistant). Up to two million colonies were plated out onto two 20cmX20cm megaplates containing 200ml L-agar plus kanamycin (30μg/ml) and a nitrocellulose filter each. Colonies were grown at 37 ^{o}C overnight, these were the master filters. Replicas of these were then made for hybridization. Firstly two 200ml L-agar plates were poured and nitrocellulose filters moistened by brief contact with the agar. The master filter was then placed colonies uppermost, on 3MM Whatman filter paper and the

upper side (colony side) of the new filter placed over this and pressed firmly down with a second piece of filter paper. Registration marks were then made using a sterile needle and the filters separated and placed on their respective plates. The master plates were kept at 4 ^{o}C whilst the replicas were incubated at 37 ^{o}C for 6-8hrs or left at 20-22^{o}C overnight.

The replicas were prepared for hybridization as follows: cells were lysed by placing the bottom of the filter on 3MM filter paper saturated in 10% SDS for 3min. Denaturation followed on a filter saturated with 1.5M sodium chloride/0.5 sodium hydroxide for 5min. Neutralization in a tray of 3M sodium chloride/0.5M Tris.HCl,pH7.5 for 5min preceeded washes in 2XSSC/0.1%SDS, 3min and 2XSSC for 3min. During the first of these washes the cell debris was wiped off with Kimwipes. The filters were air dried and baked at 80 ^{o}C for 2hrs to bind the DNA to the filters.

Hybridization and autoradiography localized the positive colonies which were then picked from the master plates. Dilutions were made in L-broth containing kanamycin and plated out onto nine cm petridishes for secondary purification. Subsequent steps were as already described. The aim was to isolate a single positive colony which would then be used for DNA isolation.

2.7 Fluorescence *In -Situ* Hybridization.

2.7.1 Biotin and Digoxigenin Nick Translation .

1µg of DNA (cosmid or plasmid) was biotinylated and 200ng of this was used for one metaphase spread. The BioNick Labelling System (Gibco BRL) was used as described by the manufacturer's protocol. Briefly, 5µl of 10X dNTP mix, 1µg DNA, sterile water to 45µl and 5µl 10X enzyme mix (DNAse 1 and DNA polymerase 1) were combined on ice in a microfuge tube. Labelling was

allowed to proceed at 16 ^{o}C for 1hr. The process essentially involves single strand nicking by DNAse 1 followed by degradation and resynthesis of new material by DNA polymerase 1 which in the process introduces labelled nucleotides into the probe.

Probes labelled with digoxigenin were used in dual colour FISH and here 100 ng of each of two probes, one biotinylated and one labelled with digoxigenin were combined after labelling, for hybridization to chromosome spreads.

Labelling with digoxigenin was carried out as follows: 2 µl of 10X salts solution (0.5 M Tris.HCl pH 7.8; 0.05 M magnesium chloride; 0.01 M beta-mercaptoethanol; 50µg/ml BSA) was mixed with 2.5 µl of each 0.5 mM stock of dATP, dCTP, dGTP, (Pharmacia) and Digoxigenin-11-dUTP (Boehringer). 1µg of DNA was then added, followed by 1µl DNAse I (0.1µg/ml) and 1µl DNA polymerase I. After centrifugating to mix, the reaction mixture was incubated at 4^{o}C overnight.

After labelling by either hapten the probes were treated as follows: 5µl of stop buffer were added and unincorporated nucleotides were separated from labelled DNA probe by column chromatography. 25µl TNE (0.2 M sodium chloride; 0.01 M Tris.HCl pH 8.0; 1mM EDTA) was added to the probe which was then loaded onto a Nick column (Pharmacia) developed in TNE. The probe was eluted with 2X400µl TNE, the second fraction was collected. The probe was then concentrated by precipitation with 100µg sonicated salmon sperm DNA (10mg/ml), 50µl 3M ammonium acetate and ice cold absolute ethanol to 1.5ml. This was placed at -20 ^{o}C for 1hr then centrifuged in a microfuge for 5 mins, the supernatant decanted and pellet dried in a freeze drier for 15 min. 100µl TE was added to dissolve the pellet thus giving a probe concentration of 180ng/20µl (the % recovery expected was 90%) 1µl of this was used to assay for biotin or digoxigenin incorporation.

2.7.2 Test for Biotin and Digoxigenin Incorporation.

1µl of probe was spotted onto a small piece of nitrocellulose filter which was baked at 80 ^{o}C for 2hrs. To test for biotin incorporation the filter was washed in 50ml TTBS (0.1M Tris.HCl pH 7.5; 0.15 M sodium chloride; 0.1% Tween 20), (3X10 min) and then immersed in 5ml of a solution prepared 30 mins earlier. This was Vectastain ABC-HRP (streptavidin-horse radish peroxidase conjugate) It was essential to prepare this solution 30 min earlier so that complex formation would take place between streptavidin and the biotinylated horse radish peroxidase.The nitrocellulose filter was immersed in this solution for 30 mins at 20-22^{o}C with agitation, (Vectastain ABC Elite Kit, Vector Labs). This step allowed the attachment of the conjugate to the biotinylated probe. Washing in TTBS followed as above. The substrate solution (DAB Peroxidase Substrate Kit ;Vector Labs) was prepared immediately before use according to the manufacturer's instructions. Briefly, DAB (diaminobenzidine) stock solution was diluted in the buffer supplied and one drop (100µl) of the Hydrogen Peroxide Solution was added and mixed then one drop (100 µl) of Nickel Solution was added and mixed. This was poured over the filter and allowed to develop in to a brown colour for 20 mins. The brown colour is the result of oxidation of the diaminobenzidine. Finally washing in water for 5min followed to remove the substrate. Probes judged to be satisfactorily labelled using this test were processed for hybridization (section 2.7.5).

To test for digoxigenin incorporation, the filter was washed in TNB (0.1M Tris.HCl pH7.5; 0.15M sodium chloride; 0.05% Boehringer blocking reagent) for 30 mins at 20-22^{o}C and then incubated in mouse anti-digoxigenin antibody peroxidase conjugate diluted in TNB (1:5000) for 30 mins. Washes followed in TNB three times for 10 mins each with agitation. Detection of peroxidase activity was achieved as above using the DAB peroxidase substrate kit.

2.7.3 Slide Preparation

Slides used for all FISH experiments were kindly provided by Dr. M. Fox. These were prepared as follows : 0.5 ml of a lymphocyte suspension (approximately 10^6 cells) from a normal male was centrifuged at 300 x g for 5min. The supernatant was decanted and the pellet resuspended in 1ml fix , (methanol:glacial acetic acid,3:1). The suspension was centrifuged again at 300 x g for 5 min and resuspended in 0.5 ml fix. Using a sterile pasteur pipette two drops/200µl of suspension (approximately 10^5 cells) were dropped onto a slide slanted at an angle such that the cell suspension would spread over as large a surface as possible. After rapid air drying (by waving the slide in the air) the metaphase spread was washed gently with 1 ml 30% aqueous acetic acid. After drying again the slides were viewed under the microscope to ensure that cells had burst and the chromosomes had spread out so that signals could easily be detected on individual chromosomes after hybridization.

2.7.4 Prehybridization.

Prehybridization of slides and competition of probe with Cot-1 DNA (human placental DNA enriched for repetitive DNA sequences by shearing, denaturing and renaturing to Cot = 1; BRL), were carried out simultaneously so that both were ready for hybridization at the same time. Slides were first dehydrated in an alcohol series , 70%, 90%, 100% for 3 min each and air dried. Each slide was then treated with 200µl RNAse (100µg/ml; 2X SSC; 37 ^oC; 1 hr) under a cover slip , in a moist chamber. This was then followed by four washes in 2X SSC in a Coplin jar. A second dehydration step followed in the above alcohol series and then the slides were air dried. Proteinase K buffer (20mM Tris HCl pH 7.4; 2mM calcium chloride) was prewarmed to 37 ^oC and the slides incubated in this for 10 min .They were then transferred to a proteinase K

solution (35µl of 50µg/ml stock) also prewarmed to 37 °C. A wash followed in PBS (Phosphate Buffered Saline, Sigma tablets pH 7) for 5 min. Slides were postfixed in 0.1M magnesium chloride prepared in 50 ml PBS with 1.3ml 37% paraformaldehyde for 10 min. Excess fixing solution was washed off in PBS for 5 min on the rocker. An alcohol series followed as earlier and air drying. Slides were then denatured in 100µl 70% formamide; 2X SSC under a coverslip for 5 min at 75 °C . The coverslip was removed and the slide immediately placed in a Coplin jar of ice cold 70% ethanol for 3 min. Lastly the slides were passed through 90% and 100% ethanol for 3 min each and then air dried.

2.7.5 Probe Preparation Using Competition with Cot-1 DNA and Hybridization.

200ng (20µl of a 10ng/µl solution) of probe DNA was combined with 10µg Cot-1 DNA, 2µl 3M ammonium acetate and 2.5 volumes absolute ethanol (80µl) and precipitated at -70 °C for at least 1 hr. After this the DNA was centrifuged for 5 min , the supernatant decanted and the pellet freeze dried for 15 min. Resuspension followed in 20µl of hybridization mix (50% formamide ;10% dextran sulphate ; 2X SSPE (0.6M sodium chloride 0.02M sodium dihydrogen phosphate; 2mM EDTA pH 7.4) and denaturation at 75 °C ; 5 min. Preannealing of the repetitive component of the probe was allowed to proceed for 30 min at 37°C. Lastly , the probe was placed on the prehybridized slide , covered with a clean coverslip and the edges sealed with Cow gum. Hybridization proceeded overnight in a moist chamber at 37 °C.

2.7.6 Probe Preparation and Hybridization without Competition .

30 ng of probe DNA were used per slide for single copy probes. This was the D1Z2 probe P1-79. The probe was combined with 20 µl of hybridization mix

and placed on a slide under a coverslip. This was sealed with Cow gum and then both probe and chromosomal DNA was denatured at 80 ^{o}C for 4min. Hybridization was allowed to proceeed overnight at 37 ^{o}C.

2.7.7 Post Hybridization Washes.

Coverslips were removed from slides and three washes were carried out in 50% formamide ; 2X SSC at 42 ^{o}C for 5 min each. Then slides were washed five times in 2X SSC at 42 ^{o}C for 2 min each wash. Signal detection then proceeded.

2.7.8 Signal Detection

2.7.8.1 Signal Detection-Single Colour FISH.

For single colour FISH using biotinylated probes the following steps were carried out. Slides were washed in 4X SSC ; 0.05% Tween 20 (Tween 20-Polyoxyethylene sorbitan monolaurate) for 5 min on a rocker. Preincubation in 4X SSC ; 5% Marvel (non-fat milk powder) followed for 20 min at 20-22^{o}C. The above blocking solution was then used to dilute stock solutions of avidin-FITC conjugate (Fluorescein isothiocyanate) and biotinylated anti-avidin (made in goat) to 5µg/ml. 100µl of the former was added to each slide under a coverslip and left for 20 min at 20-22^{o}C with protection from sunlight for all further steps. Slides were then washed in 4X SSC ; Tween 20 three times on a rocker for 5 min each. 100µl of biotinylated anti -avidin ; 4X SSC ; 5% Marvel were added to each slide under a coverslip and incubated as above for 20 min. Three washes in 4X SSC ; Tween 20 followed for 5 min each. 100µl of avidin-

FITC ; 4X SSC ; 5% Marvel were then added to the slides under a coverslip and incubated at 20-22°C as above for 20 min. One wash in 4X SSC ; Tween 20 and two in PBS were carried out for 5 min each on a rocker. Excess PBS was drained from slides and then these were mounted in 15µl of anti-fade medium (p-phenylenediamine dichloride 10mg/ml PBS) containing DAPI (4,6, diamidinophenyl indole,1µg/ml) and propidium iodide (1µg/ml) counterstains in preparation for viewing under the Confocal microscope (Nikon).

2.7.8.2 Signal Detection - Dual Colour FISH.

In dual colour FISH signal detection begins as above (section 2.7.8.1) up to the first wash in 4X SSC ; Tween 20 for 5 min. After this the slides were washed twice in TNT (10 mM Tris. HCl pH8.0; 150 mM sodium chloride; 0.05% Tween 20) for 5 min. Goat anti avidin (initial concentration-0.5mg/ml) was diluted 1:100 in TNB (100 mM Tris. HCl pH8.0; 150 mM sodium chloride; 0.05% blocking agent) and then added along with mouse anti-digoxigenin (monoclonal antibody from mouse ascites, Sigma) diluted 1:1000 in TNB to the slides under a coverslip. Incubation followed in a moist chamber for 30 min at 37 °C. Three washes followed in TNT for 5 min each. Then avidin- FITC diluted 1:100 in TNB, and TRITC (Tetramethyl rhodamine isothiocyanate) conjugated rabbit anti-mouse antibody diluted 1:1000 in TNB were added to the slides. Incubation at 37 °C in a moist chamber followed for 30 min. Washes followed three times in TNT for 5 min each. TRITC conjugated goat anti-rabbit antibody diluted 1:1000 in TNB was then added and incubation followed for 30 min at 37 °C in a moist chamber. Washing was in TNT once for 5 min and twice in PBS for 5 min. Lastly slides were drained and mounted in antifade medium as above containing only DAPI counterstain as propidium iodide drowns the red signal. The steps for dual colour detection and amplification are summarized as follows:

Step 1: (For **biotinylated** probe) add **avidin-FITC.**

Step2: (For **biotinylated** probe) add **anti-avidin.**
(For digoxigenin labelled probe) add mouse-anti-digoxigenin antibody.

Step3: (For **biotinylated** probe) add **avidin-FITC.**
(For digoxigenin labelled probe) add TRITC-rabbit anti-mouse antibody.

Step 4: (For digoxigenin labelled probe) add TRITC-goat anti-rabbit antibody.

2.8. DNA Cloning.

2.8.1 Isolation of DNA Fragments for Cloning.

1μg of fragment to be subcloned was separated from vector and other insert DNA by restriction enzyme digestion and electrophoresis and subsequently isolated from gels by placing a small strip of DEAE (Diethylaminoethyl) cellulose membrane (Schleicher & Schull) in a parallel incision in front of the band and running the fragment onto this paper. Elution of the DNA from the paper was achieved by incubation in 400μl 4M sodium chloride at 70 °C for 30 min. This solution was then dialyzed against three litres of TE to reduce the salt concentration and then precipitated with 0.1 volumes 4M sodium chloride and 2.5 volumes absolute ethanol at -70 °C for three hours. The pellet was resuspended in 20 μl TE and the concentration of the DNA estimated by comparison with standards (50ng-1μg).

2.8.2 Preparation of Vector DNA.

The vector used for subcloning of the PGD CpG island was pUC18. 5μg of this were linearized with BamH1, the enzyme inactivated by incubation at 65 °C for 10 min and then precipitated at -70 °C for 1hr with 0.1 volumes 4M sodium chloride and 2.5 volumes absolute ethanol. The pellet was then resuspended in

17µl of sterile water and 2µl of calf intestine phosphatase buffer (0.5M Tris HCl pH8 ; 10mM magnesium chloride ; 1mM zinc chloride ; 10mM spermidine) was added followed by 1µl of calf intestine phosphatase. This was incubated at 37 °C for 30 min and then a standard phenol/chloroform extraction was carried out after the volume was adjusted to 100µl by adding 80 µl of TE. A precipitation followed at -70 °C for 1 hr in 0.1 volumes of 4M sodium chloride and 2.5 volumes ethanol. The DNA was recovered by centrifugation in a microfuge for 10 min and dried in a vacuum desiccator . It was then dissolved in 40 µl TE and its concentration determined using a spectrophotometer.

2.8.3 DNA Ligation.

Ligation was carried out in 20 µl volumes of BRL ligation buffer (0.05M Tris.HCl pH 7.4; 0.01M magnesium chloride; 0.01M dithiothreitol; 1mM spermidine; 1mM ATP; 0.1mg/ml BSA). Vector and insert DNA were added at a 1:8 - 1:4 molar ratio .1µl, (9units) of T4 DNA ligase was added and the reaction mix incubated at 15 °C overnight. As a control, a ligation was also set up with vector DNA only.

2.8.4 Preparation of Competent Cells.

A 10ml overnight culture of the *E.coli* strain JM101, (genotype:SupEthi (lac-proAB) F' (traD prAB+ lacIq lacZ M15) was grown at 37°C in L-broth. 100ml of L-broth was prewarmed to 37 °C and inoculated with 1ml of the overnight culture. This was incubated at 37 °C with vigorous shaking to mid-log phase (OD600 of 0.3-0.5). The culture was then quickly chilled to arrest further growth for 10min and centrifuged in a sterile bottle at 3500 x g for 10min at 4 °C. The supernatant was discarded and the pellet resuspended in 50ml of sterile cold 50mM calcium chloride, 10mM Tris.HCl pH 8. This was left on ice for 30 min and then centrifuged at 3500 x g for 15 min at 4 °C. The pellet was

resuspended in 6.7ml of the above solution and then stored at 4 °C for 24 hrs before use for development of maximum competence

2.8.5 Transformation of Competent Cells.

50ng of ligated DNA in a 20µl volume were used to transform a 200µl aliquot of competent cells. After mixing the cells were incubated on ice for 45 min then heat shocked at 42 °C for 2 min to facilitate the uptake of DNAand finally left on ice for another 5 min. Two 2ml cultures (prewarmed) were inoculated with 20µl and 200µl of the transformed cells each and left at 37 °C for 1hr to allow expression of ampicillin resistance. The mixtures were plated out on Millipore filters and placed on L-agar containing ampicillin (50µg/ml) which would allow only transformants to grow. Recombinant identification was achieved by inclusion of Xgal (0.04mg/ml) and IPTG (0.05mg/ml) in the medium. Here the cloning of an insert in the plasmid results in insertional inactivation of the *β-galactosidase* gene which upon activation by the inducer IPTG results in no catabolism of the chromogenic substrate Xgal and therefore recombinant colonies appear white. Transformation with religated plasmid due to inefficient phosphatasing results in generation of blue colonies.

2.9. Polymerase Chain Reaction.

2.9.1 Oligonucleotide Synthesis.

Primers for PCR were synthesized on a PCR-MATE (Applied Biosystems). The primer was released from the column over a three hour period of flushing 2 ml of concentrated ammonium hydroxide solution through the column using two 10 ml syringes attached to each side. The ammonium solution was passed through the column every 20-30 minutes. After this elution the primer was transferred to a sterile bijou and the column flushed with a final 1 ml of ammonium hydroxide

solution. This 3ml eluate was incubated at 55 ^{o}C overnight and before opening the bottle it was cooled at 4 ^{o}C for 30 min. The solution was then evaporated by centrifugation under vacuum and the dried primer pellet resuspended in 0.5 ml TE. Purification followed through a 3 ml G50 sephadex chromatography column by first equilibrating the column with 15 ml TE then adding the primer to this and collecting the first fraction. The primer was collected in a second fraction by adding 1 ml TE to the column and collecting this in a separate tube. A third 1 ml fraction was also collected. The concentration of primer was determined by taking one OD_{260} as 20 μg of single stranded DNA. For conversion to molarity the sequence of the oligonucleotide was used to calculate the molecular weight.

2.9.2 Polymerase Chain Reaction Protocol.

Polymerase chain reactions were performed in 25, 50 or 100μl volumes in 0.5ml eppendorf tubes in 1X Taq polymerase buffer (Promega/NBL/Advanced Biotechnology). Reactions requiring >1.5mM magnesium chloride had appropriate amounts added from a 1M stock. Nucleotides purchased from Pharmacia (100mM stock of each) were diluted to 10X stocks, 2mM each in sterile water. 50 pmoles of each primer was used in each reaction and 0.2μl (1 unit) of Taq DNA polymerase. This was added to the reaction after the initial 5 min denaturation step. Evaporation within tubes was prevented by addition of an equal volume of sterile parraffin oil.

DNA samples were added last to the reactions. In amplification of human genomic DNA 0.1-1μg DNA was used. For hybrid DNA 1-2μg were used and 1μg of hamster or mouse DNA. For amplification from bacterial or yeast cultures, 1μl of a cell suspension was added to the reaction and the cells lysed and DNA denatured by an extended denaturation step at 95 ^{o}C (10 min). PCR screening of the a total genomic YAC Library was initiated by screening the 40

master pools in agarose plugs. These were washed for 1hr at 20-22°C in sterile water to remove detergent and EDTA and then melted at 65 °C for 30 min before a 1µl aliquot was transferred to the PCR mix. Again the initial denaturation step was carried out at 95 °C for 10 mins. A negative control was included in all PCRs to ensure that positive results were genuine.

Screening of phage cDNA library lysates was achieved by the addition of 1µl of lysate to 77µl sterile water and freezing this at -70 °C for 30 mins. After this the tubes were rapidly thawed by immediate immersion in a beaker of water at 100 °C and then continued boiling for a further 3 mins. After cooling, nucleotides, buffer and primers were added and the reaction started with addition of 1 unit of Taq polymerase.

Reamplification of PCR products with the same primers or internal primers was achieved by excising the desired band from a gel during visualization with 360nm UV light and subsequent boiling in 1ml TE. A range of volumes was then used for PCRs to see which gave optimal amplification. The second approach here was 'Band-stab' PCR. This was done by firstly preparing the PCR mix and then whilst visualizing the band of interest with long wave UV light to stab this with a sterile hypodermic needle once and then plunged this with gentle agitation into the reaction mix.

A Hybaid Thermal Cycler was used for all reactions. Finally a 10µl aliqout of PCR product was analysed on a 1.5-2% agarose gel. A phi X 174 (digested with Hae III) size marker (Gibco BRL) or 1kb ladder (Boehringer) was used to size the products.

2.9.3. PCR Screening of a YAC Library for 1p Clones.

A total genomic YAC library was screened by PCR as described above in section 2.9.2 for 1p clones. This library was constructed by Anand and others (1990). This was composed of 40 master pools each one containing the contents of nine microtitre (96 well) plates which are known as the subpools. Each master pool thus contains 864 clones. Once a positive master pool was identified the nine subpools were screened to identify the positive plate. From the positive plate it was then possible to locate the positive clone by screening in a row by row and column by column basis. The point in the plate at which the positive row and column intersect is the location of the positive YAC clone.

2.10 DNA Sequencing.

2.10.1 Purification of PCR Products for Sequencing.

The PGD PCR product to be sequenced was purified from a 3% Low melting point gel made in 1X TAE buffer. This was done with the Geneclean II Kit (Stratech) as directed by the manufacturer. The band was excised from the gel with a sterile scalpel under long wave UV light and weighed. Three volumes of saturated sodium iodide solution was then added to keep the concentration of sodium iodide at 4 molar. The gel was then melted by incubation at 55°C for 5 min and glassmilk added to adsorb the DNA, 1µl was added for each 0.5 µg of DNA. Adsorbtion was carried out on ice for 5-15 min. The silica matrix was then centrifuged briefly and the supernatant decanted off. Washing in 700µl NEW wash (a solution of Tris, sodium chloride, EDTA and absolute ethanol) followed, this was repeated three times. Elution of the DNA into 50µl TE was accomplished at 55°C for 3 min . The suspension was centrifuged for 30 seconds and the supernatant carefully removed. The concentration was assayed by running on a gel with standards (50ng-1µg)

2.10.2 Direct Sequencing of PCR Products.

Sequencing reactions were carried out using the Sequenase kit (version 2.0,USB) with minor modifications. Firstly, 4.5µl of each of the dideoxy nucleotides was placed into four labelled 0.5 µl eppendorf tubes and 0.5 µl DMSO (Dimethylsulphoxide, Sigma) added and mixed. 2.5 µl of this was then transferred into similarly labelled tubes and prewarmed at 37°C. These were used for two sequencing reactions. The annealing reaction was carried out in the presence of DMSO. Here, 6µl DNA (~500ng) was mixed with 1µl DMSO, 2µl Sequenase 5X reaction buffer and 1µl of primer (300ng). This mixture was boiled for 2 min and then plunged on ice for the addition of labelling reaction components. The following were then added : 1µl DTT, 2 µl diluted labelling mix (1:5), 0.5 µl ^{35}S - dATP, 2 µl diluted Sequenase (1:7) in dilution buffer. The labelling reaction was carried out for 5 min at 20-22°C. 3.5 µl of the latter was then added to each of the four dideoxy nucleotide mixtures whereby termination of the elongation reaction occurred whenever a dideoxy nucleotide was incorporated. These four termination reactions were allowed to proceed for 5 min at 37°C. 4µl of formamide stop dye was finally added and the tubes left on ice until loading or stored at -20 °C for one week. Just before loading, the reactions were denatured at 75 °C for 3 min and placed on ice . 3.5 µl of each reaction was loaded at the top of the gel in the order: A, C , G, T.

2.10.3 Sequencing of Cloned DNA.

Cloned DNA in the plasmid vectors pAT153 and pUC18 were sequenced using the above method with DMSO to denature DNA or by alkaline denaturation. In the latter procedure 5µg of DNA is denatured in 0.2M sodium hydroxide, 0.2mM EDTA over 30 min at 37°C. The mixture is neutralized with 0.1 volumes of 3M sodium acetate (pH 4.5) and precipitated with four volumes of ethanol at -70°C. The pellet was then washed in 70% ethanol dried and

redissolved in 7µl of distilled water. 2µl of Sequenase reaction buffer was added and 1µl (1pmol) of primer and annealing carried out by heating at 65°C for 2 min and then allowing the reaction mix to cool to 37°C, this took about 30 min.

Labelling with ^{35}S-dATP followed. 1µl of DDT, 2µl of diluted labelling mix, 0.5µl ^{35}S-dATP and 2µl of diluted sequenase were added and labelling allowed to proceed at 20-22°C for 2-5 min. 3.5µl of this was then added to each of the four prewarmed 2.5µl aliquots of dideoxynucleotides. Termination followed at 37°C for 5 min. Subsequent steps were as already described above.

2.10.4 Polyacrylamide Gel Electrophoresis of Sequencing Reaction Products.

The gel was prepared by dissolving 36.8g Urea (7M) in 29.2ml of water, 8ml of 5X TBE (0.5X) and 12ml of 40% (w/v) acrylamide and bisacrylamide solution (19:1, NBL) this gave a 6% gel. Once dissolved, the polymerization reaction was initiated by the addition of 80µl of TEMED (Bio-Rad) and 80µl of 25% ammonium persulphate (Bio-Rad). The gel was then poured into a Bio-Rad 21X50 cm Sequi-Gen apparatus which had previously had the front plate siliconized with Dimethyldichlorosilane. The gel was left at 20-22°C to set for 1-2 hrs and then was pre-run at 1800-2000 volts in 1X TBE buffer until it warmed up to 45-50°C. 3.5µl of each sample was then loaded and run for 3-7hrs. The plates were then separated and the gel carefully lifted from the back plate using a double layer of 3MM Whatman filter paper. This was covered with clingfilm and subsequently dried for 1 hour at 80 °C on a gel drier (Bio-Rad), autoradiography overnight followed.

2.11 Mammalian Cell Culture.

Cell lines were grown from stocks frozen in liquid nitrogen. The lymphoblastoid line MGL8B2 was grown in RPMI 1640 medium + 10% foetal calf serum. The Chinese Hamster cell line a23, was grown in MEM, the HCH and Taxi hybrids were grown in F12 and HAT. The hybrid Dis2.6 was grown in Eagle's Medium: MEM + 10 % foetal calf serum + non-essential amino acids.

Frozen cell stocks were quickly thawed at 37 ^{o}C and transferred into 10 ml of fresh medium in a falcon tube. The cells were then centrifuged at 150 x g for 5 min. The medium was aspirated off and replaced with 5 ml of fresh medium before being transferred into a 25cm^3 tissue culture flask. The cells were incubated at 37 ^{o}C in a 5% CO_2 incubator. The medium was replaced every 48 hrs. Cells growing in suspension (MGL8B2) were centrifuged at 150 x g for 5 mins and the medium was aspirated off. 5 ml of fresh medium was then added and the cells returned to the flask. Cells attached to the surface of the flask (a23 and Hybrids) had the old medium aspirated off and the fresh 5 ml added after.

Once cells were about 60-70% confluent they were divided into two flasks. Surface growing cells were detached as described earlier (section 2.1.3.3). 5 ml of medium was then added to each flask and when confluent these were transferred to 75cm^3 flasks.

2.12 Origin of Hybrids and Family Material .

2.12.1 Origin of Somatic Cell and Ovarian Teratoma Hybrids.

All somatic cell hybrids used in this study have been reported in published work (Carritt *et al,* 1982a; Carritt *et al,* 1992; MacGeoch *et al,* 1992). The HCH hybrids were constructed from peripheral lymphocytes and the *tk*- Chinese

hamster cell line a23. The human donor karyotype was 46,XX,t(1;10) (p36.13;q22.1). The Taxi hybrids were constructed from a fibroblast cell line and the *tk-* Chinese hamster cell line a23. The human donor karyotype was 46,XX,t(X;1) inv(X) (p21.07p11.06;p34.00). The Dis2.6 hybrid contained a small fragment of chromosome 1p and chromosome 3p, the rodent parent is also a23 (Carritt *et al*, 1882a; Carritt *et al*, 1992).The F4 (F4Sc13C1.12) hybrid contains chromosome 1p only, in a mouse background (MacGeoch *et al*, 1992).

The hybrids derived from the ovarian teratomas and their host were constructed by Carritt and others (1982b). All these hybrids contained chromosome 1. Teratoma hybrids were constructed between the *tk-* Chinese hamster cell line a23 and fibroblast cultures from the teratomas. Host hybrids were constructed between a23 and peripheral lymphocytes. The host hybrids were denoted RVL13 and RVL16. Hybrids Lucy 3 and Lucy 6 were derived from teratoma RC1(right cyst 1), hybrid LR5-15 was derived from teratoma RC5 (right cyst 5) and hybrids LLA5 and LLB3 were derived from teratoma LC1 (left cyst 1).

2.12.2 Family Material.

DNA from 40 extended three generation European families was obtained via the Centre d'Etude du Polymorphismes Humaine (CEPH) (Dausset *et al*, 1990). 5μg of DNA was used from each individual in digests.

2.13 Genetic Linkage Analysis Using the CRI-MAP Package.

CRI-MAP (version 2.4) is designed to construct large multilocus linkage maps. It uses an efficient algorithm for computing likelihoods but in doing so it does not use the population frequency of alleles to determine untyped genotypes. Instead, missing genotypes are deduced from known genotypes of ancestors or descendants. If the missing genotype at any locus includes a homozygous

genotype then the meioses from that individual are treated as uninformative. It is however stated that information loss is minimal (CRI-MAP Package Information). At the commencement of the analysis a PGD genotype data file was created This contained the identification number , sex and genotype of each CEPH member typed for the PGD BamH1 RFLP. Four different options were used from the package in linkage analysis.

1. The option 'BUILD, i.e multipoint linkage analysis was used to insert the *PGD* locus into the ordered set of loci from the CEPH Consortium map of chromosome 1 (Dracopoli *et al,* 1991). Here, the ordered loci from the chromosome 1 map were 23 markers (nine of which were haplotyped systems) which had been ordered with likelihood support of greater than or equal to1000:1. The inserted locus i.e *PGD*, was placed in each possible interval in the map and then each order was subjectet to a full maximum likelihood estimation. Build initially analysed the phase known data, for which maximum likelihood computation was much more rapid than for the full data set , and then analyzed the full data set. The order having the highest $\log_{10}$ likelihood was then found. A printout of sex-specific and sex-averaged recombination fractions was then obtained with the corresponding distance in cM for these loci which was calculated using the Kosambi mapping function.

2. Two point analyses were carried out between PGD and three other close markers from the CEPH consortium map (Dracopoli *et al,* 1991) These markers were PND, D1S71 and D1S43/D1S47. Maximum likelihood estimates of female and male recombination fractions and sex-specific lod scores were calculated.

3. 'FLIPS' was used to deduce whether likelihood support for the map derived by 'BUILD' was more or less when adjacent pairs of loci were permuted down the length of the map. This gave the relative $\log_{10}$ likelihood i.e the $\log_{10}$ likelihood of the original order minus that of the permuted order. If this difference was greater than or equal to three then it was assumed that the original order was more likely than the order with the permuted pair of loci.

4. 'CHROMPIC' gave a display of the location of likely cross over events in specified children, based upon the grandparental origin of marker loci. At the start of the analysis the programme found the maximum likelihood estimates of the recombination fractions for the specified locus order (from the option 'BUILD'), and then used this information to deduce the most probable parental phase. Disruption of this phase in the children, of course, indicates cross over events which, when multiple or closely spaced, represent possible data errors. The output from 'CHROMPIC' has been presented in a modified form, where double cross overs involving PGD and flanking markers, parental origin of the recombinant chromosome and a probability statistic for the parental phase are shown.

2.14. Description of 1p36 Polymorphisms.

2.14.1 PGD Polymorphism.

The probe pPGDH1, a 3' 600 bp cDNA clone, detects a diallelic BamHI RFLP at the *PGD* locus with variant bands at 11.0 kb and 9.5 kb, constant bands at 11 kb and 4.2 kb also occur (Kleyn, 1990). This polymorphism is shown in figure 5

2.14.2 D1S77 Polymorphism.

The probe pMCT58 detects a PvuII RFLP polymorphism at the D1S77 locus with 5 variant bands between 1.3 and 1.6 Kb (Nakamura *et al*, 1988). Three of these are shown in figure 6. Resolution of these fragments is achieved on a 1.4% agarose gel. The locus was first detected using an oligomeric sequence derived from a VNTR locus (Nakamura *et al,* 1988).

2.14. 3 D1F15S1 Polymorphism.

The probe pH3H2 detects a StuI RFLP at the 1p36 locus. The polymorphism is more complex than that of other loci due to the repetitive nature of the locus. The structure of this is illustrated in figure 7a. There are four variant bands of

3.7 kb, 2.0 kb, 1.2 kb and 0.6 kb which arise as a result of restriction site polymorphism within a repetitive sequence of 1.8 kb. The four variant bands are therefore not allelic with each other but with the 1.8 kb core. As a result, homozygotes for the variant bands cannot be distinguished from heterozygotes by simple inspection. The only accurate interpretation that can be made at this level is where all the variant Stu 1 bands are absent so a -/- phenotype actually denotes a -/- genotype (Welch *et al*, 1989). Inference of genotype can often be made in phase-known pedigrees. This polmorphism is illustrated in figure 7b.

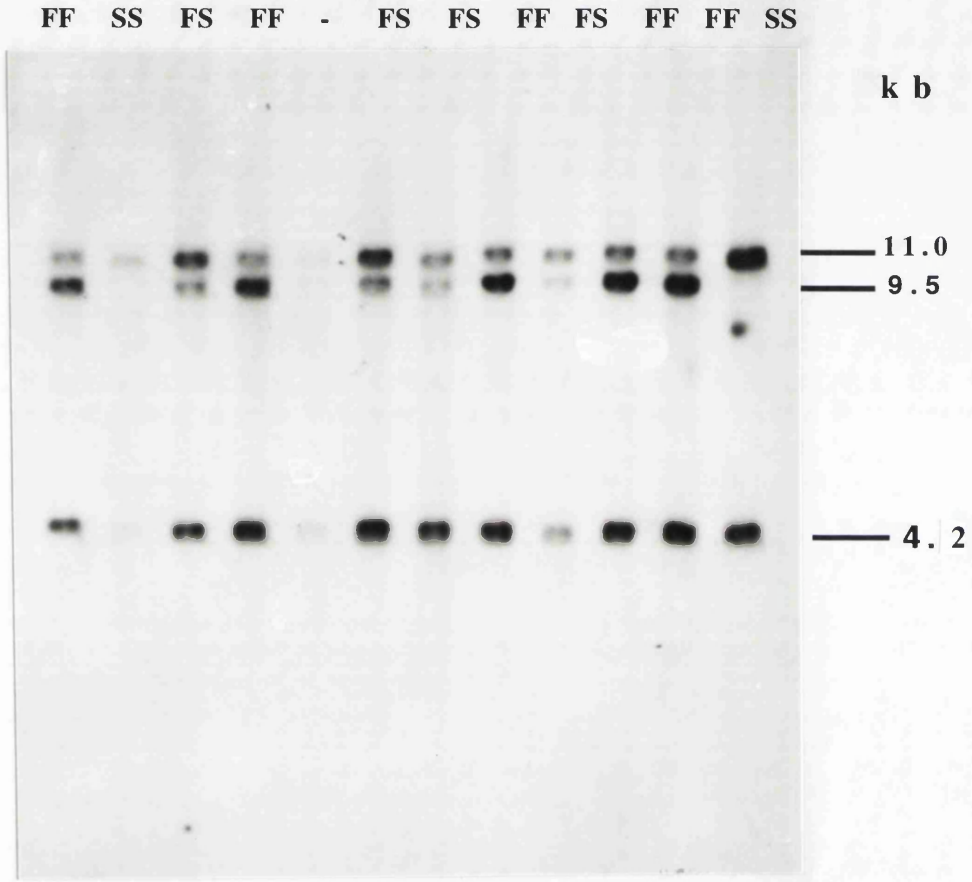


Figure 5. The PGD BamHI Polymorphism.

DNA was digested with BamHI and hybridized to the 3' PGD cDNA pPGDHI. Constant bands are detected of 4.2 and 11 kb. The variant bands are 11 and 9.5 kb. The 11 kb band is the slow allele (S), the 9.5 kb band is the fast allele (F). FF homozygotes are distinguished from FS heterozygotes by the increased intensity of the fast, 9.5 kb allele in the FF homozygote.

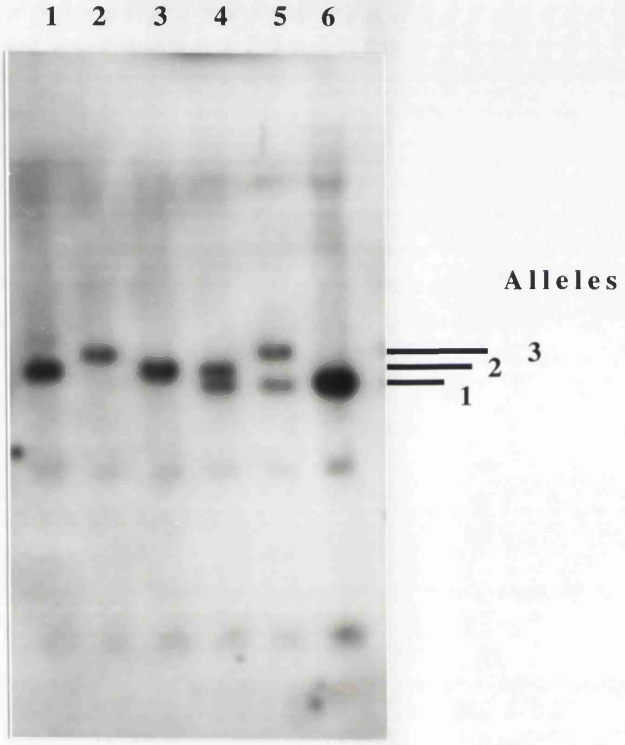
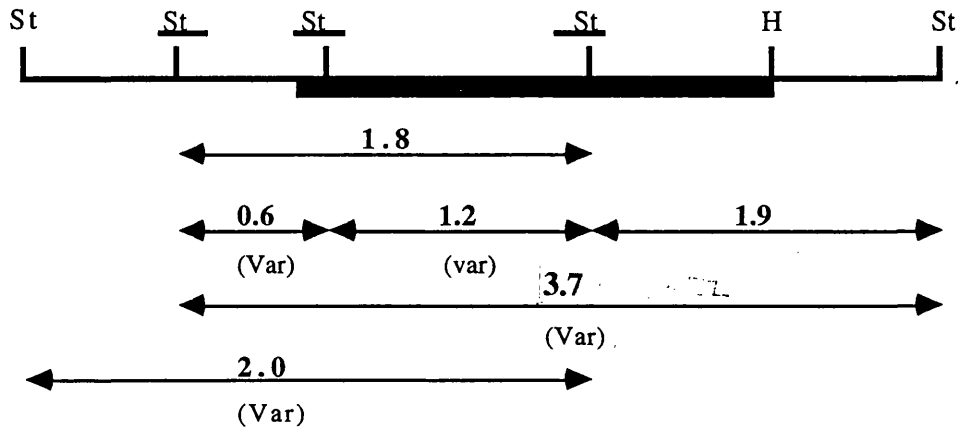


Figure 6. The D1S77 PvuII Polymorphism.

DNA was digested with PvuII, electrophoresed through a 1.4% agarose gel and hybridized to the probe MCT58. This is a VNTR polymorphism with 5 variant bands between 1.3 and 1.6 kb. Three of these alleles are shown in the above gel. The genotypes are as follows:

Lane: 1 - 2 - 2
2 - 3 - 3
3 - 2 - 2
4 - 2 - 1
5 - 3 - 1
6 - 1 - 1

Within single pedigrees PvuII fragments were assigned allele numbers with no attempt to achieve consistency between pedigrees. This allowed the parental origin of alleles to be anambiguously determined, but only using algorithms which do not use gene frequency estimates.



