The Human Y Chromosome; Towards a Physical Map in terms of Overlapping Cosmid Clones.

Submitted by

Kay Elizabeth Taylor

in

November, 1991

to

The University of London

for the degree of

Doctor of Philosophy

Research conducted at the Department of Genetics and Biometry,
Galton Laboratory, University of London

ABSTRACT

This thesis describes the construction and analysis of a cosmid library specific for the human Y chromosome. The purpose of preparing such a library is to provide a source of ordered overlapping clones for use in the identification of genes, analysis of chromosome structure, and perhaps, as a starting point for the sequencing of the chromosome.

Approximately 2,000 cosmid clones containing inserts derived from human DNA have been identified in a library prepared from a somatic cell hybrid containing the Y chromosome as the only human material.

These clones have been analysed in two principal ways. Firstly, hybridisation to various characterised DNA fragments has identified cosmids containing known repeat elements and single copy DNA sequences which have been previously assigned to regions of the chromosome by deletion mapping. Secondly, a unique 'fingerprint' has been produced for each clone; band patterns of *Hin*fI digestion products separated by polyacrylamide gel electrophoresis have been analysed by computer to detect overlapping clones. A significant proportion of the chromosome is represented in this way.

CONTENTS

LIST of FIGURES

ABBREVIATIONS

Mins	Minutes
rpm	Revolutions per minute
bp	basepairs
kb	Kilobases
Mb	Megabases
mmol	Millimoles
mM	Millimolar
ml	Millilitre
µl	Microlitre
pfu	Plaque forming units
ddH_2O	Double distilled water
PCR	Polymerase chain reaction
YAC	Yeast Artificial Chromosome
TDF	Testis determining factor
TCA	Trichloroacetic acid
DEPC	Diethylpyrocarbonate
EDTA	Ethylenediaminotetracetic acid Sodium salt
DMSO	Dimethylsulphoxide
PBS	Phosphate buffered saline
TEMED	Tetraethylmethylenediamine

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Jonathan Wolfe for convincing me that this was a feasible project and for his encouragement. Also everyone in the lab. for being friendly and supportive, especially Zanna who shared the tedious library screening with me, and Viv, Nick, Jude and Joseph.

I am grateful to Dr. J. Sulston for modifying some of his programs to be compatible with my data and to Dr. P. Little for allowing me to use his VAX computer and giving advice on data entry.

I am appreciative of Erica and Chris for coping with me writing up in their flat, and Robin who coped with that and more besides.

This thesis is for my parents who have given me more emotional and financial support than I could ever have expected of them.

CHAPTER ONE

INTRODUCTION PART 1

1.1 Cytogenetics and structure of the human Y chromosome.

The chromosomal nature of mammalian sex determination has been apparant since 1959 (Welshons *et al.*) with a gene or genes located on the Y chromosome being responsible for male development. It is only recently, however, that the sequences of several genes encoded on the human Y chromosome have been identified. For many years, before the advent of molecular techniques, the absence of meiotic recombination precluded the study of these genes by classical genetics.

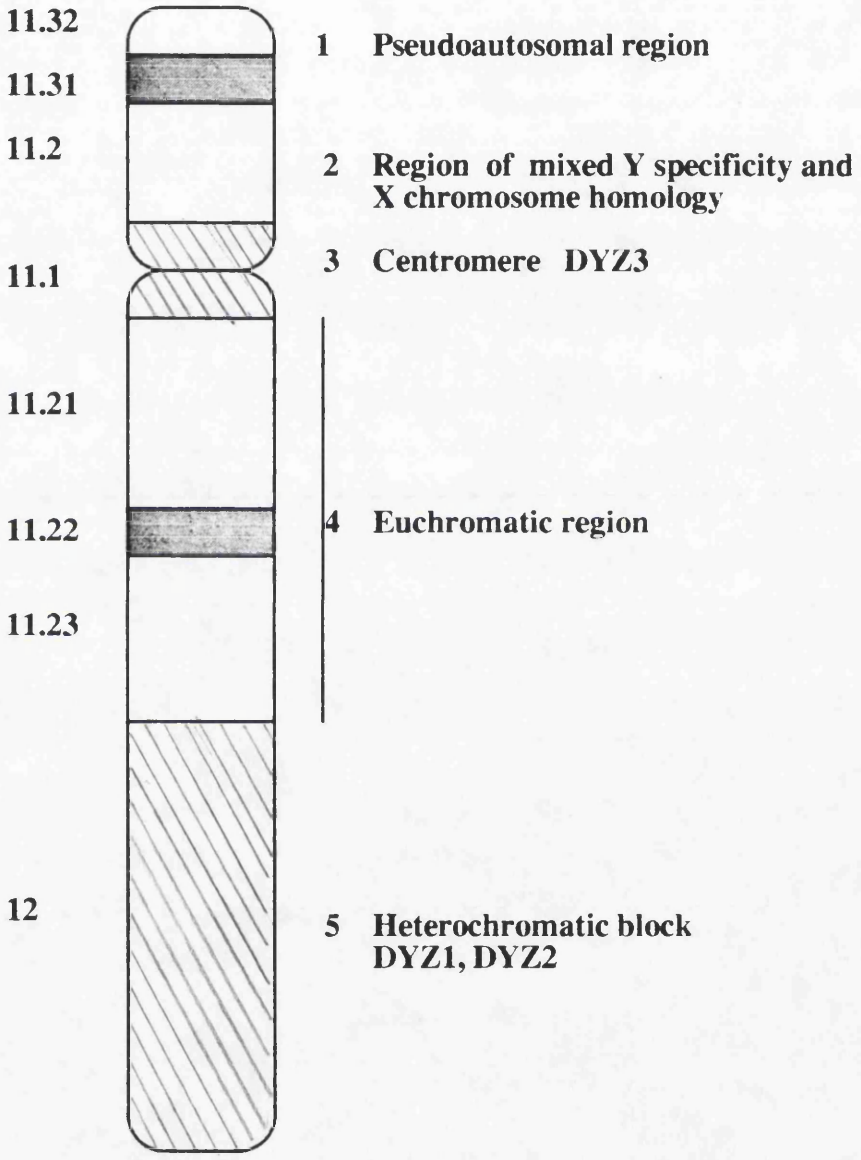
The basic characteristics of the Y chromosome have been defined by cytological techniques. It is the smallest human chromosome comprising 2% of the 32,000Mb genome, with a size in the region of 59 megabase pairs of DNA (Morton 1991).

The chromosome may be divided into five broad regions (see fig 1.1).

1) At meiosis the terminal short arm regions of the X and Y chromosomes have been shown to pair in the form of a short length of synaptonemal complex during pachytene. This pairing ensures that the chromosomes segregate correctly in the gametes, and is brought about by the interaction of homologous regions which undergo recombination. Sequences in this region therefore show only partial sex linkage and may be termed pseudoautosomal. Several X-Y homologous genes from this region have been cloned (see section 1.5)

2) Within the remaining short arm euchromatic region reside the genes *ZFY* and *SRY* , the latter being the testis-determining factor (section

Figure 1.1 Physical and functional Y chromosome regions



1.4). However, this region is not completely sex-specific as many sequences are shared between this region and the X chromosome long arm intervals Xq13-24 (section 1.6).

3) The pericentric region is heterochromatic and variable in length, containing the repeating element DYZ3 which is related to repeat elements at autosomal centromeres (see section 1.7.2).

4) The euchromatic portion of the long arm appears to encode several as yet uncloned genes including a spermatogenesis factor and the gene for the H-Y antigen (see section 1.3).

5) Differential staining highlighted a large heterochromatic region on the distal long arm covering up to 30% of the chromosome. This region is generally assumed to be genetically inert and may be highly variable in length. It is composed of two major repeating elements DYZ1 and DYZ2 (section 1.7.3).

Each of these regions have been studied and mapped to various extents using different techniques. This section of the chapter describes these regions of the chromosome and the genes encoded by them in more detail.

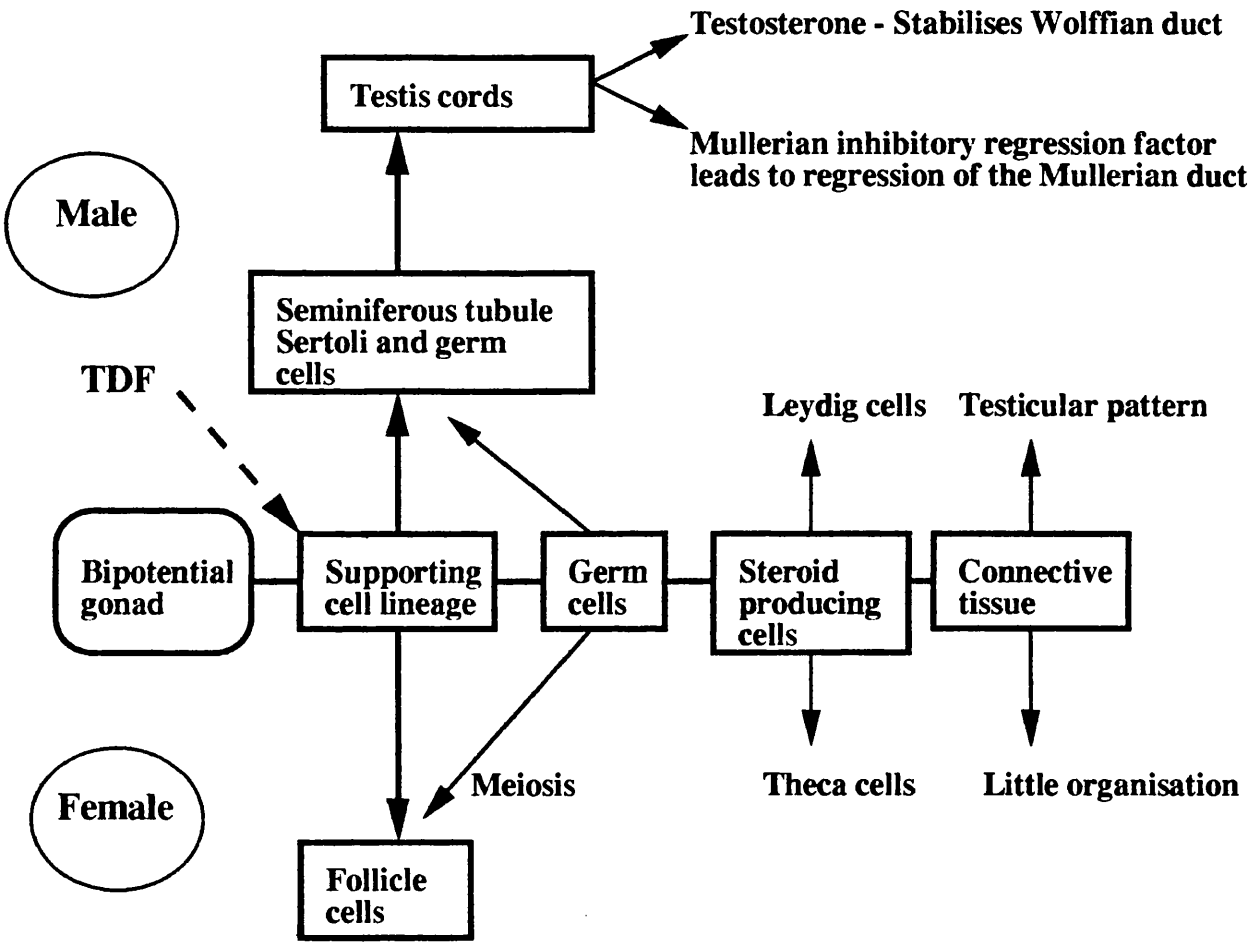
1.2 The role of the Y chromosome in sex determination

The process of sex determination is governed by different factors in different species. This may be environmental as in the case of alligators where the sex of offspring depends upon incubation temperature of the eggs (Ferguson *et al.* 1982), but is predominantly determined genetically. In *Drosophila* and *Caenorhabditis* it is the X:autosome ratio that is the underlying factor causing a 'switch' leading to the expression of a cascade of sex-determining genes.

From karyotypic analysis of male and female mammals, it has been obvious for some time that the Y chromosome plays a fundamental role in the process of sex determination. In human cases of an abnormal chromosome complement, presence of a Y chromosome leads to a male phenotype; 47,XXY Klinefelter patients are male (Jacobs and Strong 1959). The lack of the Y chromosome leads to a female phenotype in the case of Turner syndrome (45,XO), Ford *et al* 1959). This is regardless of the number of X chromosomes.

Mammalian secondary sexual characteristics are determined by the effect of hormones (Jost 1973) produced by the gonads which develop from the indifferent genital ridge. In the absence of any prior testis-determining signal the bi-potential gonad will differentiate as an ovary. Thus sex determination is dependent upon the presence or absence of a product expressed from a Y chromosome locus which induces testis development. This product is termed TDF for Testis-Determining Factor) in humans and *Tdy* (Testis Determining gene Y) in mouse. Figure 1.2.

Figure 1.2 Differentiation of the bipotential gonad



The effect of TDF expression is to direct the supporting cell lineage to form sertoli cells which associate with the germ cells, preventing them from entering meiosis which would lead to ovarian development. Seminiferous tubules are formed leading to testis development.

1.2.2 Sex reversal and the testis-determining factor

The study of individuals with chromosome abnormalities in both humans and mice has been a powerful tool in two ways; in defining a location for TDF within the Y chromosome, and in assessing candidate genes.

The phenomenon of sex reversal in mice was described by Cattenach *et al.* (1971). Mice with a female (XX) karyotype were seen to develop as phenotypic (though sterile) males due to the effect of a gene designated *Sxr*. The property was shown to be the result of a rearrangment of the Y chromosome such that a region near the centromere was duplicated and transposed to the distal end of the same chromosome (Singh & Jones, 1982). This region is then transfered to one chromatid of the X chromosome in a nonreciprocal crossover occuring during meiotic pairing of the sex chromosomes (Evans *et al.* 1982). The *Sxr* 'mutation' can be transmitted from one generation to another by XY^{sxr}carrier males which are fertile and apparantly normal.

Since this region of the chromosome is sufficient to give rise to a male phenotype in the absence of any other Y chromosome material when transposed to the X chromosome, it must contain the gene or genes required for testis-determination (McLaren 1988).

A similar sex reversal effect is seen in humans. This results, in the majority of cases, from the transfer of Y chromosome material to the X chromosome during meiosis, when the recombination event proceeds beyond the homologous pairing region and encroaches upon Y specific sequences. In this way both XY females and XX males may be produced (Page *et al.* 1987, Ferguson-Smith, 1966).

The presence of DNA derived from the Y chromosome had been detected at the molecular level in most, but not all, XX males using Y-specific single copy probes (Guellaen *et al.* 1984, Affara *et al.* 1987, Müller 1987) localising the sex-determining region to the proximal short arm of the human Y chromosome. The clinical features of those XX males described by Guellaen included reduced testis size, low testosterone levels, azoospermy and a gonadal histopathology showing testicular structure with sertoli cells but lacking germ cells. Of the four males in the study, one did not include sequences recognised by the probes tested and other studies have failed to identify Y sequences in all XX males (Affara *et al.* 1986). Similarly, of 8 XY females tested for the deletion of Y chromosome sequences by a number of probes only one showed any loss of Y material (Affara *et al.* 1987). This patient had Turner's stigmata which implies the existence of a locus on Yp, homologous to a locus on Xp, deletion of which causes Turner syndrome (see section 1.6.2). Furthermore, this patient had bilateral gonadoblastoma which is not found in classical Turner cases, but is common in XY females with dysgenetic gonads who retain part of the Y chromosome but have lost TDF function. This suggests that there is a locus on the Y chromosome (*GBY*) which is functional in a normally developed testis, but is important in the pathogenesis of gonadoblastoma in the absence of testis differention (Page 1987).Disteche *et al.* 1986, Magenis *et al.* 1984).

1.2.3 Indirect mechanisms of sex reversal

If, in some cases, sex reversal is not caused by the presence or absence of Y material (assuming that the probes used have been representative of the chromosome and that some but not all XY females may be caused by point mutations in TDF) then it is possible that an autosomally encoded testis determining factor (TDFA) exists (de la Chappelle, 1987). In some cases the presence of Y material in XX males may have been too small to be detected by the probes used (see Palmer *et al.* section 1.4.5). Such a factor is more likely to be involved in the cascade of events leading to testis determination. A mutation in an autosomal or X chromosome gene may change the function of this gene to direct testis determination.

In mice the autosomal background is known to be important in the expression patterns of the testis determining gene. When male mice of strain *Mus musculus domesticus* are crossed with C57Bl/6 females, half of the XY offspring produced are female or of an intersex nature. It is possible that the TDF gene encoded on the *M.m.domesticus* Y chromosome has a delayed onset in the C57Bl/6 background and does not properly pre-empt the action of the ovarian determinant (Palmer and Burgoyne pers comm.).

1.2.4 The H-Y antigen

The male-specific antigen H-Y is a member of a family of minor histocompatability antigens and is controlled by a gene located on the Y chromosome in both humans and mice.

Using H-Y specific T cell clones, lymphocytes from mice can be tested for expression of the H-Y antigen (Simpson *et al.* 1987). Analysis of

XX*Sxr* males showed that the H-Y antigen was present in these mice and in the absence of any other candidates was proposed as a possible TDF. However, some female mice (T16HX*Sxr*) were included in the study. These mice carry a T16H,X-autosome translocation. This X chromosome is always active causing the paternal X*Sxr* chromosome to be inactive at least during gonadogenesis allowing female development. Of nine T16HX*Sxr* females tested eight were H-Y positive indicating that the gene controlling H-Y, *Hya*, was expressed in adult life in spleen cells and perhaps other tissues. The remaining female was H-Y negative as were her non-XY progeny retaining her *Sxr* fragment, this is due to a deletion in *Sxr* termed *Sxr'* (McLaren *et al.* 1984). Since both XX*Sxr'* and XO*Sxr'* mice are male but lacking expression of the H-Y antigen, this gene could be discounted as the testis-determining factor, although it obviously maps in close proximity to *Tdy* in the mouse.

In the human, H-Y typing of XX males and XY females having inherited or lost variable portions of Yp respectively as described above, has enabled the H-Y gene to be localised to the long arm of the Y chromosome. Thus in humans, the H-Y gene is not located near TDF but is in the same region as the fertility factor (*AZF*), or in the proximal short arm (Simpson 1987).
It is possible that the H-Y antigen may be involved in the process of spermatogenesis (see below).

1.3 The Y chromosome and spermatogenesis.

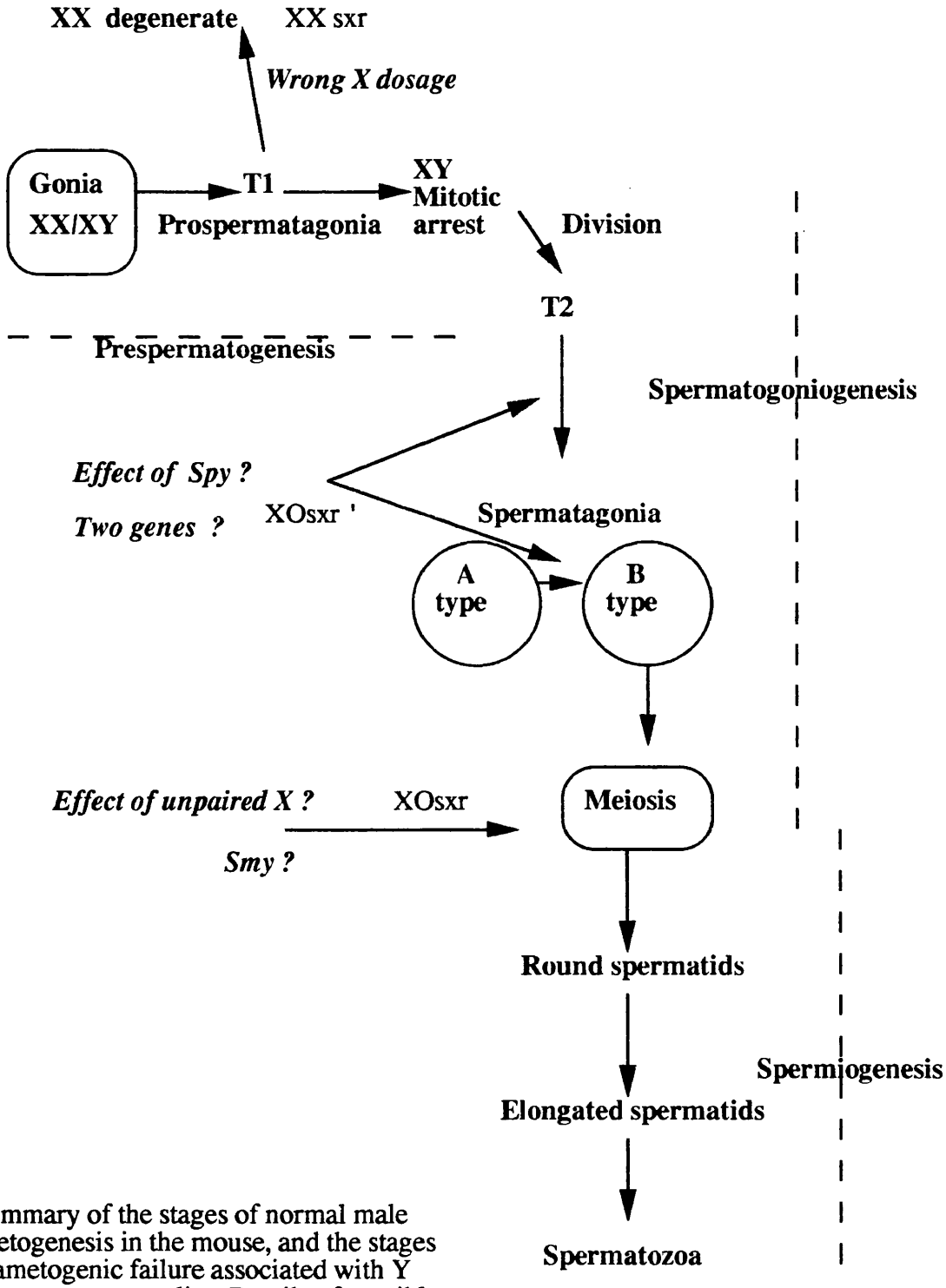
The term spermatogenesis refers to the sequence of differentiation events by which germ cells develop into mature spermatozoa.

In mice, gonia undergo mitosis until approximately two weeks post coitum at which time both male and female germ cells are identical in morphology. Mitotic arrest then occurs in male germ cells which lasts until birth when the process of spermatogoniogenesis begins. This is followed at one week post partum by a meiotic prophase and meiotic divisions forming a spermatid which develops into a mature sperm.(See figure 1.3). An analogous process occurs in the human.

The cascade of gene functions involved in this process, although under initial control by TDF, may be autosomal or X chromosome controlled. However, the study of the effect of different Y chromosome mutations in mice has suggested at least two stages at which Y chromosomal genes must play a direct part.

XX germ cells are able to enter the male pathway as T_1 prospermatogonia but degenerate perinataly due to the presence of two X chromosomes. Levy and Burgoyne (1986) studied the development of XO germ cells in a testicular environment in mosaic XO/XY males. In these cases XO spermatogonia were severly limited if present at all and only the XY spermatogonia underwent complete spermatogenesis. This implies that the germ cells require a Y chromosome and the presence of a spermatogenesis gene designated *Spy*. The location of this gene was indicated by the study of mice with the *Sxr* mutation.