Region Hybridizing to H3H2 probe.

Figure 7a. Restriction Map of D1F15S1 Locus Showing the StuI Polymorphism.

The above restriction map indicates the size of variant bands which arise as a result of the presence or absence of StuI sites in the region. These sites are underlined and the variant bands are indicated by (var) below the fragment. Constant bands of 1.8 and 1.9 kb are also shown. The map is not drawn to scale. This was presented by Welch and others, (1989).

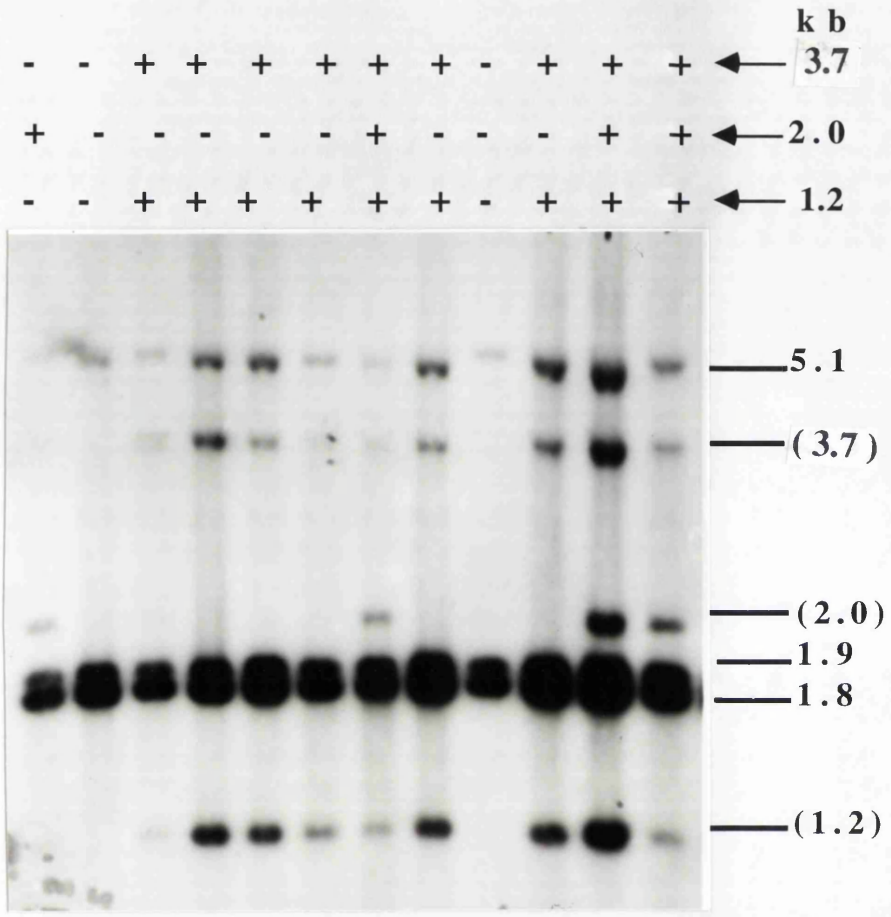


Figure 7b. The D1F15S1 StuI Polymorphism.

DNA was digested with StuI and hybridized to the probe H3H2. Variant bands of 3.7, 2.0 and 1.2 kb are detected. Constant bands at 5.1, 1.9 and 1.8 kb are also seen. The 5.1 kb fragment is from chromosome 3 and the 1.8 kb is present on both chromosomes 1 and 3 (Welch *et al,* 1989). The different phenotypes are the result of the presence or absence of certain StuI restriction sites. The variant bands are indicated in brackets.

CHAPTER 3 RESULTS.

3.1 Genetic Mapping In 1p36.

This section describes the genetic mapping studies carried out in 1p36. The markers used in the analyses were *PGD*, D1F15S1 and D1S77. *PGD* and D1F15S1 were considered important because they had not been incorporated in any of the recent genetic linkage maps of chromosome 1. This study aimed to include these two markers in the most recent map of 1p (Dracopoli *et al*, 1991). Multipoint linkage analysis was therefore carried out using the PGD RFLP on the CEPH panel and the CEPH data base which had all the genotypes for the markers incorporated in the latest chromosome 1 map. Segregation analysis was carried out using the more complex polymorphism at the D1F15S1 locus so that a location could be obtained for this. The D1S77 marker was used in the analysis of ovarian teratoma hybrids as this was the only informative marker of the group that was typed in these hybrids.

3.1.2 Mapping the *PGD* Locus by Pedigree Linkage Analysis Using the Linkage Package CRI-MAP.

The PGD RFLP described in the Methods and Materials (section 2.14.1) and illustrated in figure 5 was employed to map the *PGD* locus more accurately than its previous physical assignments. This experiment began with the initial screening of all the available CEPH parents for the identification of heterozygotes. Subsequently the whole pedigree was typed for this RFLP in those families found to have an informative parent. This gave a total of 20 families and the information was submitted to the CEPH data base. Families typed were the following: 13294, 1334, 1340, 1347, 1349, 1362, 1408, 1416, 1418, 1421, 1423, 1424, 12, 28, 884, 1332, 1346, 23, 1341 and 45. This was a total of 263 individuals, 194 were informative meioses and 66 of these were phase known. A summary of these data and a test of the segregation ratios from different mating types was carried out using a CEPH data base option. The

heterozygosity was 40%, as expected from the allele frequencies reported by Kleyn, 1990.

The segregation ratios from the three different forms of mating types obtained are in approximate agreement with the expected (Table 4). The mating of two heterozygous parents as can be seen gives the 1:2:1 progeny genotype ratio, that from homozygous and heterozygous parents is not a perfect 1:1 ratio. Ratios however, are not significantly different from the expected as shown by a chi-square test A chi-squared value of 2.97 (1d.f.) was obtained ($p < 0.1$). The progeny genotype ratios did not show any significant variation whether the father was heterozygous or the mother in the first type of mating. The expected ratio was obtained from both matings. In the third type there were no matings with a heterozygous father.

	No of	Genotype of progeny		
	Matings	11(SS)	12(FS)	22(FF)
Parental Genotypes.				
12(FS) x 11(SS)	12	30	43	0
12(FS) x 12(FS)	9	14	27	13
12(FS) x 22(FF)	3	0	8	5

Table 4. Segregation Ratios from the Different Mating Types. F/2 and S/1 refer to the two *PGD* alleles.

3.1.2.1 Multipoint Linkage Analysis I.

Multipoint linkage analysis was used to insert *PGD* in the CEPH consortium map of 1p. The analysis was carried out by the programme CRI-MAP (version 2.4). CRI-MAP was written by P. Green and described by Donis-Keller and others (1987). The 20 ordered loci from the CEPH consortium map of chromosome 1 spanned a sex averaged genetic distance of 124.1 cM, a male specific length of 85 cM and a female length of 170.5 cM, using the Kosambi mapping function. 13 of the ordered loci are shown in figure 8 from telomere to centromere with arrows marking the intervals into which *PGD* is inserted. Nine of the ordered loci from the consortium map were haplotyped systems but all of these loci were used in the likelihood calculations and the distances were not forced to zero although no recombinants were detected within these nine systems.

Two locations were suggested for *PGD* with a more distal location being more favourable and flanked by *PND* and D1S43/D1S47 The possible explanation for the placement in these two positions is that no recombinants occur between *PGD* and any of the three flanking markers of the two intervals into which it is placed..The $\log_{10}$ of the likelihood of the assignment in the distal interval (interval 1) is -784.44 and in the proximal interval (interval 2) it is -785.71. The difference between these is 1.27, the antilog of which is 18.6, thus the distal location is 18.6 times more likely than the proximal. Since this difference is below the statistically significant threshold of 1000:1 (Morton, 1955) it may be assumed that these locations are equally likely.

The ambiguous location in the CEPH consortium map of 1p was in fact a consistent occurrence throughout numerous linkage analyses carried out after thorough checking of data. It is believed here that incorrect entries and/or typings for the PGD RFLP have been eliminated by rigorous checking (i.e

repeat typing of some individuals and reading of autoradiographs and data base entry checks by two individuals simultaneously), accompanied by omission of uncertain genotypes. However, after corrections and omission of uncertain genotypes subsequent analysis failed to result in a single location as indicated above. Of course the latter result, (figure 8) and results from further analyses are from the data which were edited last, that is data which had gone through the most rigorous checking.

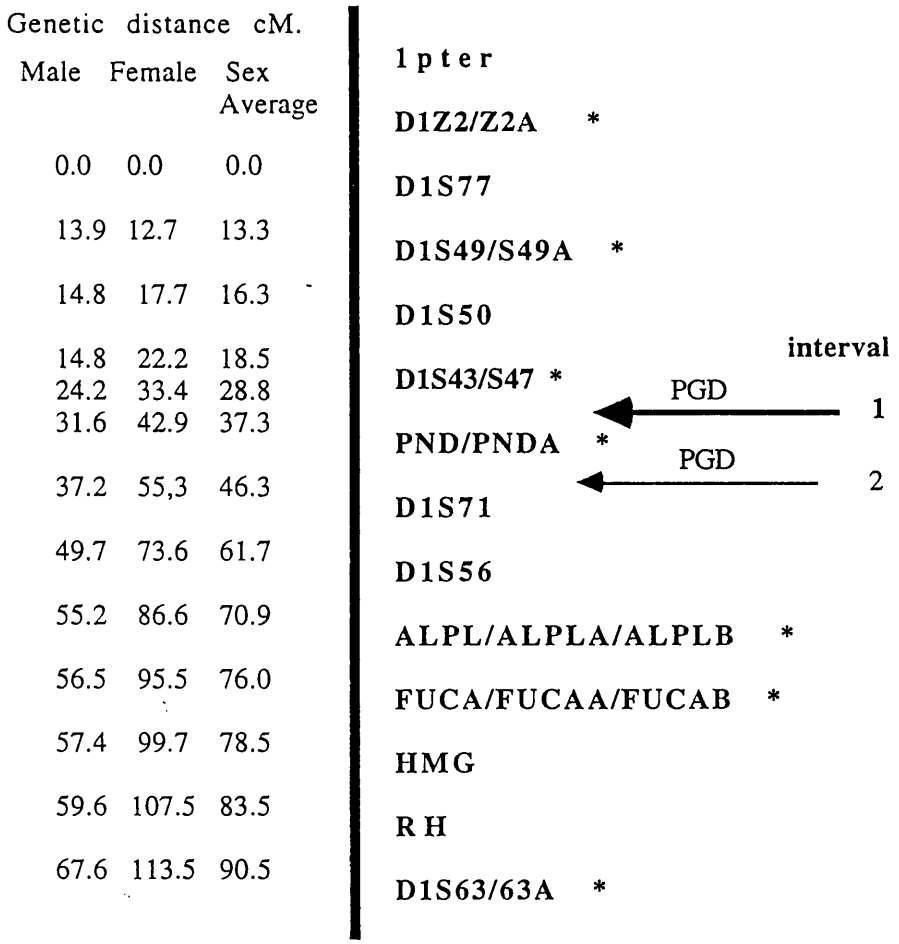


Figure 8. Location of PGD in the CEPH Consortium Map of 1p.

This map indicates the two locations in which the PGD locus is placed by multipoint linkage analysis. The assignment in interval 1 is considered to be 18.6 times more likely than in interval 2. The haplotyped markers are indicated by asterisks. The cumulative distance is also shown for the order which has PGD in interval 1.

3.1.2.2 Two Point Analysis Between *PGD* and Three 1p Markers.

Two point analysis followed between *PGD* and each of the three flanking loci of the two intervals into which *PGD* is placed in the above map (figure 8). The recombination fractions at which the lodscore was highest are presented (Table 5). From these we can see that lower recombination fractions between *PGD* and the markers D1S43 and *PND* are accompanied by higher lodscores. The recombination fraction between *PGD* and D1S71 is greater. It shows that the inability to decide in which of these two intervals *PGD* lies is not from lack of recombinants between *PGD* and the three flanking markers as speculated above.

LOCUS1	LOCUS 2	Theta (f)	Theta(m)	Lodscore
PGD	D1S43	0.09	0.15	9.16
PGD	*PND*	0.07	0.20	5.35
PGD	D1S71	0.31	0.24	1.46

Table 5. Recombination Fractions Between *PGD* and the Loci D1S43, *PND* and D1S71 at which the Maximum Lodscore is Obtained.

3.1.2.3 Multipoint Linkage Analysis II.

In order to assess the contribution that *PND* data errors may have made to the ambiguous location of *PGD* the multipoint analysis was repeated without *PND*. This analysis is referred to as multipoint linkage analysis II. Only the first 12 ordered loci of the CEPH map from 1pter were used in multipoint analysis since we now knew where *PGD* will fall. If *PND* is removed from this analysis it is found that *PGD* falls in the interval D1S43/D1S47 - D1S71. As the likelihood of this map cannot be compared with that of the map which includes *PND* , the validity of this map without *PND* was assessed using the CRI-MAP option 'flips' which gives local support for the placement of *PGD* in the CEPH consortium map. The programme 'flips' determined whether the inversion of two adjacent loci led to more significant lodscores for the order obtained. The results of this show that the location of *PGD* in the interval D1S43/D1S47-D1S71 after exclusion of *PND* from the map is more likely than that from inversion of any pair of loci. Figure 9 shows the pairs of loci inverted in the analysis. The computation of the relative likelihoods of the flipped orders is shown at the right of each inversion shown. This is calculated by subtracting the likelihood of the flipped order, denoted current order, from the likelihood of the original order shown. Where values are positive and equal to or greater than three, then it is accepted that the most likely order is the original as shown in figure 9. With respect to *PGD* and its two neighbours it can be seen from the underlined differences that their original orientation is preferred. The order with *PGD* placed proximal to D1S43/D1S47 is 7.9 X 10^5 times more likely (antilog of 5.90) than with the alternative orientation for this pair of loci. Also, the order obtained which places *PGD* distal to D1S71 is 4.4 X 10^{16} times more likely (antilog of 16.65) than that which inverts this pair of loci.

Original Order

Z2	S77	S49	S50	S43	PGD	S71	S56	ALPL	FUCA	HMG	RH	S63	Relative Likelihood of flipped order.
S77	Z2	-	-	-	-	-	-	-	-	-	-	-	3.36
-	S49	S77	-	-	-	-	-	-	-	-	-	-	10.23
-	-	S50	S49	-	-	-	-	-	-	-	-	-	4.65
-	-	-	S43	S50	-	-	-	-	-	-	-	-	19.18
-	-	-	-	PGD	S43	-	-	-	-	-	-	-	5.90
-	-	-	-	-	S71	PGD	-	-	-	-	-	-	16.65
-	-	-	-	-	-	S56	S71	-	-	-	-	-	4.60

Figure 9. Inversion of Adjacent Pairs of Loci in the Original Order (without *PND*) and Calculation of Relative Likelihoods.

This analysis indicated that the original order is the preferred order, that is placing *PGD* in between the markers D1S43/D1S47 and D1S71.

Locus key: Z2 - D1Z2, S77 - D1S77, S49 - D1S49, S50 - D1S50, S43 - D1S43, PGD - *phosphogluconate dehydrogenase* , S71- D1S71, S56 - D1S56, ALPL - *alkaline phosphatase*, FUCA - *alpha-fucasidase*, HMG - *high - mobility group protein 17* , RH - *Rhesus* blood group.

The recombination fraction between *PGD* and D1S43/D1S47 is 0.11 in females and males, whereas two point data show an excess in the male for this interval. Table 6 lists the recombination fractions and distances from this second multipoint analysis and those from the first analysis incorporating *PND* in the ordered set of loci.

The inspection of the distances from the two multipoint analyses shows that the locus *PND* is proximal to *PGD* since a greater distance separates *PGD* from the proximal locus D1S71 than that which separates *PND* from D1S71 in both male and female. Also, the combined distance from multipoint linkage analysis I which includes *PND* is in very close agreement with that from analysis II which excludes *PND* from the analysis. This is the observation in both sexes. i.e 30.7 cM (18.3 + 12.4 cM in the female) and 18.5 cM (12.5 + 6.0 in the male) for the intervals *PND* - D1S71 and *PGD* - *PND* respectively. These values compare well with 32.5 cM and 15 cM respectively for the interval *PGD* -D1S71.

	Interlocus Distances & Theta Values		Multipoint Analysis
Interval	**Theta(f)/cM**	**Theta(m)/cM**	**No.**
D1S43/S47—*PGD*	0.11, 11.1	0.11, 11.0	2
D1S43/S47—*PGD*	0.09, 9.5	0.07, 7.4	1
PGD—*PND*	0.12, 12.4	0.06, 6.0	1
PND—D1S71	0.18, 18.3	0.12, 12.5	1
PGD—D1S71	0.29, 32.5	0.15, 15	2

Table 6. List of Sex Specific Recombination Fractions and Distances from Multipoint Linkage Analyses I and II. The Kosambi mapping function was used to calculate genetic distance.

3.1.2.4. Analysis of Meioses Using the Programme CHROMPIC.

The programme Chrompic displayed the grandparental origin of chromosomes in the children in all the pedigrees typed for the PGD RFLP. The chromosomes illustrated were from the first multipoint linkage analysis which included *PND* in the map. The data frrom this were analysed by eye and the instances of double recombinations are presented in table 7 in a more simplified fashion. than that presented by the CHROMPIC output.

This programme calculates the maximum likelihood estimates of the recombination fractions for the predetermined locus order of the build option (multipoint analysis I) used earlier. These maximum likelihood recombination fractions are then used to determine and indicate the most likely grandparental origin of the chromosomes and the probability of the most likely phase and of the second best. In addition the location and number of cross over events is indicated for every chromosome

For each interval for which informative loci occur, the chrompic programme provides a list of individuals harbouring cross overs in this interval, as well as the origin of the chromosome in which crossing over occurs. This display therefore gives information on all recombination events that occur and allows detection of potential typing or entry errors by showing unlikely double or triple recombinants.

In the interval separating *PGD* from D1S43, a total of 11 recombinants are detected, six of which occur in the paternal chromosomes and five in the maternal. In the next interval, i.e *PGD -PND*, five recombinants are found and again there are more recombinants in the male meioses, here three from paternal chromosomes compared with two from maternal. The genetic distance however is greater in the female than the male in the proximal interval but almost equal

in the distal interval thus showing that the ratio in the sex difference in genetic distance is not constant. This information is derived from data in table 6. If we then inspect the chromosomes displayed, we can see that *PGD* lies in a very short region which undergoes a double cross over in eight meioses. These are illustrated in table 7.

This analysis detected eight instances of apparent double cross overs in this region in 286 opportunities. Of these, four involve flanking markers separated by 70.9 cM, 76 cM,and 52 cM (D1S77-*ALPL*, D1Z2-*FUCA*, D1S43-*ALPL*, respectively). The other four examples involve genetic distances of as little as 18 cM (D1S43-*PND*). Whether these represent residual data errors or genuine double recombinants cannot be determined. Given that these four cases are less than 1.5% of the total, it seems unlikely that they reflect the assignment of *PGD* to this interval.

Child ID no.	Parental Chromosome Origin	Probability of phase First	Second	Location of Cross Overs
142306	Maternal	1.000	0.000	D1S77^{pk} X PGD^{pk} X ALPL^{pu}
142304	Maternal	1.000	0.000	D1S77^{pk} X PGD^{pk} X ALPL^{pu}
141604	Paternal	0.997	0.000	D1S43^{pk} X PGD^{pk} X D1S56^{pu}
134908	Maternal	0.938	0.000	D1S43^{pk} X PGD^{pk} X PND^{pu}
142104	Paternal	0.881	0.060	D1S43^{pk} X PGD^{pu} X PND^{pk}
4506	Paternal	0.617	0.376	D1Z2^{pu} X PGD^{pu} X FUCA^{pu}
1329405	Maternal	0.515	0.383	D1S43^{pk} X PGD^{pu} X ALPL^{pk}
88414	Paternal	0.138	0.126	D1S43^{pk} X PGD^{pu} X D1S71^{pu}

Table 7. List of Meioses with Two Cross Over Events Encompassing *PGD.*
The above table lists the meioses in which two cross over events involving the *PGD* locus. are detected The parental origin of the recombinant chromosome is shown together with the probability of the two most liely phase Next the interval where cross over events occur is shown: 'X 'marks the location of the cross over event, 'pk' signifies a phase known marker and 'pu' a phase unknown marker. Underlined loci are those inherited from the grandmaternal chromosome those which are not underlined are grandpaternal in origin This table is a simplified summary of the output from the CHROMPIC option of the CRI-MAP package. In determining the sites of crossing over the following genetic order was assumed:

D1Z2 - D1S77 - D1S49 - D1S50 - D1S43 - PGD - PND - D1S71 - D1S56 - ALPL - FUCA

3. 1.3. Segregation Analysis of the D1F15S1 Locus in the CEPH Pedigrees.

Analysis of the D1F15S1 polymorphism is more complex than that at other polymorphic loci studied in this project as a result of the repetitive nature of the locus (Welch *et al*, 1989). 15 of the CEPH pedigrees were typed for this RFLP and the typings are listed in the Appendix (section 5.6). The only situation in which the genotype can be deduced with certainty from the phenotype is when a band is absent. This suggests homozygosity. In other situations it may be necessary to analyse the whole of a three generation pedigree so that the homozygotes and heterozygotes for the presence of a particular band can be determined and the phase of parental chromosomes may be worked out. Once genotypes are established segregation analysis follows. This is illustrated in pedigree 1334 which was found to be informative for the four loci of interest in this genetic mapping exercise (figure 10).

A number of the progeny are homozygous for the absence of one or two of the variant bands, denoted (--). The paternal phenotype is - + - which suggests that the genotype is - + - /- - - or - + -/ - + -. The grandpaternal (GF 10) phenotype + - + shows that the genotype of the father has to be - + -/- - -. This gives the phase as shown. All other markers are more easily genotyped and the D1S77 locus which is not typed in the grandmaternal chromosomes (GM 11) is partially deduced from the genotype of the father. Next the maternal phase is delineated as follows: The phenotype here is + - + and the grandpaternal (GP 12) genotype is - - -/ - - -. This immediately shows that the maternal genotype is - - -/ + - +. The grandmaternal (GM 13) phenotype is + - + and shows that the chromosome transferred to the mother was + - +. With the parental genotypes and linkage phase established the genotypes of the offspring may then be inferred and segregation of the markers followed with simultaneous detection of recombinants.

Children three, five, seven, eight and nine are all non-recombinants. Child three has the phenotype + - +, the GM 13 maternal chromosome is inherited here as shown and the GP 10 paternal chromosome which give the genotype as follows: - - -/ + - +. This segregation pattern for all four markers is shown in children five and eight. Children seven and nine also share the same parental chromosomes. Recombinant four arises as follows: The parental chromosome is that from G F 11. The maternal chromosome in child four has the phase **S 2 2** - - - , (*PGD-PND*-D1S77-D1F15S1) this means that a cross over must have occurred between the group (*PGD, PND*, D1S77) and D1F15S1 although the homozygosity at the D1S77 locus prevents confirmation of the absence of recombination between this and the other markers. It is of course highly unlikely that two cross overs would occur in such a small interval. It is therefore reasonable to assume that the detected cross over at the above interval is the only one.

Child six also shows recombination between D1F15S1 and the other three markers. Here, however the certainty of the deduced order is greater as all four loci are informative. The paternal chromosome again is in phase at all loci. The maternal chromosome phase is consistent with recombination between the group *(PGD*, D1S77, *PND)* and D1F15S1. The phase is **F 1 2** + - +, *(PGD-PND*-D1S77-D1F15S1) the latter marker coming from the GM 13 chromosome and the first three from the GP 12 chromosome. The conclusion from this analysis is that in two individuals a cross over occurs between the D1F15S1 locus and the remaining three markers.

A list of the informative meioses where recombinants were detected is shown in Table 8. Seven meioses show D1F15S1 segregating away from the other two or three markers, five meioses show D1S77 segregating away from the other three, the conclusion from these combined results is that these two are flanking

markers. Finally, the cosegregation of *PGD* and D1S77 away from *PND* gives the following order for the markers : D1S77 - *PGD* - *PND* - D1F15S1.

Pedigree I. D. no.	Origin of Recombinant Chromosome	Cross Over Location
141604	P & M	**D1S77**-(PND,PGD,D1F15S1)
141605	M	**D1S77**-(PND,PGD,D1F15S1)
133404	M	**D1F15S**1-(D1S77,PGD,PND)
133406	M	**D1F15S1**-(D1S77,PGD,PND)
142104	M	**D1S77**-(PND,PGD,D1F15S1)
140803	M	**D1S77**-(PGD,D1F15S1)
140804	M	**D1F15S1**-(D1S77,PGD)
140807	M	**D1S77**-(D1F15S1,PGD)
140809	M	**D1F15S1**-PGD
140814	P	**D1F15S1**-(D1S77,PGD)
136204	P	**D1F15S1**-(PND,PGD)
136208	M	**D1F15S1**-(PND,PGD)
141809	M	**PND**-(PGD,D1S77)

Table 8. List of Informative Meioses in the CEPH Pedigrees for D1F15S1 and Three Other 1p Markers.

The location of cross overs between D1F15S1 and the other three markers is shown, the marker segregating away from the other group is shown in bold. The order of the other markers in each meiosis is not known. The parental origin of chromosomes undergoing the recombination event is also indicated.

P- paternal chromosome, M-maternal chromosome.

CEPH Pedigree No. 1334

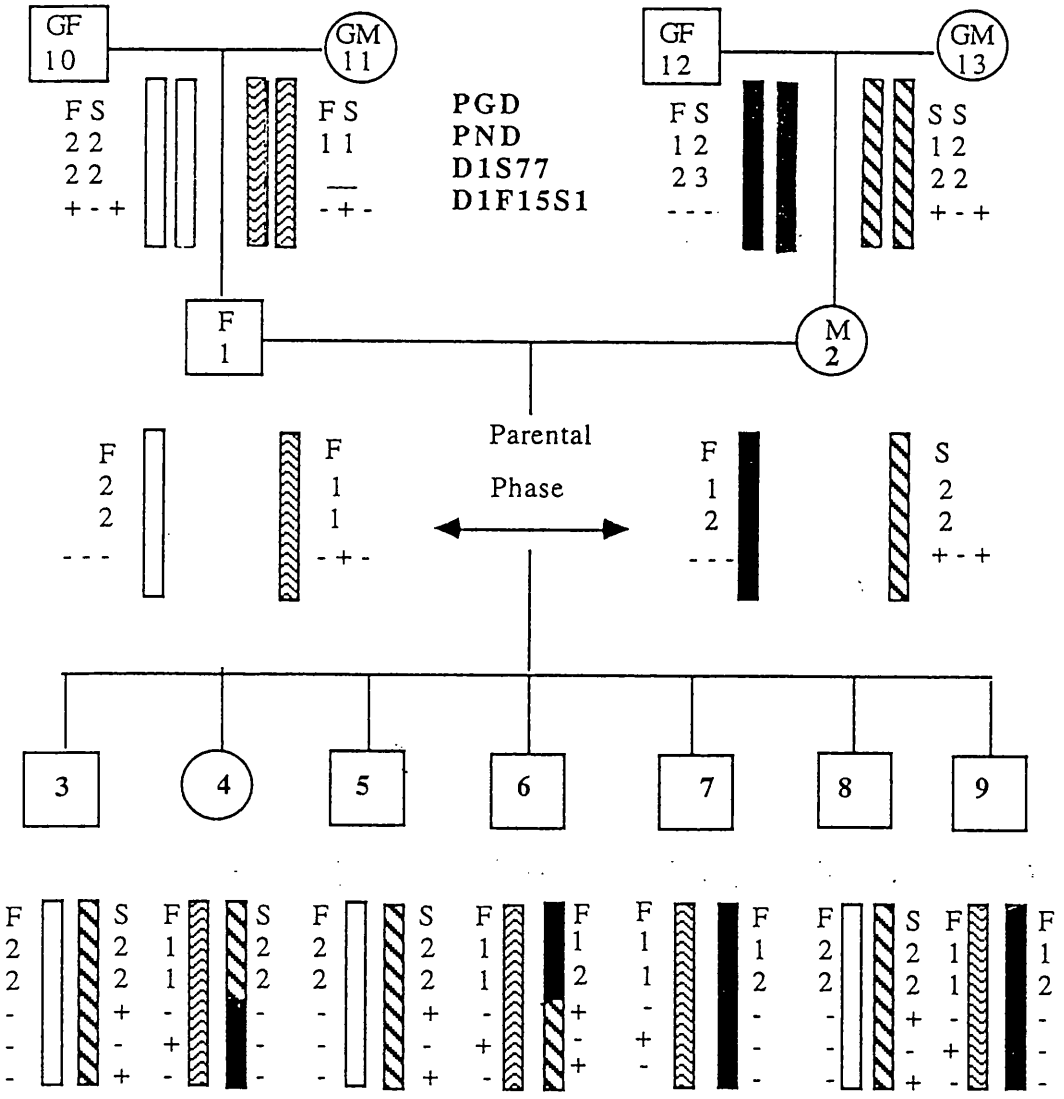


Figure 10. Segregation Analysis of 4 1p36 Markers in CEPH Pedigree 1334.

GF - Grandfather, GM - Grandmother, F - Father, M - Mother.

Grandparental phenotypes and parental and children's genotypes are shown for the D1F15S1 locus. As the D1F15S1 polymorphism is scored as fully dominant, generally genotypes cannot be inferred directly from phenoytype. It is essential to analyse the whole of a three generation pedigree, as shown above, for the inference of genotypes. Segregation analysis allows the detection of two recombinants between the locus D1F15S1 and the other three markers *PGD, PND* and D1S77. Both of these recombination events occurred during maternal meioses.

3.1.4. Genetic Mapping with Ovarian Teratoma Hybrids.

In the present study a set of ovarian teratoma hybrids were allelotyped for a set of RFLPs already used in pedigree linkage analysis in sections 3.1.2 and 3.1.3. These were the D1F15S1, D1S77, PND and the PGD RFLPs. These data were used in conjunction with information from an earlier study (carried out in 1982 by Carritt and others) for segregation analysis as a means to deduce an order for 1p markers and thereby to assess the potential of teratoma hybrids in genetic mapping.

In 1982 these ovarian teratoma hybrids had been typed for centromeric markers for four chromosomes including chromosome 1 and for chromosome 1 enzymes including PGM1 and PGD. At the time these were examined so that their origin could be deduced because unlike normal meiotic products, these cysts are diploid and in addition they differ genetically from their host (Carritt et al, 1982b).

In this earlier study the phase of markers was deduced through construction and analysis of somatic cell hybrids retaining one copy of chromosome 1 from the normal host lymphocytes. The phase of markers in the somatic cell hybrids constructed from host lymphocytes were compared with the phase of the six other hybrids each containing one of the two chromosome 1 homologues derived from the different ovarian teratomas. Once the haplotypes were known, they could be directly compared with the host phase and thus allow the inference of location of cross-over events necessary to give rise to the altered phase in the teratomas. Thus the defective mechanism of meiotic cell division giving rise to the diploid state in these cysts could be deduced.

The host was found to be heterozygous for the two enzymes, and the phenotype of one hybrid from lymphocytes of the host revealed the phase of these as :

cenS, PGM 1*1+,PGD B

cenM, PGM1*1- PGD A

With this information from the study by Carritt and others, (1982b), the aforementioned markers were typed in these same hybrids so that their location could be found and also compared with that from other mapping studies.

Lymphocyte DNA from the host was unavailable to type other markers so the hybrids derived from the host lymphocytes containing one copy of chromosome 1 each were used to determine haplotypes. Of the four markers typed, D1S77, D1F15S1, *PGD* and *PND*, only D1S77 was informative enough to allow segregation analysis to be carried out. The hybrids showed that the host must have been heterozygous at D1S77, 4-2, and that allele 2 was associated with PGM1*1+ and PGD B in the host hybrids (RVL13 and RVL 16). As the phase in these hybrids is the same, the phase of the second homologue is determined from knowledge of all the hybrid typings for D1S77. Examination of linkage phase in the teratoma hybrids and knowledge of that in the host permits the determination of recombination events during the production of these diploid cysts.

Table 9 shows the location of cross over events in each teratoma. The correct order for the markers is that indicated below the table. In an attempt to reduce the number of cross overs from two to one in each of these hybrids, each of the two markers *PGM 1* and *PGD* attain the most distal location on 1p thus giving different orders which in fact, never place D1S77 in the terminal location. This is a divergent conclusion to that of pedigree linkage analysis (section 3.1.2.).

Genotypes for 1p Markers

Teratoma	Hybrid	Cent Marker	PGM1	D1F15S1	PGD^P	PGD^R	D1S77	Cross Over Location
RC1	Lucy 3	S	1-	+-+	A	S	4	C-(PGM,PGD,D1S77)
	Lucy 6	M	1+	+-+	B	S	2	C-(PGM,PGD,D1S77)
RC5	LR5-15	S	1-	+-+	B	S	2	C-PGM-(PGD,D1S77)
LC1	LLA5	S	1+	+-+	A	S	2	(C,PGM)-PGD-D1S77
	LLB3	M	1+	+-+	B	S	4	C-(PGM,PGD)-D1S77
HOST	RVL13	S	1+	+-+	B	S	2	
	RVL16	S	1+	NT	B	S	2	

Table 9. Genotypes of 1p Markers in Hybrids from Teratomas and Location of Cross Overs.
The chromosome 1 markers found in hybrids constructed from teratomas are shown. These hybrids reconstruct the phase of markers in the teratomas from which they are derived. The phase of markers is compared with that from the host and thus allows cross over location to be determined (last column).
Linkage phase in host DNA is as follows:

cenS, PGM1*1+, PGDB, D1S77*2
cenM, PGM1*1-, PGDA, D1S77*4

PGD^P - PGD protein polymorphism
PGD^R - PGD Bam HI RFLP.
Teratomas are classified as RC (right cyst) and LC (left cyst). Hybrids Lucy 3 and Lucy 6 are drived from RC1, LR5-15 is derived from RC5, LLA5 and LLB3 are derived from LC1, and RVL13 and RVL16 are derived from host lymphocytes.

3.2. Genomic Structure and Sequence of the *PGD* Gene.

3.2.1. Genomic Structure of the *PGD* Locus.

Mapping studies at a high resolution were carried out at the *PGD* locus. These led to the identification of a CpG island which was later found to be useful in long range mapping (section 3.3.3). Also, restriction mapping of the PGD genomic clones and the identification of PGD cDNA hybridizing regions allowed the determination of exon-intron structure in these fragments (section 3.2.2.3).

Previously the 5' cDNA clone pPGDH4 was shown to hybridize to 5 EcoR1 restriction fragments of 10.5, 6, 4.5 ,3.0 and 0.75 kb. Somatic cell hybrid analysis showed that only the first three are on chromosome 1 and that the 3.0 kb fragment is from an unlinked locus. A clone which contained the 10.5 kb and the 4.5 kb fragments was isolated and these were found to be contiguous by restriction mapping and hybridization to the clone pPGDH4. The 6 kb fragment from chromosome 1 had not been found. The 10.5 and 4.5 kb fragments were subcloned and denoted pPGDE10 and pPGDE4 respectively (Kleyn, 1990).

Restriction fragments to which the cDNA hybridized were localized as shown (figure 11). Analysis of the 10.5 kb fragment (pPGDE10) led to the discovery of a cluster of restriction sites for infrequently cutting enzymes. Hybridization of pPGDH4 to these digests showed that the coding regions were all located to one side of this cluster of sites. This raised the possibility that this cluster of sites may represent a CpG island commonly found at the 5' end of housekeeping genes (Bird, 1986). The cluster was within a 1.4kb PvuII fragment (figure 11).

G+C enriched genomic sequences in the 5' region of housekeeping genes tend to be hypomethylated compared to bulk genomic DNA. To see if this region

fulfilled the criteria for an authentic CpG island genomic DNA was digested with EcoR1 and the set of methylation sensitive enzymes or 'rare cutters'. After Southern transfer, the DNA was hybridized to the 1.4 kb PvuII fragment and two fragments of approximately 6.2 kb and 4.3 kb were detected (figure 12).

The similar experiment using the probe pPGDH4 in place of the PvuII fragment is more complex due to cross hybridization to the pseudogene loci. However we can still detect a fragment at about 4.3 kb derived from pPGDE10 (figure 13). This demonstrated that the cluster of rare cutter sites is hypomethylated *in vivo* and their relative location strongly points to their presence in a CpG island. Being rich in the tetranucleotide sequence for HpaII, CpG islands are also referred to as HTF islands (HpaII tiny fragment islands). The PGD island was also found to be unique and to be cleaved by HpaII into tiny fragments (figure 14).

In addition this cluster of HpaII sites was hypomethylated in the germ line. This is shown in figure 15, where a parallel digest of lymphocyte and sperm DNA is carried out. With PvuII digestion alone, the corresponding fragment is detected in both germ line and total human DNA. As can be seen, upon cleavage with PvuII plus either MspI or its isoschizomer HpaII, the 1.4 kb PvuII fragment is lost consistent with the presence of multiple closely-spaced unmethylated 5'-CCGG-3' sites. A very faint smear is seen at the lower level of the gel representing the HTF fraction.

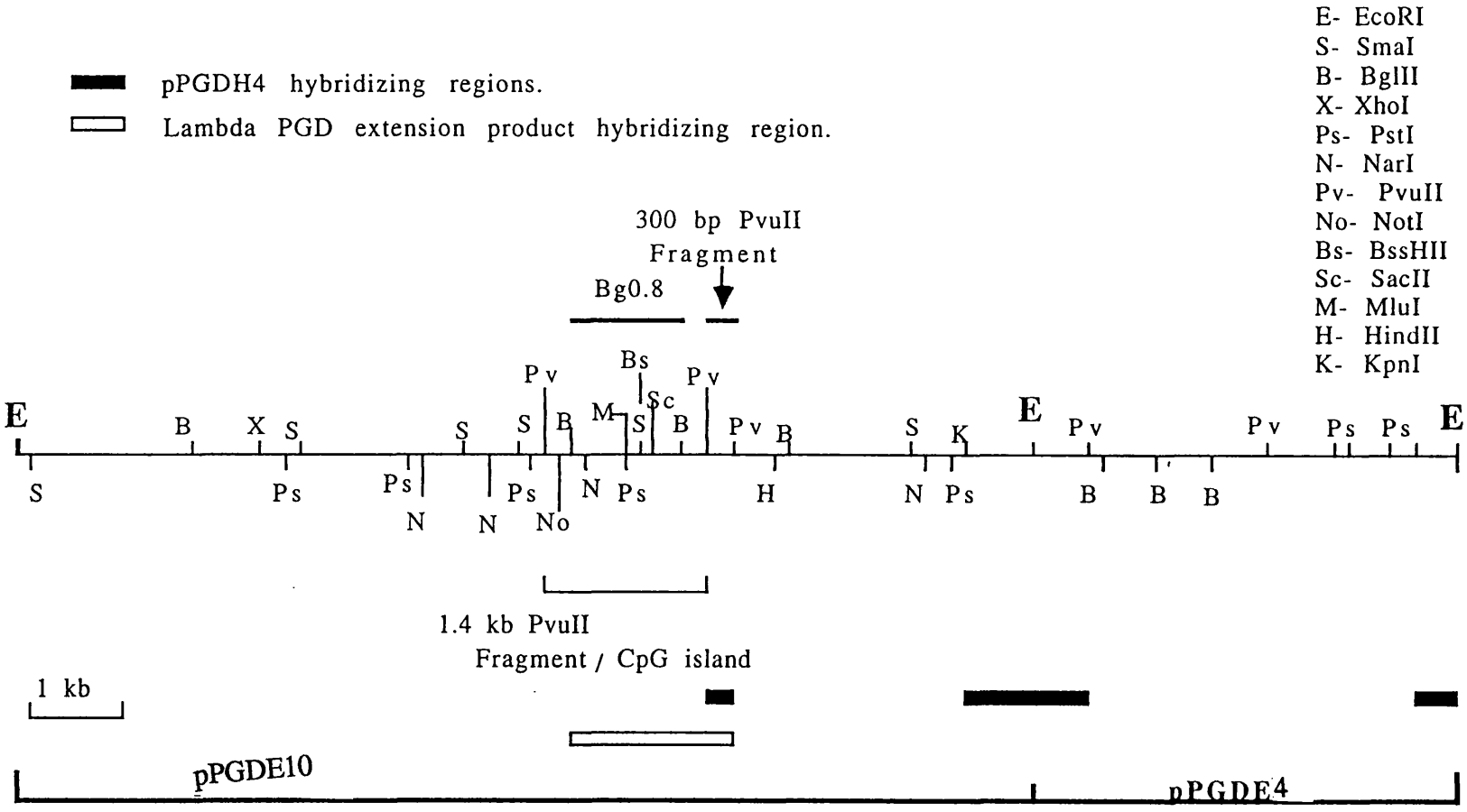


Figure 11. Restriction Map of Genomic Clones pPGDE10 and pPGDE4.