Figure 1.3 Gametogenesis in the male mouse



A summary of the stages of normal male gametogenesis in the mouse, and the stages of gametogenic failure associated with Y chromosome anomalies. Details of possible genetic interaction are given in the text and Burgoyne 1987, Sutcliffe and Burgoyne 1989.

As mentioned previously, XX*Sxr* mice are sterile and spermatogenesis is blocked at the T_1 prospermatogonia stage probably due to the two X chromosomes. However, XO*Sxr* males show all stages of spermatogenesis although the sperm produced are misshapen and few in number. Thus the *Sxr* region must contain the *Spy* gene which enables the XO*Sxr* germ cells to escape the early failure of XX germ cells.
XO *Sxr'* mice, however have very few germ cells entering the meiotic prophase, and so *Spy* can be assigned to the small region which is lost in the *Sxr'* chromosome.

XO *Sxr'* and XY*Sxr'* mice have equal numbers of germ cells at birth and equal numbers of T_1 prospermatogonia. XO *Sxr'* have fewer T_2 prospermatogonia and it is probably at the mitotic stage that *Spy* acts (Sutcliffe & Burgoyne 1989). Of the few T_2 prospermatogonia that enter the spermatogenesis pathway, a large reduction in numbers of fully A type spermatogonia compared to control samples is seen. It is possible that there is a second gene involved at this stage.

The abnormal XO*Sxr* sperm are often diploid (Cattenach *et al.* 1971) indicating omission of the second meiotic division. This may be due to non-pairing of the sex chromosomes in the absence of the Y, or due to the lack of a third gene which is not present in the *Sxr* region, but is involved in sperm morphogenesis and function, *Smy*.

Evidence in favour of meiosis impaired by incorrect chromosome pairing has been put forward by Burgoyne (1987) since XYY males have affected fertility despite presence of the Y chromosome. But $X^{Y}O$ males which have most of the Y chromosome attached to an X chromosome are less severely affected than XO*Sxr* males lending support to a further Y encoded spermiogenesis gene (*Spm*) (Eicher & Washburn,

1986).

In humans a similar spermatogenesis factor (AZF, azoospermia factor) has been postulated (Tiepolo & Zuffardi, 1976) and has been mapped to the distal region of the euchromatic portion of the Y chromosome long arm. In a study of three 45,X males involving Y-derived material translocated to the short arm of chromosomes 14 and 15, the retention of interval 6 of the long arm correlated with fertility of the patient (Andersson *et al.* 1988). This correlates with data from other studies of XYq- males (Vernaugd *et al.* 1986).

1.3.1 Is the H-Y antigen involved in spermatogenesis?

The evidence that the H-Y antigen and the spermatogenesis factor may be one and the same is based on the physical location of the two factors.

Although XO*Sxr* male mice are sterile, all stages of spermatogenesis are present and so the spermatogenesis gene, *Spy,* like *Hya,* must be located within the *Sxr* region (Cattenach *et al.* 1971). Furthermore, spermatogenesis is blocked at an early stage in the *Sxr'* deletion concurrently with the loss of H-Y expression (Burgoyne, 1986,1987). A transcript mapping to the *Sxr'* region in male mice has recently been identified (Mitchell & Bishop 1991). This gene, *Sby,* is conserved on the Y chromosome in a number of mammalian species and is transcribed in the testis. A conserved X chromosome homologue is expressed in many tissues. This gene is proposed as a candidate for *Spy.*

1.4 Genes in the sex-determining region of the Y chromosome

1.4.1 A human zinc finger gene (*ZFY*)

A human zinc finger gene has been described by Page *et al.* (1987c) and for a time, this was the best candidate for the testis determining factor. Deletion analysis of several sex-reversed individuals, two in particular, indicated a particular region of the chromosome that appeared to be both necessary and sufficient to induce testicular differentiation of the bipotential gonad.

The two individuals that appeared to represent the smallest region of the chromosome in which *TDF* could be encoded were an XX male (LGL203) and an XY female (WHT1013) who had a reciprocal translocation between her Y chromosome and chromosome 22 [46X,t(Y;22)p11.2;q11]. Y chromosome material present in the XX male but absent in WHT1013 covered a region of only 140 Kb located several kilobases proximal to the pseudoautosomal boundary.

DNA sequences throughout this region were tested for evolutionary conservation by hybridisation to filters of genomic DNA from males and females of a range of species. One fragment contained within the plasmid pDP1007 showed a high degree of evolutionary conservation, detecting in mammalian species both a Y-specific fragment and one common to both sexes. This latter fragment showed an intensity of hybridisation in females that was approximately double that of males suggesting an X chromosome location. Sequence analysis of the insert of pDP1007 identified an open reading frame of length 1.2 kb. The inferred amino acid sequence appeared to be very similar to other sequences in the database corresponding to a certain family of

eukaryotic sequence-specific binding proteins including the *Xenopus* transcription factor TF IIIA (Miller *et al.* 1985, Fairall *et al.* 1986). These proteins have a particular conformation of tandemly repeated 'fingers' formed by the interaction of a zinc ion with four amino acids, two cysteine and two histidine residues, in a tetrahedral array.

That this gene could encode a transcriptional activator was consistent with the expectations of *TDF*. The gene has been termed *ZFY* (Zinc finger gene on the Y chromosome).

If the region identified in the human was involved in testis determination then this sequence would be expected to be present in the mouse sex determining region. This could be tested using the various mouse *sxr* genotypes. A sequence in this region would be expected to be absent in females, present in one copy in XY males and XX^{sxr} males and in two copies in XY^{sxr} males. The mouse homologue of the *ZFY* sequence (*Zfy*) was used to probe a southern blot of DNA from mice described above, and the expected hybridisation patterns were observed.

The human X chromosome homologue of this sequence (*ZFX*) was cloned and localised to Xpter-p21 by hybrid mapping. This region contains a number of sequences that are similar on the X and Y chromosomes yet remain chromosome specific due to the absence of recombination outside the pseudoautosomal region. Sequence analysis of *ZFX* (Palmer *et al.* 1990) showed a high degree of homology in the predicted amino acid sequences of *ZFX* and *ZFY* (87-97%), varying slightly in different regions of the protein.

In an expression study of this sequence, transcripts of varying

length could be detected by hybridisation of a cDNA *ZFY* clone to a Northern blot of polyA+ RNA from adult testis (Palmer *et al.* 1990). Using a more sensitive technique, whereby cDNA produced by reverse transcription of RNA was amplified using the Polymerase Chain Reaction, expression of *ZFY* in several different adult and fetal male tissues could be detected. *ZFX* was expressed in all male and female adult and fetal tissues tested, and detection of *ZFX* transcripts in somatic cell hybrids containing an inactive X chromosome suggests that this gene may escape X-inactivation.

1.4.2 Homologous genes in the mouse

In contrast to other placental mammals, the mouse has two copies of *ZFY* on its Y chromosome (Mardon *et al.* 1989), these are termed *Zfy*-1 and *Zfy*-2 refering to the detection of 11Kb and 5Kb *Eco* R1 fragments by Southern blot. At reduced stringency, two other fragments are detected corresponding to loci on the X chromosome and chromosome 10 (Nagamine *et al.* 1989).

In order to determine whether both Y chromosome loci were required for testis determination, the *Sxr* series of mouse variants were examined by Southern blot analysis for the presence of the two *Eco* R1 fragments. Both sequences were localised to the sex determining region by the presence of both fragments in the *Sxr* strain. However, the *Zfy*-1 locus was shown to be present in all three *Sxr* derivatives (*Sxr*, *Sxr'* and *Sxr''* a derivative of the *Sxr'* strain that has regained H-Y antigen expression) whereas *Zfy*-2 was absent in the *Sxr'* strains. From this data it appeared that *Zfy*-1 could indeed be involved in sex determination, but *Zfy*-2 seemed to be located in the

region of the chromosome deleted in the *Sxr'* strain and is postulated to be involved in the functions of the H-Y antigen or spermatogenesis.

Analysis of *Zfy*-1 and *Zfy*-2 cDNA clones (Mardon and Page 1989) showed that they differ only by two base pair substitutions within a 135 nucleotide region. Further study of the *Zfy*-2 cDNA confirmed the similarity to the human gene: A long open reading frame encodes a protein of 783 amino acids, the carboxy-terminal half of which comprises 13 putative zinc fingers with 80% homology to human *ZFY* at the amino acid level. The amino portion, like the human protein, is highly acidic, reminiscent of eukaryotic transcriptional activators for example the yeast GAL4 and GCN4 and human glucocorticoid receptor. By analogy, these zinc finger genes would appear to acivate transcription in a sequence specific manner, targeted to particular DNA sequences by the zinc finger domain where the acidic domain may interact with the RNA polymerase II complex to initiate transcription (Ptashne 1988). The presence of a region of basic residues between the acid and zinc finger domain is consistent with a nuclear function of this gene since such regions may act as a nuclear localisation signal (Dingwall and Laskey 1986).

Both the *Zfy*-1 and *Zfy*-2 genes are expressed in adult mouse testis (Nagamine *et al.* 1989, Mardon and Page 1989, Nagamine *et al.* 1990). By PCR amplification of genomic DNA, Nagamine *et al.* showed that the products of the two genes could be distinguished due to the fact that *Zfy*-2 had an 18bp deletion compared to the *Zfy*-1 product. Interestingly, this deletion had not been reported by Mardon *et al.* The difference is due to the particular subspecies from which the Y chromosome is derived. The *Mus musculus musculus* chromosome

harbours the deletion and the *M. m. domesticus* Y does not. These two subspecies can therefore be distinguished on the basis of the nature of *Zfy*-2 and furthermore, by a *Taq* I restriction fragment length polymorphism of the *Zfy*-1 gene (Mardon *et al.* 1989)

1.4.3 *ZFY* as a candidate for the testis determining factor

The *ZFY* gene was a strong candidate for TDF for the following reasons:

1) It was identified within a region shown by deletion analysis of sex reversed individuals to be essential for human male sex determination.

2) The sequence was highly conserved throughout evolution.

3) The sequence corresponded to the sex determining region of the mouse.

4) The gene was transcribed, and the inferred protein product suggested a DNA or RNA binding potential, probably that of a transcription factor inducing the expression of a pathway or cascade of genes involved in the process of sex determination.

However, several questions remained.

1) What is the function of the X chromosome locus.

2) What is the explanation for the occurrence of human hermaphrodites, whose gonads contain ovarian and testicular tissue, when all XX hermaphrodites studied by Page *et al.* lack the *ZFY* sequence?

3) Can one assume that the few XX males lacking any detectable Y chromosome sequence including *ZFY*, arise through mutations in autosomal or X-linked genes involved in the differentiation pathway

leading to testicular differentiation, or is another Y-linked gene involved?

4) Are the expression patterns consistent with those expected of a testis determining gene; i.e. is it expressed immediately prior to the time that the developing gonads become distinguishable as ovaries or testes (at day 12 to 13 in the mouse or the sixth or seventh week of human development)?

Page suggested several models to explain possible interactions between the X and Y homologues of this gene.

a) The *ZFX* protein is not functional in gonadal sex determination which is influenced only by the presence or absence of *ZFY*.
b) The *ZFX* and *ZFY* proteins act antagonistically as positive and negative transcription factors.
c) The two proteins act together in the formation of a multimeric protein, the homodimer of which produces a female phenotype.
d) The two proteins are functionally interchangeable and sex is determined by a dosage effect as a result of X inactivation of one copy in females.

The latter hypothesis can be discounted because, as stated earlier, *ZFX* appears to escape X inactivation. The other theories were examined by Scherer *et al.* (1989). Two XY females with tandem duplications of an X short arm segment were studied and it was shown that the duplication included the *ZFX* locus. The Y chromosome in these cases appeared to be normal, suggesting that the duplication was the cause of the sex inversion, if this were true then the only consistent model is that of an antagonistic mode of action: Despite the presence of ZFY, the patients were female and a heterodimeric

protein did not lead to a male phenotype. It was suggested that the positive effect on testis development by *ZFY* is overridden when the ratio of *ZFX*: *ZFY* exceeds 1:1. However, this was obviously not the only mechanism causing sex inversion as 14 cases of 46,XY females shown to have retained the Y sex determining region were also shown not to have a *ZFX* duplication.

1.4.4 Evidence that *ZFY* is not *TDF*

Despite a high degree of sex chromosome-linked evolutionary conservation of this gene, sequences homologous to *ZFY* in marsupials were shown to be autosomal (Sinclair *et al.* 1988). This was not consistent with the observation that the Y chromosome determines the fate of the gonad in metatherian mammals.

Also, Bull *et al.* (1989) showed that although the sex determining mechanism in reptiles is environmental, mammalian *ZFY* related sequences can be identified in these organisms.

Studies of the expression of *Zfy* were carried out by Koopman *et al.* (1989) who looked at the differentiating mouse testis. The expression of a gene involved in testis differentiation would be expected to occur between 11.5 and 12.5 days *post coitum* (dpc); between the times that the genital ridges of the male and female embryos are indistinguishable and the time that differentiation can be seen to occur. Since neither *Zfy*-1 or *Zfy*-2 expression could be seen in Northern blot analysis of whole tissues (Mardon and Page 1989), fetal genital ridges and gonads were analysed to enrich for gonadal tissue. *Zfy* expression was noted only in adult testis, whereas *Zfx* transcripts could be seen in both ovary and testis and male and female genital ridge over several timepoints from 11.5dpc to adult.

However, using the more sensitive method of PCR from reverse transcribed RNA, with primers specific for *Zfy*-1 and *Zfy*-2, expression of *Zfy*-1 could be seen at 10.5dpc increasing at 12.5dpc and continuing for at least two days. *Zfy*-2 was not expressed during the fetal stages and could therefore be discounted as *Tdy*. Since the expression time of *Zfy*-1 appeared to fit that expected of *Tdy*, the cellular site of expression was studied with regard to the somatic supporting cell lineage of the genital ridge and the germ cells. It is clear that germ cells are not required for testis determination or differentiation since XX*Sxr* mice develop testes in the absence of germ cells as do mice homozygous for the mutation W^e. It was shown by PCR that neither *Zfy*-1 or *Zfy*-2 are expressed in W^e/W^e adult or fetal testis, thus assigning the site of expression to the germ cells, discounting *Zfy*-1 as Tdy and suggesting a role in male germ cell development. These findings were confirmed by Nagamine (1990): Upon examination of three mutant mouse strains giving rise to sterile adults due to blocks at different stages of spermatogenesis, *Zfy* expression in the adult testis was shown to be confined to the maturing germ cell population, being most abundant in the round spermatids. The most likely role for the *ZFY* gene products in human and mouse is in spermatogonial proliferation during spermatogenesis (see section 1.3 and Koopman, Ashworth and Lovell-Badge (1991)).

In the human, genetic evidence that *ZFY* is not the testis determining factor was provided by Palmer *et al.* (1989). Certain XX males had been shown to lack *ZFY* and this was explained by a mutation in a gene involved further along the pathway of differentiation. In these cases one would expect that no exchange of material between the X and Y chromosomes should occur. However, PCR analysis of fourteen such

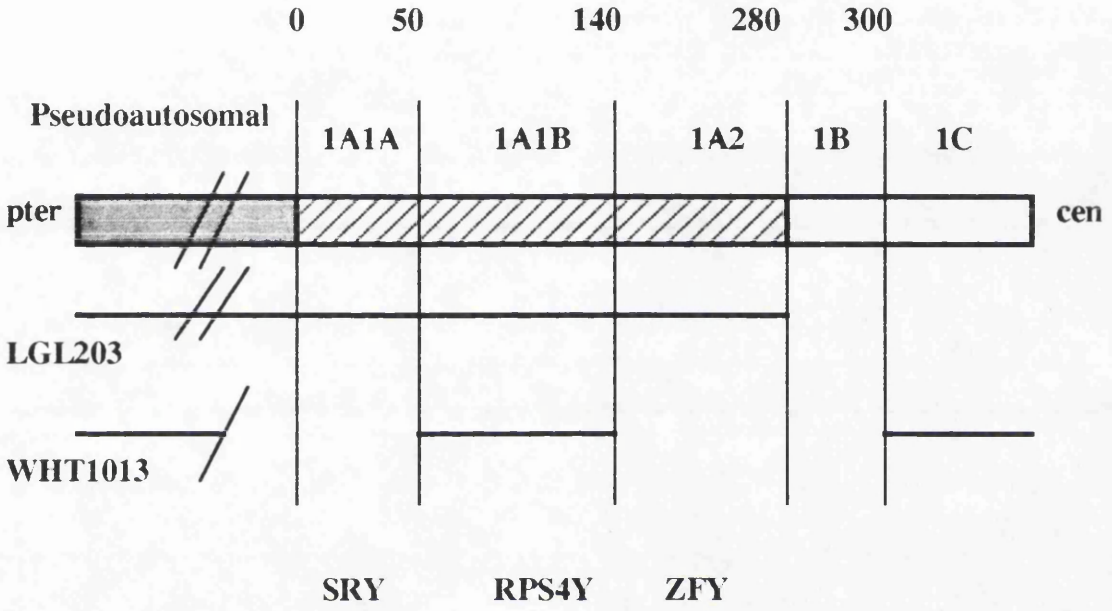
males using a pseudoautosomal boundary-specific primer in conjunction with either an X or Y specific primer showed the presence of the Y-specific side of the pseudoautosomal boundary in four cases. The region of the Y chromosome involved in this transfer was defined by various localised probes and stretched from the pseudoautosomal boundary for 60Kb towards the centromere. This was not the region proposed by Page *et al.* to contain the sex determining gene. It then transpired that the chromosomal rearrangement in the female patient studied by Page *et al.* involved a complex deletion which removed two separate portions of the chromosome, including a second region adjacent to the pseudoautosomal boundary (Page *et al.* 1990, Figure 1.4).

1.4.5 A candidate gene for the testis determining factor (*SRY*)

A 35kb region of the Y chromosome adjacent to the pseudoautosomal boundary was analysed by Sinclair *et al.* (1990). Small fragments were used to probe Southern blots of DNA from XX males lacking the ZFY sequence, and from normal human, murine and bovine males and females. One fragment showed Y specificity in all three species and was also highly conserved throughout eutherian mammals. This sequence was termed *SRY* to reflect its location in the Sex determining Region of the Y chromosome, (*Sry* in the mouse).

Sequence analysis revealed two potential open reading frames, the longer of which is orientated 5'-3' from the direction of the centromere towards the pseudoautosomal boundary. The predicted amino acid sequence shares a conserved motif of 80 amino acids with the Mc protein of the *mat3-M* gene of *Saccharomyces pombe* (Kelly *et al.* 1988). The *mat3-M* locus contains two genes *Mc and Mi* and is one of

Figure 1.4 Diagram of Y chromosome regions retained by XX male LGL203 and XY female WHT1013 [X,t(y;22)].



The complete Y chromosome short arm euchromatic region is represented at the top of the diagram. Regions retained by the XX male and XY female are represented by solid bars. Regional localisation of genes are given. Figures relate to the distance of each breakpoint from the psedoautosomal boundary. Data from Page et al. 1987c, 1990 and Fisher et al. 1990.

two alternative alleles which may be transposed to the *mat-1* locus during mating type switch. The other locus is *mat2-P* containing the genes *Pc* and *Pi*. It is possible that these genes may function as transcription factors. The conserved motif is also similar to a domain found in the nuclear non-histone proteins HMG1 and HMG2. These high mobility group (HMG) proteins are thought to play a part in chromosomal structure and gene activity (Wright and Dixon 1988). Further similarities are found in the non-histone chromosomal protein NHP6 of *Saccharomyces cerevisiae* , the yeast ARS-binding protein ABF2, and the human upstream binding factor *hUBF*, which is an RNA polymerase I transcription factor that interacts in a sequence specific manner with DNA promoter regions (Jantzen *et al.* 1990).

Expression of the *SRY* sequence was observed in Northern blot analysis which identified a transcript in adult testis but not in ovary or in male lung and kidney cell lines. This gene appears to encode a protein that could be involved in the regulation of testis development. The homologous gene in the mouse was cloned and the nature of the sequence and expression patterns were described by Gubbay *et al.* (1990). The human sequence identifies a 3.5Kb male specific *Eco*RI fragment when used to probe a Southern blot of mouse DNA, and also identifies several non Y specific fragments. Sequence analysis of the mouse homologues revealed a family of five distinct genes, one on the Y chromosome and four on autosomes, all of which have in common the conserved sequence described in the human gene. The autosomal sequences were isolated as cDNA clones from an 8.5dpc embryo library, indicating that they are expressed and may be involved in early development.

The Y chromosome sequence was shown, like *Zfy-1*, to be localised in

the minimum region of the mouse chromosome required for testis differention; that retained in XX *Sxr'* males. However, *Sry*, unlike *Zfy-1*, was not present in a particular strain of sex-reversed female mice carrying a mutant Y chromosome (Gubbay *et al.* 1990) . This mutation (*Tdy* ^{m1}) is heritable and can be complemented by a normal Y chromosome (XYY^{Tdym1} individuals are male) and also by *Sxr'*. The mutation appears to be in *Tdy* itself (Lovell-Badge and Robertson, 1990) and the fact that *Sry* is lost concurrently with this mutation suggested at least a close location to *Tdy*.

The method of PCR from reverse transcribed RNA was used to study the time and location of *Sry* expression. Amplification was seen in adult testis and 11.5dpc male urogenital ridge but not in male liver or female genital ridge of the same age. More detailed analysis (Koopman *et al.* 1990) showed *Sry* expression in 10.5 and 11.5dpc urogenital ridges and 12.5dpc testis corresponding precisely with the onset of testis differention. This expression could be seen to correlate with the genital rather than the mesonephric component of the urogenital ridge by *in situ* hybridisation. Furthermore, as expected of *Tdy*, the expression was not germ cell dependent as *W* ^{e}*/ W* ^{e} mutant mice showed amplification identical to the wild type. However, in the adult testis, expression must become germ cell dependent since amplification was not seen in XX *Sxr* and XX *Sxr'* mice which lack germ cells.

This sequence was a very strong candidate for the testis determining gene. In the human, genetic evidence equating *SRY* and *TDF* has been provided by Berta *et al.* (1990). In this study XY females were compared with normal XY males in order to find differences in the *SRY*

gene by single strand conformation polymorphism (SSCP, Orita *et al.* 1989) and by DNA sequencing. An amplification product is produced by PCR of the sample DNA using *SRY* specific oligonucleotide primers. This is then cleaved by restriction enzyme digestion and fragments separated on non-denaturing polyacrylamide gels. In 50 normal males and the majority of patients the same band pattern was seen, but two XY females had abnormal patterns. In one case the patients' father and brother had normal patterns and her mutation must be *de novo*, the second case shared an abnormal band pattern with her father from whom she inherited it. From sequence analysis, both mutations would lead to conservative amino acid changes in the conserved DNA-binding motif of the SRY protein. In the case of the *de novo* mutation, the evidence is that the mutation and sex-reversed phenotype are related. In the second case, the sex-reversal occurring only in the daughter could be dependent upon other genetic or environmental factors, or be present in addition to a second mutation in an X chromosome or autosome located sex determining gene. Other mutations within the *SRY* gene have been described (Affara *et al.*, unpublished, Jager *et al.*, 1990). However, mutations in the regions of *SRY* studied have not been identified in all of the XY females which retain these sequences. Nor do all XX males retain these sequences (Ferguson-Smith *et al.* 1991). In a study by North M.A. (1991) of 14 *ZFY* negative XX males, ten DNA samples failed to amplify in a PCR reaction using primers based on a conserved region of the *SRY* sequence. As *SRY* has been shown to be sufficient for testis determination in the mouse (see below), these results support the theory that some cases of sex-reversal may be due to mutations in a series of genes involved at additional stages of development.

1.4.6 Mice transgenic for *Sry*

Conclusive evidence that this sequence is the testis determining gene has been provided by the generation of mice transgenic for the *Sry* sequence (Koopman *et al.* 1991).

A 14kb sequence representing the conserved domain of *Sry* along with upstream and downstream regions was injected into fertile oocytes which were transferred to the oviduct of pseudopregnant females. Of 168 fetuses examined at 14 days, 2 were phenotypically male but were shown to lack a Y chromosome by the absence of *Zfy*-1 sequences, and were transgenic for multiple copies of the *Sry* sequence. Of 93 further mice allowed to develop to term, 5 were transgenic. One of which was a phenotypically normal XX male. Despite normal male development this mouse was sterile and had reduced testis weight as is characteristic of sex reversed animals with two X chromosomes or no other Y chromosome regions. Two other transgenic animals and four 14 day old fetuses were fertile females despite also containing many copies of *Sry*. This lack of sex reversal could be due to an effect of the position of integration of the transgene.

A sequence including the human *SRY* sequence was also used to produce transgenic mice. In the cases of three independent integrations no testis development was observed, although expression of *SRY* could be detected by PCR analysis of samples taken from 11.5-12dpc genital ridges of transgenic XX fetuses. There are several possible explanations for this. Either the RNA is not processed or translated correctly in the murine environment, the protein product is unstable, or, most likely, the differences in sequence between the mouse and human proteins render it unable to interact with other regulatory

genes or target proteins.

1.5 The pseudoautosomal region

Segregation of the X and Y chromosomes to individual spermatozoa in male meiosis is acheived by the pairing which occurs in the pseudoautosomal regions, this pairing may be visualised by electron microscopy and can involve up to 70% of the Y chromosome (Chandley *et al.* 1984, Ellis and Goodfellow 1989). The inheritance of randomly isolated homologous X and Y sequences in this region was studied by Cooke *et al.* 1985, and Simmler *et al.* 1985. These studies demonstrated the occurrence of a single obligatory recombination event taking place in male meiosis, as has been shown in the murine transfer of the *Sxr* region between the sex chromosomes with a recombination frequency of 50% (Evans *et al.* 1982).

Two genes have been cloned and mapped to the pseudoautosomal region of both the X and Y chromosomes, *MIC 2* encodes a cell surface antigen, and *CSF-2R* encodes the receptor for a haemopoeitic regulator. It is possible that a gene pre-disposing to schizophrenia is located in this region (Crow 1989).

Since genes in the psedoautosomal region are expressed from both chromosomes in males, it is likely that in females these genes escape X inactivation in order to retain equivalent dosage (see below).

1.5.1 A genetic map of the region

In a study of three highly polymorphic pseudoautosomal loci, several genetic properties of the pseudoautosomal region were described (Rouyer *et al.* 1986). The inheritance of the alleles of DXYS14,15 and 17 in large families confirmed that recombination does occur between the sex chromosomes. Also, the recombination frequency of the

most telomeric locus, DXYS14, approaches 50%. This is explicable by a single obligatory recombination event. No double recombinations were observed even between distant markers. Thus more distal loci co-segregate in recombinations involving the proximal ones.

Since the size of the region is estimated at 2.6 Mbp by pulsed field gel analysis (Petit *et al.* 1988), then the recombination frequency of pseudoautosomal markers in male meiosis is around twenty fold higher than in female meiosis (where 1% recombination (1 centimorgan) represents a chromosomal distance of 1.2 Mbp (Morton 1991)), and the region appears to act as a recombinational hotspot in males. A detailed genetic map may be constructed, with recombination values being additive between different markers. In accordance with a single crossover in the pseudoautosomal region, markers show a recombination frequency in direct correlation to their location with respect to the sex-specific regions.

Two further polymorphic pseudoautosomal markers have been added to this map by Page *et al.* (1987b), confirming the order suggested by Rouyer *et al.*. These two loci DXYS20 and DXYS28 were shown to be pseudoautosomal, rather than autosomal, by genetic proof in a study of the inheritance of polymorphisms in the loci in 45,XO females. The inherited polymorphism was always the same as the polymorphism exhibited by the parent from whom the XO female inherited a sex chromosome, and could therefore be shown to be sex-linked rather than linked to an autosome.

These two loci are duplicated to form blocks of tandemly repeated sequences which vary greatly in number between individuals, and also appear to mutate frequently to form further restriction sites leading

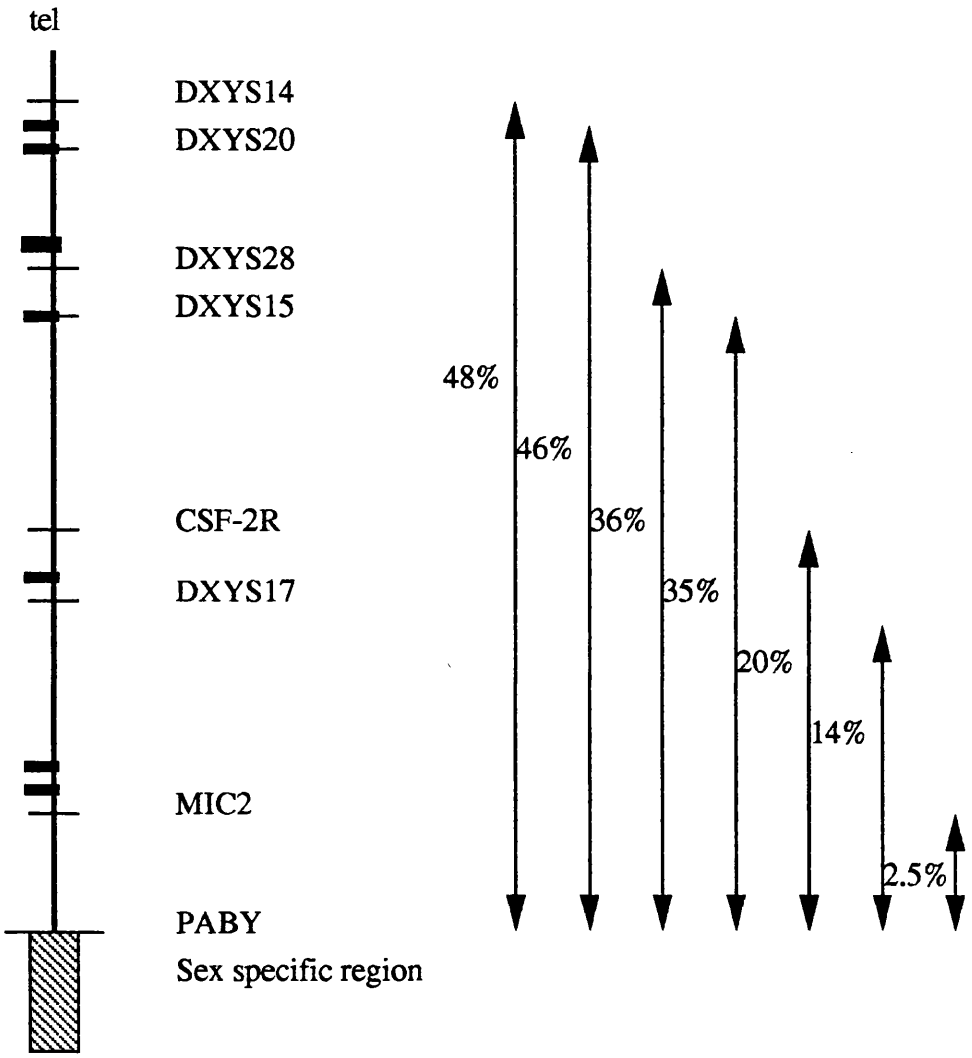
to a number of restriction fragment length polymorphisms within the block. It has been suggested (Page *et al.* 1987b) that the pseudoautosomal region is comprised mainly of such repetitive elements which may be instrumental in pairing of the sex chromosomes and recombination in male meiosis.

1.5.2 Physical mapping of the pseudoautosomal region

The genetic linkage map is in good agreement with physical maps of the region produced by Brown (1988) and Petit *et al.* (1988). A map extending 2.3 megabasepairs from the telomere to sex-chromosome specific DNA was achieved using pulsed-field gel electrophoresis (PFGE), to separate partial digestion products of DNA from Y-only somatic cell hybrids (3E7 and 853) cleaved with rare-cutting restriction enzymes. Since the two maps are correlated in terms of distance and order, this suggests that recombination is equally likely at any point in this region.

Use of rare-cutting enzymes indicates the position of CpG islands where under-methylated CpG dinucleotides are clustered. Such regions have been associated with the 5' end of genes (Bird 1986). In the pseudoautosomal region several such islands may be associated with gene loci (for example the *MIC 2* gene) but two areas of extreme density, near the telomere and around DXYS28 may reflect structural features related to telomeric functions (see figure 1.5)

Figure 1.5 Recombination frequency of markers in the pseudoautosomal region and location of CpG islands.



Recombination frequencies are given between each marker and the sex-specific region. Some of these markers have been positioned on pulsed field maps, their position is in agreement with the genetic map. Locations of CpG islands are indicated by heavy bars. (Goodfellow et al. 1987, Page et al. 1987, Brown 1988, Petit et al. 1988).

1.5.3 The pseudoautosomal boundary

The boundary region must play an important role in the separation of the two different chromosomal properties on either side of it, that is, to form a barrier proximal to which recombination between the X and Y chromosomes ceases to occur.

Using the *MIC 2* gene as the closest pseudoautosomal marker, Ellis *et al.* (1989) isolated 110Kb of DNA between *MIC 2* and DYS104, the most distal Y-specific marker, in an attempt to clone the boundary itself. Sequence analysis of the boundary highlighted the similarity between the X and Y chromosomes in the pseudoautosomal region. Only two base pair difference was seen out of 245bp sequenced. However, the insertion of an *Alu* sequence into the Y chromosome disrupts the homology, until it is partially resumed (at a level of 77%) for 225bp at the other side of the insertion. Proximal to this, the sequences become completely sex-specific. It was postulated that the *Alu* insertion had formed a new boundary distal to the more ancient boundary creating divergence through lack of recombination in this partially homologous region. The *Alu* sequence is Y-specific and is the same in human, chimp, gorilla and orangutan. However, more recent evidence from the sequence of the boundary region of Old World monkeys (baboons and macaques), which do not have the *Alu* insertion, places the boundary at the same site as the proposed insertion with sequence divergence between the X and Y occurring at this point. The possibility that the *Alu* is a more recent acquisition is confirmed by comparison of the sequence of the inserted *Alu* with the consensus *Alu*. These two elements diverge by less than 77% implying that the *Alu* insertion is more recent than other divergence events. Hence the boundary is formed by a sharp sequence diversion and was not

relocated by the *Alu* insertion. The mechanism by which this abrupt cessation of homology is achieved is not understood, but may be related to the close proximity of the testis determining factor to the boundary, such that there is a large selective disadvantage against recombination events occurring in the vicinity of the gene (Bengtsson & Goodfellow, 1989).

1.5.4 The gene *MIC 2*

The monoclonal antibody 12E7 recognises a cell surface antigen which is encoded by the gene *MIC 2* (Goodfellow 1983, Darling *et al.* 1986). This gene has been mapped to the X and Y chromosomes using antigen-antibody reactions in somatic cell hybrids (Goodfellow 1983), and has been localised to the pseudoautosomal region by the *in situ* hybridisation of a cDNA clone (Buckle *et al.* 1985).

Although the *MIC 2* cDNA has been sequenced and includes a long open reading frame, which has been shown to correspond to the antigen recognised by 12E7, no homology with any other protein is apparent and the function is unknown except for the amino terminal region which is likely to be a signal peptide directing the protein to the cell surface (Goodfellow *et al.* 1987a).

An HTF island at the 5' end of the *MIC 2* gene is very GC rich and appears to contain a cluster of unmethylated sites for rare-cutting enzymes, of which there is no difference between the active X, inactive X or Y chromosomes confirming the implication that the gene should escape X inactivation (Goodfellow *et al.* 1987a).

The 12E7 antigen expression in red blood cells is polymorphic in that the expression may be either high or low. This is associated with the

XG blood group locus on the X chromosome. *XG* ^{a} antigen positive individuals are high expressors of 12E7, and *XG* ^{a} antigen negative individuals are low expressors. Both of these genes are thought to be controlled by an X-linked *cis*-acting sequence, *XGR,* which has two alleles. *XGR* A may turn on the *XG* antigen structural locus and *MIC2* leading to expression of *XG* ^{a} and high 12E7 expression and, conversely, *XGR* B may inhibit *XG* ^{a} expression. Since *XG* ^{a} antigen negative males may be high or low expressors, it is possible that a Y-linked controlling gene also exists which could have an effect on the X-linked structural locus (Goodfellow *et al.* 1987b).

1.5.4 The gene *CSF-2R*

The gene encoding the receptor for the haemopoietic regulator, granulocyte-macrophage colony stimulating factor has been cloned (Gough *et al.* 1990). This factor GM-CSF stimulates the proliferation, differentiation and activation of granulocytes and macrophages, and may be involved in the generation of M2 acute myeloid leukaemia (AML). Since the loss of either the X or the Y chromosome is associated with 25% of AML of the M2 subtype, this is indicative of the involvement of a recessive oncogene (Ponder 1988). The loss or inactivation of this gene in both copies of a myeloid progenitor cell would be expected to generate a clone of cells unable to respond to GM-CSF, and these would be undifferentiated in phenotype as is displayed by M2 AMLs.

Gough *et al.* have mapped a cDNA clone to the X chromosome using somatic cell hybrids, and this has been sub-localised to Xp21-pter and also Ypter-p11 by *in-situ* hybridisation. Using the cDNA to probe digests of genomic DNA, two fragments have been detected. One of

these is invariant and the other is polymorphic for six different sized alleles. This polymorphism was used in family studies to prove that the sequence recombines between the sex chromosomes and is pseudoautosomal. Based on the 20% recombination frequency between the gene locus and the sex-specific region, the gene probably maps between DXYS15 and DXYS17.

1.6 Homology outside the pairing region

In addition to homology within the pseudoautosomal region, several other blocks of homology between the X and Y chromosomes have been identified. Two major regions which encode homologous genes are 1) Xp 22.1-ter and Yq11, and 2) Xq13-24 and Yp. See figure 1.6

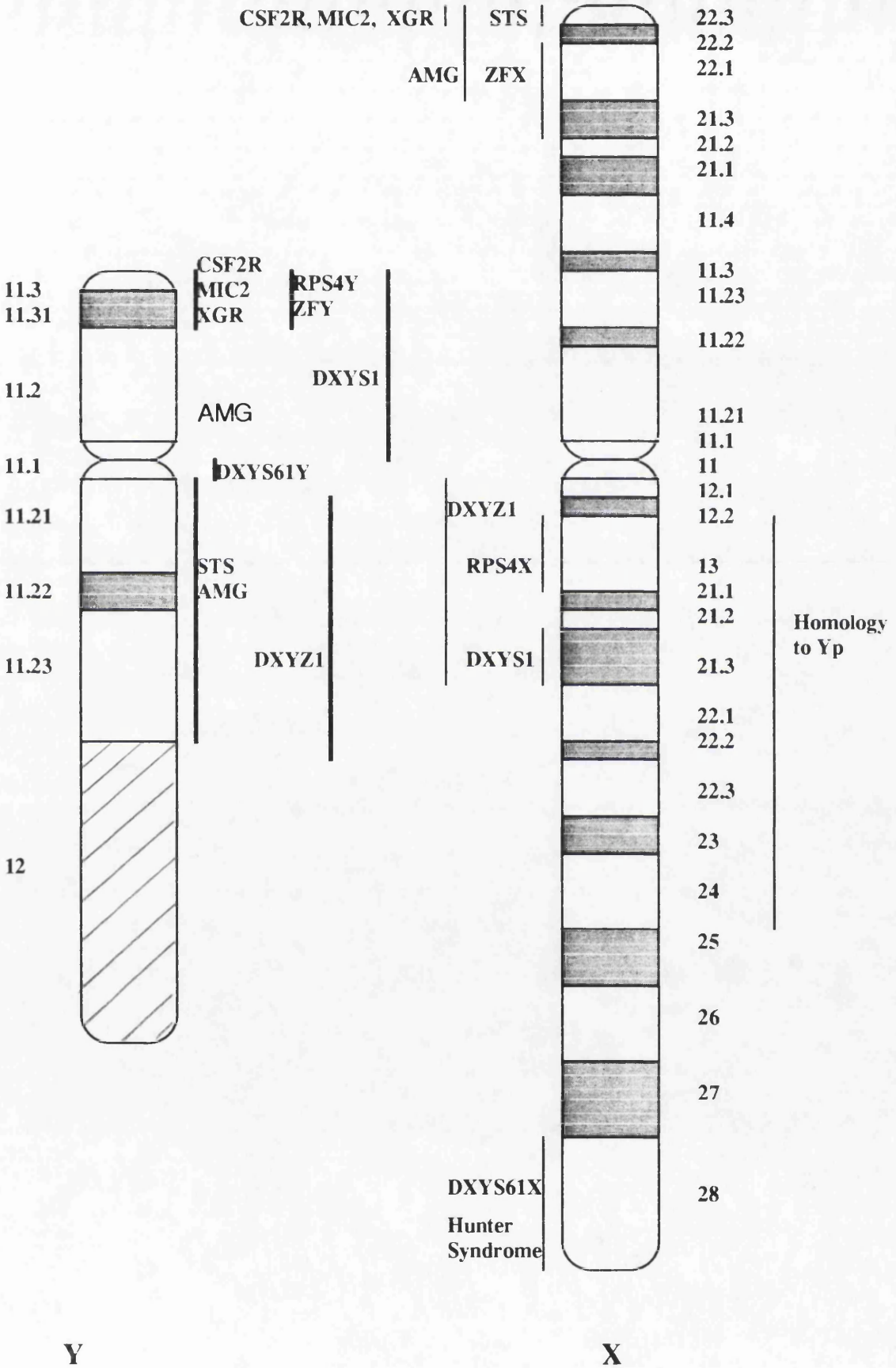
1.6.1 Short arm X : long arm Y homology

The steroid sulphatase gene (*STS*)

The steroid sulphatase gene is located on the distal X chromosome short arm in humans, close to the pseudoautosomal region, and escapes X inactivation. A deficiency of STS, a microsomal enzyme important in steroid metabolism, produces the syndrome of X-linked icthyosis which is a common congenital metabolic defect (1 in 5,000 males).

A 5' cDNA probe recognises sequences on the Y chromosome in addition to those detected on the X, although a probe corresponding to the 3' cDNA does not (Fraser *et al.* 1987). This locus (*STSP*) is assigned to Yq 11.2 by *in situ* hybridisation (Buckle *et al.* 1987) and probably represents pseudogene sequences since there is no functional *STS* gene on the Y chromosome (Yen *et al.* 1987). In contrast, in the mouse, the gene behaves in a pseudoautosomal manner (Keitges *et al.* 1985). It is expressed from both the X and Y chromosomes and appears to undergo X-Y exchange, however, these genes do not show stringent homology to the human sequences. It is likely that a pericentric inversion occuring in the human Y chromosome has removed the *STS* gene from its original pseudoautosomal location to the long arm where divergence from the X-encoded sequences has taken place.

Figure 1.6 Homology between the X and Y chromosomes



Patients with *STS* deficiency are often seen to have cytologically visible X chromosome alterations either in the form of Xp deletions or X/Y translocations. Some patients have complex groups of manifestations presumably caused by deletions involving other genes besides *STS*, (Contiguous deletion syndromes, Ballabio *et al.* 1991). Others have icthyosis only which may be caused by deletions of all or part of the *STS* gene. The relatively high occurrence of the disorder could be due to a high mutation rate, perhaps accounted for by abnormal pairing and unequal exchange between low copy number repeat elements of the X and Y chromosome. Ballabio *et al.* (1988) demonstrate an X-Y interchange occurring between Xp 22.3 and Yq 11 leading to chromosomal abnormalities including icthyosis. It is postulated that it is the homology between Xp and Yq that causes aberrant recombination events to occur. However, this does not seem to be the only reason for mutations in the *STS* locus. Shapiro *et al.* (1989) analysed 30 icthyosis patients for *STS* gene structure, *STS* - encoded mRNA and presence of inactive proteins detectable with anti-STS antiserum. By Southern blot analysis, 27 of the patients had total gene deletions whilst retaining X chromosome markers both distal (DXS283) and proximal(DXS143) to the *STS* locus. Two patients appeared to have normal *STS* sequences and a normal level of mRNA, however no STS protein was detected suggesting small mutations in the gene leading to unstable proteins. The retention of the distal marker, DXS283, argues against the deletions being caused by abnormal X-Y interchange. In addition, Shapiro *et al.* did not detect any Y chromosomal sequences on the X chromosomes of four *STS* patients using a number of probes located in distal Yp, nor did they detect loss of *STS* in any of seven XX males tested. Further studies of the nature

of the breakpoints need to be undertaken before the mechanism producing these deletions can be determined. (Gillard *et al.* 1987).

The gene encoding Amelogenin (*AMG*)

A second gene located in this particular region of homology has been described by Lau *et al.* (1989). This gene encodes amelogenin, the predominant protein in the developing enamel matrix of mammalian teeth, which may be important in the control of biomineralization (the laying down of mineral during enamel maturation). Human genetic disorders affecting the formation of tooth enamel, Amelogenesis Imperfecta (AI) show heterogeneous modes of inheritance with an indication that loci on the X and Y chromosomes may be involved. Using a mouse cDNA probe in hybridisation to somatic cell hybrids containing intact and deleted chromosomes, Lau *et al.* showed that the *AMG* sequence is located on the human X chromosome p22.1-22.3 region and a homologous region is located at proximal Yq11 (Amelogenin-like *AMGL*). The corresponding mouse sequence appears to be located exclusively on the mouse X chromosome proximal to the X:Y pairing region (Chapman *et al.* 1991).

It is not known whether the Y chromosome locus in humans is transcriptionally active or a pseudogene, however, its location in the same region as *STS* on both the X and Y chromosomes suggests that it may have been involved in the same pericentric inversion and the Y-linked sequences may have undergone some divergence. Study of the genomic DNA sequence and the inferred amino acid sequence (Nakahori *et al.* 1991a), which are 88.9% homologous at the DNA level (93% in presumed exons) and 89% homologous at the amino acid level, suggest that the gene may be active or may recently have been active. This is

based on the observation that amino acid substitutions on the Y chromosome are limited to regions where significant differences among human X, murine and bovine sequences are found. In a recent paper (Nakahori *et al.* 1991b), it is suggested that the amelogenin gene is actually situated on the human Y chromosome short arm. Evidence is given from deletion mapping studies and the fact that the locus is present in an XX male assumed to retain only the short arm of the chromosome. However, due to the possibility of complex deletion events this evidence should be treated with caution.

1.6.2 Long arm X : short arm Y homology

A second region of X:Y chromosome homology involves sequences within a large region of the X chromosome, Xq13-24, and the short arm of the Y chromosome proximal to the pseudoautosomal region. Homologous sequences in these regions have been noted by several groups (Wolfe *et al.* 1984, Bishop *et al.* 1983).