The CpG island identified is within the 1.4 kb PvuII fragment. The regions detected by pPGDH4 and the region detected by the lambda PGD extension product at the 5' end are also indicated.

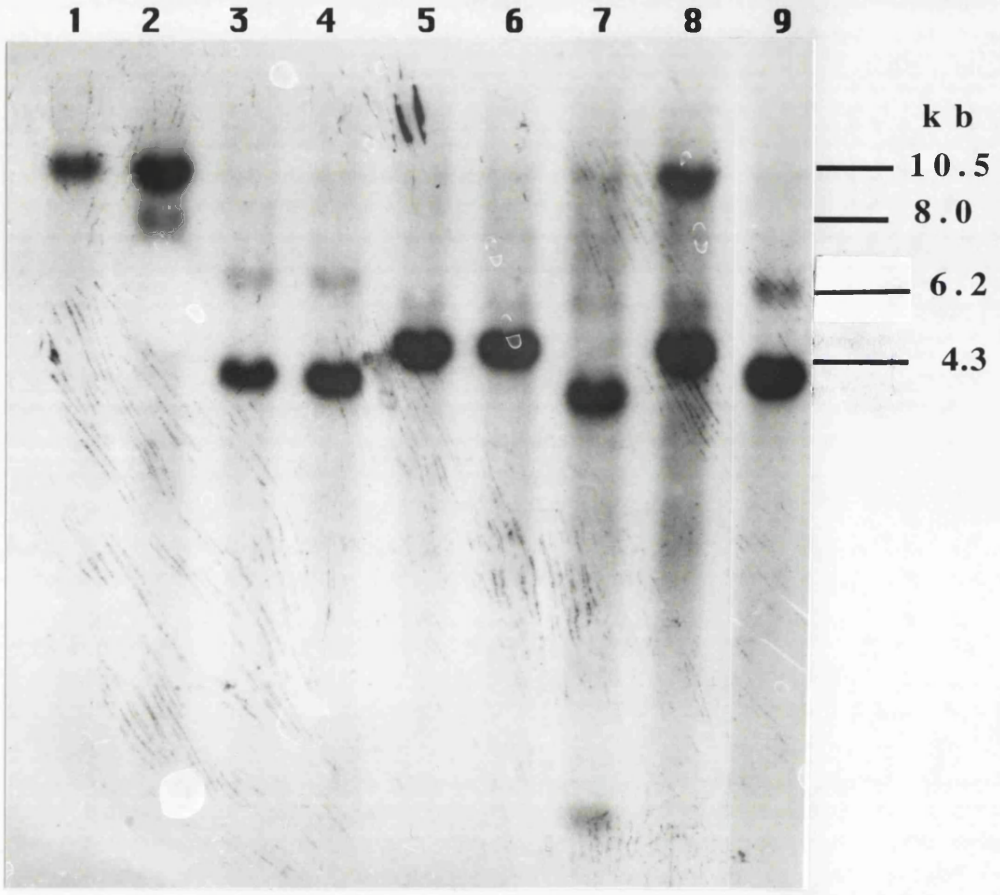


Figure 12. Hybridization of 1.4 kb PvuII Fragment to Genomic DNA Digested with EcoRI and Rare Cutter Enzymes.

Lanes: 1 - EcoRI, 2 - EcoRI & XhoI, 3 - EcoRI & BssHII, 4 - EcoRI & SacII, 5 & 6 - EcoRI & NotI, 7 - EcoRI & SmaI, 8 - EcoRI & NarI, 9 - EcoRI & MluI.

The 1.4 kb PvuII fragment used as probe in this experiment contains a cluster of rare cutter sites used in the above genomic digest. If these sites are unmethylated in genomic DNA then after cleavage and hybridization to the fragment, two fragments should be detected. This is illustrated in the above figure. The smaller fragment at about 4.3 kb is the fragment also detected by pPGD H4. The relative size of this fragment for each enzyme allowed the accurate mapping of these sites within the CpG island. In Lane 2 the presence of the 10.5 kb fragment and the weaker signal of the 8 kb fragment shows that the XhoI site present outside the island is partially methylated. The NarI digest also shows incomplete digestion probably either because of methylation or suboptimal conditions for digestion. A small fragment obtained in the SmaI digest indicates that two SmaI sites are present in the CpG island.

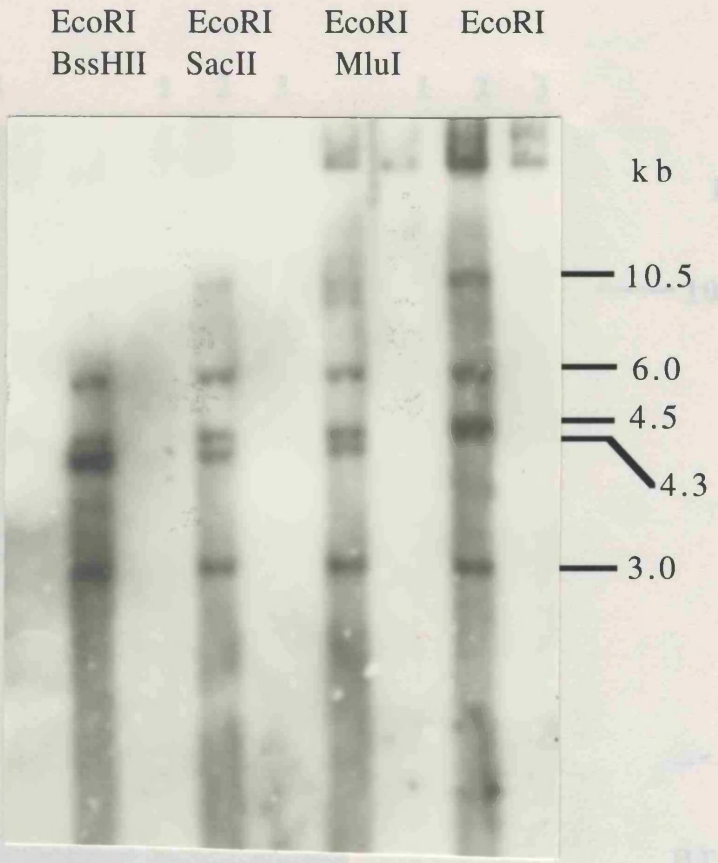


Figure 13. Hybridization of pPGDH4 to Genomic DNA Digested with EcoRI and Rare Cutter Enzymes.

When DNA is digested with EcoRI only, pPGDH4 detects 4 fragments. The 10. 5 and 4.5 kb are in the genomic clones pPGDE10 + pPGDE4 respectively, and the 6 kb fragment is the other expressed PGD sequence not yet cloned. The 3.0 kb fragment is from a PGD pseudogene (Kleyn, 1990). When digested with any of the rare cutter enzymes mapped in pPGDE10 , a 4.3 kb fragment should be detected instead of the 10.5 kb fragment. This is because the pPGDH4 cDNA hybridizes at the 3' side of the cluster of rare cutter sites which is in a 4.3 kb EcoRI-rare cutter fragment. In lanes 1 to 3 this fragment is present although the 10.5 kb is not lost completely probably as a result of incomplete digestion with the rare cutters. The experiment shows that the sites mapped in the clone pPGDE10 are unmethylated in peripheral white blood cell DNA.

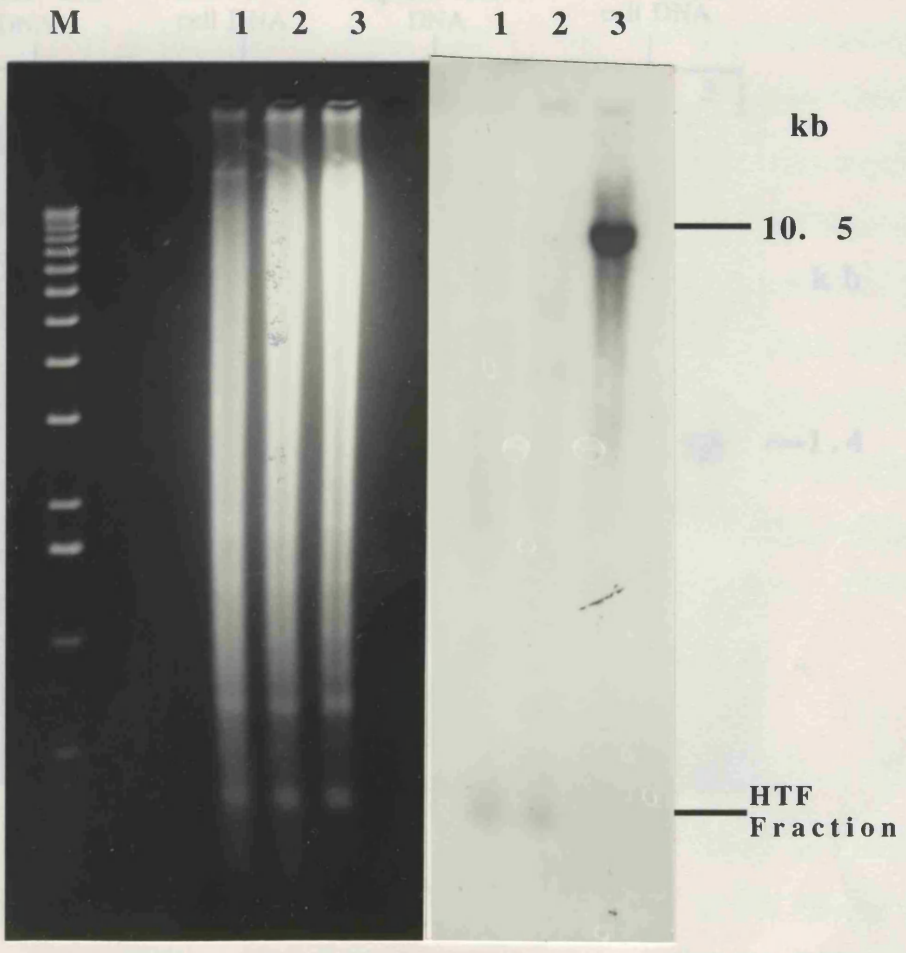


Figure 14 Ethidium Bromide Stained Gel of White Blood Cell DNA Digested with EcoR1 and Isoschizomers MspI and HpaII and Autoradiograph of Hybridization to the 1.4 kb PvuII Fragment

Lanes: M - 1 kb ladder, 1 - EcoRI & HpaII, 2 - EcoRI & MspI, 3 - EcoRI.

The 1.4 kb PvuII fragment contains the CpG island present in the 10.5 kb genomic clone pPGDE10. In lane 3 a single band is detected indicating that this is a unique sequence. In lanes 2 and 3 this band is lost indicating that the island has multiple closely spaced 5'-CCGG-3' sites cleavable by both enzymes. HpaII is a methylation sensitive enzyme thus showing that this island is also unmethylated.

Extra bands present in lane 2 are most probably from incomplete digestion by Msp1 resulting from suboptimal reaction conditions.

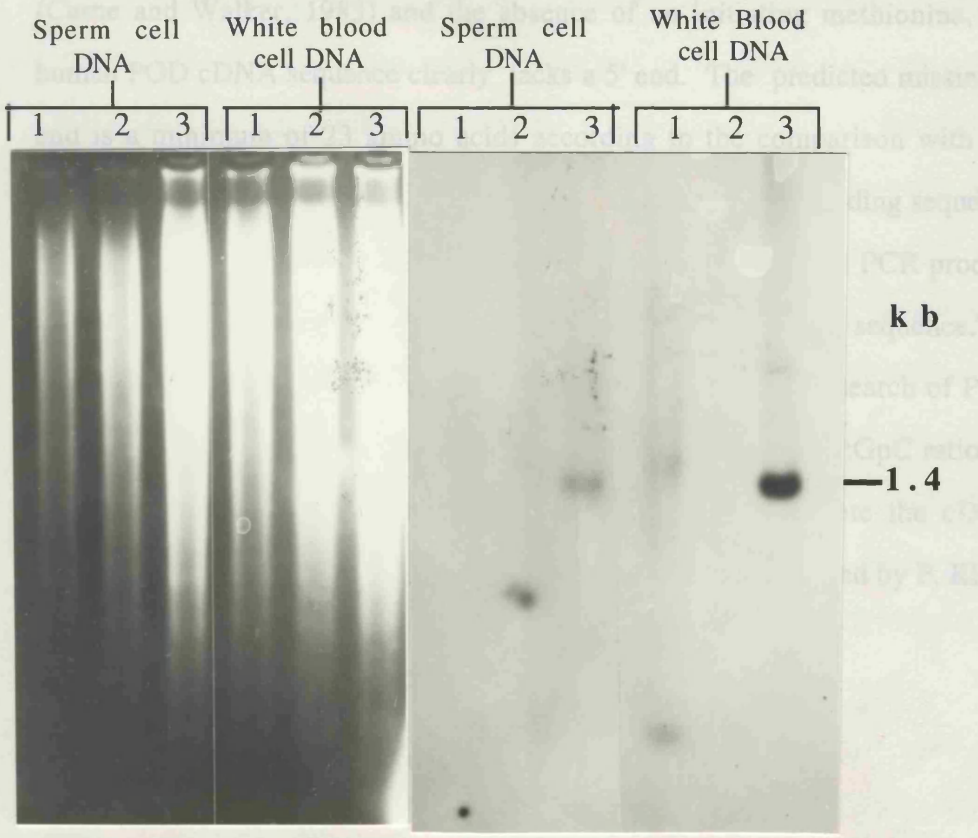


Figure 15. Ethidium Bromide Stained Gel of Germ cell and White Blood Cell DNA Digested with PvuII and Isoschizomers MspI and HpaII and Autoradiograph of Hybridization to the 1.4 kb PvuII Fragment.

Lanes: 1 - PvuII & MspI, 2 - PvuII & HpaII, 3 - PvuII.

This experiment shows that the PGD CpG island within the 1.4 kb PvuII fragment is hypomethylated in germ cell DNA as well as in white blood cells.

The two bands in the MspI/PvuII lane for WBC DNA are most probably products of incomplete digestion resulting from suboptimal reaction conditions.

3.2.2 *PGD* Gene Sequence Determination.

As a result of direct comparison of the translated human PGD cDNA sequence determined by P.Kleyn in 1990, with that of Ovine PGD amino acid sequence (Carne and Walker, 1983) and the absence of an initiating methionine, the human PGD cDNA sequence clearly lacks a 5' end. The predicted missing 5' end is a minimum of 23 amino acids according to the comparison with the Ovine amino acid sequence. In thepresent study, novel 5' PGD coding sequence was obtained by PCR amplification and direct sequencing of the PCR product showed that further homology existed with the Ovine amino acid sequence.The CpG island was also sequenced from the 5' and 3' ends for the search of PGD coding sequence and to allow the calculation of GC% and CpG:GpC ratio. In addition sequencing experiments were carried out to complete the cDNA sequence from PGD cDNA clones isolated and partially sequenced by P. Kleyn in 1990.

3.2.2.1 Isolation of PGD 5' Coding Sequence.

Initial attempts to isolate the missing sequence were based on genomic sequencing of the 5' pPGDH4 reactive clone pPGDE10. It was thought that since a CpG island was located at the 5' end of the gene within the clone pPGDE10 (figure 11) and since the 300 bp PvuII fragment to which the 5' cDNA clone hybridized was next to the island, encompassed by a 1.4 kb PvuII fragment, then sequencing of the region between the 5' end of the cDNA hybridizing fragment (300 bp PvuII fragment) and the island would give this missing 5' coding sequence. The 1.4 kb PvuII fragment contained two BglII restriction sites which gave an 800 bp BglII fragment (figure 16). This 800 bp BglII fragment was cloned into pUC9 for sequencing and denoted Bg0.8.

Figure 16 shows the region sequenced in Bg0.8 and 5' PGD clone pPGDE10 for the original search of PGD coding sequence. As can be seen, the total number of nucleotides sequenced was 684 upstream to the existing coding sequence (324+228+132) and 202 at the 5' end of the clone Bg0.8. The 684 nucleotides at the 3' end form intron one. 324 nucleotides were from the 3' end of Bg0.8 and 360 (228+132) were obtained immediately 3' to the Bg0.8 3' sequence from pPGDE10. This was achieved by designing a sequencing primer from the 3' Bg0.8 sequence obtained, such that sequence derived from the 5' genomic clone pPGDE10 with this Bg0.8 reverse primer (figure 16) would be expected to traverse the BglII and PvuII sites. The sequence obtained crossed both of these restriction sites and went into the known cDNA sequence. Translation of all this genomic sequence (886 nucleotides) and comparison with the Ovine amino acid sequence however failed to show any similarity. This suggested that the centre of the CpG island might contain the missing sequence. Attempts to sequence the Bg0.8 clone further were frustrated by artefacts caused by the very high G+C content of this fragment.

Even though no PGD coding sequence was found in this clone which encompasses the island or outside it at the 3' end, the sequence obtained was analysed so that the GC% could be deduced for the island. The sequence is presented in figure 21 and table 10 shows the GC% and CpG:GpC ratio of the parts sequenced.

The second more successful approach that was used to isolate the missing PGD 5' coding sequence, was the screening of cDNA libraries by PCR. 2 PGD cDNA primers were synthesized which on amplification gave a 200 bp product from the 5' end of the existing 5' cDNA clone pPGDH4. These primers were denoted 5+ and 5amp and their location is shown in figure 18. The role of these primers was in the initial screening of cDNA libraries. The library which was positive for this reaction was later screened with the reverse primer, i.e 5amp and each

of the vector primers so that an extension product could be identified. The sequence of these primers is given in the Appendix, section 5.4.

Since *PGD* is a ubiquitously expressed gene it should be present in any complete cDNA library screened. Those used were derived from the following cells and tissues: small cell lung cancer, normal colon mucosa, reticulocytes, human gut and T cell.

From this screening the T cell library only gave a band of the expected size. The T cell library was then tested again with the 5+ and 5amp primers but also with each of the vector primers (lambda gt10 forward and reverse primers), and the primer 5amp and one lambda primer. An extension product was obtained when using one vector primer and the primer 5amp in the T cell library and as expected, a heterogeneous smear was obtained when the vector primers were used together.

The extension product (denoted lambda PGD extension product) was then excised from the gel, and reamplified using the PGD primers (5amp and 5+) alone to confirm that the forward PGD primer sequence (5+) was within this new extension product. If this were a genuine PGD sequence, a smaller product identical to that obtained on the first screening should have been obtained. This reaction was carried out in duplicate as shown and the smaller fragment was amplified from within the extension product (figure 17).

The extension product obtained was reamplified so that it could be used as a hybridization probe on digests of the 5' genomic clone pPGDE10, the genomic fragment to which the 5' cDNA pPGDH4 hybridizes. This was essential for the following reasons: first to see if it originated from this region i.e from the same genomic fragment pPGDE10, and second, to see if any region upstream to that detected by the most 5' cDNA clone pPGDH4 was detected by this extension

product. The restriction map of pPGDE10 and the location of the regions to which the 5'cDNA pPGDH4 hybridizes is shown in figure 11. By comparing the hybridization pattern between pPGDH4 and the extension product when used as probes on pPGDE10 digests, then it is possible to answer the above questions. As shown (figure 19) in lane one there is hybridization to the 1.2 kb BglII fragment to which pPGDH4 hybridizes (figure 19), in lane two there is hybridization to the 300 bp Pvu II fragment, again a region to which pPGDH4 hybridizes. There is also weak but detectable hybridization to the 800 bp BglII fragment and to the 1.4 kb PvuII fragment. These two fragments do not hybridize to pPGDH4. This hybridization pattern overall confirms that not only is the extension product authentic PGD cDNA sequence, but that the extension seen is of a sequence originating from the genomic region 5' of the region detected by pPGDH4.

Sequencing followed to confirm the origin of this clone through the derivation of sequence that overlaps with the 5' end of pPGDH4 and to identify the missing coding region. As expected, the use of the PCR primers lambda primer and 5amp themselves in the sequencing reactions was unproductive. The use of the internal PGD primer 5+ however, the location of which is shown in figure 18, was found to give a clear sequence with overlap with the 5' end of pPGDH4 from which it was derived. This information further confirmed the nature of this product (figure 20).

Two PGD cDNA primers were then used to give sequence extending into the 5' direction from the extension product, these were named PGD1 and PGD2 and their location is also shown in figures 18 and 20. The use of PGD1 gave novel 5' sequence but PGD2 gave this novel sequence as well as some known pPGDH4 sequence thus showing that this novel sequence is continuous with and precedes pPGDH4. An analysis of the available sequence gave a region of 30 nucleotides which upon translation showed homology to the Ovine PGD

polypeptide sequence at this 5' region (figure 20). Further sequence was vector specific, thus showing that as a result of cloning a short cDNA, only partial novel 5' PGD sequence was obtained.

An extension of only 10 amino acids (30 nucleotides) has been obtained from this clone which ends prematurely. As the Ovine amino acid sequence is 23 amino acid residues longer than the sequence derived from the 5' end of pPGDH4, it suggests that 13 more amino acid residues at least, are still to be found if homology between the Ovine and human PGD sequences extends further 5'.

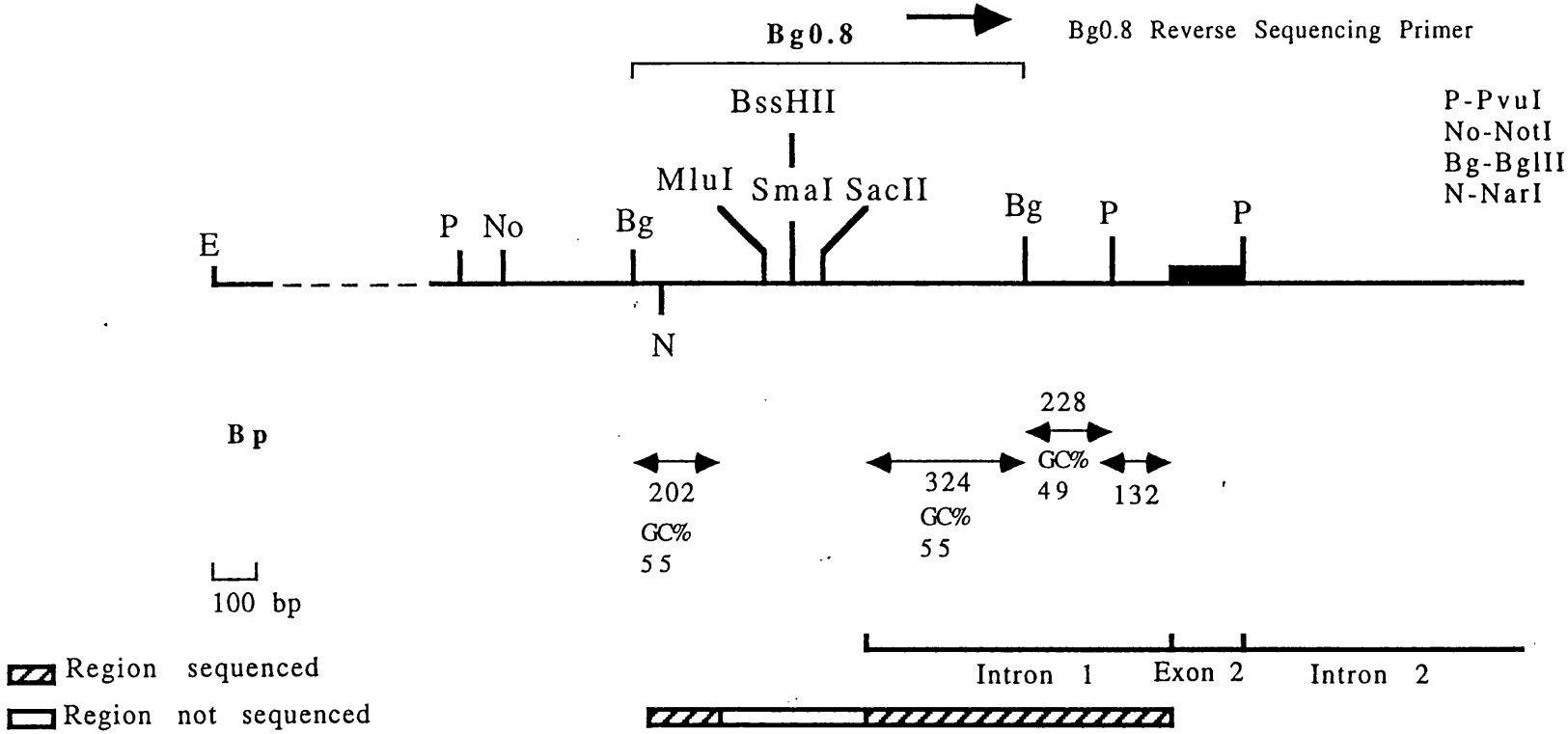


Figure 16. Detailed Restriction Map of the PGD CpG Island.

The above map shows the size of fragments sequenced at the 5' and 3' ends of the CpG island and the position of exon 2 in relation to the island. 274 basepairs not sequenced at the centre of the island are thought to contain remaining 5' PGD coding sequence, an estimated 69 nucleotides. The lambda PGD extension sequence hybridizes within the 800 bp BglII fragment.

The sequenced fragments shown at the right overlap to give a continuous sequence of 684 bp. They are shown as separate fragments simply so that the distance between the relevant endonuclease restriction sites can be shown.

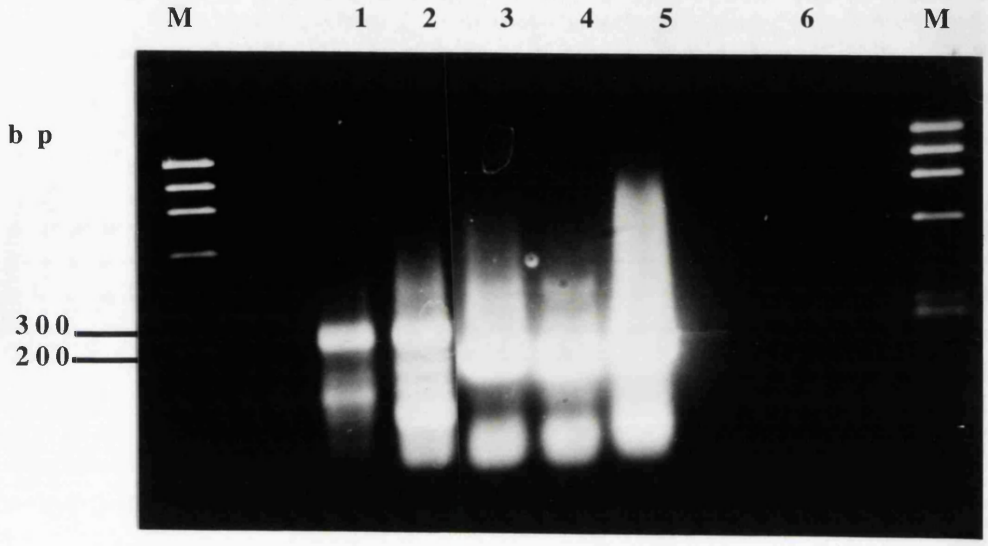
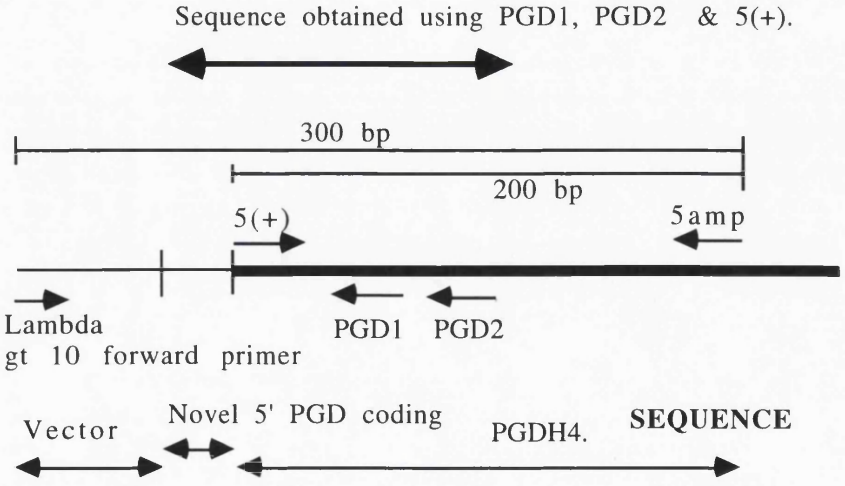


Figure 17. PCR Amplification of PGD Extension Product from a cDNA Library Followed by Semi-Nested PCR.

Lanes: M - Phi X 174 1 kb ladder,
1 & 2 - cDNA library amplified with lambda gt10 forward primer and PGD cDNA primer 5amp, product denoted lambda PGD extension product.
3 & 4 - Lambda PGD extension product amplified with PGD primers 5amp and 5(+). 5(+) sequence is nested inside vector sequence.
5 - pPGDH4 positive control, sequence amplified with PGD primers 5amp and 5(+).
6 - Negative Control

Figure 18. Location of Primers Used in the Isolation and Sequencing of the PGD Extension Product.



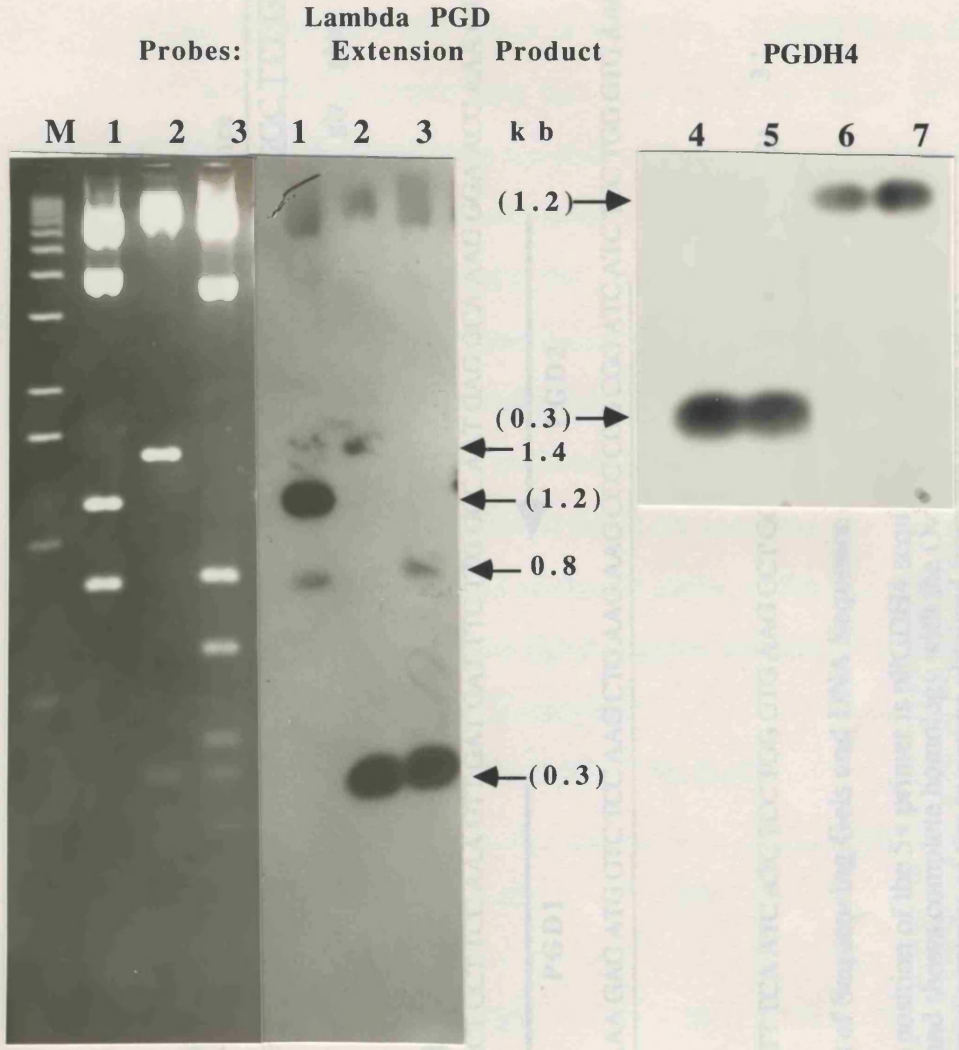


Figure 19. Ethidium Bromide Stained Gel of pPGDE10 Digests and Autoradiographs of Hybridization to pPGDH4 and Lambda PGD Extension Product.

Lanes: M - 1 kb Ladder, 1 - BglII, 2 - PvuII, 3 - PvuII & BglII.

4 & 5 - PvuII, 6 & 7 - BglII.

Hybridization of the 2 probes to the 5' genomic clone pPGDE10 shows that the extension product detects more novel 5' sequences as well as those detected by PGDH4. The fragments detected by both probes are indicated in brackets. These are the 1.2 kb BglII and 0.3 kb PvuII fragments which are adjacent to the CpG island. The fragments containing the island, the 1.4 kb PvuII fragment and that nested within the latter, the 0.8 kb BglII fragment are detected only by the extension product.

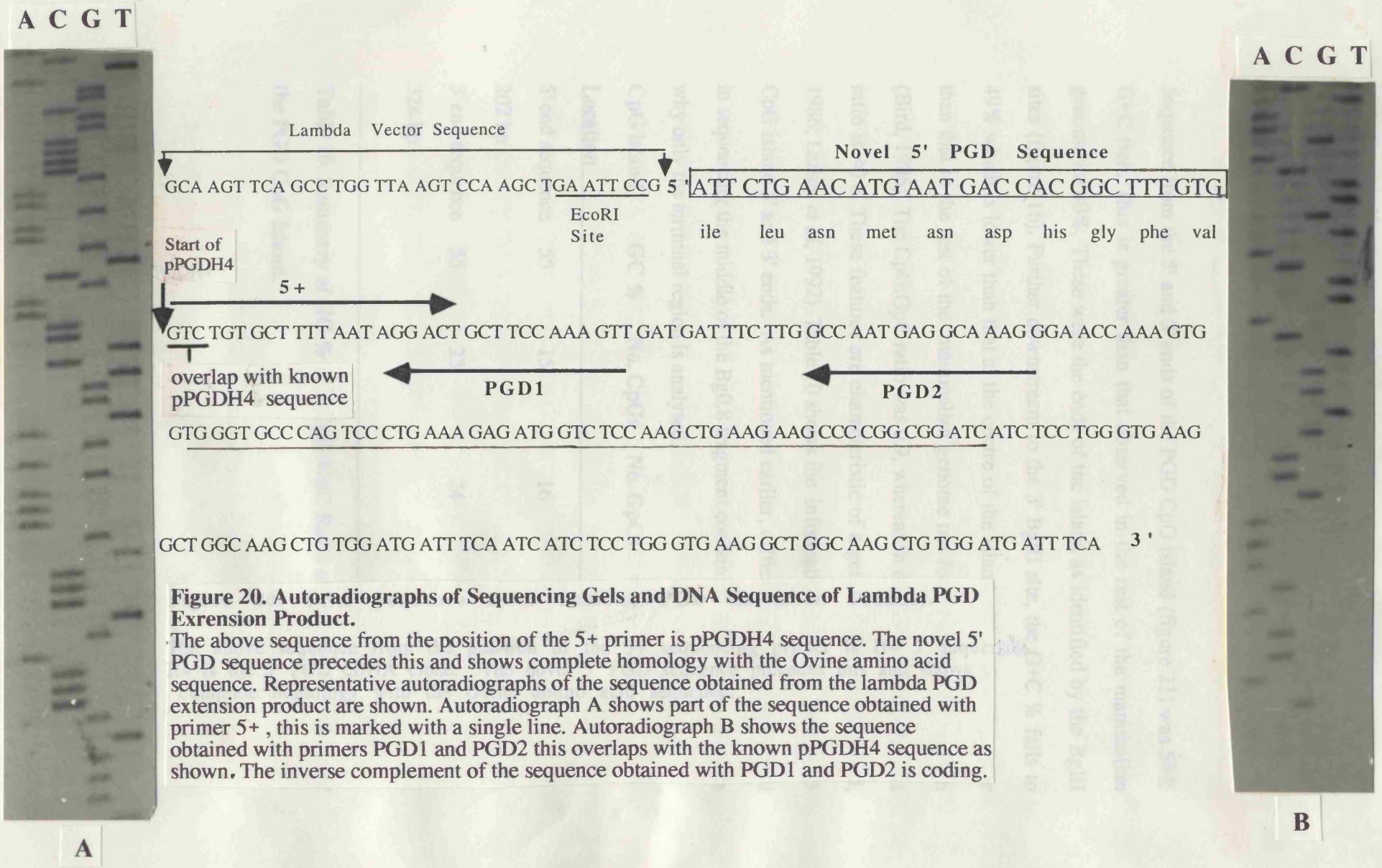


Figure 20. Autoradiographs of Sequencing Gels and DNA Sequence of Lambda PGD Exrension Product.
The above sequence from the position of the 5+ primer is pPGDH4 sequence. The novel 5' PGD sequence precedes this and shows complete homology with the Ovine amino acid sequence. Representative autoradiographs of the sequence obtained from the lambda PGD extension product are shown. Autoradiograph A shows part of the sequence obtained with primer 5+ , this is marked with a single line. Autoradiograph B shows the sequence obtained with primers PGD1 and PGD2 this overlaps with the known pPGDH4 sequence as shown. The inverse complement of the sequence obtained with PGD1 and PGD2 is coding.

3.2.2.2 Analysis of PGD CpG Island 5' and 3' DNA Sequence.

Sequence from the 5' and 3' ends of the PGD CpG island (figure 21) was 55% G+C rich. This is greater than that observed in the rest of the mammalian genome i.e 40%. These were the ends of the island as identified by the BglII sites (figure 16). Further downstream to the 3' BglII site, the G+C % falls to 49% which is lower than that at the centre of the island, although still greater than that in the rest of the mammalian genome not forming the HTF fraction (Bird, 1986). The CpG:GpC ratio was 0.9, whereas in non GC rich regions this ratio is <0.6. These features are characteristic of a genuine CpG island (Bird, 1986; Larsen *et al,* 1992). Table 10 shows the information derived for the PGD CpG island 5' and 3' ends. As mentioned earlier, difficulties were encountered in sequencing the middle of the Bg0.8 fragment containing the island and this is why only the terminal region is analysed.

CpG island Location	GC %	No. CpGs	No. GpCs	CpG: GpC Ratio
5' end sequence 202 bp	55	15	16	0.9
3' end sequence 324 bp	55	23	24	0.9

Table 10. Summary of GC % and CpG:GpC Ratio of the 5' and 3' Ends of the PGD CpG Island.

5' end of PGD CpG island -

GACCATGGTT TCCCGTACCT GGCACACATG CGGGCAGCAG TGAGCGCCAC

ACGCCACACT GTCACACTCG TGTACGCTTA CGCTCACTCA TTAGGCACCC

CAGGCTTTAC ACTTTATGCT TCCGGCGTCG TCTGTTAGCT GTGGTAAGTT

GTCCGAACGC GGCATAAAGC AAGACTCTCT CCAACGTCGA GGGCATACGA

CGAG----------202bp

---- ~274 bp not sequenced in the centre of the CpG island --

GTTGTCTTCT CTCTGTTCCC GGGCGCTTTA CGAGGCCGGC GATAGGTTTG

GAGCTTACGG GTCTCCTGGC CGTGCTTTGC TAATGTGCTC TGTTGCTGCT

CGTGGCATTT TTGTATGGAA AGGAGAAGCA CCCTGTAGGC GTGGGCGGGC

CGATCCCGAA CTTAGTCCTG CGGAGTGTGC CTGTGGGTCC GTGAGGCTTC

ACAGCCCGTA TAATGAGATC GGAGATATAT ACGTCGCTCG TGTGATATGT

TACGCACGAG ATGCAACGGG TGCGGGAGGA GAACCCTTGC TGGCTGTGTA

GAGCTCTAAC ATGATTGCCT GATG**AGATCT** GGGAGAGTGA AATGTTTCTT

CTAAATGTTA AAGTGGCTTA GACGCCAGGA TGATCAATAAC AACAGATAGA

GCCTGTTGAA CCGGCAGTTC CTGGATTTGA TTTTGAGTCC TAACACGTTG

GGTGTGGATT CTACCTGACA CACCGGGGGT AGTTGCCTTC GCCTCGGTTG

TCCGCCGTTT TCATCCTACT GAGGTGACAT AACTTTACAG TGAGCCTCCC

A**CAGCTG**GGG GAAGAAAAGG CAAAGGCAGG TCTGACCTCC CCAGAACTTG

GGTTGAATGG AAAGGTCATT GTTACTTTGG CCACAGTCTG AAAGTCTTGT

GTGTCCTGGA TCTCCTACTC AGGACTTTTG TCCTTCTAG**GTCTGTGCT**1162bp

(right splicing site,acceptor) **Overlap with pPGDH4 Exon two.**

Figure 21. PGD CpG Island 5' and 3' Sequence Leading into Exon Two.

202 bp were sequenced from the 5' end and 324 bp from the 3' end . 274 bp from the centre of the island were not sequenced. Bgl II site: AGATCT and PvuII site: CAGCTG shown in bold.