Homology at locus DXYS1

Page *et al.* (1984) have examined the location and homology of the sequence DXYS1, the first site of single copy DNA sequence homology described on the Y and X chromosomes. Although the DNA segment is homologous, an RFLP detects strictly X and Y-linked fragments suggesting that recombination does not occur between the two loci. *In situ* hybridisation has localised the sequence, as expected, outside the pseudoautosomal region to Xq13.2-q21.2 and proximal Yp.

A comparative restriction map of the X and Y chromosomes around the region detected homology over at least 36Kb of DNA, with a very low amount of divergence (0.37±0.38%). This is less than the divergence of the human and chimpanzee genomes (1-3%). It is possible either

that there is some mechanism whereby divergence of this X-Y homology is retarded, or that that the sequences represent an event which took place in recent evolutionary history. As DXYS1 is present only on the X chromosome of the great apes it is likely that the Y-linked homologous region in humans arose from a recent transposition from the X to the Y chromosome.

There is evidence for at least one gene in this region. The conditions of cleft palate and ankyloglossia often have a multifactorial origin including polygenic and environmental factors. However, in one family a single gene defect causing both of these conditions has been mapped to the X chromosome (Moore *et al.* 1987). Linkage studies show a positive lod score of above 3 at $\theta=0$, with the probe pDP34 (locus DXYS1). Although there does not appear to be any evidence of a functional Y-linked gene, cloning of the cleft palate gene may uncover homologous sequences on the Y chromosome.

Homologous ribosomal protein genes

The first cloned gene sequence, mapping to this region of homology on both the X and Y chromosomes, has recently been described. (Fisher *et al.* 1990a). A segment of single copy genomic DNA was shown to detect male-specific transcripts when hybridised to Northern blots of human poly(A)+RNA. A corresponding cDNA clone was isolated which detected the same transcripts. These are present in both the XX male (LGL203) and the X,t(Y;22) female (WHT103) (described previously in section 1.4.1) who have in common only interval 1A1B of the Y chromosome. From sequence analysis of the cDNA clone (pDP1278) and predicted amino acid sequence, the gene appears to encode a protein of 263 amino acids. The genomic organisation consists of at least five exons

spanning 23Kb or more and is contained within interval 1A1B of the Y chromosome.

A second subset of cDNA clones, which hybridised weakly to the original Y-derived genomic probe, were sequenced and found to share homology with pDP1278. The predicted amino acid sequence is the same length and 93% identical, and the nucleotide sequences are 82% identical in the coding regions but are diverged in the 5' and 3' untranslated regions. These clones were shown to be derived from sequences residing on the X chromosome (Xq13.1).

Both the Y and X chromosome-encoded predicted protein sequences are similar to ribosomal proteins in a wide range of species including archaebacteria. The rat protein RPS4 (Ribosomal Protein S4) is one of approximately 33 proteins present in the 40S ribosomal subunit (Sherton & Wool 1972). The human sequences are highly homologous to the rat protein and appear to encode isoforms of RPS4. They have been termed RPS4Y and RPS4X to reflect their chromosomal origin.

By Northern analysis, transcripts from *RPS4Y* were detected in a wide variety of male adult and fetal cell-lines and tissues. An identical sized transcript (1.0 kb) from *RPS4X* was detected in all male and female cell-lines and tissues tested, including somatic cell hybrids containing an inactive X chromosome which indicates that *RPS4X* escapes X-inactivation.

It is possible that haploinsufficiency of this gene may be involved in the Turner phenotype. Turner syndrome is a common human chromosomal disorder often associated with a 45,XO karyotype but also with structural abnormality of either sex chromosome, including

deletions of Yp observed in XY females displaying some of the abnormalities associated with Turner syndrome.

The X,t(Y;22) female does not exhibit Turner syndrome despite the Yp deletions, and it is possible that a Y-linked 'anti-Turner' gene may be located in the Y chromosome regions retained by her. *RPS4X* and *RPS4Y* are likely candidates for this factor as they represent homologous expressed genes, of which the Y chromosome homologue is located in the region described above. Furthermore, the X-linked gene would be expected to escape X-inactivation (otherwise no difference would be observed between 46,XX and 45, XO individuals). However, so far no structural abnormalities associated with a Turner phenotype have been shown to involve *RPS4X*.

1.6.3 Other homologous regions

It would appear that the homology between the X and Y chromosomes is widespread. Outside the pairing region, homology between the sex-specific regions of both chromosome short arms is signified by the presence of the genes *ZFX* and *ZFY* . There is homology between the long and short arms as described above, and further to this Cooke (1984) has described a homologous stretch of at least 1Kb between the X (Xq24-qter) and Y chromosome long arms (DXYS61Y).

Within the Y chromosome itself, there appear to be regions of homology, probably brought about by duplication events, with several probes identifying more than one locus (pO19P3/pJA36B, Shortle 1990, Arnemann *et al* 1987; GMGY46, Affara, pers. comm.). Fisher (1990b) has demonstrated the presence of a segment of DNA located in the pseudoautosomal region of both the X and Y chromosomes, and duplicated in Yq.

1.7 Repetitive elements on the Y chromosome

The Y chromosome contains several different types of repetitive elements, some shared with the autosomes and some that are chromsome-specific.

1.7.1 *Alu* and LINE elements

Repeats of the *Alu* family are present on the Y chromosome although they appear to be diverged from the genome consensus sequence (Smith *et al.* 1987) and are very under-represented in the heterochromatic block (Moyzis 1989).

LINE (Long interspersed repeat elements) on the Y chromosome, however, appear to be indistinguishable from those found elsewhere in the genome and seem to be equally abundant (Schmeckpeper *et al.*, reported in Smith *et al.* 1987).

1.7.2 Alphoid repeats (DYZ3)

The alphoid family of sequences has been found in all primates and is preferentially located at centromeres. In humans a similar repeat is observed on many autosomes and is characterised by a 340bp repeating unit on cleavage of DNA with *Eco*RI.

Wolfe *et al.* (1985) used a cloned alphoid repeat element from the human X chromosome to screen a Y chromosome cosmid library and identified clones which, when hybridised to genomic DNA digested with *Eco*RI, detected a 5.5 kb Y-specific fragment. The repeat element was shown by *in-situ* hybridisation to be located at the Y chromosome centromere. Although some hybridisation to the centromeres of several autosomes was observed, from Southern analysis these chromsomes do not seem to exhibit the 5.5 kb repeat arrangement. Tyler-Smith (review 1987) isolated the Y chromosome alphoid repeat element in a

similar way and studied the higher order organisation. The alphoid block is comprised of tandemly repeating elements of 5.7 kb in length, which are in turn made up of 170bp repeating subunits (on digestion with *Hae* III). Sequence data show that the 170bp subunits are more similar to a Y chromosome consensus sequence than they are to alphoid sequences from other chromosomes. The 5.7 kb units show little sequence divergence with restriction enzyme sites occurring in either all or non of the units.

A long range map of the alphoid block of the Y chromosome contained in the somatic cell hybrid 853 (Burk and Smith 1985) indicates that the repeats are located in a single block which is flanked on one side by complex euchromatic DNA and on the other with what may be simple sequence DNA, possibly comprised of a further repeat element.

1.7.3 Y chromosome-specific repeats, DYZ1 and DYZ2

Digestion of male genomic DNA with the enzyme *Hae*III yields two distinct bands as visualised by U.V. transillumination of ethidium-bromide stained gels. They are 3.4 and 2.1 kb in size (Cooke 1976), are not present in digests of female DNA and do not hybridise to each other (Smith *et al.* 1987). These bands correspond to the cleavage of blocks of tandemly repeated DNA. The 3.4 kb repeat is designated DYZ1 and the 2.1 kb repeat DYZ2. These elements are distributed throughout the Y chromsome long arm.

DYZ2

The Y-specific 2.1kb *Hae*III fragment is related to autosomal homologues which have a *Hae*III fragment size of 1.9kb. Both the Y-specific and autosomal sequences have a basic repeat unit of 2.4kb.

This is apparent when the 2.1kb fragment is used to probe digests of male and female DNA digested with enzymes other than *Hae*III.

DYZ2 is composed of an 800 bp sequence common to the Y chromosome and to autosomes connected with a 1.6 kb Y chromosome specific element. On the autosomes the same 800bp 'common' sequence is associated with autosome specific segments. The two segments are repeated together to form the repetitive block. In gorilla, digests of genomic DNA probed with the 2.4 kb repeat show the same pattern of hybridisation in males and females, suggesting an autosomal location for homologous sequences. This has been confirmed by *in situ* hybridisation. The hybridisation is due only to the common sequence however, and even though the human autosome specific segments are shown by hybridisation to be present in the gorilla genome, they are not associated with the common segments. These common segments appear to be capable of forming blocks of tandem repeats associated with a range of unrelated DNA segments.

DYZ1

The 3.4 kb *Hae* III repeat elements also contain Y-specific sequences interspersed with non-Y-specific sequences (Nakagome *et al.* 1991) although the arrangement is more complicated than DYZ2. In this case different 'common' sequences exist, each of which has its own autosomal specificity. Each cluster of the repeat on the Y chromosome has its own 'common' element. Linked to these are heterogeneous Y-specific and autosome-related blocks. Despite this diversity of elements within the family, each member is characterised by a conserved segment containing a *Hae* III site which ocurs at 3.4 kb intervals. Sequence analysis by Nakahori *et al.*(1986) showed that one

of these repeating units consists of 713 pentanucleotides of which 229 were TTCCA and 298 were one nucleotide substituents of it.

1.7.4 Other Y-specific repeats

Several other lower copy number repeat elements have been described. These are located on both the long and short arms of the chromosome. Locus DYZ4 (pDP105) consists of multiple Y-specific loci on Yp and Yq (Bernstein *et al.* 1987). DYZ5 (Y-190) consists of a major array of 20kb units which is up to several hundred kilobases long, and is localised to Yp. This array is variable in size between individuals (Oakey & Tyler-Smith 1990). DYZ6 (pJA36B, Arnemann *et al.* 1987) and DYZ7 (p71C2, Jakubiczka *et al.* 1989) are moderately repeated elements on Yp and DYZ8 (p21A1, Jakubiczka *et al.* 1989) represents a short alternating repeat unit on Yq.

1.8 Y Chromosome polymorphism

The Y chromosome shows a high level of polymorphism in the pseudoautosomal region but other regions are noticeably reduced in polymorphic loci.

1.8.1 The pseudoautosomal region

The first example of polymorphism in the pseudoautosomal region was demonstrated by Simmler *et al.* (1985). A subclone of cosmid 113 identifies locus DXYS15. Identical fragments are observed in digests of male and female DNA, with numerous polymorphic alleles revealed by digestion with several enzymes. The same alleles are observed in males and females as a result of the pseudoautosomal nature of the sequence.

A further subclone of the same cosmid detects the locus DXYZ2 which is a repeated sequence localised to Xp22 and Yp by deletion mapping (Simmler *et al.* 1985, Schempp *et al.* 1989). This sequence also shows some polymorphism in the number of repeated subunits.

1.8.2 Y Chromosome telomere

Telomeric sequences are known to be comprised of repeated sequences and to be highly polymorphic in the population. Cooke *et al* (1985) have isolated an homologous telomeric sequence from the X and Y chromosomes. Hybridisation of p29C1 (DXYS14) to *Bam* HI digests of genomic DNA identifies a smear of 18-23kb which is sensitive to *Bal* 31 exonuclease digestion. The sequence is highly polymorphic, probably due to variable addition of copies of the terminal repeat. In a study of 23 individuals with three restriction enzymes no two hybridisation patterns were the same.

1.8.3 Polymorphic repeat elements

Polymorphism in other repeat elements is also common. For example, the length of the long arm heterochromatic region varies greatly depending upon the number of copies of DYZ1 (McKay *et al.* 1976). Oakey and Tyler-Smith (1990) have attempted to study the genealogy of Y chromosome DNA in the same way that female lineages may be studied using mitochondrial DNA. Three repetitive probes were used to analyse polymorphism in 39 European and Asian males and three cell lines. The probes were DYZ3, DYZ5 and the *pox Y1* locus. *PoxY1* is a low copy probe linked to simple sequence DNA and identifies large, hypervariable fragments in *Bgl*II and *Xba*I digests. With a single enzyme, and using a probe recognising DYZ3, 24 different sized alleles of alphoid DNA sequences were observed. Using DYZ5 and *poxY1*, 12 and 19 different sized alleles were observed respectively. The distribution of variants does not seem to be random. The size of the alphoid region in most individuals is either 300kb or 900kb with small size variation around these sizes. The largest blocks have several sites for the enzymes *Ava*II, *Eco*O1091 and *Hin*dIII whereas the small blocks have just one site. Also, particular alleles of the DYZ5 array seem to be associated with one or another group, as do the *poxY1* alleles. It is unlikely that any selection associated with functional constraints is acting upon these regions, or that a mechanism of convergent evolution could act on so many independent features. It is more likely that the Y chromosomes in each group have descended from a single male ancestor.

From this data it is possible that the majority of men harbour one of two ancestral Y chromosomes.

1.8.4 Polymorphism at locus DYS1

A similar study has been carried out by Ngo *et al.* (1986). A probe (49f) derived from locus DYS1 on the long arm of the Y chromosome detects up to 18 fragments when hybridised to Southern blots of *Taq*I-digested male genomic DNA. Two fragments are also present in female DNA and have been assigned to an autosome. The probe detects a family of moderately repeated sequences present in restriction fragments of different sizes. Of the 18 fragments, 5 are polymorphic exhibiting variation both in size, and presence or absence of the fragment.
In family segregation studies, all males of the same family display an identical hybridisation pattern, confirming that the results are due to restriction fragment length polymorphism (RFLP). No variation was observed with other enzymes indicating that the RFLP is due to point mutations rather than deletions or insertions.

In a study of 44 unrelated individuals, each male was scored for the alternative alleles of each of the polymorphic loci. In all, 16 different combinations were seen each defining a haplotype of the Y chromosome. The haplotypes show a degree of non-random association of alleles since the observed frequency of each haplotype is very different from that expected by random segregation. Three haplotypes are more common than the others and all could be derived from another by a single mutation. In a subsequent study of natives of Papua New Guinea (Hazout & Lucotte 1986) a single haplotype was observed which was different to the 16 observed in Caucasian and Asian individuals of the initial study.

1.8.5 The Y chromosome is distinctly non-polymorphic

Two rigorous studies have been undertaken to search for further Y chromosome polymorphism.

Jakubicka *et al.* have examined 16-34 unrelated German individuals for polymorphism in 12 cloned single-and low-copy DNA sequences with 5 different restriction enzymes. No polymorphism was observed except in the low copy number repeat probe p21A1 (DYZ8). With the enzyme *TaqI* a common band pattern was observed with fragments of 11,7 and 4kb. Occasionally the 11kb band was lost. This can be attributed to a short alternating repeat of 7 and 4kb which is interrupted by the loss of one or more *TaqI* sites. Restoration of the site results in loss of the 11kb fragment.

Malaspina *et al.* (1990) have screened genomic DNA from 15 to 131 unrelated males of English or Italian origin digested with 12 different enzymes, using a range of 12 Y-specific probes. Two individuals with variants were found, one of which was shown to be inherited, but in more than 100 further DNAs examined for the particular probe/enzyme combinations, no other individuals carrying the variants were found. These two studies emphasise the very low level of Y-specific polymorphism. This could be a consequence of the lack of recombination over these regions of the chromosome since the highly polymorphic pseudoautosomal region undergoes an elevated rate of recombination.

INTRODUCTION: PART II

Molecular strategies for genome mapping

It has been necessary to employ molecular mapping strategies in the search for genes on the Y chromosome due to its unsuitability for linkage analysis.

The haploid nature of the Y chromosome in a diploid genome makes it well suited, however, for deletion mapping, and many workers have localised probes to chromosome regions using panels of patients with aberrant Y chromosomes (Vergnaud *et al.* 1986, Oosthuizen *et al.* 1990, Affara *et al.* 1987, Nakahori *et al.* 1991b). In combination with *in situ* hybridisation, deletion intervals may be positioned in broad regions along the chromosome.

With the introduction of a method for separating DNA fragments of several hundred to several thousand kilobases by pulsed field gel electrophoresis (PFGE, Schwarz & Cantor 1984) it became possible to deduce the location of sites for rare-cutting restriction enzymes and to produce physical maps of large regions of the chromosome in relation to these sites. On the Y chromosome this has been carried out particularly around the pseudoautosomal region (Brown 1988, Petit *et al.* 1988) and also in the short arm (Muller & Lalande 1990).

These methods represent a 'top down' approach to physical mapping where the long range continuity of the chromosome is maintained. However, in recent years an alternative 'bottom up' strategy has become popular particularly in organisms with small genome sizes. Using this approach, small, cloned random segments are compared to one another to detect overlaps. If a sufficiently large number of

random segments are analysed then it is possible to reconstruct the original genome in terms of overlapping clones (contigs). These are then extremely useful in the search for gene sequences and as a starting point for the sequencing of whole genomes.

The capacity of cosmid and yeast artificial chromosome (YAC) vectors to support a large insert size (40Kb and 500-900Kb respectively) makes such projects feasible (for a review of cosmid vectors see Little (1987), and of YAC vectors see Schlessinger, 1990). Cosmid clones may be introduced into a bacterial host packaged in a bacteriophage head particle. This is dependent on the presence of lambda *cos* sites inserted into a plasmid vector and has the advantage of some size selection during the *in vitro* packaging reaction. Much larger fragments may be cloned into YAC vectors which provide the centromeric and telomeric functions of a yeast chromosome along with selectable markers, thus fragments of up to one megabase may be maintained as an artificial chromosome in a yeast host strain.

A number of such projects have been established ranging from physical maps of plasmids to human chromosomes. I will describe a number of these in some detail as each has particular strengths and all are relevant to the choice of a strategy when embarking on a project such as the physical map of the human Y chromosome. Most recent results have been described at the Cold Spring Harbor Genome mapping and sequencing meeting 1991 and are to date published as abstracts of this meeting. It is to this meeting that several references apply.

1.9 Random approach mapping projects

The first projects of this type were described by Coulson *et al.* (1986) and Olson *et al.* (1986). Two different strategies were employed to examine the genomes of the nematode *Caenorhabditis elegans* and the yeast *Saccharomyces cerevisiae*.

1.9.1 *C. elegans*

With a genome size of around 100 megabases (Mb) arranged in six linkage groups, *C. elegans* is a relatively large but well studied organism with 500 known loci and many cloned segments localised to chromosome regions by *in situ* hybridisation.

In order to detect overlaps between DNA segments cloned in cosmid vectors, a unique 'fingerprint' was obtained for each clone. Initially DNA was prepared by a general small scale miniprep method from 48 clones. This was digested with the restriction enzyme *Hin*dIII in the presence of reverse transcriptase, dideoxy GTP and radiolabelled dATP, resulting in a series of fragments labelled at either end, these fragments were then cleaved with *Sau*3A1 after inactivation of the reverse transcriptase. The digestion products, a proportion of which are labelled at one end, are separated by electrophoresis on a denaturing polyacrylamide gel alongside a marker lane consisting of end-filled *Sau*3A1 digested lambda DNA. Gels were dried onto the glass plate and exposed to X-ray film. By means of a digitising unit, information about the position of bands within each clone in relation to the standard lanes is analysed by a series of computer programs (Sulston *et al.* 1988,1989). These programs compare the number of bands in common between clones and organise those with substantial overlaps (usually 50% of the clone length) into contigs.

The restriction method produces a series of fragments small enough to be resolved accurately by polyacrylamide gel electrophoresis (58-225bp). The number of fragments is important, there should be sufficient bands per clone for overlaps to be detected but not so many that the probability of bands occurring at the same position merely by chance is high. Also, since the second enzyme (*Sau*3A1) was used in the construction of the library, spurious fragments resulting from the junction of the cloned segment with the vector are not produced.

Using computer simulations to assess the progress of a project of this kind shows that the rate of contig assembly, although initially occurring at an increased rate, is slower overall than would be predicted given a completely random clone bank. This is due to the preferential cloning of some fragments and the possible 'non-clonability' of others as encountered in chromosome walking experiments.

By 1986 1,700 clones had been ordered into 860 contigs with sizes ranging from 35-350Kb, (60% of the genome), this was increased by 1987 to 17,500 cosmids assembled into 700 contigs.

At this stage it was necessary to begin a more directed approach to link known contigs together and to fill in gaps. This was achieved using yeast artificial chromosomes (Coulson *et al.* 1988). It was not feasible to apply the cosmid fingerprinting method to the YAC clones due to the size of YAC inserts and background fragments produced by small YACs. However, clones at the end of contigs and unattached clones were used to screen filters of arrayed YAC clones. These were in turn used as probes against filters containing cosmid clones, thus establishing the exact position of the YAC with respect to the

original cosmid probe and other cosmids. The two types of vector have no appreciable homology and excess hybridisation due to repeated sequences was low, making this approach relatively straightforward. This reduced the number of contigs from 700 to 346. Overlaps revealed by hybridisation were confirmed by restriction analysis in order to detect false overlaps due to repetitive elements or as a result of co-ligation of non-contiguous fragments at the stage of YAC cloning. This approach reached its practical limits when the map had been reduced to 170 contigs containing over 95Mb of DNA. Further undetected overlaps between YACs and cosmids at the end of contigs have been detected by direct sequence analysis using flanking vector primers and using the Polymerase Chain Reaction to obtain unique hybridisation probes. The map is now reduced to 90 contigs covering most coding sequences (CSH Genome mapping and sequencing 1991).

1.9.2 *Rhizobium* plasmid NGR234a

This fingerprinting strategy has also been applied to a plasmid of the *Rhizobium* species, NGR234a (Perret *et al.* 1991). The plasmid carries most of the symbiotic genes of this bacterium which associates with leguminous plants and catalyses the reduction of atmospheric nitrogen into ammonia. Following hybridisation of a total *Rhizobium* genomic library with labelled plasmid DNA, 227 positive cosmids were grouped into 14 initial contigs estimated to cover 900kb. This was surprising as the size of pNGR234a is estimated at 500kb. The possibility that dispersed chromosomal sequences homologous to pNGR234a exist was confirmed when 800 additional clones which did not hybridise to the plasmid were fingerprinted. Comparison of these fingerprints extended 8 of the original 14 contigs, which could then be assigned to the chromosome rather than the plasmid map.

Although some further overlaps were detected by increasing the amount of information available for some potentially overlapping clones by carrying out additional fingerprints using the enzymes *Bam*HI and *Eco*RI as well as *Hin*dIII, further hybridisation experiments were required to complete the map.

1.9.3 *Saccharomyces cerevisiae*

The random clone strategy described by Olson *et al.* to map the much smaller 15mb genome of *S.cerevisiae* involved fingerprinting fragments of around 15 kb cloned into a lambda vector.

In this case, DNA was digested by the restriction enzymes *Hin*dIII and *Eco*RI, and resolved by conventional agarose gel electrophoresis alongside marker lanes. An enlarged image of the ethidium bromide stained gel was projected onto the surface of a digitising tablet and fragment locations were recorded. Using this method it was possible to produce a restriction map of each clone with respect to a non-discriminated enzyme site was either *Hin*dIII or *Eco*RI, (termed RH). Accurate size resolution was maintained and vector-insert junction fragments were excluded by analysing only fragments of 400bp-7.5Kb. Computer analysis identified overlaps between clones if more than 5 of an average 8.36 bands were shared between two clones.

In the initial project, 4,946 clones were analysed, which represents 4.5 genome equivalents. 85% of these clones were positioned into 680 contigs each containing an average of 6.2 members. Most of the unpositioned clones contained too few RH fragments to meet the matching criterion.

Using a series of probes from 14 independently mapped regions to screen the random-clone data base showed that 96% of the DNA

corresponding to these regions was present, with DNA at the locus *SUP4* missing. This suggested that the cloned DNA was random and complete, the actual depth of the physical map produced was nearer 7-fold coverage than 4.5 suggesting that the yeast genome (16 chromosomes) is smaller than previously described. Physical mapping data has been linked to the genetic map by hybridisation to DNA segments from known loci. Over 150 genetic markers have been located on the physical map.

Extending the 'bottom up' approach, Link and Olson (1991) have used the subset of lambda clones included in contigs to produce a physical map in terms of *Not*I and *Sfi*I sites at a resolution of 110kb. The lambda clones were screened for the presence of *Not*I and *Sfi*I sites and then used as linking probes to demonstrate the adjacency of large fragments from individual yeast chromosomes separated by PFGE. This low resolution map is an important element in the establishment of long range continuity of the high resolution map.

Furthermore, as reported by Sgouros *et al.* (CSH genome mapping and sequencing 1991) a complete nucleotide sequence of the *S. cerevisiae* chromosome III has been achieved using contiguous lambda clones as a starting point.

1.9.4 *Eschericia coli*

A random fingerprinting aproach based on phage clones was also used to construct a physical map of the genome of *E.coli* strain W3110 (Kohara *et al.* 1987). With a genome size of approximately 4.7mb arranged as a single circular molecule, the organism was well characterised genetically. In this case detailed restriction maps were produced in order to improve the reliability of overlap recognition. An eight enzyme restriction map was produced for each clone by a partial digest procedure, followed by transfer to filters and hybridisation to a probe derived from one end of the vector. Although the labour involved is great for each clone, the resulting restriction map is an advantage. 1056 clones were analysed by this method and sorted into 70 groups with overlaps based on clones having 5 or more restriction sites in common. The data is subject to flaws if a) faint bands on the autoradiograms are missed, b) two closely located sites for a single enzyme are interpreted as a single site or c) the order of closely located sites for different enzymes is inverted.

Gaps were closed using probes derived from clones at the ends of contigs to screen a densly packed array of clones including those already assigned to contigs. The number of linked groups was reduced to 7, these could be correlated with the genetic map making use of published restriction data. The problem of random hybridisation based upon repetitive elements was overcome by fine restriction mapping of 'suspicious' clones.

A parallel study of the genome of *E. coli* K-12803 has been undertaken by Knott *et al.* (1988,1989) using cosmid clones. Random clones were digested with *Hin*fI, labelled, resolved on denaturing polyacrylamide

gels and overlaps detected by computer analysis of the band positions in each clone. 2,000 cosmid clones were grouped into 58 contigs ranging in size from 40-300kb, representing approximately 90% of the genome.

Lambda clones flanking the gaps in the map produced by Kohara *et al.* were fingerprinted using *Hinf* I and compared to the cosmid database. This additional analysis identified clones covering five of the gaps. Two further gaps were closed by hybridisation of fragments from flanking lambda clones to a library prepared in the low copy number plasmid pOU61cos. The gaps were closed, in the majority, by a single clone rather than the 18 which would be expected given the redundancy of coverage of the library compared to the genome size. This confirms that several regions are not amenable to cloning in either lambda or cosmid vectors. The low copy vector pOU61cos was essential for the cloning of 3 of the 8 gaps, and 2 more were overlapped by inserts in the Lorist B vector which has a controlled copy number (Cross and Little 1986), suggesting that cloning at a low copy number can overcome the adverse effects of gene dosage.

A large amount of sequence data has been accumulated for this organism which is being located on the physical and genetic maps (Rudd *et al.* CSH genome mapping and sequencing 1991) as a basis for the entire sequence reconstruction.

1.9.5 *Drosophila*

Three major efforts to construct a physical map of the *Drosophila* genome are currently in progress, these different approaches may serve as useful models to predict the best approach for constructing physical maps of larger genomes including human chromosomes.

Fingerprinting preselected cosmid clones

A comprehensive genetic map of around 3,800 genes have been localised to the four chromosomes, this has been linked to a cytogenetic map relating to the band patterns of the giant polytene chromosomes in the larval salivary glands of *D.melanogaster* (refs in Siden-Kiamos *et al.* 1990).

Siden-Kiamos *et al.* (1990) have constructed a total genomic cosmid library of nearly 20,000 clones in the vector Lorist6, representing a four fold coverage of the genome. The cosmids are fingerprinted using a radiolabelled *Hin*fI digestion followed by electrophoresis, digitisation and analysis with the Sulston set of programs. However, the clones are first preselected for their location within a particular chromosomal division. This is achieved by microdissection of a particular division which is then either cloned in a lambda vector and labelled by Klenow extension of the PCR-amplified insert, or is directly amplified by PCR and labelled by a random-priming reaction. The resulting probe is used to screen the cosmid library to select clones arising from that division. As each division is approximately 1% of the genome, the number of clones fingerprinted is 120-250. This approach reduces the scale of the task immensely and increases the proportion of clones assigned to contigs compared to the analysis of the whole genome using a completely random approach.

Representative clones from each contig may then be mapped to polytene chromosomes by *in situ* hybridisation to correlate them with the cytogenetic map, and terminal clones may be hybridised to the remainder of the library to detect additional overlaps. In order to overcome the problem of repetitive sequences, the division-specific probes are prepared from a sibling species *D.simulans*, which is highly homologous in sequence to *D.melanogaster* but has very few repetitive sequences in common. 470 clones have been characterised by *in situ* hybridisation, 350 of which are derived from the distal three divisions of the X chromosome (where the mapping was begun). This region is mapped in 30 contigs consisting of 220 cosmids, spanning 71% of the estimated length of the three divisions (38mb). It is expected that the remaining single clones can be attached to the map either by isolating further cosmid clones or by turning to YAC clones.

In situ hybridisation of YAC clones

A YAC cloning approach has been utilised by Garza *et al.* (1989) as a method of substantially reducing the number of clones to be analysed. A YAC containing 220kb of Drosophila DNA would span approximately ten average chromosome bands, and in theory only 1,500 clones would be required to represent three genome equivalents.

A library of 454 clones with an average insert size of 170kb was constructed and checked for randomness by hybridisation with single copy probes and with the repeat element *copia*. Restriction digests of clones containing the repeat element were compared to ensure that each clone derived from a different location in the genome. A certain amount of instability was observed in clones containing tandem repeat sequences such as ribosomal DNA but this did not seem to be a

particularly frequent problem. In an initial experiment, 58 clones were localised to polytene chromosomes by *in situ* hybridisation. Several overlaps could be detected cytologically and this has been extended to over 1000 clones (Hartl *et al.* CSH genome mapping and sequencing 1991). Confirmation of overlaps and gap-filling is to be achieved using probes specific to the ends of the cloned fragments, and also probes derived from PCR amplified microdissected fragments as described above.

The use of YAC clones may resolve some of the problems encountered in cloning sequences in lambda and cosmid vectors. Garza *et al.* have demonstrated a region of heterochromatin of the X chromosome which posed difficulties in a lambda walk but was stably cloned in a single YAC.

Oligonucleotide hybridisation

An alternative mapping strategy based upon hybridisation has been devised (Craig *et al.* 1990). This strategy has been tested on the genome of Herpes Simplex Virus type 1 (HSV-1) and has been applied to the *Drosophila* genome also. 384 cosmids from the HSV-1 genome (100 fold coverage) were spotted onto filters in a dense, ordered array and probed with 22 oligonucleotides corresponding to regions of the HSV-1 sequence. Binary information is produced for each clone based on its hybridisation pattern to the series of probes and analysis of this allows reconstruction of the genome. In the case of the Herpes Simplex Virus it was possible to identify four different isomeric forms. This method attempts to combine the high analysis rates of other hybridisation methods with the insensitivity to repeat elements which is a feature of the fingerprinting methods.

For analysis of the *Drosophila* genome, a large cosmid library (16-fold coverage) and the YAC library described above have been arrayed on filters at high density and are being hybridised in a variety of ways. These include pool hybridisations, walking protocols and oligomer fingerprinting using previously cloned single copy probes and a large number of random oligomers (Hoheisel and Lehrach CSH genome mapping and sequencing 1991). Overlaps between clones can be detected by duplication in the particular hybridisation features of each clone. As the sequence of oligonucleotide probes is known, it is possible to derive sequence information about the clones concurrently with the production of a physical map.

1.10 Mapping human chromosomes

1.10.1 Human X chromosome

Wada *et al.* (1990) are using YAC clones to map the distal region of the X chromosome long arm Xq24-qter. 127 YAC clones have been derived from the human portion of an human/hamster somatic cell hybrid containing this region. Of these, 11 strains contained more than one YAC (usually one human clone and one or more derived from hamster DNA). Of 52 characterised DNA sequences, 12 were shown by hybridisation to be present in the YACs. Four of these, which are anonymous polymorphic loci, appear to be clustered in a single YAC of 150kb which may represent a highly polymorphic region of the chromosome.

These YACs have been assigned to cytogenetic bands as have others by *in situ* hybridisation. In conjunction with the cytogenetics, clones are also analysed by a fingerprinting strategy based on a sub-population of restriction fragments in each YAC. Clones are digested

with *Eco*R1 and *Hin*dIII or with *Taq*I, fragments are separated by agarose gel electrophoresis, blotted onto filters and probed separately with the repetitive elements *Alu* and L1. Hybridising fragments are compared between clones to detect overlaps. To date, two YAC contigs of 3-4mb have been constructed which span the cytogenetic band Xq26. These encompass two genes (hypoxanthine phosphoribosyl transferase and coagulation factor IX) and several anonymous loci (Pilia *et al.* CSH genome mapping and sequencing, 1991). In a similar way, probes from the Xcen-Xp21 region have been used to screen a YAC library constructed from a human cell line (48,XXXX). YAC contigs have been identified which span several genes, notably the complete Duchenne Muscular Dystrophy region which is represented in a 36 clone contig. The minimum set of 6 overlapping YACs covering this gene span 3mb of DNA (Monaco *et al.* CSH,1991).

The merits of a fingerprinting technique based on repetitive sequences have been examined by Bellanne-Chantelot *et al.* (1991). In this case YAC clones are sub-cloned into cosmids which are then digested with a 6bp recognition site restriction enzyme. Fragments are separated, transferred to filters, and band patterns are compared between clones after hybridisation with total human DNA. This approach circumvents the problems of separating YACs from yeast chromosomal DNA (usually achieved by pulsed field gel electrophoresis), and also of handling DNA fragments greater than 50-100kb (typically in agarose blocks). A cosmid library of 3000 clones has been prepared from a yeast strain containing a 420kb YAC. 86 cosmids containing human DNA were selected by hybridisation to [α-

^{32}]P-labelled total human DNA. Cosmid DNA was digested with four different enzymes, separated by agarose gel electrophoresis against size markers, blotted and probed with radiolabelled total human DNA. Data from the autoradiograph was analysed by computer to calculate the molecular weights of each band and predict overlaps. Two large contigs were constructed which were actually seen to overlap when examined more closely. Considered as a whole, 13 minimal subset cosmids had an identical restriction map to the original YAC clone, confirming that this is a useful fingerprinting approach.

1.10.2 Human chromosome 16

Repetitive sequence fingerprinting has also been applied to human chromosome 16 (Stallings *et al.* 1990).

A cosmid library was prepared from chromosomes 16 isolated from a human/mouse somatic cell hybrid. Cosmid DNA was isolated in 96 well microtitre dishes and enzyme digests performed by robot. In this case, each clone was digested with *Eco*RI, *Hin*dIII and both enzymes together, and fragment size data from ethidium-bromide stained gels was stored in a database. These sizes were combined with data obtained when Southern blots of the gels were hybridised to consensus oligonucleotides of different repeat elements. Overlaps were detected based on the pattern of hybridisation of different fragment sizes to, say, (GT)n sequences. Even if the two clones have only a small region in common, the combination of the restriction patterns and hybridisation patterns allow overlaps to be recognised. The degree of similarity required to be confident of an overlap compared to other fingerprinting techniques is reduced from 50% to 10 or 20%.

Clones were also preselected for the presence of repeat elements before fingerprinting, resulting in a high rate and increased length

of initial contig generation. 2692 GT positive clones and 57 clones positive for the L1 repeat were fingerprinted in a total of 3145 clones. 2823 pairs of overlapping clones resulted which could be assembled into 460 contigs, the mean overlap being 43% and the minimum detected 10%. In theory, the same number of large contigs (150-200kb) could only be achieved by a totally random strategy if four times the number of clones were fingerprinted. After fingerprinting 4,000 clones, 553 contigs were constructed representing approximately 60% (54mb) of the euchromatic arms of chromosome 16. Gaps, estimated to be smaller than 65kb on average, are to be closed by screening YAC libraries with contig end-probes.

1.10.3 Chromosome 19

The physical map of chromosome 19 is being achieved by yet another method which uses fluorescence as a detection system for different size fragments resolved by electrophoresis (Carrano *et al.* 1989).

An 18bp M13 universal primer covalently linked to one of four different coloured fluorophores is annealed to an oligonucleotide designed to yield a 5' overhang complementary to a restriction enzyme site. This double stranded molecule is then ligated to products of cosmid DNA digested with the appropriate enzyme. Aliquots from four different digests labelled with different fluorochromes may be pooled, precipitated, denatured and loaded onto a denaturing polyacylamide gel. Fragments are electrophoresed through the gel into a DNA sequencer which records in a database the exact point at which each fragment passes a laser, in relation to a size marker run in the same lane. Separation between fragments is optimised as each

traverses the whole length of the gel before its size is recorded. The method has been developed using SV40 DNA and has been extended to cosmids derived from chromosome 19 (see table 1.1).

1.10.4 Chromosome 11

Short arm

A map of the short arm of chromosome 11 is being constructed by workers in the laboratory of P.Little. A pilot project (Harrison-Lavoie *et al.* 1989) involved fingerprinting cosmids by the *Hin*dIII/*Sau*3AI method of Coulson *et al.* 112 cosmids were derived from the human component of a mouse/human hybrid cell line made by chromosome mediated gene transfer (CMGT), containing 3mb of 11p. These were assembled into two major contigs covering most of the region, confirming that the method would be applicable to a larger project to produce a map of the complete short arm (20mb). Much of the contig closure was achieved by probing dot blots of the library with RNA end probes synthesised from the Sp6 or T7 RNA polymerase promoters present in the vector (Lorist X) at either side of the cloning site. Work is in progress to complete the map of the short arm using cosmid clones derived from a somatic cell hybrid enriched for 11p.

Cosmid multiplex analysis:long arm

A further hybridisation-based protocol has been devised by Evans and Lewis (1989). This strategy takes advantage of the small fraction of overlap between clones which can be detected by hybridisation methods. A genomic cosmid library is constructed and organised on filters as an ordered matrix. DNA is prepared from pools of cosmids corresponding to a row or a column of this matrix and RNA probes are

synthesised from RNA polymerase promoters in the vector. When hybridised to the filter array, the corresponding row or column should be labelled along with a few random clones. If a particular one of these random clones is also labelled by a subsequent hybridisation, then it must contain sequences in common with the clone in common between the two probe sets. The use of short probes from the the very ends of cloned DNA is expected to limit the inclusion of repetitive elements in the RNA probe and hence avoid spurious overlaps. However, probes are usually prehybridised with human DNA to ensure removal of repeat sequences.

In the application to a library prepared from a human/mouse somatic cell hybrid containing DNA from the distal long arm of chromosome 11 (11q13-11qter), 960 clones were analysed in 68 hybridisation reactions. 1099 pairs of linked clones were detected and overlap confirmed by restriction analysis and RNA probe hybridisation, these could be combined into 315 contigs. Contigs have been localised cytogenetically and are to be linked using YAC clones.

1.11 Sequence Tagged Sites

A main objective of all of these projects is to provide workers with an ordered set of clones which can be aligned with the genetic map, and analysed in more detail in terms of gene sequences or physical organisation of the DNA. The availability of such clones eliminates time consuming cloning and the redundancy of effort which is inevitable if the work is carried out in isolation by each group attempting to clone a particular sequence.

In order to provide a common identifier of a region of DNA regardless of the cloning vehicle, it is proposed that regions should be marked by a sequence tagged site (STS, Olson *et al.* 1989). Such a stretch of sequence (200-300bp) located at 2 centimorgan intervals along the genome (and related to the genetic map) will allow the identification of any region of DNA, and its retrieval using the Polymerase Chain Reaction. The production of this type of sequence data from contigs is therefore an intrinsic part of all the mapping projects described above.

Table 1.1 (Overleaf) A summary of the mapping strategies described in this section.

Table 1.1

GENOME	SIZE	STRATEGY	CLONES ANALYSED	CONTIGS	SIZE of CONTIGS	COVERAGE/ CLOSURE
C. elegans	100Mb	HindII/Sau3A F.P. Cosmids YAC hybridisation	1,700 17,500	860 700 346--175--90	35-350Kb	95%
Rhizobium plasmid pNGR234a	500Kb	HindIII/Sau3A FP. Cosmids	227	14 --6		Closed
S.cerevisiae	15Mb	Restriction maps of lambda clones, NotI/SfiI map	4,946	680-101	Average 6.2 clones At 101 contigs 122Kb	
E.coli	4.7Mb	8 enzyme restriction map of lambda clones. HinfI fingerprinting Cosmids	1,056 2,400	70 7 58	20-180Kb 40-300Kb	94% 90%
Drosophila	distal div of X Chr. 38Mb	HinfI F.P. of cosmids in division. YAC clones	20,000 454	30		71%
Human Chromosome 11	3Mb 11p 20Mb 11q 60Mb	HindII/Sau3A F.P. Cosmids Cosmid Multiplex	112 4,000 960	6--2 500 315	125-140Kb 80-300Kb	60%
Chromosome 16	Euchromatin 90Mb	Repetitive fingerprinting of cosmid clones after pre-selection for repeated seqs. 0 = 0.1-0.2	3,145 4,000	460 553	150-200Kb	54% 60%
Chromosome 19	60-70Mb	Fluorescence fingerprinting of cosmid clones	4,000 8,813	350 720	128Kb 132Kb	70%

1.12 Aims of this project

The small size of the Y chromosome, and the abundance of anonymous loci positioned by deletion mapping make it a good subject for the production of a physical map by random cloning methods.

Since the beginning of this project, a relatively large number of Y-linked gene sequences have been identified (*ZFY, SRY, CSF-2R, RPS4Y*) but other sequences are yet to be found, for example, the azoospermia factor, the gene encoding the H-Y antigen and a gene predisposing to gonadoblastoma *GBY* .

The aim of this project is to prepare a Y chromosome-specific cosmid library and, using a random fingerprinting approach, to construct a complete physical map of the chromosome in terms of overlapping cosmid clones. This will be of use both in the search for genes, the study of Y chromosome structure and evolution and ultimately, in the complete sequencing of the chromosome.

CHAPTER TWO

MATERIALS AND METHODS

2.1 Standard media and buffers.

Unless stated otherwise solutions were sterilised by autoclaving at 15psi, 121°C for 20 mins.

LB Medium

per 100ml:

Bacto-tryptone	1g
Bacto-yeast extract	0.5g
NaCl	1g
Glucose	0.1g

LB Agar

As above with 1.5g Bacto-agar

Cosmid broth

As LB medium with 1g maltose in place of glucose.

2xTY

per 100ml:

1.6g	Tryptone
1g	Yeast extract
0.5g	NaCl

Antibiotics

Kanamycin sulphate was prepared at a concentration of 25mg/ml in dH_2O and sterilised by filtration (0.22μm pore size). The working concentration was 25μg/ml. Ampicillin was prepared at 100mg/ml as above. The working concentration was 100μg/ml.

Phage storage medium

per 100ml:

NaCl	0.58g
$MgSO_4.7\ H_2O$	0.2g
1M TrisHCl pH 7.5	0.5ml
Gelatin (2% w/v in H_2O)	50μl

Phosphate buffered saline (PBS)

per 1 litre of 20X stock soln:

Na_2HPO_4	23g
KCl	4g
NaCl	160g
KH_2PO_4	4g
Na butyrate	11g

Adjust to 800ml with H_2O and autoclave, then add 200ml 0.5M EDTA.

5xTBE 0.5M Boric acid, 0.5M Tris, 10mM EDTA

1xTE 10mM tris, 1mM EDTA pH 8.0

1xTNE 100mM NaCl, 10mM Tris, 1mM EDTA pH 8.0

20xSSC 3M NaCl, 0.3M Trisodium citrate

20xSSPE 3.6M NaCl, 0.2M $Na_2HPO_4.7H_2O$, 0.02M EDTA

Gel loading buffer 0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol.

Church hybridisation solution 0.5M NaPi pH 7.2, 7% SDS, 1mM EDTA.

1M Sodium phosphate solution pH 7.2 Prepared by the titration of 1Molar solutions of $Na_2HPO_4.12H_2O$ and $NaH_2.PO_42H_2O$.

Solutions for manipulations involving RNA were prepared with solid chemicals as far as possible and dissolved in ddH_2O that had been incubated with 1% Diethylpyrocarbonate (DEPC) at 37°C overnight.

2.1.2 Suppliers of materials

Restriction Enzymes

*Sau*3AI	10u/μl	Bethesda Research Laboratories (BRL)
	40u/μl	Northumbrian Biologicals Ltd. (NBL)
*Bam*HI		Anglian
*Hind*III		NBL
*Hinf*I		NBL/New England biolabs

Other enzymes either NBL or Anglian.