3.2.2.3 Sequence Determination of PGD cDNA and Exon - Intron Structure of PGD Genomic Clones pPGDE10 and pPGDE4.

The PGD cDNA sequence of clones pPGDH4 (800 bp), pPGDH3 (918 bp) and pPGDH1(606 bp) was completed in this project. 600 bp from the 3' and 5' ends of pPGDH4 were previously sequenced by P.Kleyn (1990), as well as 660 from ends of pPGDH3. In the present study, 201 bp were therefore sequenced from the middle of pPGDH4, 240 from the middle of pPGDH3 and all of pPGDH1 (606 bp). This indicated that as previously suggested by P.Kleyn the 3' cDNA clone pPGDH1 which shares a common 3' end with pPGDH3 is mostly 3' untranslated DNA. 357 were non coding and only 249 were coding from this clone. In pPGDH3, 561 bp were coding and as in the shorter overlapping clone pPGDH1 357 bp were the 3' untranslated region. No polyadenylation signal or sequence was found here and no 5' untranslated sequence was obtained in the experiments designed to isolate further 5' cDNA sequence (section 3.2.2.1). The PGD cDNA sequence is shown in figure 22 along with that of the additional novel 5' sequence.

A comparison of the amino acid sequence obtained on translation of the PGD cDNA sequence with that of the Ovine amino acid sequence shows that there is an overall homology at the amino acid level of 83%, i.e out of a total of 455 amino acids there are 77 differences.

Having completed the sequence of the existing PGD cDNAs, the next step was to work out how much of the 5' cDNA clone pPGDH4 was actually present in the pPGDH4-reactive genomic clones pPGDE10 and pPGDE4 and simultaneously to derive the exon-intron structure of these contiguous genomic PGD clones.

Four exons were found in these two PGD genomic clones pPGDE10 and pPGDE4. pPGDE10 is 10.5 kb and pPGDE4 is 4.5 kb and figure 11 shows the regions detected by pPGDH4. Since three of these regions are at the extreme ends of the fragments, pAT153 vector primers were synthesized and used for this sequencing. The number of exons within the clones was deduced by direct comparison of the sequence obtained here with that from the 5' cDNA pPGDH4. Two exons were found in pPGDE4, one at the 5' end and one at the 3' end. In pPGDE10, one exon was found at the 3' end and one at +360 bp from the end of the CpG island i.e from the BglII restriction site which marks the 3' end of the island. The size of introns was also deduced in the process. This information on exon-intron structure is presented in table11 and illustrated in figure 23 and the different exons are marked in figure 22.

All exon-intron junctions had the consensus sequence which is important in the splicing of introns from mRNA. Exons were preceded by the splice acceptor sequence CAG and followed by the sequence GTAAGT. The percentage of junctions at which the specified base (or type of base) is found is shown in superscripts. GT.....AG is always found as can be seen from these percentages (Mount, 1982).

The consensus sequence for comparison is shown below.

EXON**I**_____INTRON_____**I**EXON

START END

$A^{64}\ G^{73}G^{100}T^{100}A^{62}A^{68}G^{84}T^{63}$ .....$6Py^{74\text{-}88}\ N\ C^{65}\ A^{100}\ G^{100}\ N$

Exon two was not sequenced, as this was rendered unnecessary by the sequence information of the region preceding the 300 bp PvuII fragment in pPGDE10 which contained this exon, as well as knowledge of the beginning of exon 3 in the cDNA pPGDH4 itself.

First the genomic sequencing of pPGDE10 carried out to locate the remaining 5' PGD coding sequence showed that the 5' end of the clone pPGDH4 is a natural 5' splice acceptor and that this must be the beginning of at least exon two since further 5' coding sequence was still missing. The start of this pPGDH4 reactive sequence in the clone pPGDE10 is 360 bp away from the CpG island as identified by the 3' BglII site (figure 16). Second, the knowledge of the location of exon three in pPGDH4 along with the fact that the 5' end is a natural 5' splice acceptor and that the 300 bp PvuII fragment was the most 5' fragment in pPGDE10 to which pPGDH4 hybridized allowed the size and precise location of exon two to be easily deduced. In brief, exon two contained the remaining 179 bp from the 5' end of pPGDH4 since a certain proportion of pPGDH4 sequence following these 179 bp (67 bp) formed exon three present at the 3' end of pPGDE10.

The reasons why this 300 bp PvuII fragment was thought to contain a single exon even though it had not been sequenced are as follows: first this 300 bp PvuII fragment shown in figure 16 has been partially sequenced. This showed that 132 non coding base pairs precede the coding sequence and a splice acceptor site is present at this junction. As the remaining coding sequence prior to exon three in pPGDH4 is 179 bp and as this can only lie within the approximately 300 bp Pvu II fragment itself, the fact that 132 bp are non coding means that only one exon of 179 bp can actually exist within this fragment. There is no space for another intron.

Finally, in summary, 439 bp of pPGDH4 are distributed within four exons in genomic clones pPGDE10 and pPGDE4. This allows us to work out the distribution of the remaining coding sequence in the EcoRI genomic fragments. PGDH4 is 800 bp and so this means that 361 bp of PGDH4 sequence must lie in the third, 6 kb EcoRI fragment , E6, previously shown to hybridize to pPGDH4 in genomic digests in section 3.2.1. The clone pPGDH1 detects a 3.5 kb EcoR1

fragment denoted E3 (kleyn, 1990). Since this clone and pPGDH3 overlap (Kleyn, 1990) then pPGDH3 must also hybridize to E3. If the whole of .the pPGDH3 sequence is contained within this 3.5 kb EcoR1 fragment then the whole genomic region encompassing the *PGD* gene would be 18.5 kb (i.e. 4.5 kb from pPGDE10+ pPGDE4 (4.5 kb) + E6 (6kb) + E3 (3.5 kb)) If, however, pPGDH3 detects additional EcoRI fragments then the region would be greater than 18.5 kb. This arrangement is summarized in figure 24.

PGD-5' Start of H4 5' ---->

----------EXON ONE----------> <----------

ATT CTG AAC ATG AAT GAC CAC GGC TTT GTG GTC TGT GCT TTT 42

I L N M N D H G F V V C A F

----------EXON TWO----------

AAT AGG ACT GTC TCC AAA GTT GAT GAT TTC TTG GCC AAT GAG 84

N R T V S K V D D F L A N E

GCA AAG GGA ACC AAA GTG GTG GGT GCC CAG TCC CTG AAA GAG 126

A K G T K V V G A Q S L K E

ATG GTC TCC AAG CTG AAG AAG CCC CGG CGG ATC ATC TCC TGG 168

M V S K L K K P R R I I S W

---><----------

GTG AAG GCT GGC AAG CTG TGG ATG ATT TCA TCG AGA AAT TGG 210

V K A G K L W M I S S R N W

----------EXON THREE----------

TAC CAT TGT TGG ATA CCT GGT GAC ATC ATC ATT GAC GGA GGA 252

Y H C W I P G D I I I D G G

----------> <----------

AAT TCT GAA TAT AGG GAC ACC ACA AGA CGG TGC CGA GAC CTC 294

N S E Y R D T T R R C R D L

----------EXON FOUR----------

AAG GCC AAG GGA ATT TTA TTT GTA GGG ACG GAT GCA TGT GGT 336

K A K G I L F V G T D A C G

GGA GAG AAT GGC ACC CTG GTA GGC CCA TCG CTC ATG CCA GGA 378

G E N G I L V G P S L M P G

---><----------

GGG AAC AAA GAA GCG TGG CCC CAC ATC AAG ACC ATC TTC CAA 420

G N K E A W P H I K T I F Q

----------EXON FIVE----------

GGC ATT GCT GCA CAA AGG TGG GGA ACT GGA GAA CCC TGC TGT 462

G I A A Q K W G T G E P C C

Figure 22. PGD cDNA and Amino Acid Sequences.(continued overleaf)

--------><u>I</u>

<u>GAC TGG G</u>TG GGA GAT GAG GGA GCA GGC CAC TTC GTG AAG ATG **504**

D W V G D E G A G H F V K M

GTG CAC AAC GGG ATA GAG TAT GGC GAC ATG CAG CTG ATC TGT **546**

V H N G I E Y G D M Q L I C

GAG GCA TAC CAC CTG ATG AAA GAC GTG CTG GCA TGG CGC AGA **588**

E A Y H L M K D V L A W R R

CGA GAT GCC CAG GCC TTT GAG GAT TGG AAT AAG ACA GAG CTA **630**

R D A Q A F E D W N K T E L

GAC TCA TTC CTG ATT GAA ATC ACA GCC AAT ATT CTC AAG TTC **672**

D S F L I E I T A N I L K F

CAA GAC ACC GAT GGC AAA CAC CTG CTG CCA AAG ATC AGG CGA **714**

Q D T D G K H L L P K I R R

GCT GCG GGG GCT AAG AAG GGC ACA GGG AAG TGG ACC GCC ATC **756**

A A G A K K G T G K W T A I

TCC GCC CTG GAA TAC GGC GTA CCC GTC ACC CTC ATT GGA GAA **798**

S A L E Y G V P V T L I G E

<u>end of H4-3'I</u>

GCT GTC TTT GCT CGG TGC TTA TCA TCT CTG AAG GAT GAG AGA **840**

A V F A R C L S S L K D E R

<u>start of H3-5'</u>>

ATT CAA GCT AGC AAA AAG CTG AAG GGT CCC CAG AAG TTC CAG **882**

I Q A S K K L K G P Q K F Q

TTT GAT GGT GAT AAG AAA TCA TTC CTG GAG GAC ATT CGG AAG **924**

F D G D K K S F L E D I R K

GCA CTC TAC GCT TCC AAG ATC ATC TCT TAC GCT CAA GGC TTT **966**

A L Y A S K I I S Y A Q G F

Figure 22. PGD cDNA and Amino Acid Sequences. (continued overleaf)

ATG CTG CTA AGG CAG GCA GCC ACC GAG TTT GGC TGG ACT CTC **1008**
M L L R Q A A T E F G W T L

AAT TAT GGT GGC ATC GCC TTG ATG TGG AGA GGG GGC TGC ATC **1050**
N Y G G I A L M W R G G C I

ATT AGA AGT GTA TTC CTA GGA AAG ATA AAG GAT GCA TTT GAT **1092**
I R S V F L G K I K D A F D

CGA AAC CCG GAA CTT CAG AAC CTC CTA CTG GAC GAC TTC TTT **1134**
R N P E L Q N L L L D D F F

Start of H1-5' >

AAG TCA GCT GTT GAA AAC TGC CAG GAC TCC TGG CGG CGG GCA **1176**
K S A V E N C Q D S W R R A

GTC AGC ACT GGG GTC CAG GCT GGC ATT CCC ATG CCC TGT TTT **1218**
V S T G V Q A G I P M P C F

ACC ACT GCC CTC TCC TTC TAT GAC GGG TAC AGA CAT GAG ATG **1260**
T T A L S F Y D G Y R H E M

CTT CCA GCC AGC CTC ATC CAG GCT CAG CGG GAT TAC TTC GGG **1302**
L P A S L I Q A Q R D Y F G

GCT CAC ACC TAT GAA CTC TTG GCC AAA CCA GGG CAG TTT ATC **1344**
A H T Y E L L A K P G Q F I

CAC ACC AAC TGG ACA GGC CAT GGT GGC ACC GTG TCA TCC TCG **1386**
H T N W T G H G G T V S S S

TCA TAC AAT GCC TGA TCA TGC TGC TCC TGT CAC CCT CCA CGA **1428**
S Y N A *

TCC CAC AGA CCA GGA CAT TCC ATG GCC TCA TGC ACT GCC ACC **1470**
TGC CCT TGC CCT ATT TTC TGT TCA GTT TTT TAA AAG TGT TGT **1512**

Figure 22. PGD cDNA and Amino Acid Sequences. (Continued Overleaf)

AAG AGA CTC CTG AGG AAG ACA CAC AGT TTA TTT GTA AAG TAG**1554**

CTC TGT GAG AGC CAC CAT GCC CTC TGC CCT TGC CTC TTG GGA **1596**

CTG ACC AGG AGC TGC TCA TGT GCG TGA GAG TGG GAA CCA TCT **1638**

CTT GCG GCA GGT TGG CTT CCG CGT GCC CCG TGT GCT GGT GCG **1680**

GTT CCC ATC ACG CAG ACA GGA AGG GTG TTT GCG CAC TCT GAT **1722**

CAA CTG GAA CCT CTG TAT CAT GCG GCT GAA TTC H3-3'/H1-3' **1755**

Figure 22. PGD cDNA and Amino Acid Sequences.

Individual exons are indicated above the sequence in broken lines. The total cDNA length is 1755 bp, 30 nucleotides are from the sequence obtained from the T cell library clone i.e the lambda PGD extension product. The remaining sequence is from cDNA clones PGDH4, the 5' cDNA and from the 3' cDNA clones pPGDH1 and pPGDH3. The open reading frame ends at position 1398 in H3.

Single Letter Amino Acid Designations.

Alanine - A	Arginine - R	Asparagine - N	Aspartic Acid - D
Cysteine - C	Cystine - C	Glycine - G	Glutamic Acid - E
Glutamine - Q	Histidine - H	Isoleucine - I	Leucine - L
Lysine - K	Methionine - M	Phenylalanine - F	Proline - P
Serine - S	Threonine - T	Tryptophan - W	Tyrosine - Y

Valine - V.

Exon/Intron no.	Exon/Intron size, bp.	Genomic Fragment
Intron	=/> 690	pPGDE10
Exon Two	**179**	**pPGDE10**
Intron	3686	pPGDE10
Exon Three	**67**	**pPGDE10**
Intron	789	pPGDE10-pPGDE4
Exon Four	**119**	**pPGDE4**
Intron	3786	pPGDE4
Exon five	**74**	**pPGDE4**
Intron	226	pPGDE4

Table 11. A Summary of the Size of Exons and Introns in PGD Genomic Clones pPGDE10 and pPGDE4.

439 bp of cDNA sequence is present in four exons in genomic clones pPGDE10 and pPGDE4. These clones are shown in figure 24 with the exons and introns marked. The clones pPGDE10 and pPGDE4 are contiguous as shown in figure 24.

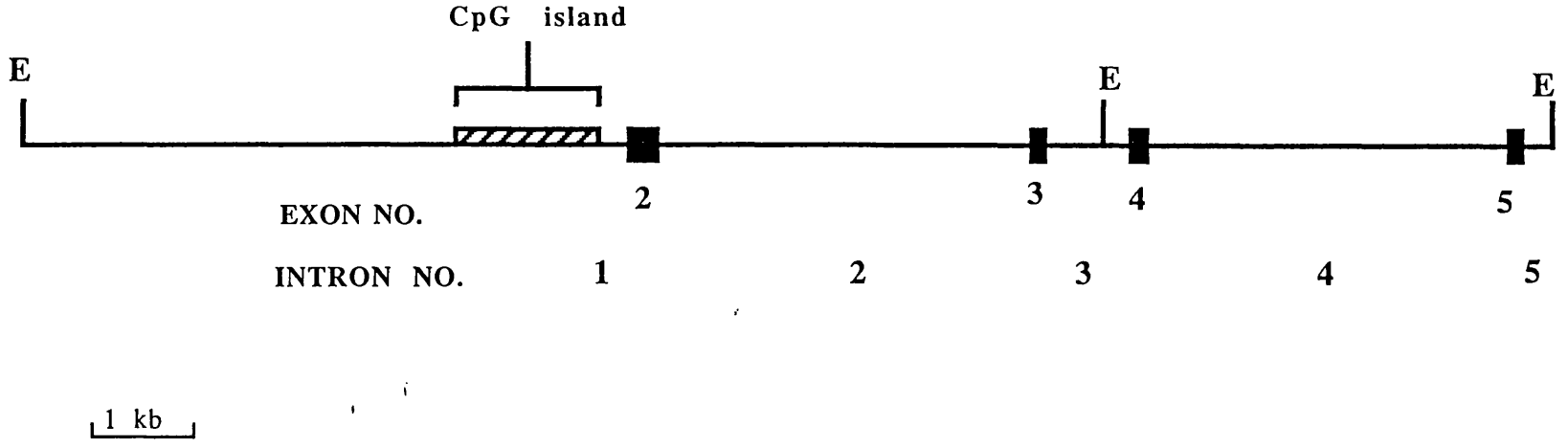


Figure 23. Exon- Intron Structure of PGD Genomic Clones pPGDE10 and pPGDE4.

The exon- intron structure of the PGD clones pPGDE10 and pPGDE4 determined by sequencing the restriction fragments which hybridized to the 5' PGD cDNA pPGDH4. Exons 2 to 5 were 179, 67, 119 and 74 bp respectively. Evidence so far suggests that exon 1 lies in the centre or the CpG island.

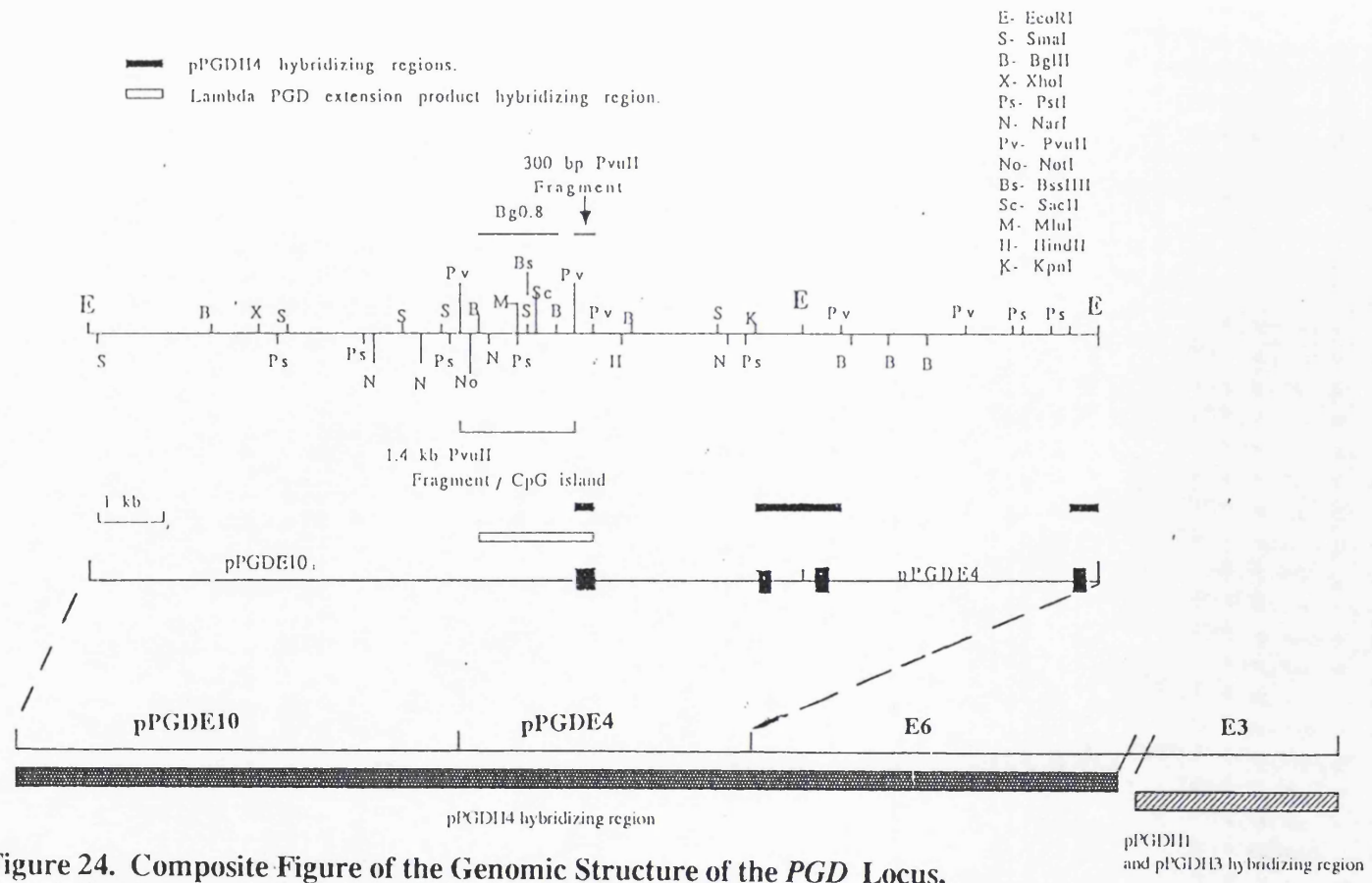


Figure 24. Composite Figure of the Genomic Structure of the *PGD* Locus, Restriction Map and Exon - Intron Organization of Genomic Clones pPGDE10 and pPGDE4. The restriction map of pPGDE10 and pPGDE4 is shown with the location of the PGD CpG island and the location of the pPGDH4 cDNA hybridizing regions, exons two-five. Exon one is presumed to lie in the centre of the PGD CpG island. The only other chromosome 1-specific EcoRI fragment detected by pPGDH4 is denoted E6 and this must be contiguous with pPGDE4. The 3' cDNA clone pPGDH1 detects a 3.5 kb EcoRI fragment denoted E3 which must also be detected by pPGDH3 this arrangement is shown. It is not known whether additional EcoRI fragments separate E6 and E3.The data on pPGDH1 is obtained from Kleyn,1990.

3.3. Physical Mapping.

Physical mapping studies were carried out using somatic cell hybrids, fluorescence *in situ* hybridization and pulsed field gel electrophoresis. The markers used in these studies were already known to be in band 1p36 and so an attempt was made to refine their location with respect to each other through various mapping techniques. In addition, a YAC was isolated by PCR from a total genomic library for one of these markers and forms a nucleation point for future contig construction in distal 1p.

3.3.1 Somatic Cell Hybrid Analysis.

The location of four markers was determined with respect to chromosome 1 translocation breakpoints and their presence in a hybrid with only a small fragment of 1p (Hybrid Dis2.6) and other chromosomal fragments was sought. These hybrids were analysed by Southern hybridization and by the polymerase chain reaction. The loci D1S77, *PND* and D1F15S1 were mapped by hybridization and *HSPG2 (heparan sulphate proteoglycan 2)* by PCR. *PND's* location was confirmed by PCR as well. The hybrids employed had breakpoints in 1p34 or 1p36.13 These loci have been assigned to 1p previously as shown in table 12. The intention of the following studies was to refine their location and obtain an order.

LOCUS	PROBE	LOCATION	METHOD	REFERENCES
PND	M13.Mp.10	1p36	SCH ISH	Yang-Feng *et al*, 1985
D1F15S1	Ch4-H3	1p36	ISH	Harper and Saunders, 1981.
D1S77	pMCT58	1p	Linkage Analysis	O'Connell *et al*, 1989
PGD	pPGDH1	1p36	ISH SCH	Kleyn ,1990 Carritt *et al*, 1982a
p58	p58cosmid	1p36	FISH SCH PCR*	Eipers *et al*, 1991
HSPG2	HSPG2cDNA	1p36.1-1p35	ISH SCH	Kallunki *et al*, 1991; Dodge *et al*, 1991.
D1Z2	P1-79	1p36.3	ISH Linkage analysis	Buroker *et al*,1987 O'Connell *et al*, 1989

Table 12: Previous Assignment of 1p36 Markers and Methods Used.
FISH - Fluorescence *In Situ* Hybridization.
ISH - *In Situ* Hybridization with Radiolabelled Probes.
PCR -Nested PCR On Microdissected Chromosome 1.
SCH- Somatic Cell Hybrid Analysis.

3.3.1.1 Southern Analysis of Somatic Cell Hybrids.

The probe pJA119 was used to detect the *PND* locus, where there is a Taq1 polymorphism with variant bands at 2.2 and 2.8 kb (Nemer *et al*, 1986). The D1S77 locus was detected by the probe pMCT58, where there is a PvuII VNTR polymorphism with 5 variant bands between 1.3-1.6 Kb (Nakamura *et al*, 1988). D1F15S1 was detected after HindIII digestion and hybridization to the probe pH3H2. This locus is a polymorphic repetitive sequence which originated as a single copy sequence at 3p21. Transposition of this sequence (20 kb) occurred to chromosome 1p36, after the divergence of man from the great apes. Duplication and amplification followed at this locus. The probe pH3H2 detects both chromosomal locations on 1 and 3 according to somatic cell analysis, with the intensity of signals at the chromosome 1 locus being consistent with six to eight copies of the sequence at 1p36. The 1p36 characteristic bands are 3.8 kb and 8kb, the 3p21 band being 2kb (Carritt *et al*, 1986; Goode *et al*, 1986; Welch *et al*, 1989). It can be seen from figure 25 that the Taxi 2.6 hybrid with both chromosomes 1 and 3 has all three bands.

It can be seen from figures 26 and 27, that both D1S77 and *PND* are present in Taxi 2.6 and HCH 5 but absent from HCH 7 and are therefore distal to both the 1p34 and 1p36.13 breakpoints. D1F15S1 lies proximal to the 1p36.13 breakpoint but distal to the 1p34 breakpoint as it is present in HCH 7 and Taxi 2.6 but absent from HCH 5 (figure 25). Thus the latter is proximal to the other two loci.

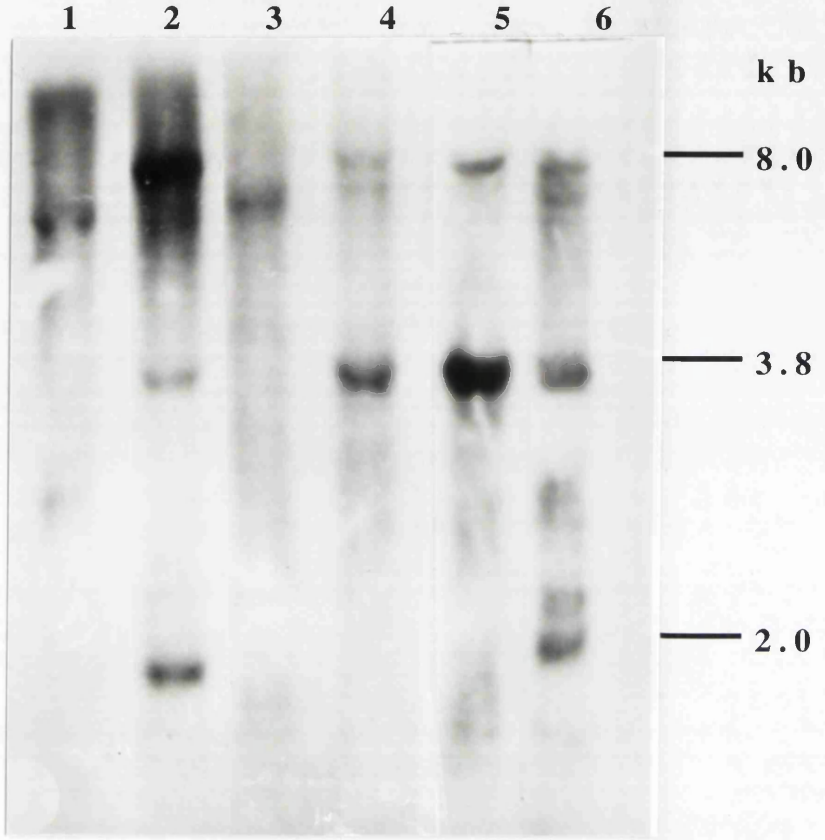


Figure 25. Screening 1p Hybrids for the D1F15S1 Locus by Southern Analysis.

Lanes:1 - a23 Chinese Hamster DNA, 2 - Total Human DNA, 3 - Hybrid HCH5,(1pter - 1p36.13), 4 & 5 - Hybrid HCH7, (1p36.13 - 1qter), 6 - Hybrid Taxi2.6 (1pter - 1p34).

DNA was digested with HindIII and hybridized to the probe H3H2. The 3.8 kb and 8 kb fragments are from chromosome 1, the 2 kb fragment is from chromosome 3 (Carritt *et al,* 1986). Hybrids containing chromosome 1 specific fragments indicate that the D1F15S1 locus is located proximal to the 1p36.13 breakpoint. A hamster- specific band is present in all hybrids. Chromosome content of hybrids taken from Carritt *et al,* 1982 a.

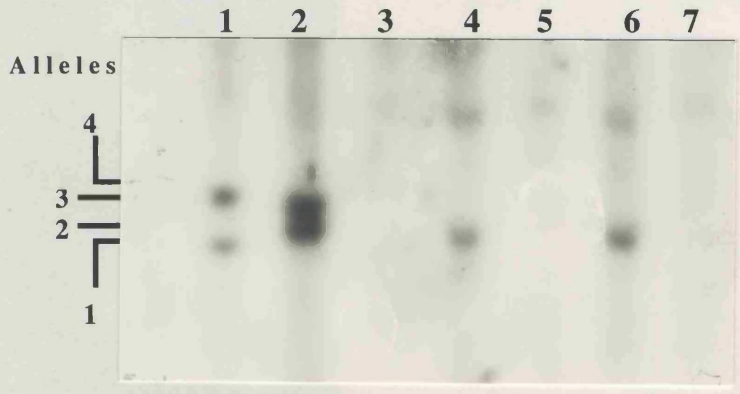


Figure 26. Screening 1p Hybrids for the D1S77 Locus by Southern Analysis.

Lanes: 1 & 2 - Total human DNA, 3 - a23- Chinese Hamster DNA, 4 - Hybrid HCH5, (1pter - 1p36.13), 5 - Hybrid HCH7, (1p36.13 - 1qter), 6 - Hybrid Taxi2.6, (1pter - 1p34), 7 - Hybrid Taxi3 (1p34 - 1qter).

DNA was digested with PvuII, electrophoresed through a 1.4% agarose gel and hybridized to the probe MCT58. This detects a VNTR polymorphism with 5 different alleles ranging in size from 1.3-1.6 kb. The two human controls in lanes 1 and 2 are genotypes 4-1 and 3-2 respectively. The two hybrids containing the D1S77 locus both have the allele denoted 2. This information places the D1S77 locus distal to the 1p36.13 breakpoint as it is present in HCH5 and absent from HCH7.
Chromosome content of hybrids is taken from Carritt *et al*, 1982 a.

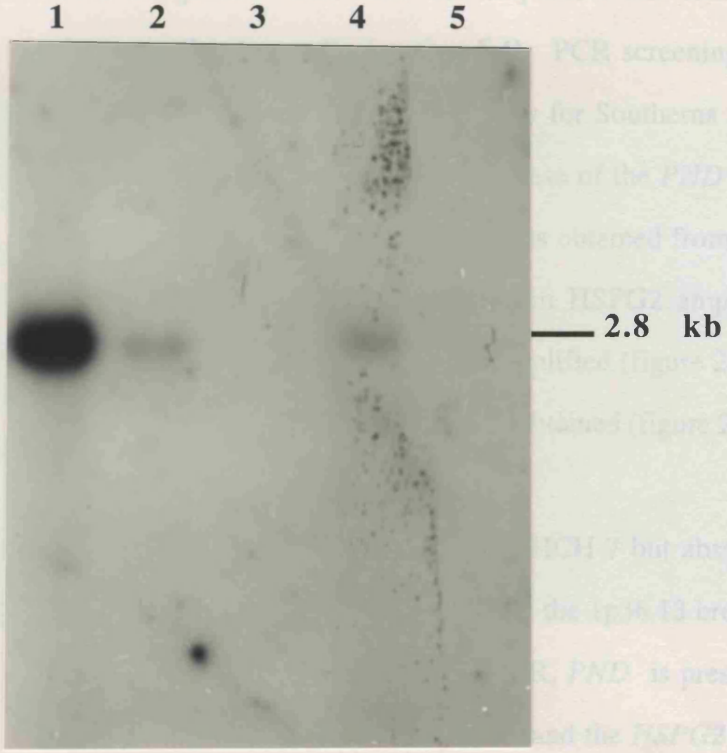


Figure 27. Screening 1p Hybrids for the PND Locus by Southern Analysis.

Lanes: 1 - Total human DNA, 2 - Hybrid HCH5, (1pter - 1p36.13), 3 - Hybrid HCH7, (1p36.13 - 1qter), 4 - Hybrid Taxi2.6, (1pter -1p34) 5 - a23 - Hamster DNA.

DNA was digested with TaqI and hybridized to the probe JA119 which detects a diallelic polymorphism. The bands are 2.8 and 2.2 kb. Above the human control is homozygous for the 2.8 kb allele. Hybrids HCH5 and Taxi2.6 are positive for this typing indicating that PND is distal to 1p36.13..

3.3.1.2 PCR Analysis of Somatic Cell Hybrids.

Primers were designed for the *HSPG2* and *PND* loci using sequence information from the genome data base. The sequences and reaction conditions used are shown in the Appendix (section 5.4). PCR screening was undertaken as this required less DNA than that necessary for Southerns and in addition the result was obtained more quickly. In the case of the *PND* locus, PCR screening of hybrids allowed confirmation of results obtained from Southern analysis of the same hybrids. The PCR product from HSPG2 amplification is 720bp. A hamster specific band of 760 bp is amplified (figure 28). The PND amplicon is 171bp in size and no hamster band is obtained (figure 29).

HSPG2 is present in TAXI 2.6 and HCH 7 but absent from TAXI 3, 8 and HCH 5 and thus is located proximal to the 1p36.13 breakpoint. As determined both by Southern analysis and by PCR, *PND* is present in HCH 5 and TAXI 2.6, and is therefore distal to 1p36.13 and the *HSPG2* locus. The results from Southern and PCR analysis are presented in table 13.

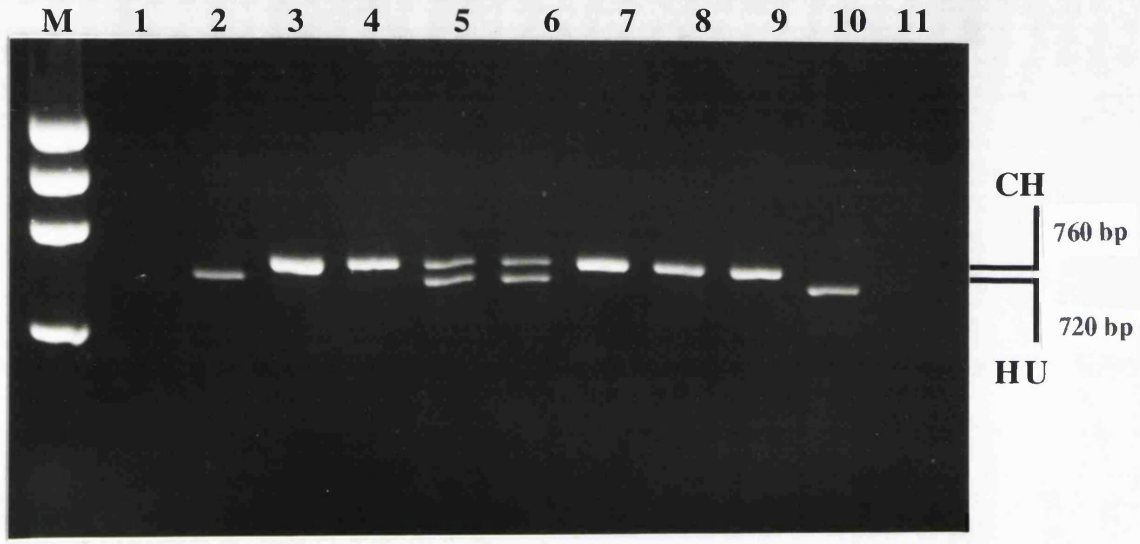


Figure 28. Screening 1p Hybrids for the HSPG2 Locus by PCR.

Lanes: M - Phi X174 1kb marker, 1 - Mouse DNA, 2 - Total Human DNA 3 - Chinese Hamster DNA, 4 - Hybrid HCH5, (1pter -1p36.13) 5 - Hybrid HCH7, (1p36.13 -1qter), 6 - Hybrid Taxi2.6, (1pter - 1p34),7 - Hybrid Taxi3, (1p34 - 1qter), 8 - Hybrid Taxi8, (1p34 -1qter) 9 - Hybrid Dis2.6, (see text) 10 - Hybrid F4, (1p), 11 - Negative Control. Hybrid F4 is a mouse - human hybrid.

CH - Chinese Hamster specific band , Hu - Human specific band. The human specific amplicon is 720 bp and the hamster specific band is 760 bp . This information indicates that the HSPG2 locus is proximal to 1p36.13. Chromosome content of hybrids is taken from Carritt *et al,* 1982 a.

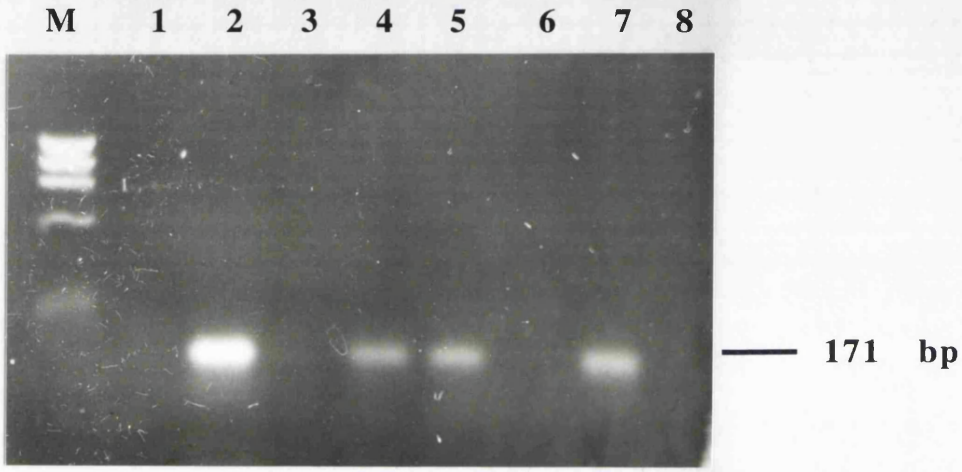


Figure 29. Screening 1p Hybrids for the PND Locus by PCR.

Lanes:

M - Phi X 174 1kb ladder, 1 - a23, Chinese Hamster DNA,
2 - Total Human DNA, 3 - Mouse DNA, 4 - Hybrid F4 (1p),
5 - Hybrid HCH5 (1pter-1p36.13), 6 - Hybrid HCH7 (1p36.13-1qter)
7 - Hybrid Taxi2.6 (1pter-1p34), 8 - Negative Control.
F4 is a mouse-human hybrid.

This information indicates that the PND locus is distal to 1p36.13.

HYBRID	Region of 1p Present	LOCUS NAME D1S77	*PND*	*HSPG2*	D1F15S1
HCH5	1pter-1p36.13	+	+	-	-
HCH7	1p36.13-1qter	-	-	+	+
TAXI2.6	1pter-1p34	+	+	+	+
TAXI3	1p34-1qter	-	NT	-	NT
TAXI8	1p34-1qter	NT	NT	-	NT
F4	1p	NT	+	+	NT
Dis2.6	~9Mb around *PGD* locus	-	-	-	-

Table 13. Markers Present in 1p Hybrids.

Analysis was carried out by PCR and Southern Hybridization.

Marker present (+), Marker absent (-).

Origin of Hybrids.

Human Cell Type Human Donor Karyotype.

HCH- Peripheral Lymphocytes, 46,XX,t(1;10) (p36.13; q22.1)

Taxi- Fibroblast Line 46,XX,t(X; 1) inv(X) (p21.07p11.06; p34.00)

(reference- Carritt et al, 1982a).

3.3.2 Ordering 1p Markers by Fluorescence *In Situ* Hybridization.

3.3.2.1 Ordering 1p Markers by Fractional Length Measurements.

In an attempt to order a set of 1p loci, FISH was used on metaphase spreads of one normal male. The loci analysed were *PGD*, D1Z2, *HSPG2*, *p58*, D1F15S1 and *PND*. Some of these were mapped with respect to translocation breakpoints in section 3.3.1 and so the aim here was to see if the order obtained at that lower level of resolution could be confirmed but also refined. Plasmid clones were already available for *PGD* and D1Z2 and a cosmid clone was available for the D1F15S1 locus. Cosmids were therefore isolated for the loci *HSPG2, p58* and *PND* from a genomic library. PCR products were used to isolate cognate cosmids for *HSPG2* and *p58*, and the probe pJA119 (Nemer *et al*, 1986) was used to isolate the PND cosmid. Cosmids were necessary as smaller clones were not as successful as larger clones in these FISH experiments. The PGD genomic clone used, pPGDE10 had a 10.5 kb insert which gave a strong signal. This was used with competition as the clone contains repetitive DNA sequences. The D1Z2 plasmid clone, p1-79 had a 900 bp insert which detects a large repeat sequence in 1p36 and so this alone gave a strong signal. The D1S77 probes used failed to give any signals in these experiments. Analysis was carried out on a confocal microscope with software allowing the enhancement and manipulation of images. For each probe the best metaphase spreads were scrutinized, the number varied between 9-24. Images were selected and stored and analysis began with enlargement of each pair of chromosomes to facilitate subsequent measurements. Usually four signals were detected corresponding to the four chromatids. Figures 30a-f show examples of the six biotinylated markers hybridized to metaphase chromosome spreads.