DNA modifying enzymes

T4 DNA ligase	Anglian
Klenow DNA Polymerase	New England Biolabs
AMV Reverse transcriptase	NBL

Kits

Random prime labelling kit	Amersham International
T7 RNA polymerase kit	Amersham International

Chemicals,reagents and membrane

Hybond N	Amersham
Zetaprobe	Bio Rad
[α-^{32}P]dCTP, [α-^{32}P]UTP	Amersham
X-Ray film	Kodak X-omat AR
	Amersham Hyperfilm MP
	Amersham Hyperfilm β max
Agarose	Sigma
Bacterial media	Difco
Tissue Culture	Flow
General Chemicals and reagents	BDH

2.2 General Techniques

2.2.1 Plasmid and Cosmid isolation

Small scale (miniprep) Maniatis *et al.* (1982)

A 2ml bacterial culture was grown overnight in L-broth with the appropriate antibiotic selection. 1ml of this was pelleted in an eppendorf tube by microcentrifugation for 2 minutes. A further 1ml was then pelleted in the same tube. The pellet was resuspended in 100μl of miniprep buffer (50mM glucose, 10mM EDTA, 25mM Tris.HCl pH 8.0). Following incubation for 5 minutes at room temperature, 200μl of solution II (0.2M NaOH, 1% SDS) were added and the tube was inverted to mix the contents, after 5 minutes on ice, lysis could be seen to have occurred. The bacterial chromosomal DNA was precipitated by the addition of 150μl of solution III (per 100ml: 60ml 5M K-acetate, 11.5ml glacial acetic acid. Final pH4.8) and incubation in ice for 10 minutes. The debris was removed by centrifugation for 10 minutes in a microfuge and the supernatant was transferred to a clean tube. The cloned DNA was precipitated by the addition of 50μl 5M NaCl and 900μl absolute ethanol. After 30 mins at 20°C, and 10 minutes centrifugation, the DNA pellet was rinsed in 70% ethanol, air dried and resuspended in 10-20μl TE.

large scale (maxiprep)

Up to 1mg of DNA was prepared by a scaled up version of the miniprep method. Cells from a 400ml overnight culture were harvested by centrifugation in two aliquots (6000rpm, 10 mins, 4°C), and each pellet was resuspended in 10ml miniprep buffer with the addition of lysosyme (2mg/ml). After 5 mins at room temperature 20ml of solution II was added and gently mixed and then, after a further 5-10 mins on

ice, 15ml of solution III were added, thoroughly mixed and allowed to stand on ice for 10 mins. The mixture was centrifuged (8,000rpm, 10mins, 4°C) and DNA was precipitated from the supernatant with 0.6 volumes of isopropanol for 10 mins at room temperature. After centrifugation at 8,000rpm for 10 mins, the pelleted nucleic acids were resuspended in 10.7ml T.E. To this were added 0.4ml 1M Tris base, 0.4ml 0.5M EDTA, 12.7g Caesium Chloride and 1.2ml Ethidium Bromide (10 mg/ml). This solution was transferred into polyallomer ultra-centrifuge tubes and spun for 16 hours in a vertical rotor at 45,000rpm. The region of the density gradient containing the cosmid or plasmid DNA was visualised under long wave UV light and was collected using a syringe. Ethidium bromide was removed by extraction with caesium chloride saturated isopropanol and the CsCl was removed by dialysis against TE before ethanol precipitation of the DNA.

2.2.2 Preparation of DNA from Genomic and Hybrid cell lines.

The cell lines GM1416B and OXEN were cultured in flasks in RPMI 1640 medium supplemented with 10% v/v Foetal calf serum (FCS), 3μg/ml glutamine, 0.2% w/v $NaHCO_3$ and 100 units of Penicillin and Streptomycin. The medium was brought to a slightly alkaline pH by adding a soluton of $NaOH/NaHCO_3$. Cells were grown in suspension at 37°C with 5-8% CO_2, and were diluted as became necessary by pelleting followed by resuspension in a larger volume of medium.

Other cell lines (3E7, 7/2, etc.) were grown as attached monolayers in Dulbecco's modified Eagle's medium (DMEM), with supplements as above but without CO_2. If HMT selection was required (Hor19X, cf108) 100μm hypoxanthine, 10μm methotrexate and 10μm thymidine were also added to the culture medium. When the cells became confluent they

were washed in PBS, and released from the flask surface by incubation at 37°C for 2-5 mins with 0.25% trypsin in PBS. Washing in medium containing Foetal calf serum inhibited the trypsin and the cells were then divided or transferred to a larger flask.

Cells were stored in 1ml aliquots at -70°C in freezing vials in 95% v/v FCS, 5% v/v DMSO.

DNA preparation: Cells from four large flasks were washed in ice cold PBS, trypsinised, pelleted and resuspended in TE. To this was added 10 volumes of lysing solution (0.5M EDTA, 100μg/ml Proteinase K, 0.5% Sarcosyl) and incubation was carried out for 3 hours at 37°C. Protein debris was removed by extraction with phenol:$CHCl_3$:isoamylalcohol (25:24:1) three times, and the solution was dialysed against 50mM TrisCl pH8.0, 10mM EDTA, 10mM NaCl. The nucleic acids were precipitated with ethanol, resuspended in TE, and RNA was removed by a 3 hour incubation with 50μg/ml RNAse followed by a further phenol/$CHCl_3$ extraction. Before ethanol precipitation, contaminating phenol was removed by extraction with $CHCl_3$. The spooled DNA was rinsed in 70% ethanol, air dried, and resuspended in TE.

2.2.3 Extraction of genomic DNA from human placenta.

1cm cubes of tissue frozen in liquid nitrogen were crushed to a fine powder using a mortar and pestle. To this was added 5ml of lysing buffer (10mg Proteinase K, 2ml 10% SDS in 50ml TE[10mM Tris pH7.5, 10mM EDTA]) and the solution was incubated for 1 hour on a daisy wheel, and then at 37°C overnight. The solution in each tube was diluted with 5mls of TE and was extracted with Phenol/$ChCl_3$ at least three times. RNase (10μl of 10mg/ml stock) was added, and, after 30 mins at 37°C, Proteinase K was added (1mg per tube) and incubation

was carried out for a further 30 mins at 37°C. The Phenol/$ChCl_3$ extraction was repeated, followed by a $CnCl_3$ extraction and the DNA was precipitated with ethanol at -20°C overnight. The high molecular weight DNA was spooled out, rinsed in 70% ethanol and dried before resuspension in TE.

2.2.4 **Nucleic acid concentration:** This was determined by spectrophotometry at 260nm (Maniatis, Fritsch & Sambrook 1982).

2.3 DNA Analysis

2.3.1 Restriction enzyme digestion

Restriction enzyme digestion of cosmid or plasmid DNA was typically carried out in a reaction volume of 15μl, with 0.5-1μg of DNA. Reaction buffer supplied by the enzyme manufacturer or prepared according to Maniatis *et al.* (1982) was included along with 750mM Spermidine and 4 units of enzyme per μg of DNA. Incubation was at 37°C (unless specified otherwise, for example, *Bss*H II = 50°C, *Taq* I = 65°C) for at least 2 hours. For the digestion of genomic and hybrid DNAs, a reaction similar to that described above was set up but the reaction volume was increased to 50-100μl and incubation was continued overnight.

2.3.2 Agarose gel electrophoresis.

DNA fragments were separated on agarose gels prepared in 1xTBE with 1xTBE as the running buffer. Ethidium bromide was incorporated into the gel at 0.5μg/ml. Samples were loaded with the addition of 0.1 vol of 10x loading buffer and electrophoresis was carried out depending upon the sample. Visualisation was by UV transillumination.

2.3.3 Isolation of fragments for radioactive labelling.

The appropriate DNA digest was separated on a gel prepared using low-melting point agarose and the required fragment excised and weighed. To this was added sterile distilled water at a ratio of 3ml of H_2O per gram of gel. After boiling, the denatured DNA could be used in random-prime oligolabelling reactions (usually 26μl,50ng DNA per reaction).

2.4 Radioactive labelling of Nucleic acid

2.4.1 Multiprime labelling of DNA probes.

A commercial random prime kit was used. 20-50ng of DNA were boiled in a volume of 28μl with dH_2O to denature the DNA. After chilling on ice, 10μl of buffer 1 (containing dATP, dGTP, dTTP and reaction buffer) were added followed by 5μl of buffer 2 (containing random hexanucleotide primers). 40 Ci radiolabeled nucleotide [α- ^{32}P] dCTP (10 Ci/μl) were then added to the reaction mix followed by 2μl (2 units) DNA polymerase 1 'Klenow' fragment. Incubation was for at least 3 hours at room temperature, after which time the reaction was

terminated by the addition of 50μl of TNE/0.1% SDS. Unincorporated nucleotides were removed using a Sephadex G50 column. Sephadex suspended in TNE/0.1% SDS was compacted into a 1ml syringe by centrifugation (1,500 rpm,2 mins), the sample was layered onto this and the spin repeated. The probe DNA was contained 1n the eluate. Prior to hybridisation the probe was denatured by boiling in the presence of 1mg of sonicated herring sperm DNA, and cooled on ice.

2.4.2 **End labelling of DNA probes.**

In the case of short oligonucleotides such as the repeat oligo $(CA)_9$, the γ-phosphate of ATP was attached to the 5'terminus of the molecule using T4 Polynucleotide Kinase. The reaction was as follows, 10pmol of DNA in 4.5μl TE was incubated with 1.5μl of buffer supplied with tne enzyme (Amersham), 2μl [γ-^{32}P] ATP (10μCi/μl), 2μl enzyme (5u/μl) and 5μl dH_2O. The reaction was carried out at 37°C for 50 mins and then terminated by incubation at 65°C for 5 mins. Unincorporated nucleotides were removed by centrifugation through a G25 sepnadex column which had been equilibriated in 3xSSC.

2.4.3 <u>Preparation of labelled RNA probes</u>

Cosmid DNA previously digested with the restriction enzyme *Rsa* I, was incubated with the following reagents mostly included in a T7 Paired-promoter system supplied by Amersham. To 4μl of 5 x transcription buffer were added at room temperature, 1μl of freshly prepared 0.2M DTT, 1μl (20u) Human placental ribonuclease inhibitor (HPRI), 1.5μl of nucleotide mix ATP:GTP:CTP (1:1:1), 2μg template DNA, 10μl (12.5 M) [α-^{32}P]UTP and water to 20μl. 4 units of T7 RNA polymerase

were added and incubation was at 37°C for 1 hour. The RNA product was precipitated by the addition of 4μl of 5M ammonium acetate and 40μl of ethanol at -20°C overnight and then resuspended in 40μl TE, 10μl Vanadyl ribonucleoside complex (VRC), 50μl 20xSSPE.

Prior to hybridisation, the probe was incubated with 1mg of denatured human DNA at 65°C for 1 hour in order to compete out any repetitive elements. Hybridisation to nylon filters was carried out in Church solution including 10mM VRC.

2.5 DNA transfer to membrane by capillary blotting. (Southern 1975)
Digested cosmid or genomic DNA was transferred from agarose gels to nylon membranes by capillary blotting. DNA was depurinated to reduce fragment length by immersion of the gel in 0.25M HCl for 15 mins, this solution was replaced by denaturation solution (0.5M NaOH, 1.5M NaCl) for 30 mins and then neutralising solution (1.5M NaCl, 0.5M Tris-Cl pH 7.5, 1mM EDTA) for a further 30 mins. Transfer was allowed to proceed overnight in 10xSSC after which time the filter was rinsed in 2xSSC, air dried, and baked for 2 hours at 80°C. Filters were hybridised with radiolabelled probe as described.

2.6 DNA:DNA, RNA:DNA hybridisation
Typically, filters were prehybridised for at least one hour at 65°C in a large volume of Church solution. The denatured probe was added in a small volume of the same solution, and hybridisation was permitted to occur at 65°C for around 16 hours. The probe-containing solution was then discarded and the filters rinsed in 2 x SSC/0.1% SDS for 20 mins at 65°C, followed by 0.2 x SSC/0.1% SDS for 20 mins at 65°C. Filters were sealed in plastic and exposed to X-ray film at -70°C.

2.7 Preparation of the Human Y Chromosome cosmid library.

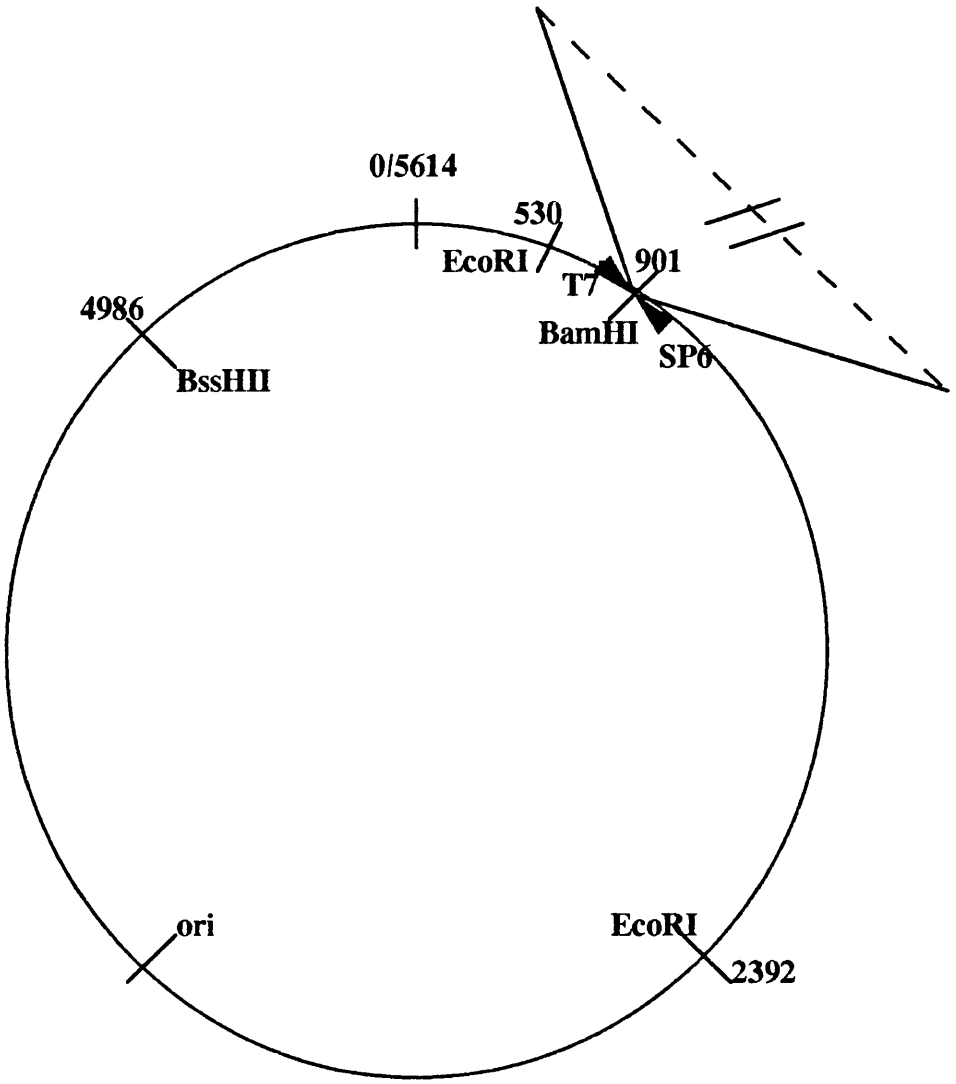
High molecular weight DNA was prepared by Jonathan Wolfe from the somatic cell hybrid 3E7 (Marcus *et al.* 1976), which contains the human Y Chromosome as the only obvious human material on a mouse background. This was partially digested with the restriction enzyme *Sau*3A1, and size fractionated by sodium chloride density gradient. Fragments of 35-45 Kb were ligated at 14°C overnight to the cosmid vector Lorist B (Cross and Little 1986). Vector arms were prepared according to the conditions described by P. Little (1987). A diagram of the vector is shown in figure 2.1. Approximately 1µg of the 3E7 DNA was ligated to an equal mass of vector in a 10µl reaction volume. The experiments described in this paragraph were carried out by Dr. Jonathan Wolfe.

2.7.1 Packaging reaction.

The linear cosmid molecules were packaged into bacteriophage lambda head particles in the following reaction: 10µl of the ligation reaction were mixed with 70µl of buffer A (20mM Tris.Cl pH8.0, 1mM EDTA, 5mM $mgCl_2$, 0.05% v/v ß-mercaptoethanol) and 10µl of buffer Q (6mM Tris.Cl pH7.5, 18mM $MgCl_2$, 60mM spermidine pH7.5, 15mM ATP pH7.6, 0.2% v/v ß-mercaptoethanol)and this was divided into 9µl aliquots.

To each aliquot were added the packaging extracts; 4µl of sonicated extract and 5µl of freeze-thaw lysate, the packaging extracts were prepared according to Manniatis *et al.* by Steven Hall and were stored at -70°C and used directly from liquid nitrogen. After one hour at 30°C the contents of the tubes were pooled and 1.8ml of phage storage medium was added. The resulting phage were then used to infect the recombination deficient strain of *E.coli* ED8767 (Murray *et al.* 1977).

Figure 2.1 Diagram of the cosmid vector Lorist B (Cross & Little 1986).



Sites for the initiation of transcription by RNA polymerases SP6 and T7 are indicated by arrows.
The vector confers resistance to kanamycin.

2.7.2 The recipient bacterial strain.

The recipient bacteria were tested for retention of their *RecA*- phenotype before use by exposure to ultra-violet light. The mutation which renders this strain unable to repair damage caused by UV also reduces the incidence of internal deletion events in cosmids which occur as a result of recombination between direct repeat elements. A clone shown to be sensitive to UV was grown overnight in cosmid broth with 0.2% maltose in order to promote the production of maltose receptor molecules to which the phage particles can adhere. The culture was centrifuged at 4,000rpm for 5 minutes in a bench centrifuge and the pellet was resuspended in one half volume 10mM $MgSO_4$. 2ml of this was added to 2ml of the packaged cosmid mixture and, after incubation at 30°C for 20 minutes, 16ml of LB broth was added. Incubation was continued for a further 45 minutes in order to allow the infected bacteria to express the antibiotic resistance genes, after this time the bacteria were plated onto agar plates containing kanamycin.

2.7.3 Control packaging reaction.

Prior to the experiment described above, control reactions were carried out using 10ng of cosmid DNA that had been prepared by maxipreparation. For the packaging reaction one tenth of the volumes stated were used, and 20μl of this were added to 20μl of bacteria. At this stage different combinations of packaging extracts and also different cultures of *ED8767* could be tested in order to acheive the reaction conditions of highest efficiency. After the addition of LB broth and incubation at 30°C, 160μl of the cells were plated onto LB agar containing kanamycin. The number of colonies apparant after incubation overnight at 37°C could be used to calculate the expected

number of colonies that could be achieved from the ligation reaction. This determined the number of 22x22cm nylon filters required in order to give a calculated density of 10-20,000 colonies per filter.

2.7.4 Preparing library filters.

The appropriate volume of the culture (usually 7ml) was plated onto Hybond N filters prewet in sterile LB broth, under gentle suction. Usually three filters were necessary for the plating of each ligation aliquot. The filters were transferred onto large agar plates containing kanamycin and colonies were allowed to grow at 37°C overnight and then three replica litts were taken from each plate. The orientation of the filter was recorded using a series of pin holes which were transferred from the master filter to each replica. One of these was grown on agar containing 30% glycerol and was then frozen at -70°C. The other replicas and the master filter were allowed to regenerate at 37°C for a few hours. The master filter was stored at 4°C. DNA from colonies on the replica filters was attached to the membrane by denaturation in alkali (1.5M NaCl, 0.5M NaOH), followed by neutralisation (1.5M NaCl, 0.5M Tris pH8.0) and removal of bacterial debris by washing in 2XSSC. Filters were baked for two hours at 80°C.

2.7.5 Screening for cosmids containing human DNA sequences.

In order to ascertain whether each individual clone contained an insert derived from human or mouse DNA, the filters were hybridised to radiolabelled total human DNA prepared from male placenta. Filters were prehybridised in Church solution (0.5M NaPi pH7.2, 7% SDS, 1mM EDTA) for longer then one hour and were hybridised in a

smaller volume of the same solution with the addition of the DNA probe labelled with [α-^{32}P] dCTP, at 10^6 cpm/ml for approximately 16 hours. Filters were washed in Church wash solution (0.04M NaPi,1% SDS) with 2 x 10 minute washes at room temperature followed by 2 x 20 minute washes at 65°C. Autoradiography was for 18 hours at -70°C, using Kodak x-omat film with intensifying screens. Clones giving a positive signal on both duplicate filters were picked from the master filter onto gridded membranes in triplicate. Two of these membranes were prepared for hybridisation as above and one was hybridised with a probe of total human DNA and one with total mouse DNA. Clones hybridising to human and not to mouse DNA were picked into LB broth with kanamycin, grown overnight at 37°C and stored at -70°C with 15% glycerol.

2.7.6 Representation of the library in ordered arrays.

Each clone was then picked by hand onto a single nylon filter (22x22cm) in an array of 40x40 clones. This master filter was used to generate 20 replica filters, which could be used for hybridisation to various DNA probes. After each hybridisation, bound probe was removed by washing the filters in 0.4M NaOH at 45°C for 30 minutes, and then in [0.1 x SSC, 0.1% SDS, 0.2M trisHCl pH7.5] for 30 mins at 45°C. Replica filters were also prepared by robot with the assistance of Gunter Zehetner and Dan Nizetic at the ICRF. In this case 16,000 clones were represented on filters or 7x11 cm.

Futhermore, DNA was prepared from cultures stored in 96 well microtitre trays by the method of Gibson and Sulston (1987, see below). DNA from a single micropreparation was used to produce 3 or 4 dot blot filters. DNA was denatured by the addition of 300μl of

0.4M NaOH, 10mM EDTA for 10 minutes at room temperature. 100µl or 75µl of this was then transferred by multichannel pipette to a dot blot apparatus fitted with Zetaprobe membrane (Bio-rad) which had been pre-wet in distilled water. After suction via vacuum, the wells were washed with a further 300µl of the denaturing solution which was also taken across the membrane under vacuum. Membranes were then moved diagonally within the apparatus to allow the application of DNA from a further microtitre tray to the same filter. Filters were then washed in 2xSSC, then allowed to air dry without touching, and were baked for 1 hour at 80°C. Prehybridisation of these filters was carried out under the following conditions; 5xSSC, 0.033mg/ml sonicated herring sperm DNA, 2% SDS, 10xDenharts solution, at 65°C for at least three hours. Hybridisation was carried out in the same solution with the addition of 5% Dextran sulphate. Filters were washed for 15 minutes in 2xSSC/0.1% SDS at room temperature and then twice in 0.5xSSC/0.1 %SDS at 65°C usually for 20 minutes. Removal of bound probe was achieved by washing the filters twice in 0.1xSSC/0.5% SDS at 95°C.

2.8 cDNA first strand synthesis

Poly $(A)^+$ RNA was prepared by Steve Jeremiah from a human adult testis. 5μg of this was used as a template for complementary DNA synthesis in the following reaction mix; 20μl 5x Reverse transcriptase buffer (BRL), 5μl 0.6M ß-mercaptoethanol, 10μl oligo dT_{18} (1mg/ml), 1μl each of dATP, dCTP, dGTP, dTTP at 100mM, 20 Ci [α-^{32}P] dCTP and 100units of AMV reverse transcriptase (BRL). This was incubated at 42°C for 1 hour, after which time 2μl were removed for TCA precipitation. The remainder was boiled for 3 mins, placed on ice for 5mins, spun for 2 min and the supernatant transferred to a new tube.

To this was added 10μl of freshly prepared 1M NaOH. Base hydrolysis was carried out at 68°C for 20 mins. After cooling to room temperature, the solution was neutralised with 10μl 1M HCl and 10μl 1M Tris(pH7.5). Free nucleotides were removed by running the sample through a Sephadex G50 column, the volume was reduced to 100μl and the DNA precipitated using ethanol. The material was resuspended in dH_2O and used for labelling by the standard random hexamer method.

The amount of cDNA synthesised was calculated by TCA (Trichloroacetic Acid) precipitation: 2μl sample was added to 18μl TE, 2μl of this was spotted onto a Whatman GF/C glass fibre disc and set aside. To the remainder were added 1μl 0.5M EDTA, 1μl 10mg/ml herring sperm DNA and 1ml of ice cold 10% TCA. This was mixed and left on ice for 15 mins. The precipitate was collected by filtering through a GF/C disc and both discs were Cerenkov counted. The percentage of radiolabel incorporated was calculated and used to calculate the mass of DNA synthesised.

2.9 Preparation of DNA coupled to cellulose

This protocol devised by Ute Hochgeschwender and Miles Brennan is adapted from 'Brison, O., Ardeshir, F. and Stark, G.R. (1982) General method for cloning amplified DNA by differential screening with genomic probes. *Mol Cell Biol.* 2 578-587' and references contained therein.