In each case measurement began at 1pter, passing through the signal and continuing through to 1qter. This was displayed on the monitor as a graph with pixel intensity on the y axis and distance along the chromosome on the x axis in microns. From this, one could see the signal intensity along the chromosome. In addition a total length measurement of the chromosome is obtained and the distance at which the signal peak occurs from 1pter. The fractional length is then calculated by dividing the distance from 1pter to the signal over the total length of the chromosome and is denoted as flpter. The mean and standard deviation were then calculated and tabulated (Table 14). From the mean results the following order is deduced from telomere to centromere:

TEL - D1Z2 - p58 - PGD - PND - D1F15S1 - HSPG2 - CEN

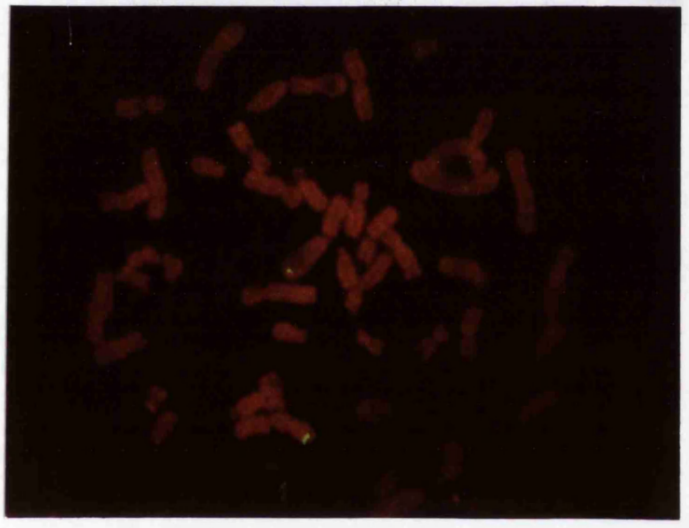
The standard deviation from the mean however, suggests that the position of these loci as determined by fractional length measurements can actually fluctuate greatly with the only confident marker assignment being that for the most probable distal locus D1Z2. If we compare this order with that from other methods we see that there is some degree of support for this order.

First, analysis of the available somatic cell hybrids with breakpoints in 1p34 and 1p36.13 place *PGD* and *PND* distal to both breakpoints, whereas D1F15S1 and *HSPG2* are both distal to the 1p34 breakpoint but proximal to the 1p36 breakpoint (see section 3.3.1). This latter orientation supports location of the proximal pair (D1F15S1 and *HSPG2*) with respect to the distal pair (*PGD* and *PND*). Second, segregation analysis in pedigrees places D1F15S1 proximal to both *PND* and *PGD* (see section 3.1.3)

Figure 30a.

Locus
D1F15S1

Probe
H3H2 cosmid.



Locus
PND

Probe
PND cosmid

Figure 30b.

Figure 30. Localization of 1p Probes by FISH on Metaphase Chromosome Spreads.

Probes were labelled with Biotin and amplified with FITC (Fluorescein isothiocyanate) conjugated avidin. Chromosomes were counterstained with DAPI and propidium iodide present in the anti fade medium. The position of the probe on the chromosome was estimated by measuring the distance from 1pter to the signal, and the total length of the chromosome. The distance from 1pter to the signal was then divided by total length and expressed as fractional length (flpter) in microns. Several images were analysed for the 6 probes and the mean value was used to order these from 1pter.

Figure 30c.

Locus
PGD

Probe
pPGDE10

Figure 30d.

Locus
D1Z2

Probe
p1-79

Figure 30 continued.

Figure 30e.

Locus
p58

Probe
p58 cosmid.

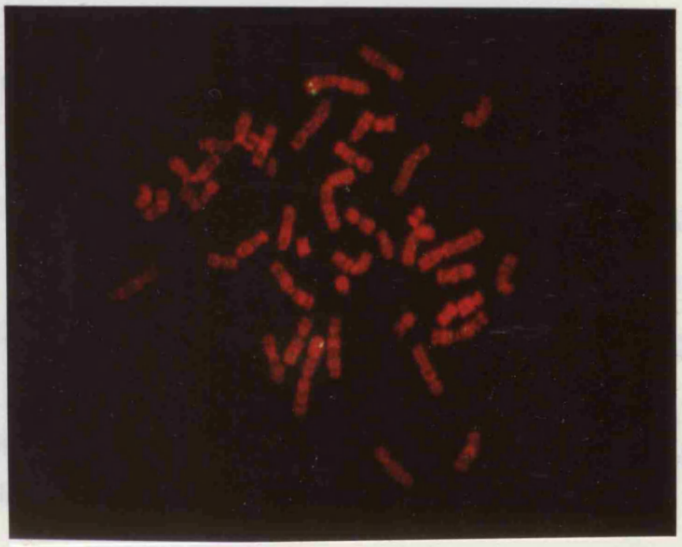


Figure 30f.

Locus
HSPG2

Probe
HSPG2 cosmid.

Figure 30 continued.

LOCUS NAME	PROBE NAME	MEAN (flpter) (Micrometers)	STANDARD DEVIATION (±)	S.E.M (±)	n
HSPG2	C.HSPG2	0.100	0.01	$3X10^{-4}$	23
D1F15S1	C.H3H2	O.095	0.04	$5X10^{-4}$	15
PND	C.JA119	0.059	0.02	$1X10^{-4}$	14
PGD	pE10	0.045	0.02	$4X10^{-5}$	9
p58	C.p58	0.026	0.02	$5X10^{-4}$	9
D1Z2	p1-79	0.016	$6X10^{-3}$	$8X10^{-6}$	24

Table 14 Order of 1p Markers from Measurement of Signal Location after FISH Experiments.

Probes were biotinylated cosmids and plasmids hybridized to metaphase chromosome spreads. Probes are listed in their established order.
n - number of chromosomes measured.
S.E.M - standard error of the mean.
flpter - fractional length measurement.

The order deduced from mean flpter values is :

TEL - D1Z2 - p58 - PGD - PND - D1F15S1 - HSPG2 - CEN

3.3.2.2 Ordering 1p Markers by Dual Colour Fluorescence *In Situ* Hybridization.

This set of experiments was carried out in collaboration with Dr. M. Fox. Dual colour fluorescence in general confirmed the order obtained for the markers above. The result from dual FISH with *HSPG2* and D1F15S1 however, places D1F15S1 proximal to *HSPG2*. This is not at all unexpected since the flpter difference was only 0.005 micrometers. All other orders were confirmed and are tabulated (Table 15). Of particular significance here, is the order of the loci *PGD* and *PND*, which is the same as that favoured by multipoint linkage analysis. In addition the proximal location of D1F15S1 with respect to *PGD* and *PND* is in agreement with the order inferred from segregation analysis of these loci in the CEPH pedigrees (see section 3.1.3). Figures 31a-g show examples of pairs of markers hybridized to prometaphase chromosome spreads.

A summary of the somatic cell hybrid and FISH mapping data is presented in figure 32. This indicates that the order of markers determined by FISH is in agreement with the division of two groups of markers in the hybrids.

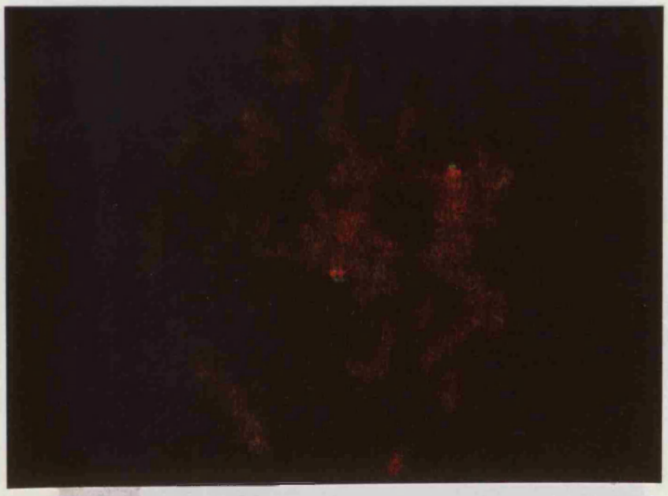


Figure 31a.

Probe/ Locus	Label
c. p58/ p58	Biotin
cJA119/ PND	Digoxigenin

Order 1pter → cen

p58 – PND

Probe/ Locus	Label
c.p58/ p58	Biotin
c.H3H2/ D1F15S1	Digoxigenin

Order 1pter → cen

p58 – D1F15S1

Figure 31b.

Figure 31. Ordering 1p Markers by Dual Colour Fluorescence *In Situ* Hybridization.

Prometaphase chromosome spreads were used to order 1p markers labelled with biotin and digoxigenin. Biotinylated probes were amplified with FITC conjugated avidin and probes labelled with digoxigenin were amplified with TRITC conjugated sheep anti-digoxigenin antibody. Chromosomes were counterstained with DAPI only.

Probe/ Locus	Label
pPGDE10 PGD	Biotin
cH3H2/ D1F15S1	Digoxigenin

Order 1pter ▶ cen

PGD - D1F15S1

Figure 31c.

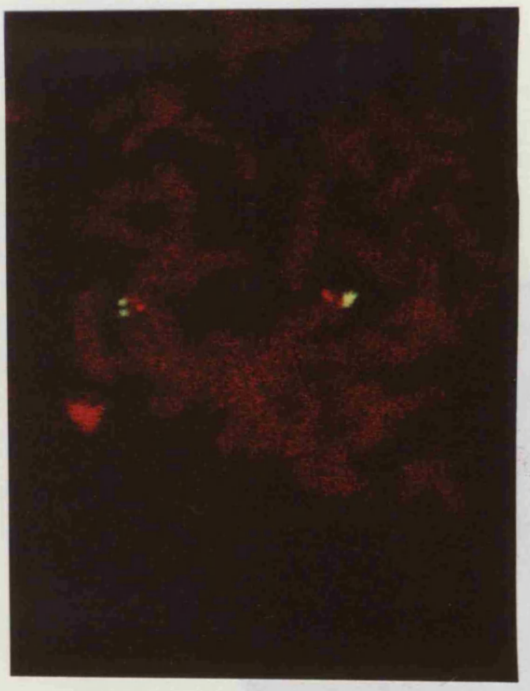
Probe/ Locus	Label
pPGDE10 PGD	Biotin
c.JA119/ PND	Digoxigenin

Order 1pter ▶ cen

PGD - PND

Figure 31d.

Figure 31 continued.

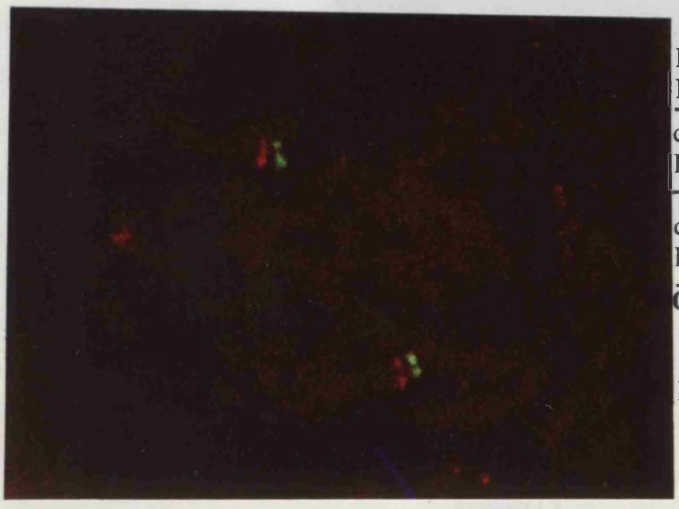


Probe/ Locus	Label
cJA119/ PND	Digoxigenin
P1-79/ D1Z2	Biotin

Order 1pter → cen

D1Z2 – PND

Figure 31e.



Probe/ Locus	Label
cH3H2/ D1F15S1	Digoxigenin
cHSPG2/ HSPG2	Biotin

Order 1pter → cen

HSPG2 – D1F15S1

Figure 31f.

In figure 31f the telomere is not clearly visible in the photograph but on visual inspection on the monitor the telomere was visible.

Figure 31 continued.

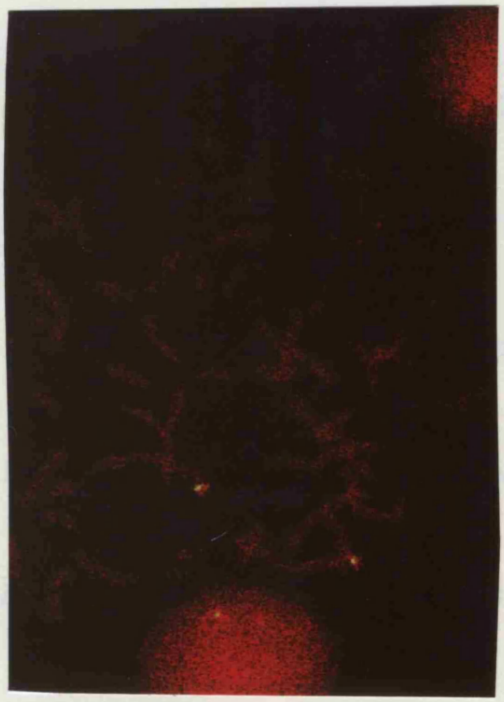


Figure 31g.

Probe/ Locus	Label
cH3H2/ D1F15S1	Biotin
cJA119/ PND	Digoxigenin

Order 1pter→cen

PND – D1F15S1

Figure 31 continued.

Locus 1/Probe Biotin.	Locus 2/ Probe Digoxigenin.	n	Locus Order: 1pter-cen
PGD/pPGDE10	D1F15S1/c.H3H2	10	*PGD*- D1F15S1
PGD/pPGDE10	*PND*/cJA119	10	*PGD-PND*
D1F15S1/c.H3H2	*PND*/cJA119	10	*PND*-D1F15S1
HSPG2/cHSPG2	D1F15S1/c.H3H2	10	*HSPG2*-D1F15S1
PND/c.JA119	D1Z2/P1-79	10	D1Z2-PND
p58/c.p58	*PND*/c.JA119	10	*p58-PND*
p58/c.p58	D1F15S1/c.H3H2.	10	*p58*-D1F15S1

Table 15. Order of Pairs of Loci Mapped by Dual Colour FISH on Prometaphase Chromosome Spreads.

The order determined using the information from dual colour fluorescence and flpter measurements is as follows:
TEL- D1Z2 - p58 - PGD - PND - HSPG2 - D1F15S1 - CEN
The difference to that from flpter measurements alone, is the inversion of the two proximal loci *HSPG2* and D1F15S1.

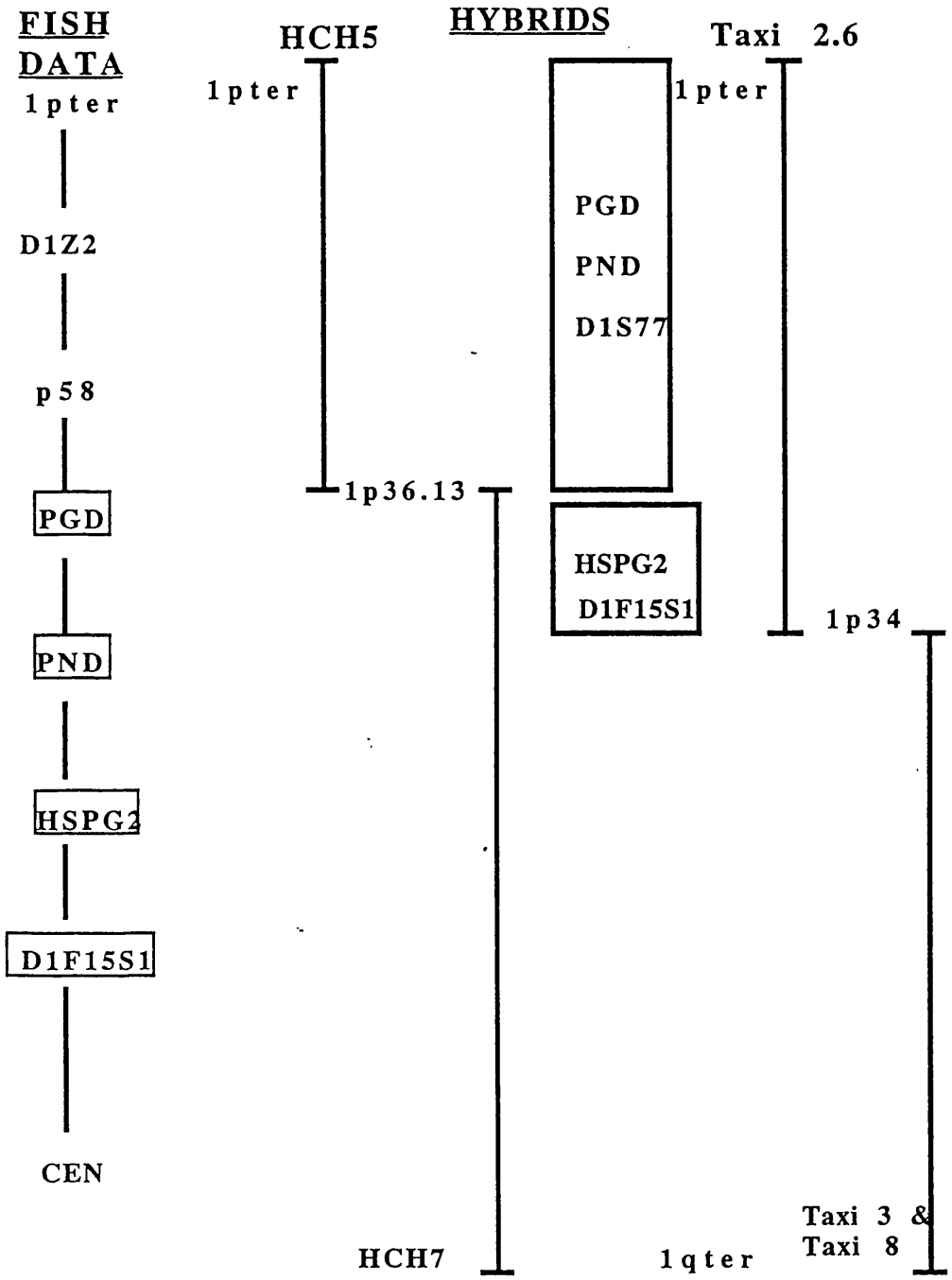


Figure 32. Mapping 1p36 Markers by Somatic Cell Hybrid Analysis and Fluorescence *In Situ* Hybridization.

All markers were mapped with respect to translocation breakpoints in the hybrids by Southern analysis and/ or by PCR. The PGD locus was mapped in these hybrids previously (Carritt *et al,* 1982a, Kleyn, 1990,). FISH experiments gave the order of markers by dual labelling and by measurement of signals from the 1p terminus.

3.3.3 Long Range Mapping by Pulsed Field Gel Electrophoresis.

Long range mapping by pulsed field gel electrophoresis was carried out using some of the markers already analysed by FISH and somatic cell hybrids, as well as two others previously mapped to 1p36. The aim here was to investigate the physical distance separating them and this was done through sequential hybridization of probes to PFGE filters. Long range maps were constructed for some of these loci and an idea of the minimum distance between various markers was obtained.

3.3.3.1.1 Long Range Mapping Around The *PGD* locus.

The presence of a CpG island at the 5' end of the *PGD* gene described in section 3.2 and shown in figures 11 & 16, facilitated the construction of a long range map around the *PGD* locus. Enzymes known to cleave genomic DNA in the *PGD* gene were used in this first, then others were also employed to expand the map and to look for physical linkage with other markers by sequential hybridization with other probes. Tables 16 and 17 list the size of restriction fragments detected by PGD and other 1p36 probes.

Due to the existence of at least two pseudogenes detected by the PGD cDNA clone pPGDH4 it was necessary when constructing the pulsed field map to employ locus specific reagents . These were the 1.4 kb PvuII fragment referred to earlier and a somatic cell hybrid (HCH 5). The probe pPGDH4 was also used in these experiments.

The enzymes cleaving the 5' PGD genomic clone pPGDE10 and genomic DNA around the *PGD* locus were BssHII, SacII, NotI, MluI, NarI and SmaI. The presence of these sites within a CpG island, the presence of a probe immediately

3' to the island pPGDH4 and a 1.4 kb PvuII fragment which spans the island, allowed restriction mapping bidirectionally from this island. Cleavage of high molecular weight DNA with these enzymes alone or in double digests allowed the construction of a 1.8 Mb map around the region (figure 36). The enzymes BssHII and SacII produce chromosome 1-specific fragments of 50 kb as these occur in both the chromosome 1-specific hybrid and genomic DNA. 200 kb fragments which were also detected appeared to be pseudogene fragments. The 50 kb fragments begin in the island and end downstream since they hybridize to pPGDH4.

In NruI digests a doublet exists in the 680-750 kb range and in MluI digests a 360 kb single fragment is detected. The BssHII site is nested within the NarI, NruI and the MluI fragments as the BssHII product is constantly obtained in double digests of BssHII and any of these enzymes. Double digests with NruI and MluI produce two fragments, one is the MluI specific fragment at 360 kb and the second is a new fragment at 550 kb. Since the NruI sites must lie outside the 360 kb fragment produced by MluI cleavage it follows that derivation of a 550 kb fragment is by partial digestion with MluI 3' to the island (the site in the CpG island would be expected to be permanently hypomethylated *in vivo*) and cleavage with NruI further 3' to the MluI site appearing at 360 kb from the island. This suggests that the NruI site is 550 kb downstream to the island and that if the 680 kb fragment is chromosome 1 specific an NruI site must lie 130 kb upstream to the island. If the 750 kb fragment is of chromosome 1 origin, then the site is 200 kb upstream.

In NarI digests three fragments are detected by pPGDH4. The smallest of these is the more intense and remains after double digests with NotI and MluI, whereas the larger NotI and MluI fragments are lost. This suggests that the NarI site is also nested within these two fragments.

Two NotI fragments are detected by pPGDH4. In double digests with MluI and NotI the MluI 360 kb fragment is lost whilst both the upper and lower NotI fragments remain (figure 33). Since it is certain that the MluI fragment is chromosome 1 - specific it is reasonable to assume that the smaller of these NotI fragments is nested in the MluI fragment, that is, as with others, its 5' end is in the CpG island and its 3' end is internal to the 360 kb MluI fragment. To test this a hybrid was used (HCH5) which had previously been shown to contain only chromosome 1-specific PGD sequences (Kleyn, 1990). It was anticipated that only one of these fragments would be detected by pPGDH4 in this hybrid, namely the smaller of these. However, it is apparent that both fragments are also in the hybrid (figure 33). This is unexpected since this hybrid does not contain genomic sequences detected by pPGDH4 other than those in the 1pter-1p36.13 region (Kleyn,1990). This raises the possibility that there may be a second *PGD*-like locus on chromosome 1. To identify the fragment corresponding to the expressed locus, the 1.4 kb PvuII fragment was used on the same filter. It is seen here (figure 33) that the lower band corresponds to the expressed locus.

3.3.3.1.2 Methylation Differences in Cell Line and Peripheral White Blood Cell DNA.

When using peripheral white blood (PWB) cell DNA for map construction it was found that certain PGD fragments differed in size to the ones detected in the corresponding lymphoid cell line MGL8B2 (Table 16). An apparent hypomethylation in the PWB DNA relative to the cell line was seen. Specifically, this was a property of the NotI and MluI sites in the vicinity of the *PGD* locus. A direct comparison was therefore carried out on the same gel (figure 34). As can be seen, two MluI fragments are detected by pPGDH4 . Both are detected by the 1.4 kb PvuII fragment suggesting that they are both from the expressed locus and that one is a partial. Second, a smaller NotI fragment is obtained (figure 34) and this is of chromosome 1 origin according

to hybridization to the 1.4 kb PvuII fragment. This is very surprising since NotI sites are on average 1 Mb apart and the finding that this produces one of the smallest fragments in the region is unexpected.

A second possibility other than a methylation difference from cell line and PWB DNA, is a polymorphism resulting from variation in methylation pattern or a restriction site polymorphism caused by a point mutation. To investigate this possibility, three different peripheral white blood cell DNA samples were analysed with the same enzymes and probes. As can be seen (figure 35) interindividual concordance in hybridization pattern suggests that a methylation difference between cell line and PWB DNA actually exists at MluI and NotI sites. Other sites were however, constant in both sources. The similarity in hybridization pattern after cleavage with other enzymes of cell line and PWB DNA allowed the extension of the map using PWB DNA.

A PvuI site 100 kb downstream to the island is present in both cell line and PWB cell DNA. In single digests with PvuI a band is detected at about 535 kb and one at about the 1 Mb level. In double digests with PvuI and the two enzymes known to cleave in the CpG island, NotI and MluI, a 100 kb product hybridizes to pPGDH4 and the 1.4 kb PvuII fragment. From the map we can deduce that NotI and PvuI sites are coincident 3' to the island at 100 kb. In MluI and PvuI double digests the same 100kb band is obtained with a loss of both MluI specific products, thus the PvuI site does occur 100 kb downstream from the CpG island. From this we can also position the first PvuI site 5' to the island at 435 kb.

In addition the 1.4 kb PvuII fragment detects an 800 kb MluI fragment not detected by pPGDH4 which means that a restriction site has been located 800 kb upstream of the island. In total this provides a large scale restriction map of 1.8 Mb around the *PGD* locus.

In an effort to detect physical linkage between *PGD* and markers known to lie in the same chromosomal band, sequential hybridization was carried out with other probes. The loci investigated were *ENO1*, D1S77, *p58, PND, HSPG2.* None of these appeared to cohybridize to a convincing set of fragments to suggest that they were both located within such fragments (Table 17). It is possible that different enzymes and/or partial digest experiments may have detected linkage, it is nevertheless clear, that at least 915 kb separate *PGD* from the above loci.

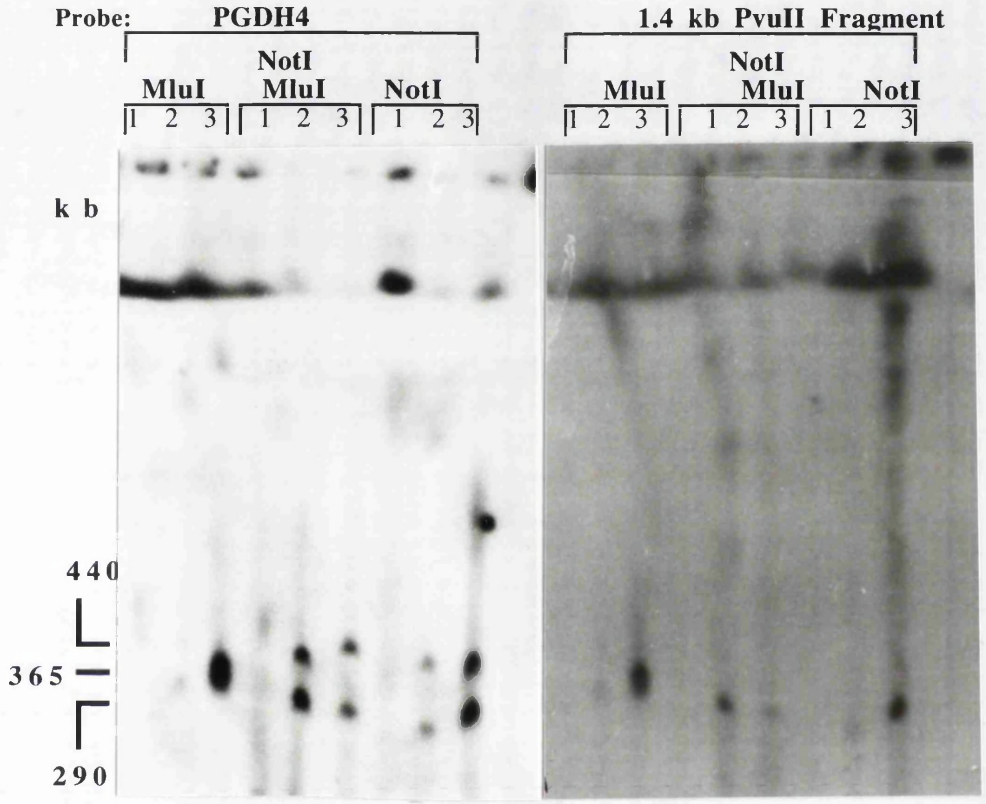


Figure 33. Autoradiographs of a PFG Blot Sequentially Hybridized to pPGDH4 and the 1.4 kb PvuII Fragment.

Lanes: 1 - a23 Chinese Hamster DNA, 2 - Hybrid HCH5
3 - Lymphoblastoid Cell Line MGL8B2.

PFG electrophoresis was carried out using a 90 second pulse time for 40 hrs. These autoradiographs show that the NotI fragment corresponding to the expressed PGD locus is the smaller 290 kb fragment as this is the one detected by the locus specific 1.4 kb PvuII fragment. The second NotI fragment may be from a PGD-like sequence on chromosome 1 as this is present in the hybrid HCH5.

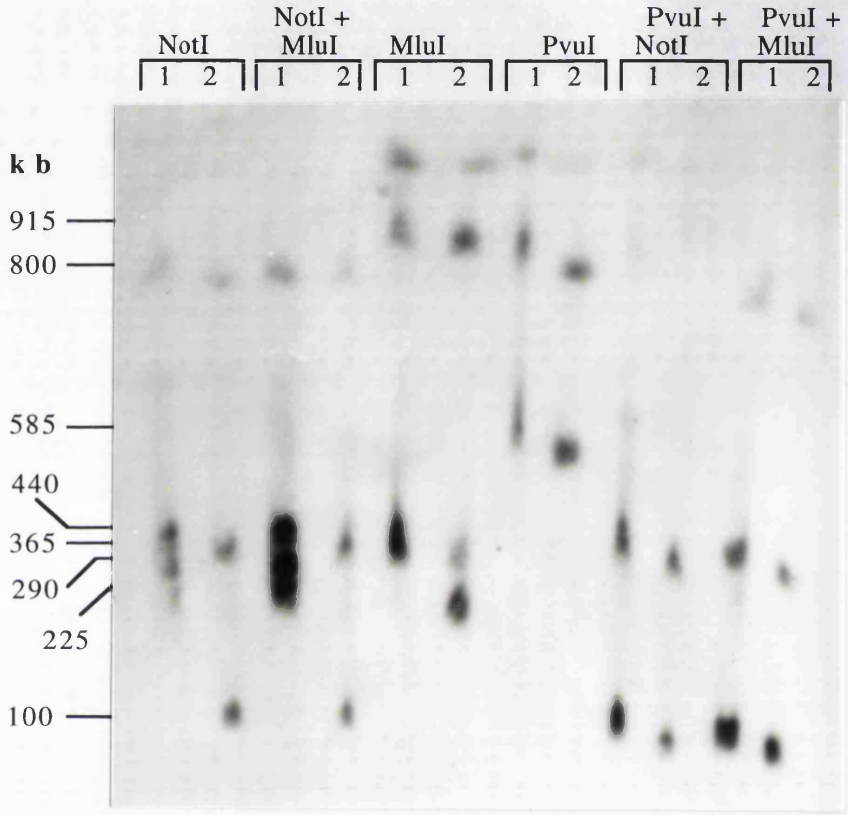


Figure 34. Autoradiograph of PFG Blot Hybridized to the PGD cDNA pPGDH4 Showing the Methylation Difference in Cell Line and PWB Cell DNA.

Lanes: 1 - Lymphoblastoid Cell Line DNA, MGL8B2,
2 - Peripheral White Blood Cell DNA.

PFG electrophoresis was carried out using a ramped programme starting at 20 seconds and going up to 90 seconds over 12 hrs, then going back down to 20 seconds over the next 12 hrs. The experiment here shows the methylation difference in PWB cell DNA and the corresponding cell line DNA. This involves NotI and MluI sites around the PGD locus. It is apparent that PWB cell DNA is hypomethylated in comparison to cell line DNA. Other restriction sites are the same in both DNA sources. The small variation in migration of fragments in PvuI digests is most probably the result of variation in DNA concentration in the plugs digested.

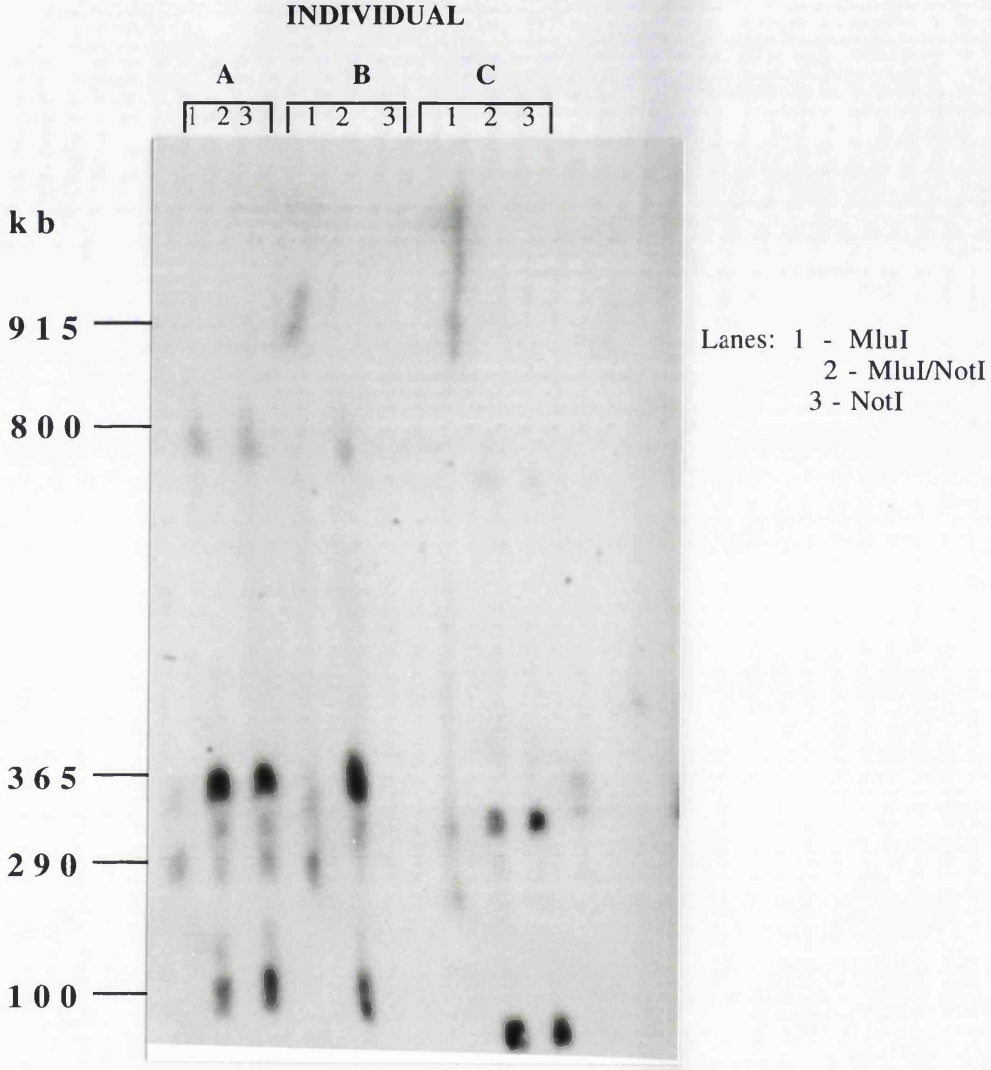
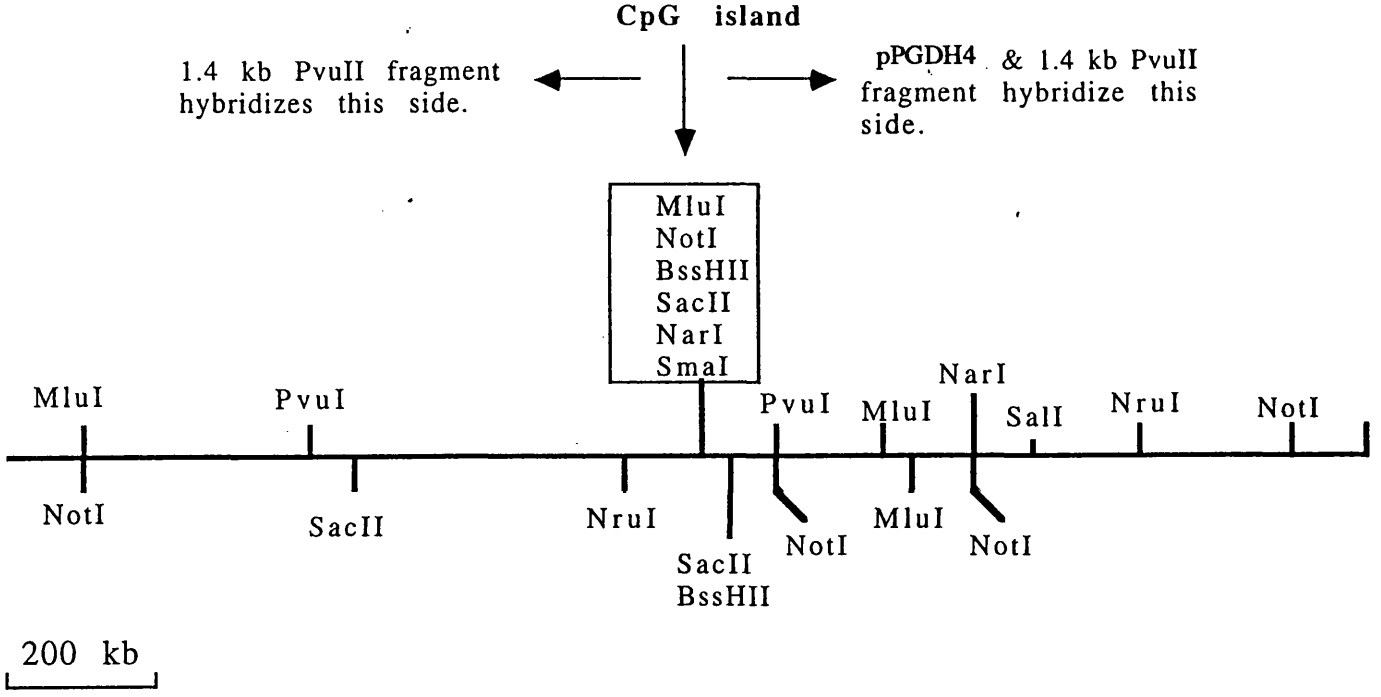


Figure 35. Autoradiograph of a PFG Blot Hybridized to the PGD cDNA pPGDH4 Showing the Invariable Methylation Pattern in Different Individuals.

PFG electrophoresis was carried out using a ramped programme starting at 20 seconds and going up to 120 seconds over 18 hrs, maintening a constant pulse time of 120 seconds for 30 hrs, followed by an 18 hr phase whereby the pulse time was increased to 400 seconds. This experiment shows that in 3 different individuals the same hybridization pattern is obtained for NotI and MluI digests when hybridized to the PGD probe H4. PWB cell DNA was used from each individual.

Figure 36. Long Range Map Around the PGD Locus.

The map around the locus spans 1.8 Mb. The smallest fragments detected by pPGDH4 are 50 kb SacII and BssHII restriction products. The CpG island is shown as a cluster of methylation sensitive restriction enzyme sites. The map was constructed using both peripheral white blood cell DNA and lymphoblastoid cell line DNA.



Size of Fragments Detected by PGDH4/PV1.4 on Pulsed Field Gels. (Kb)

Enzyme	Lymphoid Line (MGL8B2) & Chromosome-1 specific Hybrid (HCH5)*	PWB Cells.
MluI	365, 915	290, 365, (800), 915
NotI	250, 290, 400, 915	100, 365, (800), 915
NotI/ MluI	250, 290, 400, 915	100, 365, (800), 915
PvuI	585, 915	585, 915
PvuI/ NotI	100, 365	100, 365
PvuI/ MluI	100, 365	100, 365
SalI	440	440

Table 16. Size of Restriction Fragments Detected by PGD Probes on Pulsed Field Gels.

The size of fragments detected in the lymphoid line MGL8B2, chromosome-1 specific hybrid HCH5 and PWB cell DNA by the PGD cDNA probe pPGDH4 and the locus specific 1.4 kb PvuII fragment (PV1.4). As can be seen, some variation in fragment sizes of certain enzymes occurs. The most profound of these is detectable upon NotI digestion, in the form of hypomethylation at these sites in the PWB cell relative to the cell lines. Fragments underlined are those to which pPGDH4 and the PvuII fragment cohybridize; fragments in brackets are those to which the PvuII fragment only hybridizes. Those fragments in plain text are probable pseudogene fragments.

* Only the first three sets of enzymes were used in the analysis of HCH5 DNA.

Sizes of Fragments Detected by Probes on Pulsed Field Gels (kb)

ENZYME	pPGDH4	PV1.4	ENO 1	PND	p58	HSPG2	D1S77
MluI	365,290 915	800	750	1.1Mb	80,40	225,332	180
NotI	100, 290	800	550	295	70,80	450	180
NotI/ MluI	100, 290 915	800	360	295	40	225,322 450	180
PvuI	585	585	1 Mb	295	-	-	50x2
PvuI/ NotI	100	100	360	295	-	-	50x2
PvuI/ MluI	100	100	-	-	-	-	50x2
NruI	680,750	-	400	295	40,375	-	-
SalI	440	440	-	375	-	-	610
SacII	50	50	-	-	-	160,280	-
BssHII	50	50,450	-	225	40,375	160,320	-
BssHII/ NruI	50	50	-	225	40,375	-	-
NarI	220	-	220,280				
NarI/ NotI	220	-	260				
NarI/ MluI BssHII	50						
MluI NarI NotI SacII/ MluI NarI NotI	50						

Table 17. Size of Restriction Fragments Detected by 1p Probes on Pulsed Field Gels.
Table showing the size of fragments detected by the various 1p probes. PGD probe PGDH4 detects unlinked loci but in the table above only those originating from the 1p locus are listed i.e those detected by the 1.4 kb PvuII fragment (denoted PV1.4) also or present in the 1p hybrids. In the PV1.4 column, fragments hybridizing to this probe only are shown i.e those upstream to the island for the first 3 rows.

3.3.3.2. Long Range Mapping Around the *ENO1* Locus.

The *ENO1* gene, encodes the glycolytic enzyme alpha enolase (EC 4.2.1.11) which is present in most tissues. *ENO1* is a member of a gene family (Giallongo *et al*, 1986). The use of the alpha-enolase cDNA as a probe was thought to present complications in the interpretation of results due to other cross hybridizing members of the *ENO* gene family. However, no complications were encountered in map construction around this locus. Map construction was undertaken at the *ENO1* locus on chromosome 1 in an attempt to determine the physical distance between this locus and *PGD*. Sequential hybridization to the same filters showed a similarity in the size of Nar1 fragments detected by both probes but a difference in the size of other fragments detected and the PGD map indicate that such close physical linkage does not exist between these two loci.

The alpha enolase cDNA (Giallongo *et al*, 1986).was used as a probe on PFGE filters, an example is shown in figure 37. The alpha-enolase cDNA was found to hybridize to a 750 kb Mlu1 fragment reduced to a 360 kb product upon digestion with Mlu1 and Not1. Not1 alone gives a 550 kb product, therefore one Not1 site occurs within the larger Mlu1 fragment 360 kb inwards from one end. Two Nar1 fragments are detected at 220 kb and 280 kb. The two bands are present in double digests with Mlu1 suggesting that three Nar1 sites may be present within the Mlu1 fragment. In Nar1 plus Not1 digests the 280 kb fragment is reduced in size and the 220 kb is lost completely. This suggests that one Not1 site is within the Nar1 fragment to which the enolase cDNA hybridizes. Digests with Pvu1 produce a fragment one Mb in size. Double digests with Pvu1 and Not1 produce a 360 kb fragment, this suggests that the PvuI and MluI sites are coincident. The map is illustrated (figure 38) obviously

to confirm the location of all sites relative to each other more digests would be necessary. Here, the main purpose was to investigate the physical relationship of these loci and in the process the best map possible from these data was constructed. In the case of the *ENO1* locus, 1040 kb has been restriction mapped around it (figure 38).

3.3.3.3. Long Range Mapping Around the *HSPG2* Locus.

This map was constructed using the HSPG2 PCR product as a probe. In digests with Not1 a 450 kb fragment is obtained. In Mlu1 digests, fragments of 225 kb and 322 kb are obtained. In double digests, the larger Not1 fragment is lost whilst both Mlu1 fragments remain (figure 39). This means that the Mlu1 fragments are derived from a partial digest and are within the Not1 fragment. Since the smallest fragment detected by the probe is 225 kb and not a fragment whose size is indicative of the difference in size between the two Mlu1 fragments, it means that the *HSPG2* locus is within the 225 kb fragment. The map is shown in figure 40.

In Nae1, SacII, and BssHII digests, two fragments are detected for each enzyme. The intensity of each is similar so if these are not partials, they may represent a cluster of rare cutter sites in a CpG island. Looking for this in the probe itself would confirm this.

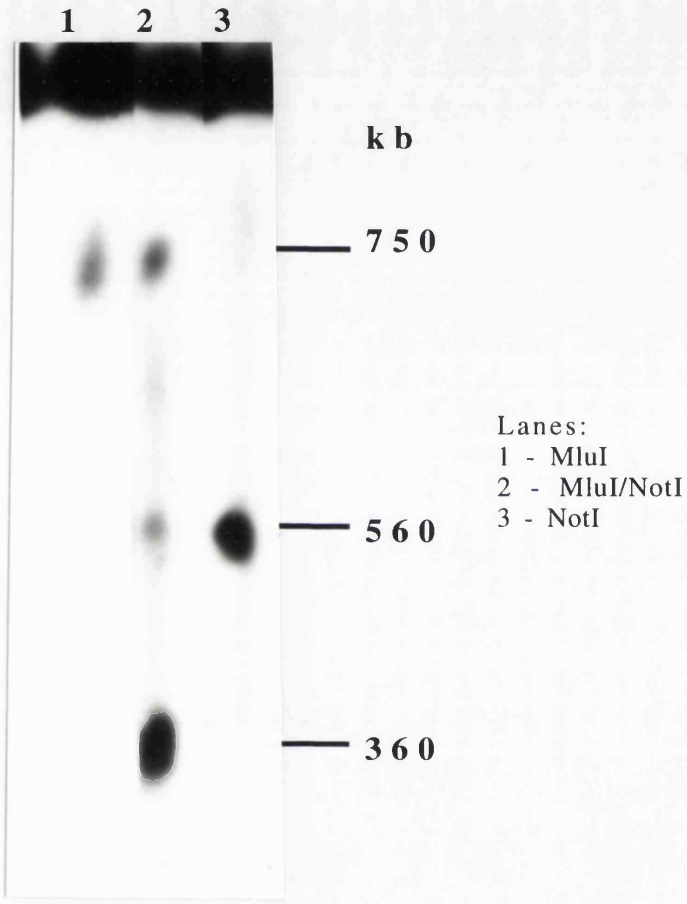


Figure 37. Autoradiograph of a PFG Blot Hybridized to the Alpha-Enolase cDNA Probe.

PFG electrophoresis was carried out using a basic programme, pulse time was 90 seconds, run time was 40 hrs. Lymphoblastoid cell line DNA was used. Digests show that a NotI site is nested within the 750 kb MluI fragment such that in double digests a 360 kb product is obtained. Partial digest products are seen in lane 2.

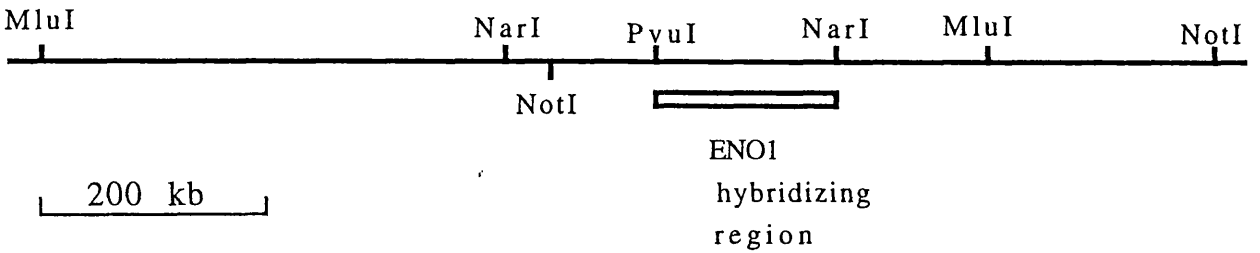


Figure 38. Long Range Map Around the ENO1 Locus.

The map around the locus spans 1040 kb . The cDNA probe is shown to hybridize within 150 kb flanked by PvuI and NarI restriction sites.

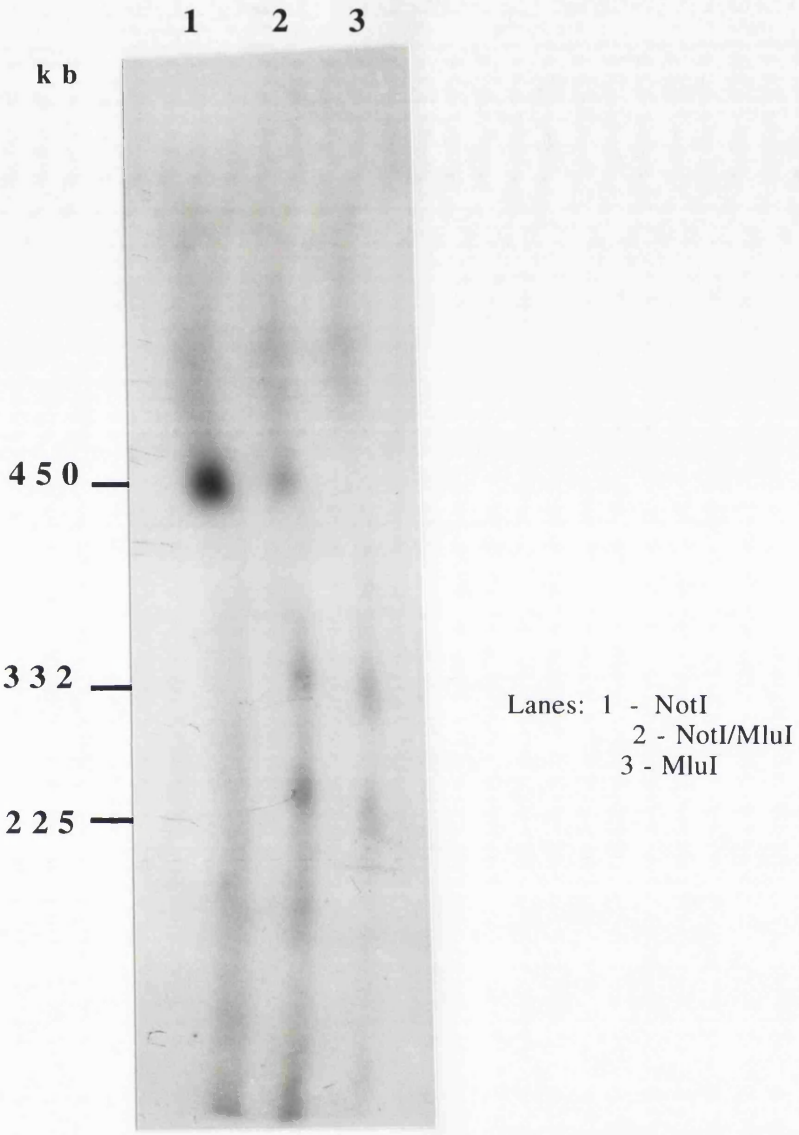


Figure 39. Autoradiograph of a PFG Blot Hybridized to the HSPG2 Probe.

PFG electrophoresis was carried out using a ramped programme starting at 15 seconds and going up to 45 seconds over 11 hrs, then going back down to 15 seconds over 11 hrs. This gel shows that the 3 MluI sites are nested in the larger 450 kb NotI fragment. PWB cell DNA was used in digests. In lane 2 partial digest products are seen.

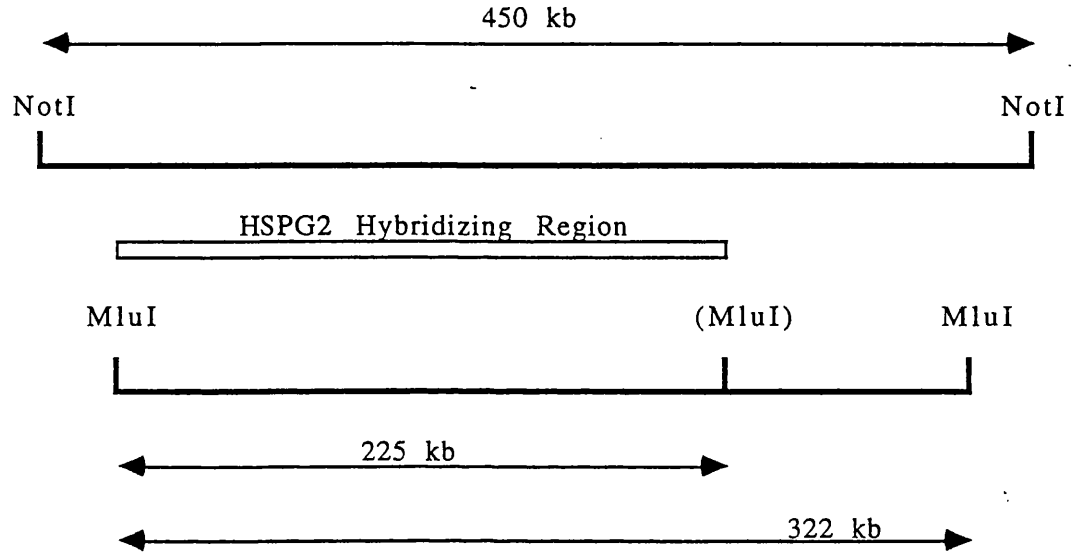


Figure 40. Long Range Map Around the HSPG2 Locus.

The exact location of the MluI fragments within the larger NotI fragment has not been determined. The location of the HSPG2 locus is thought to be within the 225 kb MluI fragment. The partially methylated MluI site is indicated in brackets.

3.3.4 Isolation of Yeast Artificial Chromosomes from 1p36.

Isolation of YAC clones for the loci already mapped by other techniques in this project was considered an important step as this would allow the eventual linking of these markers by contig construction. Since FISH (section 3.3.2) and somatic cell hybrids (section 3.3.1) have already suggested an order for these markers in 1p36, then some guidance for YAC contig construction was already available.

Some of the markers mapped by linkage analysis (section 3.1.2), FISH (section 3.3.2), somatic cell hybrids (section 3.3.1) and analysed by PFGE (section 3.3.3), were used in screening the total genomic YAC library constructed by Anand and others (1990), for cognate clones. Initially, PCR based screening was used in an attempt to isolate clones for *PGD, PND, ENO1* and *p58*. One YAC was isolated for the *p58* locus, but the other three PCRs were unsuccessful in YAC isolation. A probe 'cocktail' (containing radioactively labelled DNA for the *p58, PGD, PND* and *ENO1* loci) used later on gridded filters of the YAC library gave several strong signals one of which was the p58 position. Other positive colonies were not assessed in this project but are currently being assessed by Carritt and others.

The p58 YAC was isolated using two sets of primers from the p58 cDNA sequence reported by Eipers and others (1991) by nested PCR of the YAC library (figure 41). Primers one and three were derived from an intron/exon boundary to provide locus specificity, necessary because of the reported existence of a related locus in chromosome 15. Primers one and two are used in the first reaction and primers three and four in the second. The relative location of these primers is indicated in figure 41. Reaction conditions and primer sequences are given in the Appendix, section 5.4.

To confirm that this was an authentic clone, colony filter hybridization was also carried out with the PCR product as probe (figure 42). A single positive signal was obtained in the expected position on the gridded filter. The YAC isolated was 120 kb in size and shown to hybridize to the p58 PCR product (figure 43).

Indirect evidence suggests that this YAC originates from 1p36. This being the assignment of the cosmid isolated by hybridization to the p58 PCR product to band 1p36 by FISH (see section 3.3.2). The primers used to generate the probe are the same set as those used to isolate the YAC.

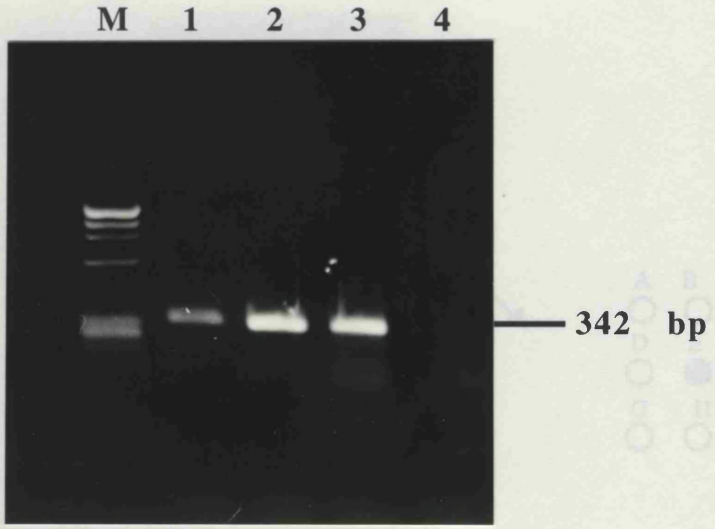
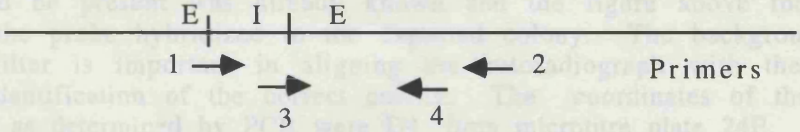


Figure 41. Isolation of p58 YAC Clone by PCR Amplification.

Lanes: M - phi X174 1kb ladder,
1 - p58 amplicon from yeast cell suspension of positive clone 24E , D4.
(1 microlitre of a yeast cell suspension was amplified after suspending a small colony of cells in 200 microlitres of sterile distilled water).
2 - p58 amplicon from 100 ng of p58 YAC DNA.
3 - p58 amplicon from Total human genomic DNA (5 ng).
4 - Negative Control.

Nested PCR was carried out to amplify the p58 gene locus as described by Eipers and others (1991). To screen the genomic YAC library, the sensitivity of the p58 PCR was assessed so that amplification would be detected with as little as 1 ng of template DNA. The sensitivity achieved was high, a titration showed that a strong band was obtained even with 32 pg of genomic DNA. The clone isolated was tested by amplification of the locus from yeast cells after growth for two days and then from the DNA isolated from the clone. Both gave a band of expected size, 342 bp.. The relative location of the primers used is shown below. Part of the sequence of primers 1 and 3 is derived from intronic p58 sequence to provide locus specificity.



E - Exon, I- Intron. The PCR product is 342 bp in size.

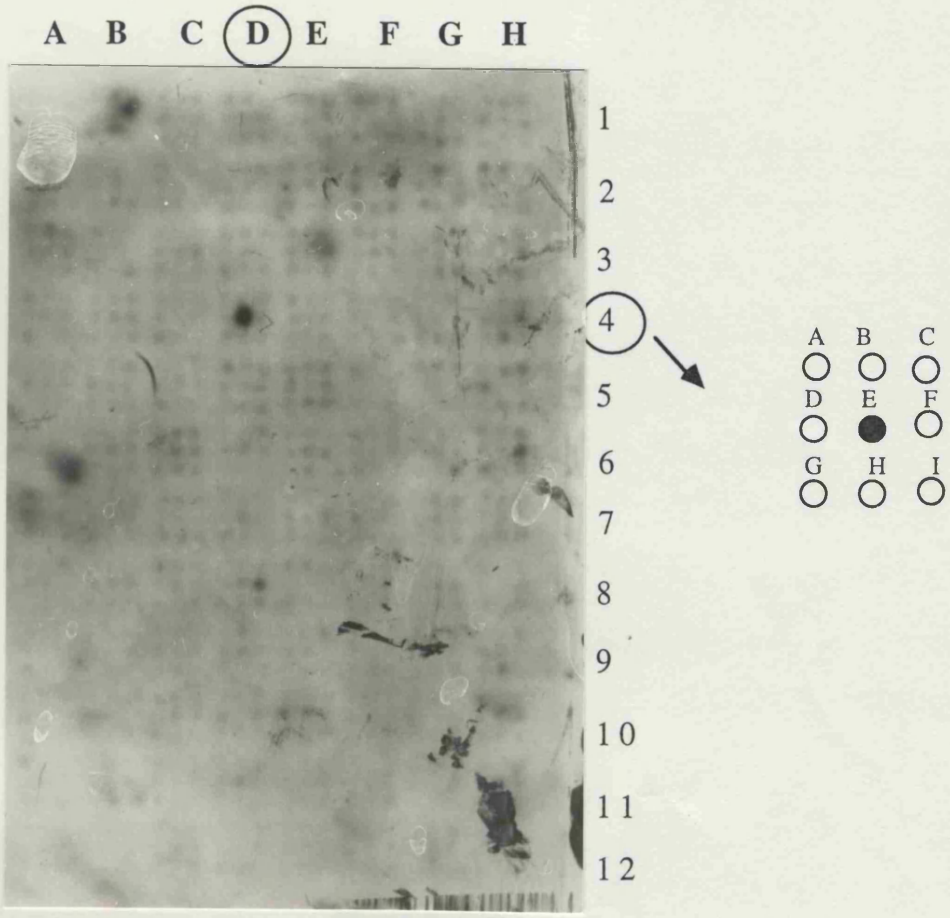


Figure 42. Hybridization of p58 Probe to Gridded YAC Colony Filter.

This experiment was carried out to confirm the authenticity of the YAC clone isolated by PCR. The position at which a positive signal should be present was already known and the figure above indicates that the probe hybridized to the expected colony. The background on the filter is important in aligning the autoradiograph with the filter for identification of the correct colony. The coordinates of the positive clone as determined by PCR were D4, from microtitre plate 24E. On the gridded filter this corresponds to the position of the positive signal.

Figure 43. Hybridization of p58 Probe to PFG Blot of p58 YAC DNA.

Lanes: 1 - p58 YAC clone, 2 - p58 YAC clone, 3 - Delta 39 kb ladder.

PFG electrophoresis was carried out using a 50 second pulse time for 40 hrs. A Delta 39 kb lambda phage ladder was used as a marker. The YAC is 117 kb.

CHAPTER 4 DISCUSSION.

DISCUSSION

In this project physical and genetic mapping has resulted in the establishment of one order for markers residing in a region on chromosome 1 which is of importance in clinical genetics. This region, 1pter-1p36, has been shown by other authors (Martinsson *et al*, 1989; Laureys *et al*, 1990; Biegel *et al*, 1993; Dracopoli *et al*, 1989; Bale *et al*, 1989) to be a target for molecular abnormalities associated with genetic defects. Thus by refining the map which encompasses this region, and generating resources such as region specific YAC and cosmid clones, future studies in positional cloning of genes directly involved in disease will be greatly facilitated. In addition, the resources and data accumulated here, may be of significance in comparative mapping. Once the physical maps constructed here have been linked, a precise relationship between genetic and physical distance can be established. An estimate for this is given later.

Collectively, the genetic and physical mapping data from this thesis support one order for the loci used in the analyses. This illustrates the importance of several mapping strategies in the construction of a detailed and reliable map. Individual techniques offer a different resolution range. Hybrids bearing translocations have allowed the division of groups of markers across breakpoints whilst single and dual colour FISH has allowed the order of markers to be determined. Linkage analysis has given an order and simultaneously an interlocus distance for all loci in the map.

4.1 PHYSICAL MAPPING in 1p36

4.1.1 Fluorescence *In Situ* Hybridization and Somatic Cell Hybrid Analysis of Markers in 1p36.

One order was deduced from physical mapping using fluorescence *in situ* hybridization measurements. This result was partly confirmed by dual colour FISH, the only change being the inversion of the two proximal loci *HSPG2* and D1F15S1. The difference in the distance obtained by measurement for each locus from the telomere was very small (0.005 microns) and the standard deviation for the two locations from their mean shows that D1F15S1 may be the more proximal locus.

The lack of dual FISH data for the pair of loci *PND* and *HSPG2,* prevented the confirmation of the order obtained by fractional length measurement, but the results from somatic cell hybrid analysis, which places *PND* distal and *HSPG2* proximal to the 1p36.13 breakpoint, allows them to be orientated. Here it is seen that two different mapping strategies complement and corroborate each other.

In the present study, *HSPG2* is assigned a location distal to D1F15S1 and proximal to *PND* by fluorescence *in situ* hybridization. *PND* and D1F15S1 were two of the 48 loci in a high resolution cytogenetic map of 1p (Van Roy *et al*, 1993). Their relative location is the same as that obtained in this project as well as that of *PGD* and D1Z2, which were also included in both of these independent FISH experiments.

Polymorphism in the heterochromatic bands has been demonstrated (reviewed by Sumner, 1982). As a result of this phenomenon the use of metaphase spreads from a single individual was essential. This was an important way of eliminating

the variation that may be caused by polymorphism of the heterochromatic region at the centromere of chromosome 1.

The original localization of the *HSPG2* gene to 1p36 by *in situ* hybridization and somatic cell hybrid analysis (Kallunki *et al*, 1991; Dodge *et al*, 1991) was accompanied by the detection of three RFLPs at this locus, involving Taq1, EcoR1 (Dodge *et al*, 1991) and BamH1 sites (Kallunki *et al*, 1991). Typing of the CEPH pedigrees for the most informative of these will allow the inclusion of this locus in the map of chromosome 1, and refine its location with respect to the markers mapped in this project.

The somatic cell hybrid Dis2.6 apparently retains a small number of unrelated human chromosomal fragments on a chinese hamster background (Carritt *et al*, 1982a). The best characterized of these originating from chromosomes 1 (Carritt *et al*, 1982a) and 3 (Carritt *et al*, 1992). This hybrid is therefore a potentially useful mapping resource for these regions. Thus the presence of *PGD* (Carritt *et al*, 1982; Kleyn, 1990) and the absence of *HSPG2* and *PND* (section 3.3.1) and the absence of D1S47 (N. Dracopoli, 1990, personal communication), may be used to estimate the size of the 1p fragment in the hybrid. Since the present estimate of the genetic distance between *PGD* and its flanking markers D1S47 and *PND* is 18 cM (9 cM for each interval), then the 1p fragment should in theory span something less than this genetic distance. Based on the physical to genetic distance ratio of 500 kb being equivalent to one cM, (section 4.3) then this means that there would be a maximum of nine Mb of 1p in Dis2.6. This makes the hybrid a potentially valuable source of probes for this chromosomal region. The value of this type of resource generation has already been demonstrated for radiation reduced hybrids (Zoghbi *et al*, 1991; Monaco *et al*, 1991) and also for Dis2.6 in the 3p21 chromosomal region (Carritt *et al*, 1992). Repeat element mediated PCR would be an efficient way to saturate more fully the region specifically around *PGD*.

4.1.2. Long Range Mapping Around The *PGD* Locus.

Construction of the long range map around the *PGD* locus was facilitated by the identification of a CpG island in the 5' PGD genomic clone pPGDE10. Probing PFGE blots with a locus specific 1.4 kb Pvu II fragment containing the island, allowed the expansion of the map bidirectionally from the CpG island.

The CpG island detected at the 5' end of the *PGD* gene had all the characteristics typical of other islands (Bird *et al*, 1986; Larsen *et al*, 1992). It was found to be G+C rich (55%) in the region sequenced, to have a ratio of CpG:GpC dinucleotide equal to 0.9, to be single copy, hypomethylated in both germ line and somatic cell DNA and to possess a cluster of rare cutter sites. This also conforms to the general trend of association of CpG islands with constitutively expressed genes.

When constructing long range maps it is important to know the origin of any variation detected between tissue and cell lines, or between DNA from normal individuals and those with a genetic disorder. Any variation detected by PFGE may be the result of a silent polymorphism, as in that detected by a marker *(MET)* linked to the *CF* locus (Julier and White, 1988), methylation of certain restriction sites in fragile-X patients (Vincent *et al,* 1991), or methylation variation occurring in CpG rich regions which is associated with gene expression in cell culture (Antequera *et al,* 1990). It has been shown that patterns of DNA methylation are concordant over randomly chosen sequences in the human genome within genetically distinct populations (Behn-Krappa *et al*, 1991), indicating apparent homogeneity in methylation patterns. A similar observation was made in the analysis of the sequences of the *tumour necrosis factor* genes *(TNF)* in normal individuals (Kochanek *et al*, 1990).

In the present study, peripheral white blood cell DNA was found to give different restriction enzyme cleavage patterns in comparison to cell line DNA. Larger restriction fragments were obtained with certain methylation sensitive enzymes in cell lines when compared to PWB cell DNA around the *PGD* locus.

Experiments were carried out to determine the origin of this variation in the two DNA sources and whether it involved the PGD CpG island. The hybridization pattern of cell line DNA and that of the 5' genomic clone pPGDE10 on digestion with EcoR1 and the rare cutters, followed by hybridization to the 5' PGD cDNA clone pPGDH4, or the 1.4 kb Pvu II fragment was the same. This indicates that methylation of the PGD CpG island does not occur. The lack of methylation of the PGD CpG island is consistent with the fact that *PGD* is a housekeeping gene crucial in the generation of reducing power, and therefore it is likely that it remains active in both PWB and tissue culture cells.

As Not 1 sites (GCGGCCGC) occur almost exclusively in CpG islands i.e 89% of the time (Lindsay and Bird, 1987), the smaller Not1 fragment detected in PWB cell DNA may indicate the existence of another island which is normally methylated in cell lines and demethylated in PWB cell DNA. Also the smallest fragments detected by PGD probes are 50 kb on cleavage with SacII and BssHII, which occur in CpG islands 74% of the time (Lindsay and Bird, 1987), then it may be postulated that a CpG island may occur at the position identified by at least two rare cutter sites 50 kb away from the PGD CpG island.

A high degree of polymorphism is created by the presence, or absence of MspI sites whose recognition sequence is 5'-CCGG-3' (Barker *et al*, 1984). Here methylation of the sequence does not prevent cleavage but through methylation to give 5-methylcytosine, and deamination to give thymine, the restriction site is lost (Coulondre *et al*, 1978). Three different individuals were tested (one male, two females) to see if the difference in enzyme cleavage sites was a restriction

site polymorphism. Although a larger number of individuals should perhaps have been analysed, the similarity of the cleavage maps is consistent with a methylation difference rather than a restriction site polymorphism.

The methylation status in the PWB cell DNA represents the natural state and this is perhaps the best approach to the detection of CpG islands in gene mapping. Cell lines would permit the identification of physical linkage with greater ease, however, as a result of *de novo* methylation of dispensable genes in culture, thus creating larger fragments. This methylation difference was only investigated around the *PGD* locus and it was not the intention in this project to analyse methylation difference in cell line and PWB cell DNA for all loci investigated. Since the use of PWB cell DNA as a source of high molecular DNA was more beneficial for many reasons, it was used throughout the later long range mapping experiments.

4.1.3 Long Range Mapping Around Other 1p36 Loci.

Other loci (*p58, PND, HSPG2*, D1S77, *ENO1)* analysed by PFGE did not cohybridize to any fragments. The absence of physical linkage was not necessarily indicative of excessively large distances between the loci, since there appeared to be an unusually high density of sites for the long range mapping enzymes used here. For example, fragments ranging from 40-80 kb were detected by some of the probes.

The small fragments detected by some of the probes here are often very similar in size as those produced by other enzymes suggesting their coincident occurrence, possibly in a CpG island. This is most obvious in the case of the *PND* locus where similar sized fragments are obtained for NotI, PvuI and NruI, suggesting that two CpG islands flank the *pronatriodilatin* locus. This is also true for some enzyme sites around the *p58* gene, notably, BssHII and NruI in

double digests. This is a ubiquitously expressed cell division control- related protein kinase whose gene structure has been deduced (Eipers *et al*, 1992). The presence of a CpG island however, was not reported .

Also, Saccone and others (1992) suggested that there is a high G+C content in the distal part of 1p. In their study they show that the telomeric and subtelomeric R bands of 1p have a very high G+C content. This region is part of the isochore family H3 (defined as compositionally homogeneous and the G+C richest 3% of the genome) which is shown in their study to reside on the terminal and subterminal R bands. This fraction of DNA contains 50-70% of single copy DNA. Isochore family H3 is also the region which is most transcriptionally active i.e having the highest gene density, (28%) (Mouchiroud *et al*, 1991) and so it is not surprising that this is also reflected in a high CpG island concentration (Aissani and Bernardi,,1991).

4.1.4 Isolation of YAC Clones for 1p36 Loci.

The different long range maps could perhaps be linked by experiments involving partial digests with rare cutting enzymes. Several procedures for achieving partial digestion have been published, including Mg^{++} cation titration (Albertson *et al,* 1989) and site-specific methylation (Hanish *et al*, 1991). In the present study varying the time of digestion and the enzyme concentration were both tried without success. This difficulty in controlling the reaction is undoubtedly related to the fact that the substrate DNA is embedded in agarose.

An alternative way to link markers physically is through the isolation and characterization of large Yeast artificial chromosomes (YACs). One serious artifact potentially limits the utility of YACs in analytical experiments. This is chimeric YAC formation resulting either from coligation or from *in vivo*

recombination (Bronson *et al*, 1991). Nevertheless, the advantages of working with single clones up to one Mb in size have been illustrated by numerous publications (Nelson *et al.* 1991; Sheiro *et al*, 1991; Anand *et al*, 1991; Eliceiri *et al*, 1991). Therefore, YAC library screening was initiated in the closing stages of this project.

The p58 YAC isolated, although not itself assigned to lp, is thought to originate from 1p from indirect evidence. The primers used to isolate the YAC generated the PCR product which was used to isolate the p58 cosmid assigned to 1p36 by FISH. The YAC was found to be 120 kb in size, quite small in comparison to the size of high molecular weight fragments clonable in YAC vectors (Burke *et al,* 1987).

To date only seven YACs have been derived from 1pter-1p36.11 (Dreisen *et al,* 1991). These range in size from 90-290 kb, collectively covering 1.2 Mb of the region although none are overlapping. This therefore requires walking experiments to isolate further YACs to link these. The p58 YAC and the above may be considered as nucleation points instrumental in YAC contig construction in 1p36, end probes from these can initiate large scale walking bidirectionally. A YAC contig constructed around the *p58* locus already known to lie within band lp36, which encompasses all the other markers employed in this study, will be an efficient and direct way to a systematic linking and ordering of these probes. It would be combined with the direct measure of physical distance and generation of anonymous probes. Some of these may be polymorphic and of use in genetic mapping of hereditary diseases associated with 1p aberrations or already showing linkage to 1p markers. In the process, potential coding sequences may be identified through CpG islands and in the final stages, the hybridization of the contig to cDNA filters would identify all expressed sequences originating from this region irrespective of their association with a CpG island.

4.2 GENETIC MAPPING in 1p36.

4.2.1 Pedigree Linkage Analysis-Incorporation of *PGD* in the Genetic Map of 1p.

Twenty of the CEPH pedigrees previously used to construct a consortium map of chromosome 1 (Dracopoli *et al,* 1991) were typed for the PGD RFLP in an attempt to locate this marker on the map. Prior to the publication of this map and that of O'Connell *et al* (1989), and Dracopoli *et al* (1988), the consensus genetic map of 1p was as follows:

CEN-RH-FUCA-D1S56-D1S71-PND-D1S47/43-D1S50-D1S49-1pter
(Donis-Keller *et al,* 1987).

In the present study, *PGD* was located by genetic mapping in the interval between the markers D1S43/47 and D1S71.(section 3.1). However, the orientation in here with respect to *PND* was not established unequivocally. The preferred location by multipoint analysis is that established by fractional length measurements and dual colour FISH experiments (section 3.3.2).

One possible reason for the assignment of *PGD* to two almost equally likely locations, is the lack of recombinants between *PGD* and any of the three markers which flank the two intervals. However, two point analyses between *PGD* and each of the flanking markers show that *PGD* recombines with all of them. Recombination fractions betwen *PGD* and (D1S43 and *PND*) markers defining the distal of the two intervals, are lower with higher lod scores than that between *PGD* and the more proximal marker, D1S71. It seems more likely that typing errors at the *PGD* locus, or at any of the other loci in the map give rise to false double recombinants. Although errors in the PGD data can never be completely ruled out, the data were rigorously and repeatedly checked. Other

data in the public data base could not be similarly checked. Errors in other typings may come to light with the introduction of *PGD* in the analysis.

By analysing individual chromosomes, the option CHROMPIC indicates where double cross over events occur. There were eight families in which double cross overs in the region were detected, four of which were over a short interval. It would have been possible to perform another analysis excluding these families but considerable data loss would have been associated with this. Obviously physical data were essential in this situation as they gave more direct and conclusive information.

Analysis of more meioses using this RFLP may be considered in the future. This might be done using the remaining 20 of the CEPH pedigrees since only 40 have been assessed here. Alternatively, a more informative marker from this locus may be required. A genomic phage clone was isolated using the 3' PGD cDNA clone pPGDHI, and shown to possesss a CA repeat by hybridization to a GT oligonucleotide. If this is found to be polymorphic it may be used to type these families.

Two point data and the results from multipoint analysis shed some light on the ratio of male to female recombination, as well as the variation in this ratio in different intervals. First multipoint analysis indicates that female recombination at every interval is greater than in the male with a variable ratio from 1.1-2.0. This was also the observation in the construction of the chromosome 1 map (Dracopoli *et al,* 1991). Two point data however, although giving comparable results as to the location of *PGD*, give contrasting results with respect to genetic distance and the ratio of these values in the different sexes. In fact the two point analyses suggest greater recombination in the male between *PGD* and its adjacent loci D1S43 and *PND.*

A recent unpublished study shows that using the 3' PGD cDNA clone pPGDH1 as a biotinylated probe on prometaphase chromosome spreads, *PGD* maps to two regions within 1p36. One signal is obtained within band 1p36.33-1p36.31 and a more proximal signal is seen at 1p36.13-1p36.12. A probe for the *PND* locus is also used in this high resolution mapping endeavour and surprisingly its assignment is between these two PGD signals (Van Roy *et al*, 1993). While it is most unlikely that the *PND* locus is contained within the expressed *PGD* gene, there may be more than one locus on chromosome 1 which hybridizes to the 3' PGD cDNA clone pPGDH1. The BamHI RFLP detected by this clone is a diallelic, codominant polymorphism with Mendelian autosomal inheritance. This indicates that it is the result of a mutation at a single site, which does not recombine with the expressed *PGD* locus and is in linkage disequilibrium with it.

The second related locus may be a pseudogene. It was shown earlier that at least two *PGD* pseudogenes exist (Kleyn, 1990). Somatic cell hybrid analysis indicated that the hybrid HCH5 does not contain *PGD* related sequences which segregate independently of the expressed *PGD* gene. Even so, in the current study in this hybrid the 5' PGD probe pPGDH4 detects two Not1 fragments on PFGE. Only one of these represents the expressed gene on chromosome 1 as is shown when the locus specific 1.4 kb PvuII fragment is hybridized to Not1 digested DNA. This leads to speculation that the second Not1 fragment is that from the pseudogene locus which is present in HCH5 and whose EcoR1 restriction fragment must comigrate with that from the expressed locus. For confirmation and further information on the nature of this homologous sequence, microdissection of these two regions followed by PCR amplification using pPGDH1 specific primers and comparison of the sequence could be used to answer these questions.

Pseudogenes exist in two different forms. Some have sequences corresponding to exons and introns, and are rendered inactive by mutations that prevent expression. Others are processed, i.e showing homology to mature mRNA transcripts up to the cap site and often have a poly(A) tail. Their similarity to mRNA has led to the suggestion that they are derived by reverse transcription (reviewed by Vanin, 1985; Wagner, 1986).

1p36 is a region where virus integration frequently occurs ((Romani *et al*, 1990) and this is perhaps where one of the *PGD* pseudogenes has integrated. A survey of the location of integration of adenovirus/simian virus 40 recombinant constructs showed a preferential integration at 1p36 a region said to be highly recombinogenic (Romani *et al*, 1990).

In earlier *in situ* experiments with both of the PGD cDNA clones there was no evidence of a dual signal on one chromatid (Kleyn, 1990). This was the observation in FISH experiments using the 10.5 kb genomic clone pPGDE10, which hybridizes to the 5' cDNA clone pPGDH4 but not to the 3' clone pPGDH1. pPGDE10 was used as a probe on prometaphase spreads in dual FISH experiments and there was no sign of a double signal on one chromatid.

The above finding, i.e the presence of just one signal on 1p when pPGDE10 is used as probe, indicates that if this PGD related sequence is a *PGD* pseudogene then only the 3' PGD cDNA clone shows homology to it. This would imply that the hypothetical 1p pseudogene is a truncated version of the expressed gene and therefore not a processed pseudogene. If it were processed it should also have been detected by the 5' PGD cDNA clone pPGDH4 and the 5' genomic clone pPGDE10. This conclusion is unexpected because processed pseudogenes are typically derived from constitutively expressed genes (Wagner, 1986).

The failure to detect a double signal with pPGDH1 in earlier *in situ* experiments may be explained by the fact that radioactively labelled probe was used. The signal is spread out as a result of long exposure and it may camouflage the result seen by Van Roy and others, (1993.) In addition the advanced technology afforded by FISH and the use of prometaphase chromosome spreads must have contributed to the greater resolving power of the above study.

It is noteworthy that the mapping studies with the anonymous DNA sequence (D1S90) were associated with the same difficulties as those encountered in the present study. The inability to map this locus by multipoint analysis is thought to be the consequence of duplication of the locus as shown by *in situ* hybridization on proximal 1q (Dracopoli *et al,* 1991). Unidentified typing errors may of course be equally responsible and the duplication may not encompass the polymorphism itself.