100mg of m-Aminobenzyloxymethyl cellulose (Sigma) was mixed with an ammoniacal solution of copper hydroxide and ammonium hydroxide. To the dissolved and heated material was added sulphuric acid, until the solution changed pH from 12 to 6 at which point a colour change from dark to light blue occurred and the cellulose precipitated. The cellulose was collected by centrifugation and rinsed in ice-cold H_2O. The pellet was resuspended in 1.2M HCl and, whilst stirring, sodium nitrite solution was added. Excess HNO_2 was removed by the addition of solid urea and the cellulose was collected by centrifugation, washed in water, sodium acetate buffer and sodium acetate buffer containing DMSO.

To 1mg of high molecular weight DNA that had been boiled for 2 min and cooled on ice, were added cold sodium acetate buffer and DMSO, the cellulose was resuspended in this solution and the mixture was incubated for 24 hours at room temperature. The pellet was washed, resuspended in TE and stored at 4°C.

2.10 Hybridisation of DNA probes with DNA-cellulose

The ratio of probe DNA to Cellulose-bound DNA is 1:500. The required amount of DNA cellulose was washed in TE and resuspended in a prehybridisation solution [50% formamide, 0.75M NaCl, 50mM sodium phosphate buffer, 5mM EDTA, 0.1% SDS, 5xDenharts solution (1% wt/vol of each Ficoll, Polyvinylpyrrolidone, Bovine serum albumin), 100μg/ml denatured herring sperm DNA, 1% wt/vol glycine] and incubated at 42°C for 2 hours. The hybridisation solution was the same but without glycine and with 1xDenharts solution. The probe DNA was melted in 1ml of hybridisation solution at 80°C for 10 min and the DNA-cellulose was collected by centrifugation from the prehybridisation solution and resuspended with the probe DNA. This mixture was heated to 80°C for 2 mins and incubated on a shaking platform for 60 hours at 37°C. Every 12 hours the cellulose was pelleted, the probe DNA denatured by heating to 80°C for 2 min and then added back to the cellulose. After 60 hours the DNA cellulose was precipitated and the supernatant used as a probe in a hybridisation solution containing 50% formamide at 42°C.

Hybridisation solution = 50% formamide, 4xSSPE [20x = 3.6M NaCl, 0.2M $Na_2HPO_4.7H_2O$, 0.02M EDTA], 1% SDS, 0.5% BLOTTO [10% = 10g nonfat powdered milk, 0.2g sodium azide/100ml], 0.5 mg/ml carrier DNA.

2.11 A DNA probe prepared from a foetal testis cDNA library

A lambda gt11 cDNA library was prepared from foetal testis RNA by Y. Edwards and J. Wolfe. Insert sequences were amplified by the polymerase chain reaction (Saiki *et al.* 1988) in the following protocol according to Wong *et al.* 1989.

The sequences of the forward and reverse oligonucleotide primers were as follows;

5'> GGT GGC GAC GAC TCC TGG AGC CCG < 3' Forward

5'> TTG ACA CCA GAC CAA CTG GTA ATG < 3' Reverse

The reaction conditions were;

1μl of cDNA library (titre 2.3×10^{14}pfu/ml)
1μl of each forward and reverse primer (1μM final concentration)
1μl dNTP mixture (20mM stock dCTP, dGTP, dATP, dTTP)
10μl PCR buffer (166mM $(NH_4)_2SO_4$, 0.67M TrisCl pH8.8, 67mM $MgCl_2$, 100mM βMercaptoethanol, 67μM EDTA, 1.7mg Bovine Serum Albumin)
10μl DMSO
75μl H_2O

After denaturation (94°C for 9 mins), 2.5 units of Taq DNA polymerase (Cetus) were added and the amplification cycles carried out;

A) 9 cycles of 49° 2.5 mins
70° 4.0 mins
94° 1.5 mins

B) 1 cycle of 49° 2.5 mins
70° 4.0 mins

A further 2.5 units of Taq polymerase were added along with the primers (1μM) in 1xPCR buffer and after denaturing for 1.5 min at 94°C, 15 cycles of the above set A were carried out followed by 1 cycle of set B.

The products were cleaned and labelled by random-prime oligolabelling and used to probe the Y chromosome library. The small amount of PCR product suggested that the PCR reaction was not efficient from lambda gt11. In this case, the probe was not representative of the whole content of the library and this could account for the fact that only one positive cosmid was identified.

2.12 Micropreparation of cosmid DNA in microtitre plates

(Gibson & Sulston 1987)

This method facilitates the preparation of a small quantity of cosmid DNA from large numbers of cultures. Two microtitre plates of 96 wells can be easily handled simultaneously.

100μl overnight cultures were grown in 2xTY medium with kanamycin in the microtitre trays. These were spun in a plate centrifuge (2500rpm, 2 mins) and the medium thrown off. The plates were vortexed gently to resuspend the pellet and 25μl of solution I was added to each well. Plates were incubated at room temperature for 5 mins. 25μl of solution II were then added, mixed, and, after incubation on ice for 5 mins, 25μl of solution III were added, mixed vigorously, and incubation was carried out on ice for a further 10 mins. (Solutions I, II and III are as described in the miniprep method).

The plates were spun (3000rpm, 5 mins) and 70μl of supernatant were removed to fresh wells containing 100μl of isopropanol. Cosmid DNA was precipitated for 30 mins at -20°C. Plates were spun (3000rpm, 5 mins), the supernatant thrown off and the drained pellets were resuspended in 25μl ddH_2O and 25μl 4.4M LiCl. Plates were placed at 4°C for at least 1 hour to allow precipitation of RNA and chromosomal debris to occur.

After this time, the plates were spun (3000 rpm, 5 mins) and 50μl of supernatant was transferred to new wells containing 100μl of isopropanol. After the cosmid DNA had been precipitated at -20°C for at least one hour and usually overnight, the plates were centrifuged as before, the pellets drained and dried at 42°C and then resuspended in 7.5μl of TE.

2.13 Cosmid fingerprinting technique

The total volume of DNA prepared as above was digested with the enzyme *Hin*fI in the following reaction; 7.5μl DNA
1μl ddH_2O
1μl 10xbuffer (NEB)
0.5μl *Hin*fI (10 units/μl)

Digestion was for 3 hours at 37°C.

The fragments were labelled in the following way; To each well was added in a pre-prepared mixture: 2μl TE
1.5μl dATP (20 M stock)
0.3μl Klenow (5 units/μl)
1μl [α-^{32}P] dCTP (0.2 Ci/μl in TE)

The reaction was allowed to proceed for 15 mins at room temperature, and was terminated by the addition of 2μl of gel loading solution.

The *Hin*fI-digested lambda standard was prepared as follows; 20μl (10μg) of lambda DNA (cI857ind1Sam7 supplied by New England Biolabs) was digested in 4μl of 10x buffer supplied with the enzyme by New England Biolabs, 2μl (20 units) *Hin*fI and 14μl dH_2O for 3 hours at 37°C. The fragments were labelled by diluting the digestion reaction with 30μl of TE and adding 4μl of [α-^{35}S] dATP (1000mCi/mmol) and 4μl (20 units) of Klenow DNA polymerase. After incubation at room temperature for 15 mins, the reaction was further diluted with 3x volume of TE and terminated with 28μl of gel loading solution.

2.14 Polyacrylamide gel electrophoresis

Gels were run in a 21 x 40cm apparatus (Bio Rad). The back plate was prepared by rinsing in silane solution and the gel was bonded to the front plate using a small amount of the solution [3ml 96% Ethanol, 5µl (methacryloxy)-propyltrimethoxysilane, 50µl 10% Acetic acid]. The non-denaturing gel mixture was;

6ml 10xTBE
6ml 40% Acrylamide:bis-acrylamide (19:1)
48 ml dH_2O

A casting gel was prepared using 20ml of this mixture with the addition of 140µl 25% Ammonium persulphate (APS) and 100µl of TEMED. The remaining solution was polymerised with 90µl of APS and 70µl of TEMED. Gels were run in 1xTBE and pre-heated to 45°C before the loading of 4.8 l of sample. 3µl of the lambda standard was loaded every 6 lanes. Running time was for 1nour and 10mins at 50 Watts with the temperature maintained at 50°C. After this time the unincorporated nucleotides had run to the very bottom of the gel. After running the gel was dried for 45-60 mins at 80°C and then exposed to Amersham Hyperfilm ß max overnight at room temperature.

CHAPTER THREE

RESULTS PART 1

3.1 The human Y chromosome cosmid library

I have prepared an arrayed cosmid library which is, in the main part, specific for the human Y chromosome.

This was achieved by first isolating clones from a library constructed using DNA from the somatic cell hybrid 3E7. From analysis of its karyotype, this cell line contains several rearranged human Y chromosomes as the only obvious human material in a mouse background (Marcus *et al.* 1976). Radiolabelled male genomic DNA was hybridised to lysed colony filters in order to identify clones containing human rather than mouse inserts. Having screened in the region of 700,000 cosmid clones (24 separate platings of the library, each consisting of two to three filters containing 10,000-20,000 clones), I have identified approximately 1,600 clones containing human inserts. This is 0.2% of the number of clones screened. The karyotype of the cell line was not tested immediately before the library preparation, but from the paper by Marcus *et al.* it is expected that the line is diploid with four Y chromosomes per cell. In this case, the expected percentage of human clones is 2%. It is possible that many of the cells have not retained this number of human chromosomes resulting in a lower percentage of human clones in this library.

An example of a typical screen of around 20,000 colonies is shown in figure 3.1. Clones which hybridised to the probe on both duplicate filters were picked onto gridded filters.

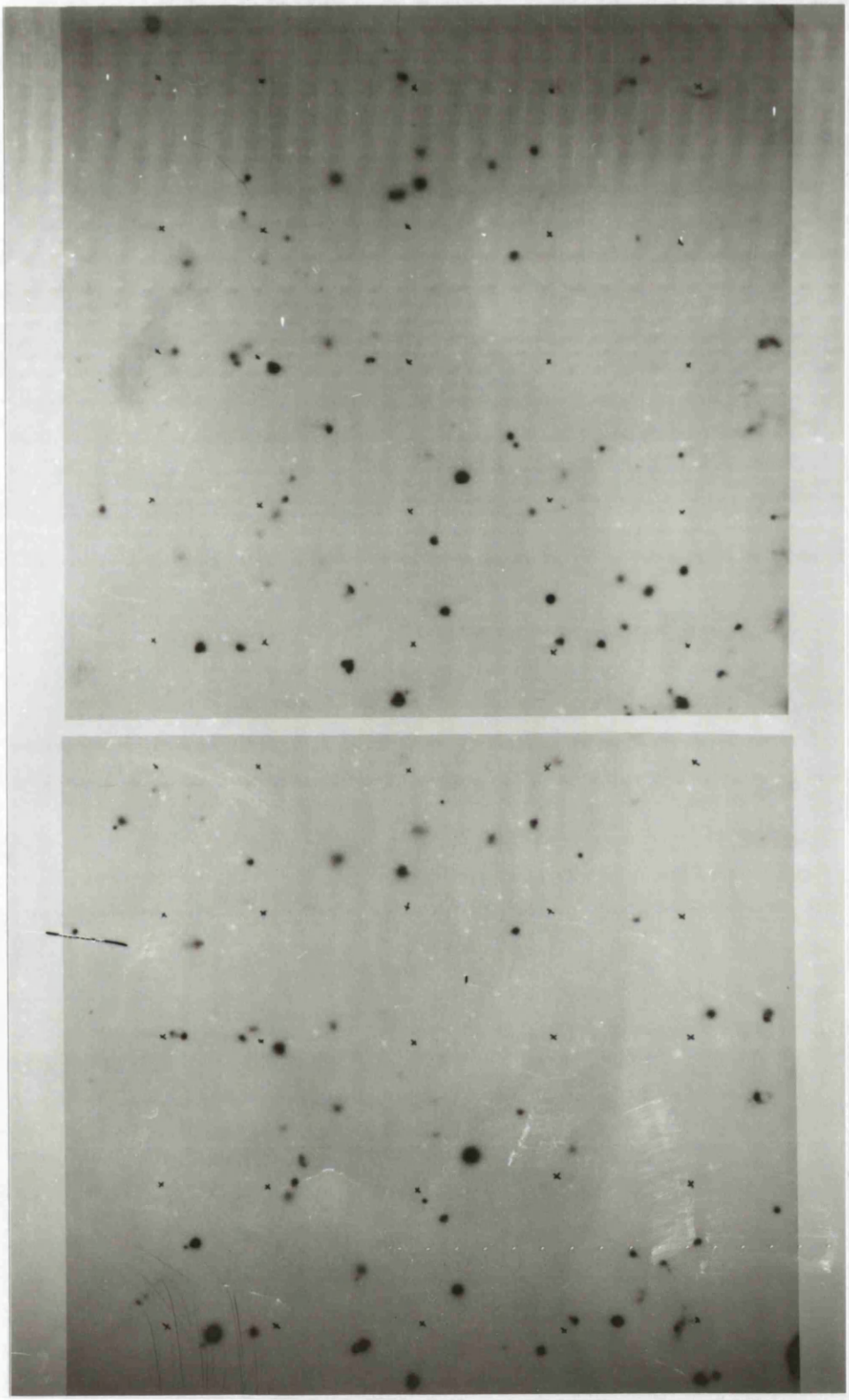


Figure 3.1 Primary screening of the 3E7 cosmid library with radiolabelled total human male DNA.

Duplicate filters are shown. Clones containing inserts derived from human material may be identified by their hybridisation to the probe. Registration pinpoints (marked by small crosses) are made in the filter to allow location of each clone on the master filter.

Hybridisation conditions were as described in section 2.7.5. Filters were washed in Church wash solution at 65°C and exposed to X-Ray film overnight.

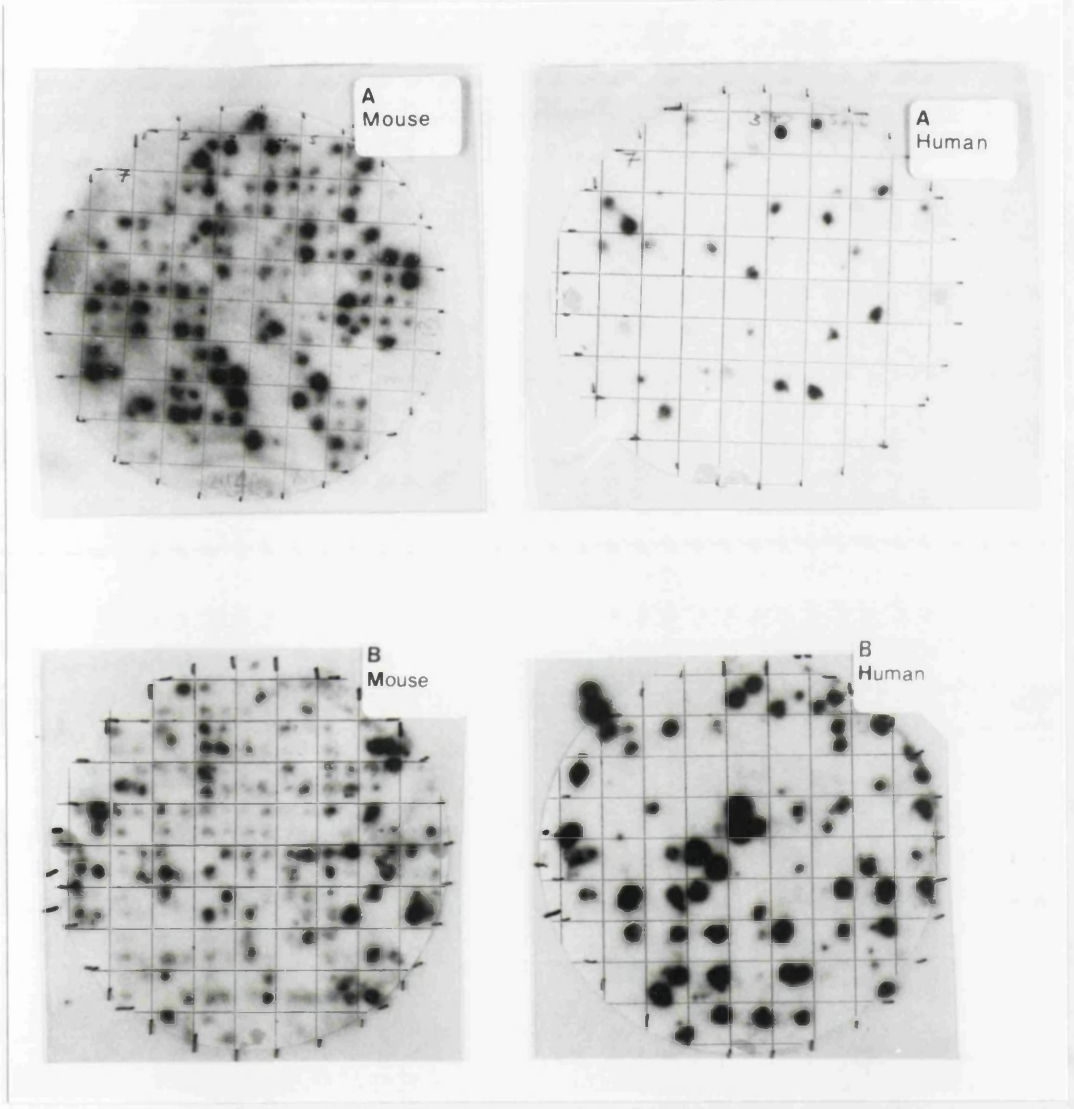


Figure 3.2 Secondary screening of clones with radiolabelled total human and total mouse DNA.

Two sets of duplicate filters are shown (**A** and **B**), each square contains four clones picked around each original signal. In each case one filter was probed with radiolabelled total human DNA and the other with radiolabelled total mouse DNA, as indicated on the figure. On these particular filters the percentage of *squares* containing a cosmid with a human insert is 47% and 78% on filters A and B respectively.

To ensure that each positive clone was picked it was usual to select four colonies in the region of the signal to be screened a second time. Duplicates of these filters were hybridised separately to radiolabelled human male and mouse DNA (figure 3.2). Colonies that were positive for the human probe only were stored at -70°C. Approximately 90% of the clones picked at the first screen were deemed to contain human inserts at the second stage. These 1,600 clones were added to those already prepared by J. Wolfe and S. Hall making a total of 1,728.

3.1.1 Hybridisation to known sequences

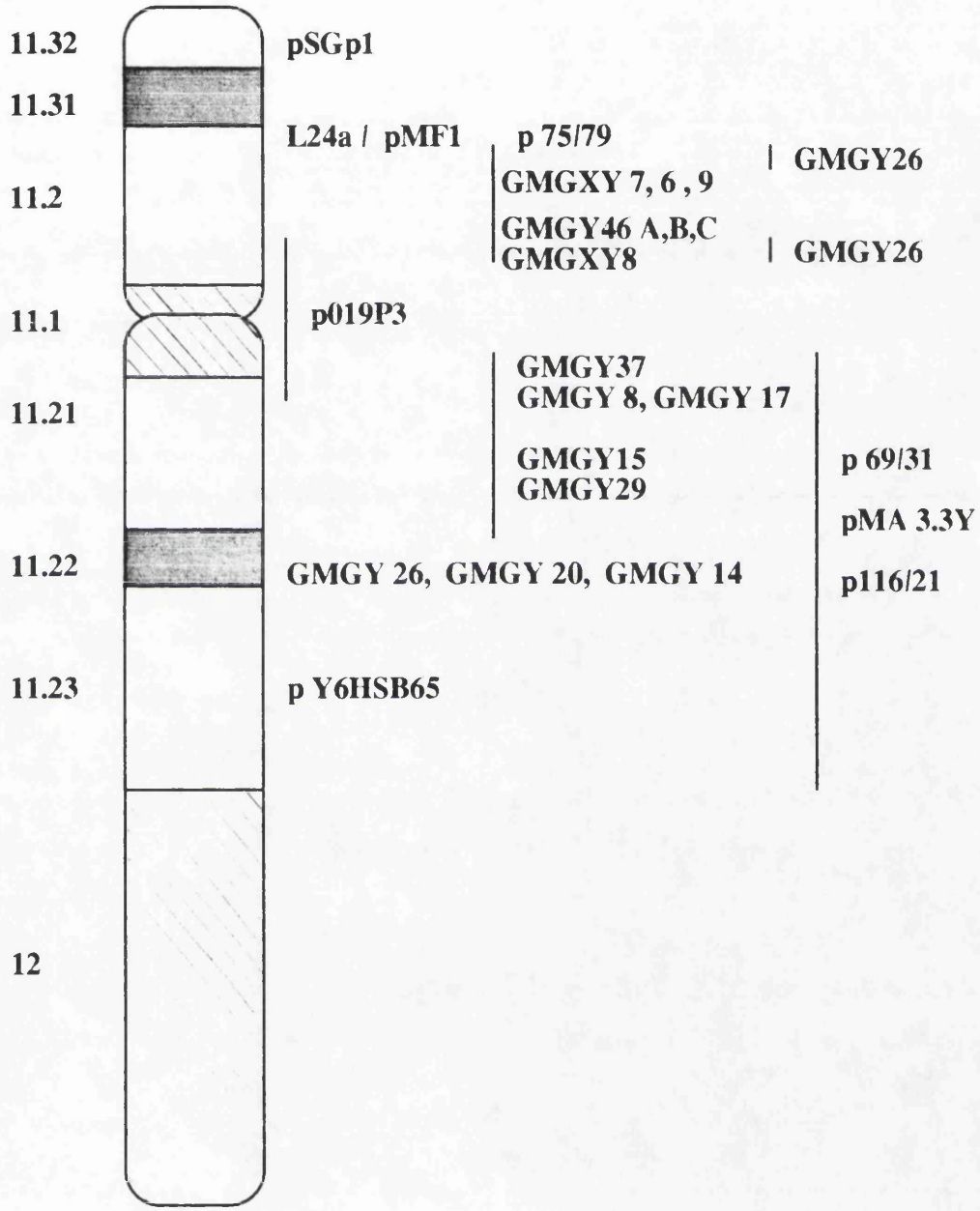
Having obtained this number of clones, it was necessary to characterise them. This was achieved by carrying out a series of hybridisation experiments using probes located along the length of the chromosome (see figure 3.3). These results are summarised in table 3.1.

Repetitive elements

To determine the proportion of clones containing repetitive elements, I hybridised radiolabelled DNA segments corresponding to the elements DYZ1, DYZ2 and DYZ3 to dot blots and lysed colony arrays on filters. Results are shown in figures 3.4, and 3.5 and table 3.1. In total, these repetitive elements are expected to cover up to 20Mb of the chromosome (DYZ1 containing regions may be as large as 13Mb (Schmid 1990), DYZ3 block sizes range from 0.3 to 0.9Mb (Oakey & Tyler-Smith 1990) and DYZ2 regions may cover 7.5Mb, (Smith *et al.* 1987)). However, DYZ1 repeats in particular appear to be under-represented in this library. This is difficult to explain. If the

units of a repeated block do not contain a restriction site for the enzyme used in the cloning process, then these regions will be too large to clone. However, the repeating unit of DYZ1 and DYZ2 do contain *Sau*3AI sites. It has been noted however, that tandem repeats often undergo extensive deletion events even when perpetuated in a recombination deficient host (Wyman & Wertman 1987, Yokobata *et al.* 1990) and it is likely that such tandemly repeated regions would be unstable in cosmid clones. This phenomenon has also been observed by Wolfe *et al.* (1984) and Bishop *et al.* (1982). Clones containing this repeated element can be easily recognised in the fingerprinting analysis by their deleted nature (see section 4.3).