4.2.2 Pedigree Linkage Analysis-Incorporation of D1F15S1 in the Genetic Map of 1p.

Segregation analysis of the D1F15S1 locus and the loci *PND, PGD* and D1S77 led to the confident assignment of D1F15S1 outside this group of loci. D1S77 was located at the other end of the cluster. Cosegregation of *PGD* and D1S77 away from *PND* gives the order: D1S77-*PGD*-*PND*-D1F15S1. The orientation with respect to the centromere is obtained from both somatic cell hybrid analysis and FISH studies. First, D1F15S1 is found to be proximal to the 1p36.13 breakpoint in the hybrid HCH7 whilst D1S77 and *PND* are shown to be distal in this thesis; *PGD* is also distal to this breakpoint (Carritt *et al,* 1982a; Kleyn, 1990). Also, this order is confirmed by FISH studies, which order the loci *PGD, PND* and D1F15S1 from 1pter to the centromere. The genetic maps of chromosome 1 (O'Connell *et al,* 1989; Dracopoli *et al,* 1991) corroborate

these findings. The meioses analysed conformed to the general trend of greater recombination in female meiosis. Out of 14 cross overs, 11 took place in maternal meiosis and three in paternal.

4.2.3 Genetic Mapping with Ovarian Teratoma Hybrids.

Ovarian teratomas employed in centromere-related mapping of 1p markers did not give the same order as that derived by pedigree linkage analysis. The order of *PGM 1* and *PGD* is known to be correct but one teratoma here (RC5) suggests that *PGM 1* may be the more distal locus. Analysis of hybrid LLA5 suggests that *PGD* is distal to D1S77 which is contrary to the conclusion of pedigree analysis (section 3.1.2). i.e linkage analysis places D1S77 distal to *PGD*. Also, the physical mapping study by Van Roy and others (1993), places D1S77 distal to *PGD*.

The discrepancies between centromere based mapping using teratoma hybrids and conventional pedigree linkage analysis cast doubt on the potential of teratomas for centromere based mapping. They confirm the earlier findings that a) the teratomas arose following a failure of meiosis I and a precocious meiosis II, b) complex interchromatid exchanges occur during this process and c) maps obtained by reducing the number of exchanges required during teratoma formation may not be correct (Carritt *et al,* 1982b).

4.3 The Relationship Between Genetic and Physical Distance in 1p36.

The frequency of chiasmata in this distal region of chromosome 1 is greater than any where else on 1p (Hulten *et al* , 1982). For 1p 40% of crossing over is concentrated in the terminal 20% of each arm in males. The total genetic length of chromosome 1 in males was found to be 194.3cM. 1p was found by chiasma counts to be 99cM in length with 1p36 having a genetic length that was 46% of

the total 1p length i.e 45.3 cM The genetic length of this part of 1p is found to be approximately the same by the most recent of 1p maps using multipoint linkage analysis (Dracopoli *et al,* 1991). The loci *PND* and *PGD* are in band 1p36 and the sex averaged distance from the most distal locus, D1Z2, to *PND* is 37.2 cM. When *PGD* is incorporated in this map this distance is 46.3 cM, quite close to the size of 1p36 derived from chiasma counts. Thus an estimate can be made of the relationship between physical and genetic length in this region and the highly recombinogenic nature more accurately defined. The physical length of chromosome 1 is 252 Mb accounting for 8.4% of autosomal length (Paris Conference, 1971), 1p is 123.48 Mb and 1p36 (18% of 1p) is 22.23 Mb. This gives a ratio of almost exactly two cM per Mb based on either chiasma counts or linkage analysis.

If this is a uniform relationship across 1p36, then for *PGD* and *PND*, being nine cM apart, the corresponding physical distance would be about 4.5 Mb. From PFGE studies using these two loci the conclusion is that a minimum of 900 kb separate these two loci. So although these approximations are only tentative they are not divergent in the final conclusion drawn. In the course of this study none of the loci were physically linked, precluding a more realistic value for this highly variable ratio of genetic and physical distance.

Two cM per Mb is twice the genetic:physical distance ratio which is obtained by dividing the total genetic size of the human genome by its physical size. Other studies have shown that chromosomal termini are highly recombinogenic (Romani *et al,* 1990; Rynditch *et al,* 1991) and that the genetic:physical distance ratio may be as high as one cM being equivalent to 50 kb in distal chromosomal regions (Burmeister *et al,* 1991b). In addition to being highly recombinogenic, the PFGE experiments demonstrated in section 3.3.3 show the region to be rich in its G+C content and probably, in the frequency of CpG islands. This is also

consistent with the observations established for 1p36 by Saccone and others 1992).

It has been suggested that there is a causal relationship between the presence of transcribed regions and recombination (Thomas and Rothstein, 1991). Certainly, the recombinogenic nature may not just be attributable to transcription, as the preferential location of VNTRs at the proterminal regions of chromosomes (Royle *et al,* 1988) may also influence the recombination rate by virtue of the similarity of the VNTR core sequence to the Chi recombination sequence of *E.coli* and phage lambda. This has been been shown to stimulate homologous recombination (Wahls *et al,* 1990). In 1p36 at least two VNTRs have been mapped, D1Z2 and D1S77 (O'Connell *et al,* 1989; Dracopoli *et al,* 1991).

4.4 Comparative Mapping in 1p36 and Mouse Chromosome 4.

In the mouse genome loci on chromosome 4 show synteny and linkage conservation to 1p loci (see introduction, section 1.2). An addition to the group is the locus *HSPG2* mapped to chromosome 4 by segregation analysis in the mouse (Chakravati *et al,* 1991) and 1p36 in man (Kallunki *et al,* 1991; Dodge *et al,* 1991)..

This conserved group in man and mouse harbours the loci *PND* and *PGD.* The loci *Pgd* and *Pnd* in the mouse are found to be one cM apart with the latter being the more distal locus (Davissson *et al,* 1991). In the present study cumulative evidence from physical and genetic mapping suggests that *PGD* is the distal locus. In the mouse the distance obtained between these two syntenic loci may not have been from a direct segregation analysis involving the pair of loci in the same experiment, even if this were the case the small distance between these does not imply that the order on 1p is inaccurate especially in the light of dual FISH information in this project.

4.5 PGD cDNA Sequence and Exon Intron Organization.

At the commencement of this thesis, some DNA sequence of the PGD cDNA was available (Kleyn, 1990). In this project this cDNA sequence was completed and the distribution of 439 bp of this within the PGD genomic clones pPGDE10 and pPGDE4 was determined by sequencing the 5' cDNA pPGDH4 reactive fragments within these clones. This was divided into four exons, two were present in each clone. Comparison of this cDNA sequence, after translation, with the Ovine PGD amino acid sequence derived by Carne and Walker in 1983, showed a high degree of homology of 83%. Divergence from the Ovine sequence is detected between the last four residues sequenced. The human sequence continues with an open reading frame for 12 amino acid residues beyond the COOH terminus of the sheep PGD sequence.

The 5' coding sequence of human *PGD* gene is still incomplete with only ten amino acid residues obtained from the clone isolated by PCR from the T-cell cDNA library. The fact that the usual 5' control sequences are absent and the initiating methionine has not been found means that this is not the full 5' sequence. This is also suggested by comparison with the Ovine sequence which has additional 5' coding sequence. In addition the codon presumed to have been interrupted by cloning of the PGD extension product appears to have been the same as that in the sheep, i.e a leucine.

Sequencing of the PGD extension product with the 5+ primer, which gives pPGDH4 sequence, and PGD primers 1 and 2 shows that the new sequence is continuous with the known sequence. In addition, the hybridization of the PCR extension product to the same fragments in the 5' PGD genomic clone pPGDE10 as the 5' cDNA pPGDH4, as well as to novel more 5' regions confirms that it is a bona fide PGD sequence.

Genomic sequencing to find the missing 5' coding and untranslated sequence has shown that it is absent from the 202 bp at the 5' end of the PGD CpG island and from the 324 bp at the 3' end of the island. This leaves a remaining internal region of 274 bp which was found difficult to sequence because of its secondary structure. It is assumed that because of the hybridization pattern of this extension product that the additional ten codons lie in these 274 bp along with the remaining 5' sequence. Since numerous sequencing strategies and 5' RACE have been unsuccessful in this endeavour, it would perhaps be appropriate to consider screening other cDNA libraries not yet screened.

The elucidation of the intron exon structure in the clones pPGDE4, (4,5 kb) and pPGDE10, (10.5kb), indicates that 439 bp from the cDNA pPGDH4, approximately one third of the coding sequence is distributed in these two clones in four exons with a size range between 67-179 bp. The *PGD* gene is presumed to be at least 18.5 kb according to the size of fragments detected by 5' and 3' cDNA clones, this is similar in size (18 kb) to the gene encoding the other NADP-dependent oxidoreductase in the pentose phosphate pathway, *G6PD*. The *G6PD* gene is divided into 13 exons whose size range is between 12-236 bp, the enzyme subunit is 58 kD (Martini *et al,* 1986). This gene is also associated with a 5' CpG island which overlaps the start of the first exon (Larsen *et al,* 1992). With a common enzymatic role, it was thought that some sequence similarity may exist between them. However, a comparison of the DNA and amino acid sequence shows no similarity. In fact it was thought that the *G6PD* gene structure could be used as a means to predict the number of exons that the PGD cDNAs are likely to be divided into. So far there is no similarity in the exon-intron structure either. This means that their common enzymatic role is probably attributable to a similar conformation of the proteins which results irrespective of dissimilar primary protein structure.

4.6 The Potential Role of 1p36 Loci in Disease.

From this mapping study certain aspects of disease involvement of 1p36 loci may be tackled. This refers to the potential role of *p58* in neoplasia and *PND* in essential hypertension.

The mapping of *p58* in this project relative to other 1p36 markers was stimulated by its initial assignment to 1p36 (Eipers *et al,* 1991) and by its function in cell division. As a result of the reports on cytogenetic abnormalities and/or LOH of markers in 1p36 in malignant disease such as neuroblastoma (Weith *et al,* 1989; Martinsson *et al,* 1989; Fong *et al,* 1989); cutaneous malignant melanoma (Dracopoli *et al,* 1989); breast cancer (Genuardi *et al,* 1989); colon cancer (Gebhart *et al,* 1992) and MEN 2A (Mathew *et al,* 1987), it seems that the loss of gene(s) in this region may lead, or contribute to tumourigenesis.

PGD and *PND* are just two of the markers which are lost in some of these tumours and the *p58* locus has been assigned by FISH at a location distal to these in the present study. Analysis of these same tumour DNA samples for loss of this gene through development of a polymorphism, or gene dosage analysis would be a first step. The attempt to reverse the transformed phenotype of tumour cell lines using p58 constructs such as the p58 YAC isolated here, may shed light on the exact, if any, involvement of *p58* in any of these tumours. Linkage analysis in affected pedigrees with a polymorphic marker at this locus would also be valuable procedure in this investigation. Knowledge from FISH experiments here should direct the search for molecular lesions distal to *PGD* where *p58* lies, a candidate tumour suppressor whose inactivation may lead to malignancy.

The PGD protein polymorphism was employed in sib pair linkage analysis to investigate linkage between 39 quantitative traits related to essential hypertension and 25 genetic marker loci (Wilson *et al,* 1991).

Sib pair analysis identifies single loci with major effects or minor effects, oligogenic/multiple loci that affect single traits, and single loci that have pleiotropic effects. Sib pair linkage analyses are broadly divided into two forms. Those based on identity by state (IBS) relationships of the sib pairs at the marker loci (Penrose, 1953), and those based on the sib pairs' identity by descent (IBD) relationships (Haseman and Elston, 1972). In the latter, information from both sibs' and parents' marker phenotypes is used to estimate the proportion of alleles each sib pair shares at the marker loci. The rationale behind this analysis is as follows: if a marker is linked to a putative locus responsible for some of the phenotypic variation of a trait, the alleles at the marker and putative locus should cosegregate. Sibs with a high proportion of alleles identified by descent for the marker alleles should also have the same alleles at the putative linked locus and the difference between the sib's phenotypes should be small. A large difference in the sib's phenotypes would be associated with a small proportion of alleles identified by descent at the marker locus and of the shared alleles at the putative linked locus.

In the essential hypertension study which included two other 1p loci, *RH* and *PGM1*, a marker - phenotype matrix showed that *PGD* or a putative locus linked to it, was associated with mean fifth phase diastolic blood pressure, one of the traits related to hypertension. PGD's biochemical role probably excludes it from being involved in this trait, but another gene close to *PGD* is implicated. *PND* is one candidate in 1p36.

PND is the precursor of atrial natriuretic factors (ANF), which have natriuretic and diuretic activity, and the ability to relax precontracted vascular smooth

muscle (Currie *et al,* 1983). The latter function raised the possibility that PND may be involved in the regulation of blood pressure. Also a deficiency in this protein in degenerating hearts of hamsters is the cause of inadequate renal sodium and water excretion, as well as the syndrome of congestive heart failure. No significant difference was observed in the arterial blood pressure however (Chimoskey *et al,* 1984).

Linkage of this locus to hypertension and of the *renin* locus on 1q, has been investigated and excluded by sib pair analysis (IBS) in normally distributed blood pressure groups (N. Carter, personal communication, 1991)

The genetic distance between *PGD* and *PND* established in this thesis does not exclude the existence of a locus predisposing to hypertension which is linked to *PGD* but not to *PND.* In addition, in humans hypertension is a risk factor in the development of cardiovascular disease whereas in hamsters an increase in blood pressure is not observed in the myopathic strain (Chimoskey *et al,* 1984). Thus it may be that the human disease is not an exact reflection of the hamster disease.

Another candidate gene for hypertension was the *Na^+/H^+ antiporter (APNH).* This is involved in pH homeostasis and operates in cells as a major H^+-extruding system. This locus was mapped by linkage analysis 3cM proximal to *RH* (Lifton *et al,* 1990). A test for linkage of *APNH* to genes predisposing to hypertension by analysis of hypertensive sib pairs, showed that mean allele sharing at *APNH* was not greater than expected from random assortment in hypertensive sibs (Lifton *et al,* 1991). This finding is similar to that of Wilson and others (1992) where there is no linkage to *RH,* a locus mapped 3 cM proximal to *APNH* (Lifton *et al, 1*991).

Studies in mouse models sensitive to hypertension with salt administration show cosegregation of the *renin* gene with hypertension, suggesting that it is directly or closely linked to one of the genes regulating blood pressure (Rapp *et al*, 1989). Further studies show that in hypertensive rats a candidate locus maps to a region homologous to human chromosome 17q which contains the *ACE1* gene, encoding the angiotensin1-converting enzyme (Hilbert *et al*, 1991). The latter works synergistically with renin so this explains the results of Rapp and others (1989). Current evidence shows no linkage between *ACE1* in humans and hypertension, but there is linkage between the disease and the *angiotensinogen* gene itself (Jeunemaitre *et al*, 1992).

CHAPTER 5 APPENDIX

5.1 STOCK SOLUTIONS

Solutions Used in DNA Isolation and Buffers.

Lysozyme Buffer.

50mM Glucose

25mM Tris. HCl. pH8

10mM EDTA

(no lysozyme added)

Alkaline SDS.

0.2N Sodium hydroxide

1% SDS

Acid Potassium Acetate

3M Potassium acetate

11.5% Glacial Acetic Acid.

Yeast Resuspension Buffer.

1.2M Sorbitol

10mM Tris. HCl pH 7.5

20mM EDTA

Yeast Lysis Buffer.

100mM EDTA

10mM Tris. HCl pH 7.5

1% Lithium dodecyl sulphate

2.5M Potassium Phosphate Buffer pH 8.0.

A 2.5M $K_2HPO_4.3H2O$

B 2.5M $KH_2PO_4.3H2O$

100ml of A was adjusted with 18ml of B to pH 8.0.

TE.

10mM Tris. HCl pH 8.0

1mM EDTA

STE.

100mM Sodium chloride

10mM Tris HCl. pH 8.0

10mM EDTA

20X SSC.

3M Sodium Chloride

0.3M Sodium Citrate

pH 7-7.5

5X TAE Electrophoresis Buffer.

0.2M Tris.HCl pH8.0

0.1M Sodium Acetate

5mM EDTA

5X TBE Electrophoresis Buffer.

0.45M Tris.HCl pH8.0

0.45M Boric Acid

10mM EDTA

20X Denhardt's Solution.

4g/L Bovine Serum Albumin

4g/l PVP360

4g/L Ficoll

0.4M Phosphate Buffer pH6.5 (prepared using 1M Na_2PO_4 and 1M NaH_2PO_4 adjusted to pH 6.5 and then 400ml used in 1L of above solution.)

Salmon Sperm DNA.

Salmon sperm DNA was dissolved at 5mg/ml in deionized water by boiling for 10 min. The volume was adjusted to 50ml and then the DNA was quenched on ice. It was then sheared by sonication for 10 min.

5.2. BACTERIAL and YEAST CELL CULTURE MEDIA.

Bacterial Culture Media

L-broth.

1% Tryptone

0.5% Yeast Extract

0.5% Sodium chloride

0.1% Glucose

L-agar.

1% Tryptone

0.5% Yeast Extract

0.5% Sodium chloride

1.5% Agar

0.1% Glucose

Yeast Culture Medium.

S.D medium.(synthetic dextrose minimal medium)

7g/L Yeast Nitrogen Base (without amino acids)

20g/L Glucose

55mg/L adenine and tyrosine

Sterilized by autoclaving.

Before use 56ml/L 20% Casamino acids were added, filtered first.

This gives selection under TRP and URA.

5.3 MOLECULAR WEIGHT MARKERS

PhiX 174 (kb) (x HaeIII)	1 kb Ladder.	Saccharomyces Cerevisiae (YNN295)Chromosomes, kb
1.35	12.2	1900
1.08	11.2	1640
0.87	10.2	1120 & 1100
0.60	9.2	945
0.31	8.1	915
0.28	7.1	815
0.27	6.1	785
0.23	5.1	745
0.19	4.1	680
0.12	3.1	610
0.07	2.0	555
	1.6	450
	1.02	375
	0.52	295
	0.39	225
	0.34	
	0.30	
	0.21	

50 kb and 39 kb Lambda concatemers were also used in PFGE. The 39 kb lambda concatemers were of a restricted mutant phage DNA with a monomer size of 39 kb. Both were purchased from Promega.

5.4 PCR PRIMER SEQUENCES and PCR CONDITIONS.

Locus	Primers	PCR Product Size	^0C	$[Mg^{2+}]$ mM
HSPG2	GCGTCAGCAGCCAGTGTGAT GTGGGTTCACCAGGGCAAAG	700	60	1.5
PND	TGGGTGTTGGGGCAGAAACT GGGCAGGATGGACAGGATTG	171	52	2.0
PGD	GATGATGTCACCAGGTATCC GTCTGTGCTTTTAATAGGAC	200	50	2.0
p58*	TGAGTGCCCTGGGCTTGG (1) TTCCTTGGCACCAAGCAG (2)		50	1.5
	GGGTCCACATGGGAGCCT (3) CTCTGGGGCGCGGTACCA (4)	342	50	6.0

^0C- Annealing Temparature.

The annealing temparature is calculated as follows: Tm - 5 ^0C

Tm = 2 X (A+T) + 4 X (G+C)

* Primers 1 and 2 were used in a PCR which was run for 20 cycles. A 5 microlitre aliquot was then transferred into a second reaction mix with nested primers 3 and 4.

10% DMSO was added in both of the p58 reactions.

5.5 DNA SEQUENCING PRIMERS

Vector Primers.	Sequence 5'->3'
Lambda gt l0 (forward) 24mer	AGCAAGTTCAGCCTGGTTAAG
pBR322 EcoRl Site (forward) l5mer	GTATCACGAGGCCCT
pBR322 EcoRl Site (reverse) l5 mer	GATAAGCTGTCAAAC

PGD Primers.	Sequence 5'->3'
5AMP (reverse) 20 mer	GATGATGTCACCAGGTATCC
5(+) (forward) 20mer	GTCTGTGCTTTAATAGGAC
PGD 1 (reverse) l7 mer	AACTTTGGAGACAGTCC
PGD 2 (reverse) 16 mer	CCTTTGCCTCATTGGC
E4 internal primer (forward) l5 mer	GCATAATGAACATGG
El0 internal primer (forward)	GGCACTACAGGCGCAT

5.6 D1F15S1 PHENOTYPES FOR CEPH PEDIGREES

ID no.	Phenotype	ID no.	Phenotype
1201	+ - +	142101	+ - +
1202	+ + +	142102	+ + +
1203	+ + +	142103	+ - +
1204	+ - +	142104	- + -
1205	- - -	142106	+ - +
1206	+ - +	142107	+ - +
1207	+ + +	142108	+ - +
1208	+ - +	142109	+ + +
1209	+ - +	142110	- + -
1210	+ - +	142111	+ - +
1211	+ - +	142112	+ - +
1212	- - -	142113	+ + +
1213	+ + +	142113	+ + +
		142114	+ + +
2801	- + -		
2802	+ + +		
2803	- + -	142301	+ - +
2804	+ + +	142302	+ - +
2805	+ + +	142303	+ - +
2807	+ - +	142304	+ - +
2808	+ - +	142306	+ - +
2809	- + -	142308	+ - +
		142309	+ - +
88401 - 88417		142310	+ - +
(All + + +)			
		142401	- + -
133401	- + -	142402	+ + +
133402	+ - +	142403	- + -
133403	+ - +	142404	- + -
133405	+ - +	142405	+ + +
133406	+ + +	142406	- - -
133407	- + -	142407	+ + +
133408	+ - +	142408	- - -
133409	- + -	142409	- + -
133410	+ - +	142410	- - -
133411	- + -	142412	- - -

133412	- - -
133413	+ - +
134001	+ + +
134002	+ + +
134003	+ + +
134004	+ + +
134005	+ - +
134006	+ + +
134007	+ + +
134009	+ + +
134010	+ - +
134011	+ - +
134012	+ + +
134012	+ + +
134701	+ + +
134702	- + -
134703	+ + +
134704	- + -
134705	+ + +
134706	+ + +
134707	- + -
134708	+ + +
134709	+ + +
134710	- - -
134711	- - -
134712	+ + +
134713	- + -
134714	- + -
134715	+ + +
134716	+ + +

134901 - 134908
(All + + +)

142413	+ + +
142414	+ - +
1329401	+ + +
1329402	+ - +
1329405	+ - +
1329406	+ - +
1329407	+ + +
1329408	- + -
1329409	+ + +
1329410	+ - +
1329412	+ - +
136201	+ + +
136202	- + -
136203	+ - +
136204	+ + +
136205	- + -
136206	- + -
136208	+ - +
136209	+ + +
136210	+ - +
136211	- + -
136212	+ - +
136214	- + -
136215	+ + +
136216	- - -
136217	- + -

140801 - + -
140802 + + +
140803 - + -
140804 + - +
140806 - + -
140807 - + -
140808 + - +
140809 + - +
140810 - + +
140811 - - -
140812 + - +
140813 - + -
140814 + - +

141601 + + +
141602 - + -
141603 + + +
141604 - + -
141605 + + +
141607 - + -
141609 - + -
141610 + + +
141611 + - +
141612 + + +
141614 + + +
141615 + + +
141616 + + +

141801 + + +
141802 + - +
141803 + + +
141805 + + +
141806 + - +
141807 + + +
141809 + + +

141810 + + +
141811 + + +
141812 + + +
141813 + - +
141814 + - +

REFERENCES

Abbott C. and Povey S. (1991). Development of Human Chromosome-Specific PCR Primers for Characterization of Somatic Cell Hybrids. GENOMICS 9; 73-77.

Abidi F.E., Wada M., Little R.D. and Schlessinger D. (1990). Yeast Artificial Chromosomes Containing Human Xq24-Xq28 DNA: Library Construction and Representation of Probe Sequences. GENOMICS 7; 363-376.

Aissani B. and Bernardi G. (1991). CpG islands: features and distribution in the genomes of vertebrates. Gene, 106; 173-183.

Albertson H.M., Le Paslier D., Abderrahim H., Dausset J., Cann H., and Cohen D. (1989). Improved control of partial DNA restriction enzyme digest in agarose using limiting concentrations of Mg++. Nucleic Acids Research 17: 808.

Anand R., Ogilvie D.J., Butler R., Riley J.H., Finniear R.S., Powell S.J., Smith J.C. and Markham A.F. (1991). A Yeast Artificial Chromosome Contig Encompassing the Cystic Fibrosis Locus. GENOMICS 9; 124-130.

Anand R., Riley J. H., Butler R., Smith J. C. and Markham A. F. (1990). A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. Nucleic Acids Research, 18 No. 8; 1951-1956.

Antequera F., Boyes J.and Bird A. (1990). High Levels of De Novo Methylation and Altered Chromatin Structure at CpG Islands in Cell Lines. Cell 62; 503-514.

Arenstoff H.P., Kandpal R.P., Baskaran N., Parimoo S., Tanaka Y., Kitajima S., Yasukochi Y. and Weissman S.M. (1991). Construction and Characterization of a NotI-BsuE Linking Library from the Human X Chromosome. GENOMICS 11; 115-123.

Baldini A., Ross M., Nizetic D., Vatcheva R., Lindsay E.A., Lehrach H. and Siniscalco M. (1992). Chromosomal Assignment of Human YAC Clones by Fluorescence *in Situ* Hybridization: Use of Single-Yeast-Colony PCR and Multiple Labeling. GENOMICS 14; 181-184.

Bale S.J., Chakravarti A. and Greene M.H. (1986). Cutaneous Malignant Melanoma and Familial Dysplastic Nevi: Evidence for Autosomal Dominance and Pleiotropy. Am. J. Hum. Genet. 38; 188-196.

Bale S. J., Dracopoli N.C., TuckerM. A., Clark W. H., Fraser M.C., Stanger B. Z., Green P., Donis-Keller H., Housman D. E., and Greene M. H. (1989). Mapping the Gene for Hereditary Cutaneous Malignant Melanoma-Dysplastic Nevus to Chromosome lp. The New England Journal of Medicine. 320; 1367-1371.

Barker D., Schafer M., and White R. (1984). Restriction Sites Containing CpG Show a Higher Frequency of Polymorphism in Human DNA. Cell, 36; 131-138.

Behn-Krappa A., Holker I., De Silva U.S. and Doerfler W. (1991). Patterns of DNA Methylation Are Indistinguishable in Different Individuals over a Wide Range of Human DNA Sequences. GENOMICS 11; 1-7.

Biegel J.A., White P. S., Marshall H.N., Fujimori M., Zackai E. H., Scher C. D., Brodeur G. M., Emanuel B.S. (1993). Constitutional lp36 Deletion in a Child with Neuroblastoma. Am. J. Hum. Genet. 52: 176-182.

Bird A., Taggart M., Frommer M., Miller O. J., and Macleod D. (1985). A Fraction of the Mouse Genome That is Derived from Islands of Nonmethylated, CpG-Rich DNA. Cell. 40: 91-99.

Bird A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature 321; 209-213.

Bird A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends in Genetics , 3, No.12; 342-347.

Botstein D., White R .L., Skolnick M. and Davis R. W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. Am. J. Hum. Genet. 32; 314-331.

Bowman J. E., Carson P. E. , Frischer H. and de Garay A. L. (1966). Genetics of Starch-Gel Electrophoretic Variants of Human 6-Phosphogluconic Dehydrogenase: Population and Family Studies in the United States and in Mexico. Nature 210; 811-813.

Breen M., Arveiler B., Murrray I., Gosden J.R. and Porteous D.J. (1992). YAC Mapping by FISH Using Alu-PCR-Generated Probes. GENOMICS 13; 723-730.

Brockdorff N., Montague M., Smith S. and Rastan S. (1990). Construction and Analysis of Linking Libraries from the Mouse X Chromosome. GENOMICS 7; 573-578.

Brodeur G.M. and Fong C. (1989). Molecular Biology and Genetics of Human Neuroblastoma. Cancer Genet Cytogenet 41; 153-174.

Bronson S.K., Pei J., Taillon-Miller P., Chorney M.J., Geraghty D. E. (1991). Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. Proc. Natl. Acad. Sci. USA 88:1676-1680.

Brook J.D., Zemelman B.V., Hadingham K., Siciliano M.J., Crow S., Harley H.G., Rundle S.A., Buxton J., Johnson K., Almond J.W., Housman D.E., and Shaw D.J. (1992). Radiation-Reduced Hybrids for the Myotonic Dystrophy Locus. GENOMICS 13; 243-250.

Brooks-Wilson A.R., Goodfellow P.N., Povey S., Nevanlinna H.A., De Jong P.J. and Goodfellow P.J. (1990). Rapid Cloning and Characterization of New Chromosome 10 DNA Markers by Alu Element-Mediated PCR. GENOMICS 7; 614-620.

Brownstein B. H. , Silverman G.A. , Little R.D. , Burke D.T. , Korsmeyer S.J. , Schlessinger D. , Olson M.V. (1989) . Isolation of Single Copy Human Genes from a Library of Yeast Artificial Chromosome Clones.Science 244; 1348-1351.

Bunnell B., Heath L. S., Adams D. E., Lahti J. M., and Kidd V. J. (1990). Elevated expression of a p58 protein kinase leads to changes in the CHO cell cycle. Proc. Nat. Acad. Sci USA. 87; 7467-7471.

Burke D.T. , Carle G.F. and Olson M.V. (1987). Cloning of Large Segments of Exogenous DNA into Yeast by Means of Artificial Chromosome Vectors. Science 236; 806-812.

Burmeister M., diSibio G., Cox D. R. and Myers R.M. (1991a). Identification of polymorphisms by genomic denaturing gradient gel electrophoresis: application to the proximal region of human chromosome 21. Nucleic Acids Research, 19 No.7: 1475-1481.

Burmeister N., Kim S., Price E.R., De Lange T., Tantravahi U., Myers R.M. and Cox D.R. (1991b). A Map of the Distal Region of the Long Arm of Human Chromosome 21 Constructed by Radiation Hybrid Mapping and Pulsed-Field Gel Electrophoresis. GENOMICS 9; 19-30.

Buroker N., Bestwick R., Haight G., Magenis R. E., and Litt M. (1987). A hypervariable repeated sequence on human chromosome lp36. Hum Genet 77; 175-181.

Butler R. , Ogilvie D.J. , Elvin P. , Riley J.H., Finniear R.S., Slynn G., Morten J.E.N., Markham A.F. and Anand R. (1992). Walking, Cloning, and Mapping with Yeast Artificial Chromosomes: A Contig Encompassing D21S13 and D21S16. GENOMICS 12; 42-51.

Cangiano G., Ameer H., Waterston R. and La Volpe A. (1990). Use of repetetive DNA probes as physical mapping strategy in *Caenorhabditis elegans*. Nucleic Acids Research, 18, No. 17; 5077-5081.

Cannon-Albright L.A., Goldgar D.E., Wright E.C., Turco A., Jost M., Meyer L.J., Piepkorn M., Zone J.J. and Skolnick M.H. (1990). Evidence against the Reported Linkage of the Cutaneous Melanoma-Dysplastic Nevus Syndrome Locus to Chromosome lp36. Am. J. Hum. Genet. 46; 912-918.

Carle G. F. and Olson M.V. (1985). An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA 82:3756-3760.

Carle G. F., Frank M., Olson M. V. (1986). Electrophoretic Separations of Large DNA Molecules by Periodic Inversion of the Electric Field. Science, 232: 65-68.

Carne A. and Walker J. E. (1983). Amino Acid Sequence of Ovine 6-Phosphogluconate Dehydrogenase. THE JOURNAL OF BIOLOGICAL CHEMISTRY, 258 No. 21: 12895-12899.

Carritt B., King J. and Welch H.M. (1982a). Gene order and localization of enzyme loci on the short arm of chromosome 1. Ann. Hum. Genet. 46; 329-335.

Carritt B., Parrington J.M., Welch H.M. and Povey S. (1982b). Diverse origins of multiple ovarian teratomas in a single individual. Proc. Natl. Acad. Sci. USA 79; 7400-7404.

Carritt B., Welch H.M. and Parry-Jones N.J. (1986). Sequences Homologous to the Human DlSl Locus Present on Human Chromosome 3. Am J Hum Genet 38; 428-436.

Carritt B., Kok K., van der Berg A., Osinga J., Pilz A., Hofska R.M.W., Rabbitts P.H., Gulati K., and Buys C.H.C.M. (1992). A gene for human chromosome region 3p21 with reduced expression in small cell lung cancer. Cancer. Res. 52: 1536-1541.

Chakravarti S ., Phillips S.L ., Hassell J.R. (1991) Assignment of the Perlican (Heparan-Sulphate-proteoglycan) gene to mouse chromosome 4. Mammalian Genome 1; 270-272.

Chen L.C., Dollbaum C. and Smith H.S. (1989). Loss of heterozygosity on chromosome lq in human breast cancer. Proc. Natl. Acad. Sci. USA; 7204-7207.

Chimoskey J.E., Spielman W.S., Brandt M.A. , Heidemann S.R. (1984). Cardiac Atria of BIO 14.6 Hamsters Are Deficient in Natriuretic Factor. Science 223; 820-822.

Chodirker B. N., Evans J. A., Lewis M., Coghlan G., Belcher E., Philipps S., Seargeant L. E., Sus C., and Greenberg C. R. (1987). Infantile Hypophosphatasia - Linkage with the RH Locus. GENOMICS 1; 280-282.

Christiansen H. and Lampert F. (1988). Tumour karyotype discriminates between good and bad prognostic outcome in neuroblastoma. Br. J. Cancer 57; 121-126.

Chu G., Vollrath D., Davis R. W. (1986). Separation of Large DNA Molecules by Contour-Clamped Homogeneous Electric Fields. Science , 234: 1582-1585.

Collins F.S. and Weissman S.M. (1984). Directional cloning of DNA fragments at a large distance from an initial probe: A circularization method. Proc. Natl. Acad. Sci. USA 81; 6812-6816.

Collins A., Keats B.J., Dracopoli N., Shields D.C., and Morton N.E. (1992). Integration of gene maps: Chromosome 1. Proc. Natl. Acad. Sci, USA 89; 4598-4602.

Conboy J., Mohandas N., Tchernia G. and Kan Y.W. (1986). Molecular basis of hereditary elliptocytosis due to protein 4.1 deficiency. New Eng. J. Med. 315; 680-685.

Cook P.J.L. and Robson E.B. (1974). Segregation of genetic markers in families with chromosome polymorphisms and structural rearrangements involving chromosome 1. Ann. Hum. Genet. Lond. 37; 261-274.

Cook P.J.L., Noades J.E., Newton M.S. and Mey R.D. (1977). On the orientation of the Rh: El linkage group. Ann. Hum. Genet. Lond. 41; 157-162.

Coulondre C. & Miller J.H., Farabaugh P.J. & Gilbert W. (1978). Molecular basis of base substitution hotspots in Escherichia coli. Nature 274; 775-780.

Coulson A., Sulston J., Brenner S. and Karn J. (1986). Toward a physical map of the genome of the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA 83; 7821-7825.

Coulson A., Waterston R., Kiff J., Sulston J. and Kohara Y. (1988).Genome linking with yeast artificial chromosomes. Nature 335; 184-186.

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.

Currie M. G., Geller D. M., Cole B. R., Boylan J. G., Yusheng W., Holmberg S. W., Needleman P. (1983). Bioactive Cardiac Substances: Potent Vasorelaxant Activity in Mammalian Atria. Science, 221 ; 71-73.

D'Urso M., Zucchi I., Ciccodicola A., Palmieri G., Abidi F.E. and Schlessinger D. (1990). Human Glucose-6-phosphate Dehydrogenase Gene Carried on a Yeast Artificial Chromosome Encodes Active Enzyme in Monkey Cells. GENOMICS 7; 531-534.

Darby J. K., Willems P. J., Nakashima P., Johnsen J., Ferrell R. E., Wijsman E. M., Gerhard D. S., Dracopoli N. C., Housman D., Henke J., Fowler M. L., Shows T. B., O'Brien J.S., and Cavalli-Sforza L.L. (1988). Restriction Analysis of the Structural alpha-L-Fucosidase Gene and its Linkage to Fucosidosis. Am. J. Hum. Genet. 43; 749-755.

Darling S. M. and Abbott C. M. (1992). Mouse models of human single gene disorders. BioEssays 14; 359-366.

Dausset J., Cann H., Cohen D., Lathrop M., Lalouel. J-M., and White R. (1990). Centre d'Etude du Polymorphism Humaine (CEPH): GENOMICS 6 575-577.

Davisson M.T., Lalley P.A., Peters J., Doolittle D.P., Hillyard A.L., and Searle A.G. (1991). Report of the comparative committee for human, mouse and other rodents. Cytogenet Cell Genet 58: 1152-1189.

Dean M., Lucas-Derse S., Bolos A., O'Brien S.J., Kirkness E.F., Fraser C.M. and Goldman D. (1991). Genetic Mapping of the beta 1 GABA Receptor Gene to Human Chromosome 4, Using a Tetranucleotide Repeat Polymorphism. Am. J. Hum. Genet. 49; 621-626.

Dodge G. R., Kovalszky I., Chu M.L., Hassell J. R., McBride O.W., Yi H. F., Iozzo R. V. (1991). Heparan Sulfate Proteoglycan of Human Colon: Partial Molecular Cloning, Cellular Expression, and Mapping of the Gene (HSPG2) to the Short Arm of Human Arm of Human Chromosome 1. GENOMICS: 10, 673-680.

Donis-Keller H., Green P., Helms C., Cartinhour S., Weiffenbach B., Stephens K., Keith T.P., Bowden D.W., Smith D.R., Lander E.S., Botstein D., Akots G., Rediker K.S., Gravius T., Brown V.A., Rising M.B., Parker C., Powers J.A., Watt D.E., Kauffman E.R., Bricker A., Phipps P., Muller-Kahle H., Fulten T.R., Schumm J.W., Braman J.C., Knowlton R.G., Barker D.F., Crooks S.M., Lincoln S.E ., Daly M.J. and Abrahamson J. (1987). A Genetic Linkage Map of the Human Genome. Cell 51; 319-337.

Dracopoli N. C., Rose E., Whitfield G. K., Guidon P. T., Bale S. J., Chance P. A., Kourides I. A. and Housman D. E. (1988). Two Thyroid Hormone Regulated Genes, the beta-Subunits of Nerve Growth Factor (NGFB) and Thyroid Stimulating Hormone (TSHB) , Are Located Less Than 310 kb Apart in Both Human and Mouse Genomes. GENOMICS 3, 161-167.

Dracopoli N.C., Stanger B.Z., Ito C.Y., Call K.M., Lincoln S.E., Lander E.S. and Housman D.E. (1988). A Genetic Linkage Map of 27 Loci from PND To FY on the Short Arm of Human Chromosome I. Am. J. Hum. Genet. 43; 471-475.

Dracopoli N.C., Harnett P., Bale S.J., Stanger B.Z., Tucker M.A., Housman D.E. and Kefford R.F. (1989). Loss of alleles from the distal short arm of chromosome 1 occurs late in melanoma tumor progression. Proc. Natl. Acad. Sci. USA 86; 4614-4618.

Dracopoli N.C., O'Connell P., Elsner T.I., Lalouel J.M., White R.L., Buetow K.H., Nishimura D.Y., Murray J.C., Helms C., Mishra S.K., Donis-KellerH., Hall J.M., LeeM.K., King M.C., Attwood J., Morton N.E., Robson E.B., Mahtani M., Willard H.F., Royle N.J., Patel I., Jeffreys A.J., Verga V., Jenkins T., Weber J.L., Mitchell A.L and Bale A.E. (1991). The CEPH Consortium Linkage Map of 1p. GENOMICS 9; 686-700.

Driesen M. S., Dauwerse J. G., Wapenaar M. C., Meershoek E. J., Mollevanger P., Chen K. L., Fischbeck K. H. and Van Ommen G. J. B. (1991). Generation and Fluorescent in Situ Hybridization Mapping of Yeast Artificial Chromosomes of 1p, 17p, 17q, and 19q from a Hybrid Cell Line by High Density Screening of an Amplified Library. GENOMICS 11.; 1079-1087.

Drumm M.L., Smith C.L., Dean M., Cole J.L., Iannuzzi M.C. and Collins F.S. (1988). Physical Mapping of the Cystic Fibrosis Region by Pulsed-Field Gel Electrophoresis. GENOMICS 2; 346-354.

Eaton J.W., and Brewer G.J. (1974). in " The Red Blood Cell ". (Surgenor D.M., ed). Academic Press, New York.

Eaton J.W., Boraas M., and Etkin N.L. (1972). in " Haemoglobin and Red Cell Structure and Function" (Brewer G., ed). p121. Plenum Press, New York.

Economou E.P., Bergen A.W., Warren A.C. and Antonarakis S.E. (1990). The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc. Natl. Acad. Sci. USA 87; 2951-2954.

Eipers P. G., Lahti J. M. and Kidd V. J. (1992). Structure and Expression of the Human p58clk-1 Protein Kinase Chromosomal Gene. GENOMICS 13; 613-621.

Eipers P.G., Barnoski B.L., Han J., Carroll A.J. and Kidd V.J. (1991). Localization of the Expressed Human p58 Protein Kinase Chromosomal Gene to Chromosome lp36 and a Highly Related Sequence to Chromosome 15. GENOMICS 11; 621-629.

Eliceiri B., Labella T., Hagico Y., Srivastava A., Schlessinger D., Pilia G., Palmieri G. and D'Urso M. (1991). Stable integration and expression in mouse cells of yeast artificial chromosomes harboring human genes. Proc. Natl. Acad. Sci. USA 88;2179-2183.

Evans G.A. and Lewis K.A. (1989). Physical mapping of complex genomes by cosmid multiplex analysis. Proc. Natl. Acad. Sci. USA 86; 5030-5034.

Fan J.B., Grothues D. and Smith C.L. (1991). Alignment of Sfi I sites with the Not I restriction map of Schizosaccharomyces pombe genome. Nucleic Acids Research, 19, No. 22; 6289-6294.

Feingold J.M., Ogden S.D. and Denny C.T. (1990). Streamlined approach to creating yeast artificial chromosome libraries from specialized cell sources. Proc. Natl. Acad. Sci. USA 87; 8637-8641.

Fildes R. A., Parr C. W. (1963). Human Red-cell Phosphogluconate Dehydrogenases. Nature 200; 890-891.

Fong C., Dracopoli N.C., White P.S., Merrill P.T., Griffith R.C., Housman D.E. and Brodeur G.M. (1989). Loss of heterozygosity for the short arm of chromosome 1 in human neuroblastomas: Correlation with N-myc amplification. Proc. Natl. Acad. Sci. USA 86; 3753-3757.

Fountain J. W., Karayiorgou M., Taruscio D., Graw S. L., Buckler A. J., Ward D.C., Dracopoli N.C. and Housman D. E. (1992). Genetic and Physical Map of the Interferon Region on Chromosome 9p. GENOMICS, 14; 105-112.

Fountain J.W., Karayiorgou M., Graw S.L., Buckler A.J., Taruscio D., Ward D.C., Ernstoff M. S., Kirkwood J.M., Andersen L.B., Collins F.S., Dracopoli N.C., Housman D.E. (1992). Chromosome 9p Involvement in Melanoma. Am. J. Hum. Genet. Suppl. 49;45.

Gebhart E., Rau D., Neubauer S., and Dingermann T. (1992). Chromosome 1 in human colorectal cancer. Hum. Genet. 90: 188-190.

Gemmill R.M., Coyle-Morris J.F., McPeek F.D., Ware-Uribe L.F., and Hecht F. (1987). Construction of Long-Range Restriction Maps in Human DNA Using Pulsed Field Gel Electrophoresis. Gene Anal Techn 4:119-131.

Genuardi M., Tsihira H., Anderson D.E. and Saunders G.F. (1989). Distal Deletion of Chromosome ip in Ductal Carcinoma of the Breast. Am. J. Hum. Genet. 45; 73-82.

Giallongo A., Feo S., Moore R., Croce C. M., and Showe L. C. (1986). Molecular cloning and nucleotide sequence of a full-length cDNA for human alpha enolase. Proc. Natl. Acad. Sci. USA 83; 6741-6745.

Giblett E.R., Chen S.H., Anderson J.E and Lewis M. (1973). A family study suggesting genetic linkage of phosphopyruvate hydratase (enolase) to the RH blood group system. Human Gene Mapping. 1: 91-92.

Gilbert F., Balaban G., Moorhead P., Bianchi D. and Schlesinger H. (1982). Abnormalities of Chromosome lp in Human Neuroblastoma Tumors and Cell Lines. Cancer Genetics and Cytogenetics 7; 33-42.

Glock G.E. and McLean P. (1954). Levels of Enzymes of the Direct Oxidative Pathway of Carbohydrate Metabolism in Mammalian Tissues and Tumours. Biochemical Journal 56; 171-175.

Glock G. E. and McLean P. (1955). A Preliminary Investigation of the Hormonal Control of the Hexose Monophosphate Oxidative Pathway. Biochemical Journal 6l; 390-396.

Goode M.E., VanTuinen P., Ledbetter D.H. and Daiger S.P. (1986). The Anonymous Polymorphic DNA Clone DlSl, Previously Mapped to Human Chromosome lp36 by in Situ Hybridization, Is from Chromosome 3 and Is Duplicated on Chromosome l. Am J Hum Genet. 38; 437-446.

Goradia T.M., Stanton, JR., V.P., Cui X., Aburatani H., Li H., Lange K., Housman D.E. and Arnheim N. (1991). Ordering Three DNA Polymorphisms on Human Chromosome 3 by Sperm Typing. GENOMICS 10; 748-755.

Goss S.J. & Harris H. (1975). New method for mapping genes in human chromosomes. Nature 255; 680-684.

Green E. D. , Riethman H.C. , Dutchik J.E. and Olson M.V. (1990). Detection and Characterization of Chimeric Yeast Artificial-Chromosome Clones. GENOMICS 11; 658-669.

Green E.D. and Olson M.V. (1990). Chromosomal Region of the Cystic Fibrosis Gene in Yeast Artificial Chromosomes: A Model for Human Genome Mapping. Science 250; 94-98.

Green E.D., Mohr R.M., Idol J.R., Jones M. , Buckingham J.M. , Deaven L.L. , Moyzis R.K. and Olson M.V. (1991). Systematic Generation of Sequence-Tagged Sites for Physical Mapping of Human Chromosomes: Application to the Mapping of Human Chromosome 7 Using Yeast Artificial Chromosomes. GENOMICS 11; 548-564.

Greene M.H., Goldin L.R., Clark W.H., Lovrien E., Kraemer K.H., Tucker M.A., Elder D.E., Fraser M.C. and Rowe S. (1983). Familial cutaneous malignant melanoma: Autosomal dominant trait possibly linked to the Rh locus. Proc. Natl. Acad. Sci. USA 80; 6071-6075.

Gusella J. F., Wexler N. S., Conneally P. M., Naylor S. L., Anderson M. A., Tanzi R. E., Watkins P. C., Ottina K., Wallace M.R., Sakuchi A. Y., Young A. B., Shoulson I., Bonilla E., and Martin J. B. (1983) A polymorphic DNA marker genetically linked to Huntington's disease. Nature 306. 234-238.

Guzzetta V., De Oca-Luna R.M., Lupski J.R. and Patel P.I. (1991). Isolation of Region-Specific and Polymorphic Markers from Chromosome 17 by Restricted Alu Polymerase Chain Reaction. GENOMICS 9; 31-36.

Haldane J. B. S. (1919). The Combination of Linkage Values, and the Calculation of Distances between the Loci of Linked Factors. Journal of Genetics 8; 299-309.

Hamada H. & Kakunaga T. (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature 298; 396-398.

Hanish J., Rebelsky M., McClelland M. and Westbrook C. (1991). Application of Methylase-Limited Partial Notl Cleavage for a Long-Range Restriction Map of the Human ABL Locus. GENOMICS 10; 681-685.

Harper M.E., and Saunders G. F. (1981). Localization of single copy sequences on G-banded human chromosomes by in situ hybridization. Chromosoma, 83: 431-439.

Haseman J.K., Elston R.C. (1972). The estimation of linkage between a quantitative trait and a marker locus. Behav. Genet. 2: 3-19.

Hazan J., Dubay C., Pankowiak M.P., Becuwe N. and Weissenbach J. (1992). A Genetic Linkage Map of Human Chromosome 20 Composed Entirely of Microsatellite Markers. GENOMICS 12; 183-189.

Higgins M.J., Turmel C., Noolandi J., Neumann P.E. and Lalande M. (1990). Construction of the physical map for three loci in chromosome band 13q14: Comparison to the genetic map. Proc. Natl. Acad. Sci. USA 87; 3415-3419.

Hilbert P., Lindpaintner K., Beckmann J. S., Serikawa T., Soubrier F., Dubay G., Cartwright P., De Gouyon B., Julier C., Takahasi S., Vincent N., Ganten D., Georges M. and Lathrop G. M. (1991). Chromosomal mapping of two genetic loci associated with blood pressure regulation in hereditary hypertensive rats. Nature 353; 521-529.

Horecker B. L., Smyrniotis P. Z. and Seegmiller J. E. (1951). The Enzymatic Conversion of 6-Phosphogluconate to Ribulose-5-Phosphate and Ribose-5-Phosphate. Journal of Biological Chemistry 193; 383-395.

Hulten M.A., Palmer R.W. and Laurie D.A. (1982). Chiasma derived genetic maps and recombination fractions:chromosome 1. Ann. Hum. Genet. 46; 167-175.

Hunt J.D. and Tereba A. (1990). Molecular Evaluation of Abnormalities of the Short Arm of Chromosome 1 in Neuroblastoma. Genes, Chromosomes & Cancer 2; 137-146.

Huxley C., Hagino Y., Schlessinger D. and Olson (1991). The Human HPRT Gene on a Yeast Artificial Chromosome Is Functional When Transferred to Mouse Cells by Cell Fusion. GENOMICS 9; 742-750.

Imai T. and Olson M.V. (1990). Second-Generation Approach to the Construction of Yeast Artificial-Chromosome Libraries. GENOMICS 8 ; 297-303.

Jarvela I., Schleutker J., Haataja L., Santavuori P., Puhakka L., Manninen T., Palotie A., Sandkuul L. A., Renlund M., White R., Aula P., and Peltonen L. (1991). Infantile Form of Nueronal Ceroid Lipofuscinosis (CLNl) Maps to the Short Arm of Chromosome 1. GENOMICS 9; 170-173.

Jeffreys A.J., Wilson V. 7 Thein S.L. (1985). Hypervariable 'minisatellite' regions in human DNA. Nature 314; 67-72.

Jeffreys A.J., Neumann R. and Wilson V. (1990). Repeat Unit Sequence Variation in Minisatellites: A Novel Source of DNA Polymorphism for Studying Variation and Mutation by Single Molecule Analysis. Cell, 60; 473-485.

Jeffreys A.J., MacLeod A., Tamaki K., Neil D.L. & Monckton D.G. (1991). Minisatellite repeat coding as a digital approach to DNA typing. Nature 354; 204-209.

Jenkins R.N., Osbornee-Lawrence S .L., Sinclair A. K., Eddy R. L., Byers M. G., Shows T. B., Duby A. D. (1990). Structure and Chromosomal Location of the Human

Gene Encoding Cartilage Matrix Protein. The Journal of Biological Chemistry, 265, No.32; 19624-19631.

Jeunemaitre X., Soubrier F., Kotelevtsev Y. V., Lifton R. P., Williams C. S., Charru A., Hunt S. C., Hopkins P. N., Williams R. R., Lalouel j., and Corvol P. (1992). Molecular Basis of Human Hypertension : Role of Angiotensinogen. Cell 71: 169-180.

John S.W.M., Weitzner G., Rozen R., and Scriver C.R. (1990). A rapid procedure for extracting genomic DNA from leukocytes. Nucleic..Acids. .Research. 19 No.2: 408.

Julier C. and White R. (1988). Detection of a Notl Polymorphism with the pmetH Probe by Pulsed-Field Gel Electrophoresis. Am. J. Hum. Genet. 42; 45-48.

Kallunki P., Eddy R.L., Byers M. G., Kestila M., Shows T. B., and Tryggvason K. (1991). Cloning of Human Heparan Sulfate Proteoglycan Core Protein, Assignment of the Gene (HSPG2) to 1p36.1 – p35 and Identification of a BamHI Restriction Fragment Length Polymorphism. GENOMICS 11; 389-396.

Kefford R.F., Salmon J., Shaw H.M., Donald J.A. and McCarthy W.H. (1991). Hereditary Melanoma in Australia, Variable Association with Dysplastic Nevi and Absence of Genetic Linkage to Chromosome lp. Cancer Genet Cytogenet 51; 45-55.

King B.J. and Cook P.J.L. (1981). Glucose dehydrogenase polymorphism in man. Ann. Hum. Genet. 45; 129-134.

Kingsmore S.F., Watson M.L., Howard T.A. and Seldin M.F. (1989). A 6000 kb segment of chromosome 1 is conserved in human and mouse. The EMBO Journal 8, No. 13; 4073-4080.

Kingsmore S.F., Moseley W.S., Watson M.L., Sabina R.L., Holmes E.W. and Seldin M.F. (1990). Long-Range Restriction Site Mapping of a Syntenic Segment Conserved between Human Chromosome l and Mouse Chromosome 3. GENOMICS 7; 75-83.

Kleyn P.W. (1990). Isolation of the Human 6-Phosphogluconate Dehydrogenase Gene, a Genetic Marker in Man. Ph.D Thesis. University of London.

Kline A.D., Rojas K., Mewar R., Moshinsky D. and Overhauser J. (1992). Somatic Cell Hybrid Deletion Map of Human Chromosome l8. GENOMICS 13; 1-6.

Kmiec E.B. and Holloman W.K. (1986). Homologous Pairing of DNA Molecules by Ustilago Recl Protein Is Promoted by Sequences of Z-DNA. Cell, 44; 545-554.

Knudson A. G. (1971). Mutation and Cancer: Statistical Study of Retinoblastoma. Proc. Nat. Acad. Sci. USA 68, No.4; 820-823.

Kobori J.A., Strauss E., Minard K., Hood L. (1986). Molecular Analysis of the Hotspot of Recombination in the Murine Major Histocompatibility Complex. Science 234; 173-179.

Kochanek S., Toth M., Dehmel A., Renz D. and Doerfler W. (1990). Interindividual concordance of methylation profiles in human genes for tumor necrosis factors a and b. Proc. Natl. Acad. Sci. USA 87; 8830-8834.

Kosambi D. D. (1944). The Estimation of Map Distances from Recombination Values. Ann. Eugen. (Lond) 12; 172-175.

Kushner J. P., Barbuto A. J. and Lee G. R. (1976). An inherited enzymatic defect in porphyria cutanea tarda: decreased uroporphyrinogen decarboxylase activity. J. Clin. Invest. 58; 1089-1097.

Kwiatkowski D.J., Henske E.P., Weimer K., Ozelius L., Gusella J.F. and Haines J. (1992). Construction of a GT Polymorphism Map of Human 9q. GENOMICS 12; 229-240.

Lalley P. A., Francke U. and Minna J. D. (1978). Homologous genes for enolase, phosphogluconate dehydrogenase, phosphoglucomutase, and adenylate kinase are syntenic on mouse chromosome 4 and human chromosome lp. Proc. Natl. Acad. Sci. USA 75, 2382-2386.

Larin Z., Monaco A.P., and Lehrach H. (1991). Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc. Natl. Acad. Sci. USA 88; 4123-4127.

Larsen F., Gundersen G., Lopez R. and Prydz H. (1992). CpG Islands as Gene Markers in the Human Genome. GENOMICS 13; 1095-1107.

Laureys G., Speleman F., Opdenakker G., Benoit Y., and Leroy J. (1990). Constitutional Translocation t(1;17) (p36;q12-21) in a Patient With Neuroblastoma. Genes, Chromosomes & Cancer 2: 252-254.

Laurie D. A., Hulten M., and Jones G. H. (1981). Chiasma frequency and distribution in a sample of human males: chromosomes 1, 2, and 9. Cytogenet. Cell Genet. 31; 153-166.

Laurie D. A., and Hulten M. A. (1985). Further studies on chiasma distribution and interference in the human male. Ann. Hum. Genet. 49: 203-214.

Lawrence J.B., Singer R.H., McNeil J.A. (1990). Interphase and Metaphase Resolution of Different Distances Within the Human Dystrophin Gene. Science 249; 928-931.

Lengauer C., Green E.D. and Cremer T. (1992). Fluorescence in Situ Hybridization of YAC Clones after Alu-PCR Amplification. GENOMICS 13; 826-828.

Levine P., Vogel P., Katzin E.M., and Burnham L., (1941). Pathogenesis of Erythroblastosis Fetalis: Statistical Evidence. Science. 94; 371-372.

Li H., Gyllensten U.B., Cui X., Saiki R.K., Erlich H.A. & Arnheim N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells. Nature 335; 414-417.

Li S. R., Baroni M. G., Oelbaum R. S., Stock J. and Galton D. J. (1988). Association of genetic variant of the glucose transporter with non-insulin-dependent diabetes mellitus. Lancet II; 368-370.

Lichter P., Tang C.J.C., Call K., Hermanson G., Evans G.A., Housman D., Ward D.C. (1990). High-Resolution Mapping of Human Chromosome 11 by *in Situ* Hybridization with Cosmid Clones. Science, 247; 64-68.

Lifton R. P., Sardet C., Pouyssegur J., and Lalouel J.M. (1990). Cloning of the Human Genomic Amiloride-Sensitive Na+/H+ Antiporter Gene, Identification of Genetic Polymorphisms, and Localization on the Genetic Map of Chromosome lp. GENOMICS 7; 131-135.

Lifton R. P., Hunt S. C., Williams R. R., Pouyssegur J., and Lalouel J. (1991). Exclusion of the Na+-H+ Antiporter As a Candidate Gene in Human Essential Hypertension. Hypertension 17: 8-14.

Lindsay S. & Bird A.P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327; 336-338.

Link A.J. and Olson M.V. (1991). Physical Map of the Saccharomyces cerevisiae Genome at ll0-Kilobase Resolution. Genetics 127: 681-698.

Litt M. and Luty J.A. (1989). A Hypervariable Microsatellite Revealed by In Vitro Amplification of a Dinucleotide Repeat within the Cardiac Muscle Actin Gene. Am. J. Hum. Genet. 44; 397-401.

MacGeoch C., Mitchell C.J., Carritt B., Avent N.D., Ridgwell K., Tanner M. J. A., and Spurr N.K. Assignment of the chromosomal locus of the human 30-kDal Rh (Rhesus) blood group antigen related protein (Rh30A) to chromosome region 1p36.13-p34. Cytogenet. Cell Genet. 39 : 261-263.

Martini G., Toniolo D., Vulliamy T., Luzzatto L., Dono R., Viglietto G., Paonessa G., D'Urso M., and Persico M.G. (1986). Structural analysis of the X-linked gene encoding human glucose 6-phosphate dehydrogenase. The EMBO Journal 5, No.8; 1849-1855.

Martinsson T., Weith A., Cziepluch C. and Schwab M. (1989). Chromosome I Deletions in Human Neuroblastomas: Generation and Fine Mapping of Microclones From the Distal Ip Region. Genes, Chromosomes & Cancer 1; 67-78.

Mathew C.G.P., Smith B.A., Thorpe K., Wong Z., Royle N.J., Jeffreys A.J. and Ponder B.A.J. (1987). Deletion of genes on chromosome 1 in endocrine neoplasia. Nature 328; 524-526.

McCormick M.K., Campbell E., Deaven L., and Moyzis R. (1993). Low-frequency chimeric yeast artificial chromosome libraries from flow-sorted human chromosomes 16 and 21. Proc. Natl. Acad. Sci. USA 90:1063-1067.

Mendez M. J., Klapholz S., Brownstein B. H., and Gemmill R. M. (1991). Rapid Screening of a YAC Library by Pulsed-Field Gel Southern Blot Analysis of Pooled YAC Clones. GENOMICS, 10; 661-665.

Miksicek R. J. and Towle H. C. (1982). Changes in the Rates of Synthesis and Messenger RNA Levels of Hepatic Glucose-6-phosphate and 6-Phosphogluconate Dehydrogenases Following Induction by Diet or Thyroid Hormone. The Journal of Biological Chemistry 257; No. 19; 11829-11835.

Miksicek R.J. and Towle H.C. (1983). Use of a Cloned cDNA Sequence to Measure Changes in 6-Phosphogluconate Dehydrogenase mRNA Levels Caused by Thyroid Hormone and Dietary Carbohydrate. The Journal of Biological Chemistry, 258, No. 15; 9575-9579.

Mitchell B., Haigis E., Steinmann B. and Gitzelmann R. (1975). Reversal of UDP-galactose 4-epimerase deficiency of human leukocytes in culture. Proc. Nat. Acad. Sci.USA 72; 5026-5030.

Monaco A.P., Lam V.M.S., Zehetner G., Lennon G.G., Douglas C., Nizetic D., Goodfellow P.N. and 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.

Monckton D.G. and Jeffreys A.J. (1991). Minisatellite "Isoallele" Discrimination in Pseudohomozygotes by Single Molecule PCR and Variant Repeat Mapping. GENOMICS 11; 465-467.

Morton N.E. (1955). Sequential Tests for the Detection of Linkage. Am. J. Hum. Genet. 7; 277-318.

Moseley W.S. and Seldin M.F. (1989). Definition of Mouse Chromosome 1 and 3 Gene Linkage Groups That Are Conserved on Human Chromosome 1: Evidence That a Conserved Linkage Group Spans the Centromere of Human Chromosome 1. GENOMICS 5; 899-905.

Mouchiroud D., D'Onofrio G., AissaniB., Macaya G., Gautier C. and Bernardi G. (1991). The distribution of genes in the human genome. Gene, 100; 181-187.

Mount S.M. (1982). A catalogue of splice junction sequences. Nucleic Acids Reasearch. 10: 459-472.

Nakamura Y., Leppert M., O'Connell P., Wolff R., Holm T., Culver M., Martin C., Fujimoto E., Hoff M., Kumlin E., White R. (1987). Variable Number of Tandem Repeat (VNTR) Markers for Human Gene Mapping. Science 235; 1616-1622.

Nakamura Y., Carlson M., Krapcho K. and White R. (1988). Isolation and mapping of a polymorphic DNA sequence (pMCT58) on chromosome lp (DIS77). Nucleic Acids Research 16 No.19; 9367.

Nancarrow D.J., Palmer J.M., Walters M.K., Kerr B.M., Hafner G.J., Garske L., McLeod G.R. and Hayward N.K. (1992). Exclusion of the Familial Melanoma Locus (MLM) from the PND/D1S47 and MYCL1 Regions of Chromosome Arm lp in 7 Australian Pedigrees. GENOMICS 12; 18-25.

Nelson D.L., Ledbetter S.A., Corbo L., Victoria M.F., Ramirez-Solis R., Webster T.D., Ledbetter D.H. and Caskey C.T. (1989). Alu polymerase chain reaction: A method for rapid isolation of juman-specific sequences form complex DNA sources. Proc. Natl. Acad. Sci. USA 86; 6686-6690.

Nelson D.L., Ballabio A., Victoria M.F., Pieretti M., Bies R.D., Gibbs R.A., Maley J.A., Chinault A.G., Webster T.D. (1991). Alu-primed polymerase chain reaction for regional assignment of l10 yeast artificial chromosome clones from the human X chromosome; Identification of clones associated with a disease locus. Proc. Natl. Acad. Sci. USA 88; 6157-6161.

Nemer M., Sirois D. and Drouin J. (1986). Taql polymorphism at the 3' end of the human pronatriodilatin gene (hPND). Nucleic Acids Research l4 No. 21; 8697.

Nguyen C., Poustka A.M., Djabali M., Roux D., Mattei J. F., Lehrach H. and Jordan B. R. (1989). Large-Scale Mapping and Chromosome Jumping in the q27 Region of the Human X Chromosome. GENOMICS 5; 298-303.

NIH/CEPH Collaborative Mapping Group. (1992). A Comprehensive Genetic Linkage Map of the Human Genome. Science 258; 67-78.

Nizetic D., Zehetner G., Monaco A.P., Gellen L., Young B.D. and Lehrach H. (1991). Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: Their potential use as reference libraries. Proc. Natl. Acad. Sci. USA 88; 3233-3237.

O' Connell P.O., Lathrop G.M., Nakamura Y., Leppert M.L., Ardinger R.H., Murray J.L., Lalouel J.M. and White R. (1989). Twenty-Eight Loci Form a Continuous Linkage Map of Markers for Human Chromosome 1. GENOMICS 4; 12-20.

Olson M., Hood L., Cantor C, Botstein D. (1989). A Common Language for Physical Mapping of the Human Genome. Science, 245; 1434-1440.

Orita M., Iwahana H., Kanazawa H., Hayashi K. and Sekiya T. (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. Proc. Natl. Acad. Sci. USA 86; 2766-2770.

Orita M., Suzuki Y., Sekiya T. and Hayashi K. (1989b). Rapid and Sensitive Detection of Point Mutations and DNA Polymorphisms Using the Polymerase Chain Reaction. GENOMICS 5; 874-879.

Orita M., Sekiya T. and Hayashi K. (1990). DNA Sequence Polymorphisms in Alu Repeats. GENOMICS 8; 271-278.

Ott J., Linder D., McCaw B.K., Lovrien E. W., Hecht F. (1976). Estimating distances from the centromere by means of benign ovarian teratomas in man. Ann. Hum. Genet. Lond. 40. 191-196.

Paris Conference. (1971). Standardization in human cytogenetics. Cytogenet. Cell Genet. 15: 201-238.

Parr C. W. (1966). Erythrocyte Phosphogluconate Dehydrogenase Polymorphism. Nature 210; 487-489.

Pavan W.J. , Hieter P. and Reeves R.H. (1990). Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. Proc. Natl. Acad. Sci. USA 87; 1300-1304.

Penrose L. S. (1953). The General Purpose Sib-Pair Linkage Test. Ann. Eugenics 18: 120-124.

Perez-Perez C., Taliano V., Mouro I., Huet M., Salat-Marti A., Martinez A., Rouger P. and Cartron J. P. (1992). Spanish Rh-null family caused by a silent Rh gene: hematological, serological, and biochemical studies. Am. J. Hemat. 40; 306-312.

Petit C., Levilliers J., and Weissenbach J. (1988). Physical mapping of the human pseudo-autosomal region; comparison with genetic linkage map. The EMBO Journal 7 No.8: 2369-2376.

Pinkel D., Straume T. and J.W. Gray (1986). Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc. Natl. Acad. Sci. USA 83; 2934-2938.

Pinkel D., Landegent J., Collins C., Fuscoe J., Segraves R., Lucas J. and Gray J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: Detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. USA 85; 9138-9142.

Poustka A.M. and Lehrach H. (1986). Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. Trends in Genetics 2; 174-179.

Poustka A., Pohl T.M., Barlow P., Frischauf A.M. & Lehrach H. (1987). Construction and use of human chromosome jumping libraries from NotI-digested DNA. Nature 325; 353-355.

Poustka A.M., Lehrach H., Williamson R. and Bates G. (1988). A Long-Range Restriction Map Encompassing the Cystic Fibrosis Locus and Its Closely Linked Genetic Markers. GENOMICS 2; 337-345.

Povey S. and Parrington J. M. (1986). Chromosome 1 in relation to human disease. Journal of Medical Genetics, 23:107-115.

Rapp J. P., Wang S., Dene H. (1989). A Genetic Polymorphism in the Renin Gene of Dahl Rats Cosegregates with Blood Pressure. Science, 243; 542-544.

Renwick J. H. (1971). The Mapping of Human Chromosomes. Ann. Rev. Genet., 5: 81-113.

Renwick J.H. and Schulze J. (1965). Male and female recombination fractions for the nail-patella: ABO linkage in man. Ann. Hum. Genet. Lond. 28; 379-392.

Richards R.I., Holman K., Lane S., Sutherland G.R. and Callen D.F. (1991). Human Chromosome 16 Physical Map: Mapping of Somatic Cell Hybrids Using Multiplex PCR Deletion Analysis of Sequence Tagged Sites. GENOMICS 10; 1047-1052.

Riethman H.C., Moyzis R.K., Meyne J., Burke D. T. and Olson M.V. (1989). Cloning human telomeric DNA fragments into Saccharomyces cerevisiae using a yeast-artificial-chromosome vector. Proc. Natl. Acad. Sci. USA 86; 6240-6244.

Riley J., Butler R., Ogilvie D., Finniear R., Jenner D., Powell S., Anand R., Smith J.C. and Markham A.F. (1990). A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. Nucleic Acids Research. 18; 2887-2890.

Ritke M.K., Shah R., Valentine M., Douglass E.C. and Tereba A. (1989). Molecular analysis of chromosome 1 abnormalities in neuroblastoma. Cytogenet Cell Genet 50; 84-90.

Rognstad R. and Katz J. (1979). Effects of 2, 4-Dihydroxybutyrate on Lipogenesis in Rat Hepatocytes. The Journal of Biological Chemistry, 254, No.23; 11969-11972.

Romani M., De Ambrosis A., Alhadeff B., Purrello M., Gluzman Y., and Siniscalco M. (1990). Preferential integration of the Ad5/SV40 hybrid virus at the highly recombinogenic human chromosomal site 1p36. Gene 95; 231-240.

Rommens J.H., Jannuzzi M.C., Kerem B.S., Drumm M.L., Melmer G., Dean M., Rozmahel R., Cole J.L., Kennedy D., Hidaka N., Zsiga M., Buchwald M., Riordan J.R., Tsui L.C.., Collins F.S. (1989). Identification of the Cystic Fibrosis Gene: Chromosome Walking and Jumping. Science 245;1059-1056.

Rouleau G.A., Bazanowski A., Gusella J.F. and Haines J.L. (1990). A Genetic Map of Chromosome 1: Comparison of Different Data Sets and Linkage Programs. GENOMICS 7; 313-318.

Rouyer F., Simmler M., Johnsson C., Vergnaud G., Cooke H. J. and Weissenbach J. (1986). A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. Nature 319; 291-295.

Royle N. J., Clarkson R., Wong Z., and Jeffreys A.J. (1987). Preferential localization of hypervariable minisatellites near human telomeres. Human Gene Mapping 9 (1987). Cytogenet. Cell Genet. 46: 685.

Royle N. J., Clarkson R. E., Wong Z., and Jeffreys A. J. (1988). Clustering of Hypervariable Minisatellites in the Proterminal Regions of Human Autosomes. GENOMICS 3: 352-360.

Rynditch A., Kadi F., Geryk J., Zoubak S., Svoboda J., and Bernardi G. (1991). The isopycnic, compartmentalized integration of Rous sarcoma virus sequences. Gene, 106; 165-172.

Saccone S., De Sario A., Della Valle G., and Bernardi G. (1992). The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes. Genetics 89; 4913-4917.

Saito A., Abad J.P., Wang D., Ohki M., Cantor C.R. and Smith C.L. (1991). Construction and Characterization of a NotI Linking Library of Human Chromosome 21. GENOMICS 10; 618-630.

Schorderet D.F. and Gartler S.M. (1992). Analysis of CpG suppression in methylated and nonmethylated species. Proc. Natl. Acad. Sci. USA 89; 957-961.

Schwartz D.C., and Cantor C. R. (1984). Separation of Yeast Chromosome-Sized DNAs by Pulsed Field Gradient Gel Electrophoresis. Cell, 37: 67-75.

Searle A. G., Peters J., Lyon M. F., Hall J. G., Evans E. P., Edwards J. H. and Buckle V. J. (1989). Chromosome maps of man and mouse. IV. Ann. Hum. Genet. 53; 89-140.

Sefton L., Kelsey G., Kearney P., Povey S., and Wolfe J. (1990). A Physical Map of the Human PI and AACT Genes. GENOMICS 7: 382-388.

Sheiro J.H. , McCormick M.K. , Antonarakis S.E. and Hieter P. (1991).Yeast Artificial Chromosome Vectors for Efficient Clone Manipulation and Mapping . GENOMICS 10; 505-508.

Sherman S.L., King J., Robson E.B. and Yee S. (1984). A revised map of chromosome 1. Ann. Hum. Genet. 48; 243-251.

Shows T. B., Eddy R. L., Byers M. G., Fukushima Y., Dehaven C. R., Murray J. C. and Bell G. I. (1987). Polymorphic human glucose transporter gene (GLUT) is on chromosome 1p31.3-p35. Diabetes 36; 546-549.

Shukla H., Gillespie G.A., Srivastava R., Collins F. and Chorney M.J. (1991). A Class I Jumping Clone Places the HLA-G Gene Approximately 100 Kilobases from HLA-H within the HLA-A Subregion of the Human MHC. GENOMICS 10; 905-914.

Smith G.R. (1983). Chi Hotspots of Generalized Recombination. Cell, 34; 709-710.

Southern E.M. (1975). Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis. J. Mol. Biol. , 98: 503-517.

Stallings R.L., Torney D.C., Hildebrand C.E., Longmire J.L., Deaven L.L., Jett J.H., Doggett N.A. and Moyzis R.K. (1990). Physical mapping of human chromosomes by repetitive sequence fingerprinting. Proc. Natl. Acad. Sci. USA 87; 6218-6222.

Steinlein O., Fischer C., Keil R., Smigrodzki R. and Vogel F. (1992). D20S19, linked to low voltage EEG, benign neonatal convulsions, and Fanconi anaemia, maps to a region of enhanced recombination and is localized between CpG islands. Human Molecular Genetics 1, No.5; 325-329.

Sumner A.T. (1982). The Nature and Mechanisms of Chromosome Banding. Cancer Genetics and Cytogenetics. 6: 59-87.

Swartz M.N., Trautner T.A. and Kornberg A. (1962). Enzymatic Synthesis of Deoxyribonucleic Acid. THE JOURNAL OF BIOLOGICAL CHEMISTRY 237, No. 6; 1961-1967.

Tashian R. E., Brewer G. J., Lehmann H., Davies D. A. and Rucknagel D. L. (1967). Genetic Variability in Some Serum and Erythrocyte Enzymes, Hemoglobin, and the Urinary Excretion of beta-Aminoisobutyric Acid. Am. J. Hum. Genet. 19, No. 4; 524-531.

Tepperman H. M., and Tepperman J. (1963). On the Response of Hepatic Glucose-6-Phosphate Dehydrogenase Activity to changes in Diet Composition and Food Intake Pattern. Adv. Enz. Reg 1: 121-135.

Tepperman H.M. and Tepperman J. (1965). Effect of saturated fat diets on rat liver NADP-linked enzymes. American Journal Physiology, 209: 773-779.

Thomas B.J. and Rothstein R. (1991). Sex, Maps, and Imprinting. Cell, 64; 1-3.

Tomlinson J. E., Nakayama R. and Holten D. (1988). Repression of Pentose Phosphate Pathway Dehydrogenase Synthesis and mRNA by Dietary Fat in Rats. J. Nutr. 118; 408-415.

Trask B. J. (1991). Fluorescence *in situ* hybridization: applications in cytogenetics and gene mapping. Trends in Genetics. , 7 No.5: 149-154.

Traver C.N., Klapholz S., Hyman R.W. and Davis R.W. (1989). Rapid screening of a human genomic library in yeast artificial chromosomes for single-copy sequences. Proc. Natl. Acad. Sci. USA 86; 5898-5902.

Uematsu Y., Kiefer H., Schulze R., Fischer-Lindahl K., and Steinmetz M. (1986). Molecular characterization of a meiotic recombinational hotspot enhancing homologous equal crossing over. The EMBO Journal, 5: 2123-2129.

Van Camp G., Van Hul W., Backhovens H., Stinissen P., Wehnert A., Patterson D., Vandenberghe A., and Van Broeckhoven C. (1990). Physical Mapping of Chromosome 21 DNA Markers in Alzheimer's Disease Region Using Somatic Cell Hybrids. Somatic Cell and Molecular Genetics. 16, No.3; 241-249.

Van der Schroeff J. G., Nijenhuis L. E., Meera Khan P., Bernini L. F., Schreuder G. M. T., Van Loghem E., Volkers W. S. and Went L. N. (1984). Genetic linkage between erythrokeratodermia variabilis and Rh locus. Hum. Genet. 68; 165-168.

Van Roy N., Laureys G., Versteeg R., Opdenakker G., Speleman F. (1993). High Resolution Fluorescence Mapping of DNA Markers Previously Assigned to the Short Arm of Human Chromosome 1. (Unpublished).

Vanin E. F. (1985). Processed pseudogenes: Characteristics and Evolution. Ann. Rev. Genet. 19: 253-272.

Verloes A., Massart B., Jossa V., Langhendries J. P., Paquot J. P., Koulischer L. (1991). Neuroblastoma in a Dwarfed Newborn Possible Clue to the Chromosomal Localization of the Gene for Achondroplasia Ann Genet, 34 , No. 1; 25-26.

Vincent A., Heitz D., Petit C., Kretz., Oberle I. & Mandel J.L. (1991). Abnormal pattern detected in fragile-X patients by pulsed-field gel electrophoresis. Nature 349; 624-626.

Wada M., Little R. D., Abidi F., Porta G., Labella T., Cooper T., Della Valle G., D'Urso M., and Schlessinger D. (1990). Human Xq24-Xq28: Approaches to Mapping with Yeast Artificial Chromosomes. Am. J. Hum. Genet. 46; 95-106.

Wagner M. (1986). A consideration of the origin of processed pseudogenes. Trends in Genetics 2; 134-136.

Wahl M.G., Lewis K.A., Ruiz J.C., Rothenberg B., Zhao J. and Evans G.A. (1987). Cosmid vectors for rapid genomic walking, restriction mapping, and gene transfer. Proc. Natl. Acad. Sci. USA 84; 2160-2164.

Wahls W.P., Wallace L.J. and Moore P.D. (1990). Hypervariable Minisatellite DNA Is a Hotspot for Homologous Recombination in Human Cells. Cell, 60: 95-103.

Wallace M.R., Fountain J.W., Brereton A.M. and Collins F.S. (1989). Direct construction of a chromosome-specific NotI linking library from flow-sorted chromosomes. Nucleic Acids Research 17, No. 4; 1665-1677.

Weber J.L. and May P.E. (1989). Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain Reaction. Am. J. Hum. Genet. 44; 388-396.

Weber J.L. (1990). Informativeness of Human (dC-dA)n . (dG-dT)n Polymorphisms. GENOMICS 7; 524-530.

Weiss M.C. and Green H. (1967). Human-Mouse Hybrid Cell Lines Containing Partial Complements of Human Chromosomes and Functioning Human Genes. Proc. Natl. Acad. Sci. 58; 1104-1110.

Weissenbach J., Gyapay G., Dib C., Vignal A., Morissette J., Millasseau P., Vaysseix G. & Lathrop M. (1992). A second-generation linkage map of the human genome. Nature, 359; 794-801.

Weith A., Martinsson T., Cziepluch C., Bruderlein S., Amler L. C., Berthold F. and Schwab M. (1989). Neuroblastoma Concensus Deletion Maps to lp36.1-2. Genes, Chromosomes & Cancer 1: 159-166.

Welch H.M., Darby J.K., Pilz A.J., Ko C.M. and Carritt B. (1989). Transposition, Amplification, and Divergence in the Origin of the DNF15 Loci, a Polymorphic Repetitive Sequence Family on Chromosomes 1 and 3. GENOMICS 5; 423-430.

Wevrick R. and Willard H.F. (1991). Physical map of the centromeric region of human chromosome 7: relationship between two distinct alpha satellite arrays. Nucleic Acids Research, 19, No.9; 2295-2301.

White R., Leppert M., Bishop D.T., Barker D., Berkowitz J., Brown C., Callahan P., Holm T. & Jerominski L. (1985). Construction of linkage maps with DNA markers for human chromosomes. Nature 313; 100-105.

White R. and Lalouel J.M. (1987). Investigation of Genetic Linkage in Human Families. Advances In Human Genetics 16; 121-228.

Williamson R., Bowcock A., Kidd K., Pearson P., Schmidtke J., Ceverha P., Chipperfield M., Cooper D.N., Coutelle C., Hewitt J., Klinger K., Langley K., Beckmann J., Tolley M. and Maidak B. (1991). Report of the DNA committee and catalogues of cloned and mapped genes, markers formatted for PCR and DNA polymorphisms. Cytogenet. Cell. Genet. 58; 1190-1832.

Wilson A.F., Elston R.C., Tran L.D. and Siervogel R.M. (1991). Use of the Robust Sib-Pair Method to Screen for Single-Locus, Multiple-Locus, and Pleiotropic Effects: Application to Traits Related to Hypertension. Am. J. Hum. Genet. 48; 862-872.

Yang-Feng T.L., Floyd-Smith G., Nemer M., Drouin J. and Francke U. (1985). The Pronatriodilatin Gene Is Located on the Distal Short Arm of Human Chromosome 1 and on Mouse Chromosome 4. Am J Hum Genet 37; 1117-1128.

Zabarovsky E.R., Boldog F., ErlandssonR., Kashuba V.I., Allikmets R.L., Marcsek Z., Kisselev L.L., Stanbridge E., Klein G., Sumegi J. and Winberg G. (1991). New Strategy for Mapping the Human Genome Based on a Novel Procedure for Construction of Jumping Libraries. GENOMICS 11; 1030-1039.

Zhang L., Cui X., Schmitt K., Hubert R., Navidi W. and Arnheim N. (1992). Whole genome amplification from a single cell: Implications for genetic analysis. Proc. Natl. Acad. Sci. USA 89: 5847-5851.

Zoghbi H.Y., McCall A.E. and Leborgne-Demarquoy F. (1991). Sixty-five Radiation Hybrids for the Short Arm of Human Chromosome 6: Their Value as a Mapping Panel and as a Source for Rapid Isolation of New Probes Using Repeat Element-Mediated PCR. GENOMICS 9; 713-720.

Zuliani G. and Hobbs H.H. (1990). A High Frequency of Length Polymorphisms in Repeated Sequences Adjacent to Alu Sequences. Am. J. Hum. Genet. 46; 963-969.