Figure 3.3 Regional assignment of DNA probes



Assignments from Affara et al. 1987, and HGM 10.5

TABLE 3.1.

LOCUS	PROBE	REGIONAL ASSIGNMENT	REFERENCE	NUMBER OF COSMIDS
Highly repeated sequences				
DYZ1	pY3.4	q12	Cooke 1976	107 (of 960)
DYZ2	pYH2.1	q12	Cooke 1976	5
DYZ3	cY84	cen	Wolfe *et al.* 1985	51 (of 1728)
	$(CA)_9$			99
Low copy number repeats				
DYS56	GMGY46	p11.2	Affara *et al.* 1987	16
DYS77	GMGY26	p11.2 + q11.2	Kwok *et al.* 1987	64
DYS67	GMGY8	q11.2	Kwok *et al.* 1987	67
DYS73	GMGY20	q11.2	Kwok *et al.* 1987	23
DXYS25Y	p75/79	p11	Wolfe *et al.* 1984	16
DYS20	p69/31	q11.2	Wolfe *et al.* 1984	11
DYS21	p116/21	q11-qter	Wolfe *et al.* 1984	23
Single copy loci				
DYS62	GMGY15	q11.2	Kwok *et al.* 1987	3
DYS87	GMGY37	q11.2	Kwok *et al.* 1987	3
DYS80	GMGY29	q11.2	Kwok *et al.* 1987	6
DYS76	GMGY17	q11.2	Kwok *et al.* 1987	3
DYS118	GMGY14	q11.2	Kwok *et al.* 1987	3
DXYS32Y	GMGXY6	p11.2	Affara *et al.* 1987	2
DXYS34Y	GMGXY8	p11.2	Affara *et al.* 1987	6
	pMA3.3Y	q	Burk *et al.* 1985	7
	pYHSB65	q11.23	Vogt *et al.* 1991	5
	CI.1-2		A. O'Reilly pers. comm.	5

TABLE 3.1 Continued.

LOCUS	PROBE	REGIONAL ASSIGNMENT	REFERENCE	NUMBER OF COSMIDS
Gene loci				
ZFY	L24a	P11.3	North *et al.* 1991	1
	pMF1	P11.3	Palmer *et al.* 1991	1
MIC2	pSGp1	P11.3	Goodfellow *et al.* HGM8	-
Hunter syndrome	HFB2/E	X-Linked	G. Camerino pers. comm.	4
TSPY	pO19P3	p11.2? q11.21?	Shortle 1990	28
ASSP6	pAS1		Daiger *et al.* 1982	-

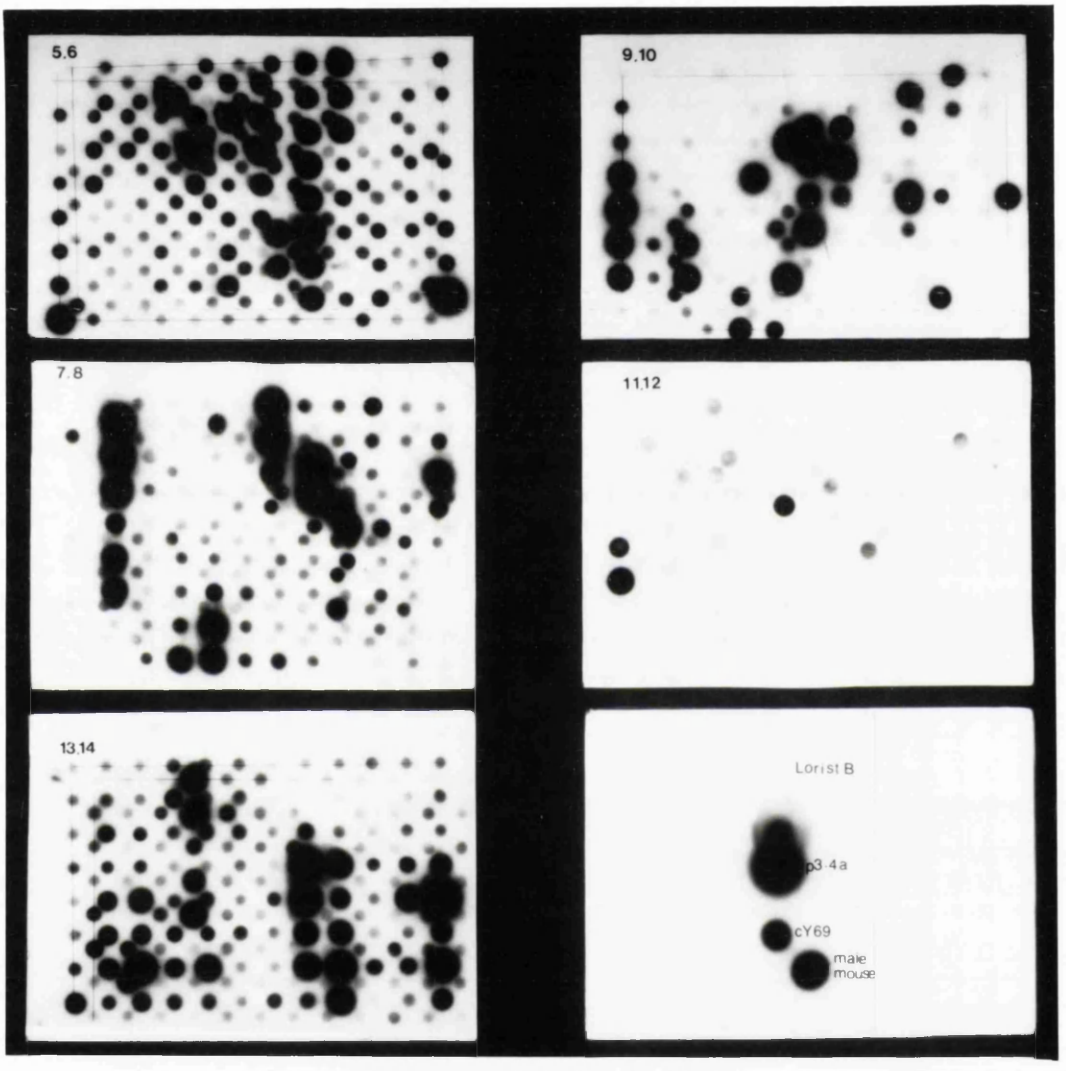


Figure 3.4 Hybridisation of probe pY3.4a (DYZ1) to DNA dot blots.

The first 960 clones picked from the library have been screened in this experiment. Each filter contains DNA prepared from clones arrayed in microtitre trays, each tray is labelled with a number from 5 to 12. DNA from two trays has been fixed on each filter and these are offset by moving the filter diagonally in the dot blot apparatus after addition of the first set of samples.

The filters were hybridised as described in section 2.7.6 and washed in 0.1XSSC/0.1%SDS for 30 mins at 65°C.

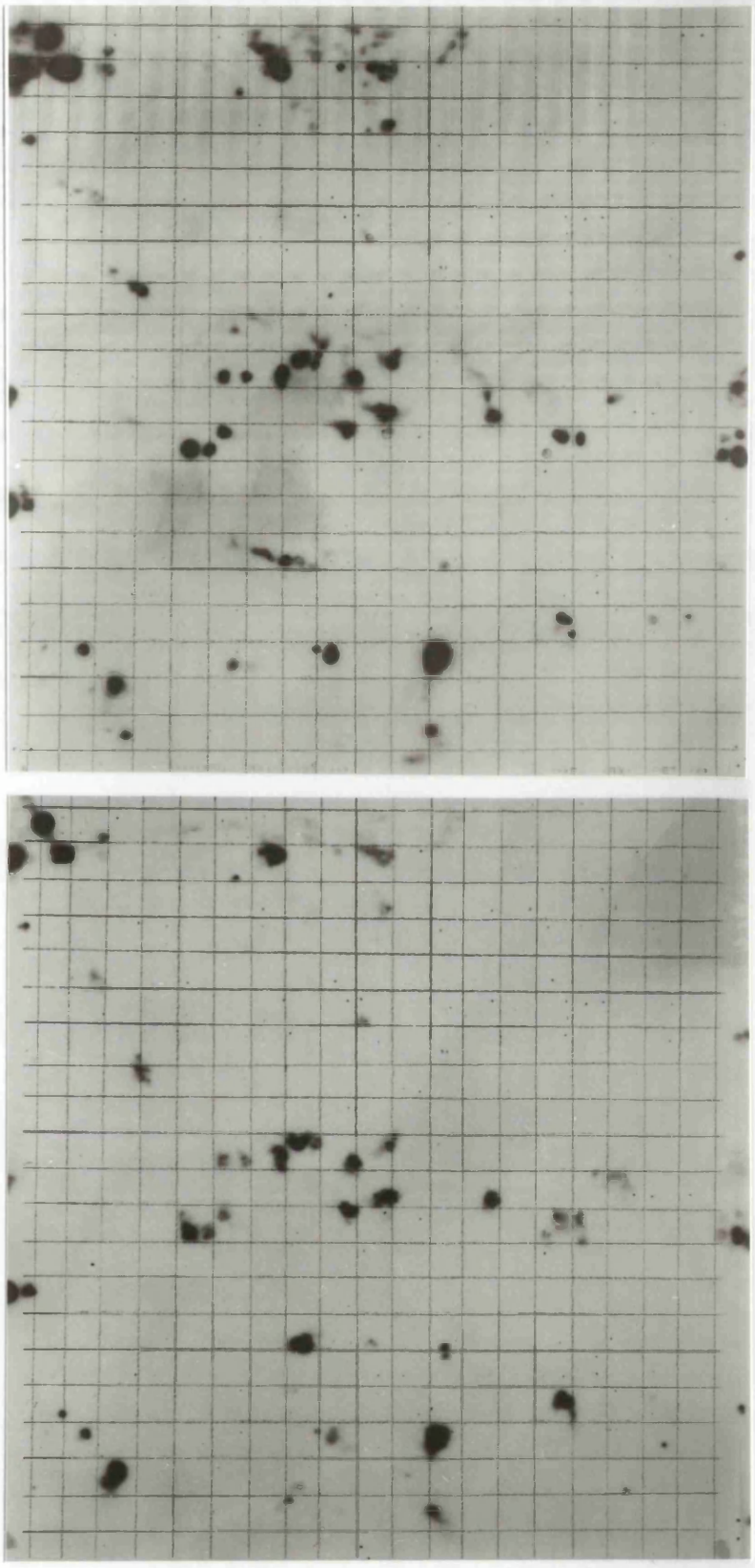


Figure 3.5 Identification of cosmid clones containing DYZ3 loci.

These duplicate filters of lysed colony arrays were probed with a 5.5kb *Eco*RI fragment from cosmid cY84 and washed as described(2.7.5).

Gene probes and anonymous DNA segments

I have also screened the library with a number of single and low-copy DNA sequences. The results are tabulated (Table 3.1, see also appendix). Representative hybridisation results are shown in figures 3.6 to 3.10.

The number of clones hybridising to a variety of single-copy sequences suggests that the library represents, on average, a three-fold coverage of the chromosome. As the repeated regions are under-represented, the apparent size of the chromosome is reduced from 59Mb to around 30Mb or less. Such a depth of coverage is therefore possible with the number of cosmids in the library.

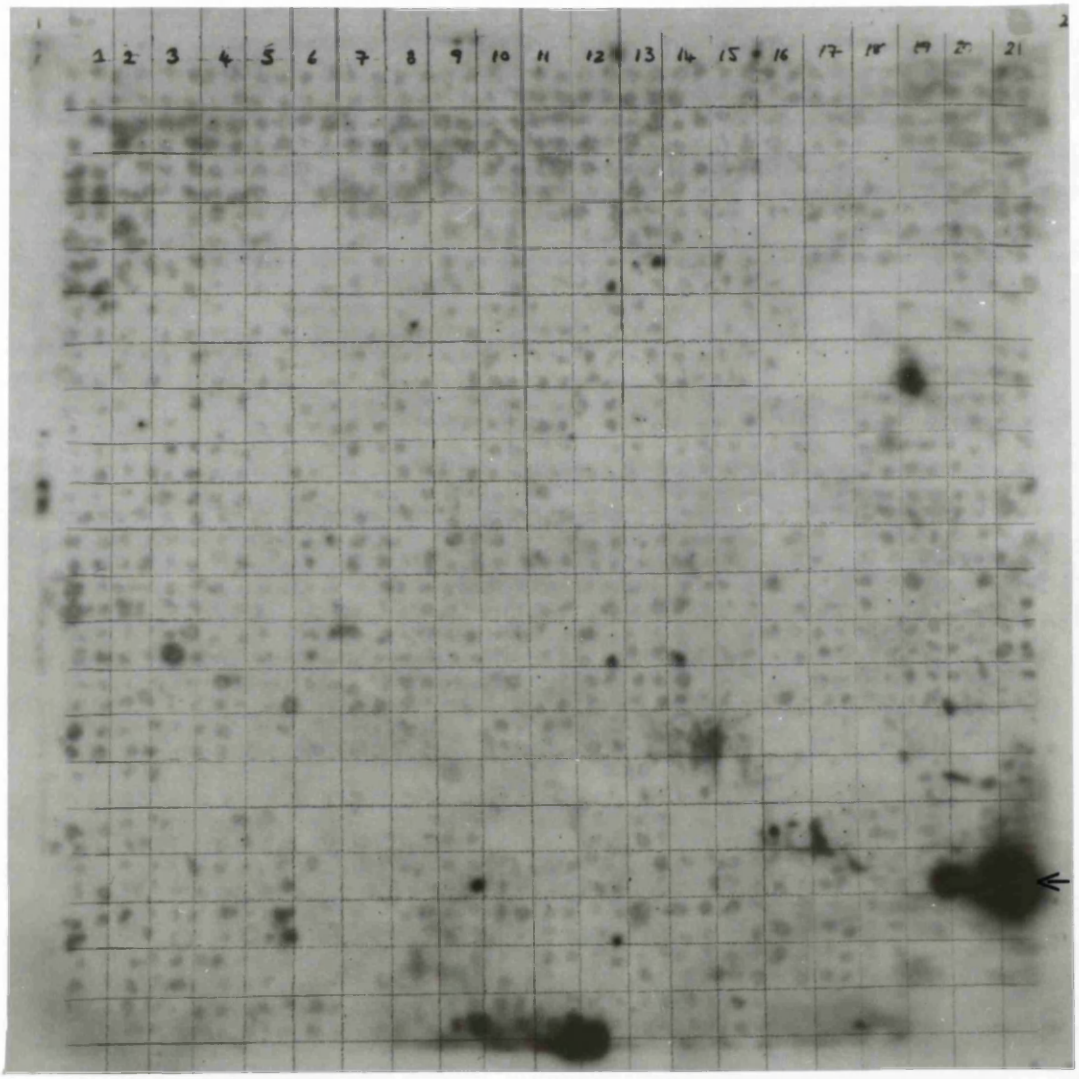


Figure 3.6 Hybridisation of probe pMF1 to a lysed colony array.

This cDNA clone contains the zinc finger region of the *ZFY* gene. The positive clone 2.26,37 was the only clone to show positive hybridisation upon a secondary screen. Hybridisation conditions were as for Hybond N filters (2.7.5).

Clone 2.26, 37 is marked with an arrow.

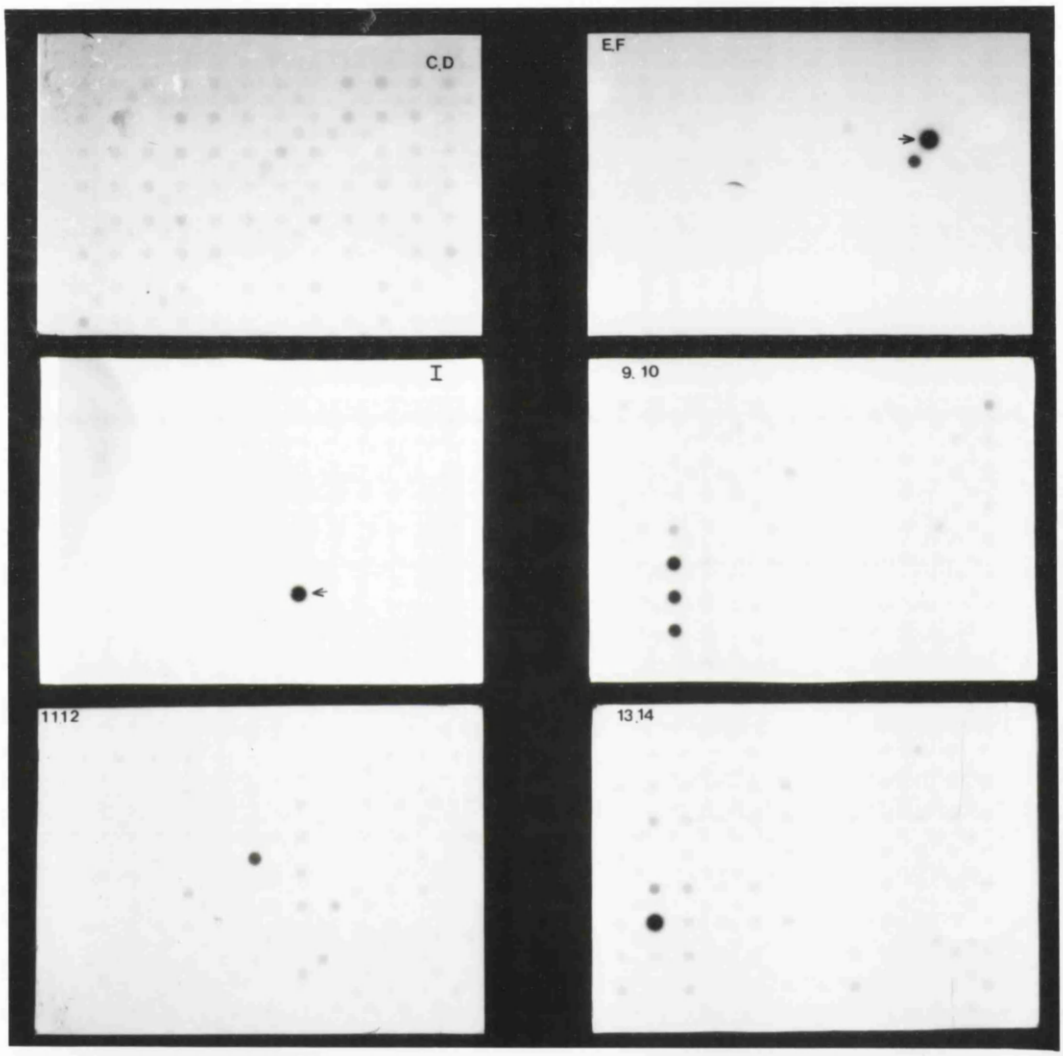


Figure 3.7 Hybridisation of probe L24A to DNA dot blots.

The filters presented contain DNA from 1056 clones, which were dot blotted as described (Pge 115). In this case microtitre trays were labelled with either a number or letter.

The cDNA probe L24A contains most of the *ZFY* gene. Clone 2,26,37 as identified by pMF1 (an independent cDNA clone also containing the *ZFY* gene) is present on two different filters (marked with an arrow). Other hybridising clones appear to contain sequences with homology to vector sequences (see section 3.1.2).

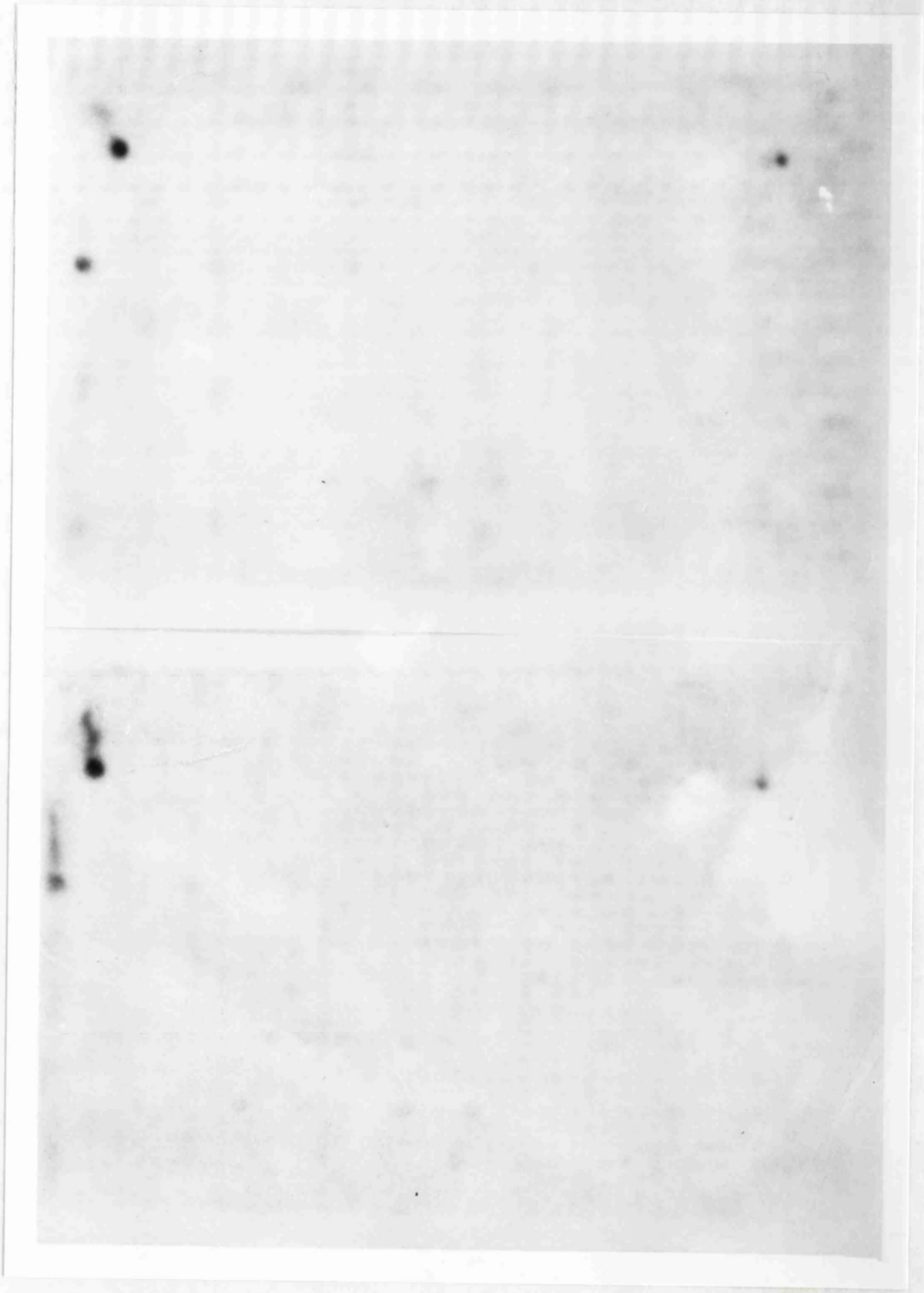


Figure 3.8 Hybridisation of the single copy probe p69/31 to lysed colonies gridded by robot.

Three hybridising colonies are observed on duplicate filters. Conditions were as described in section 2.7.5

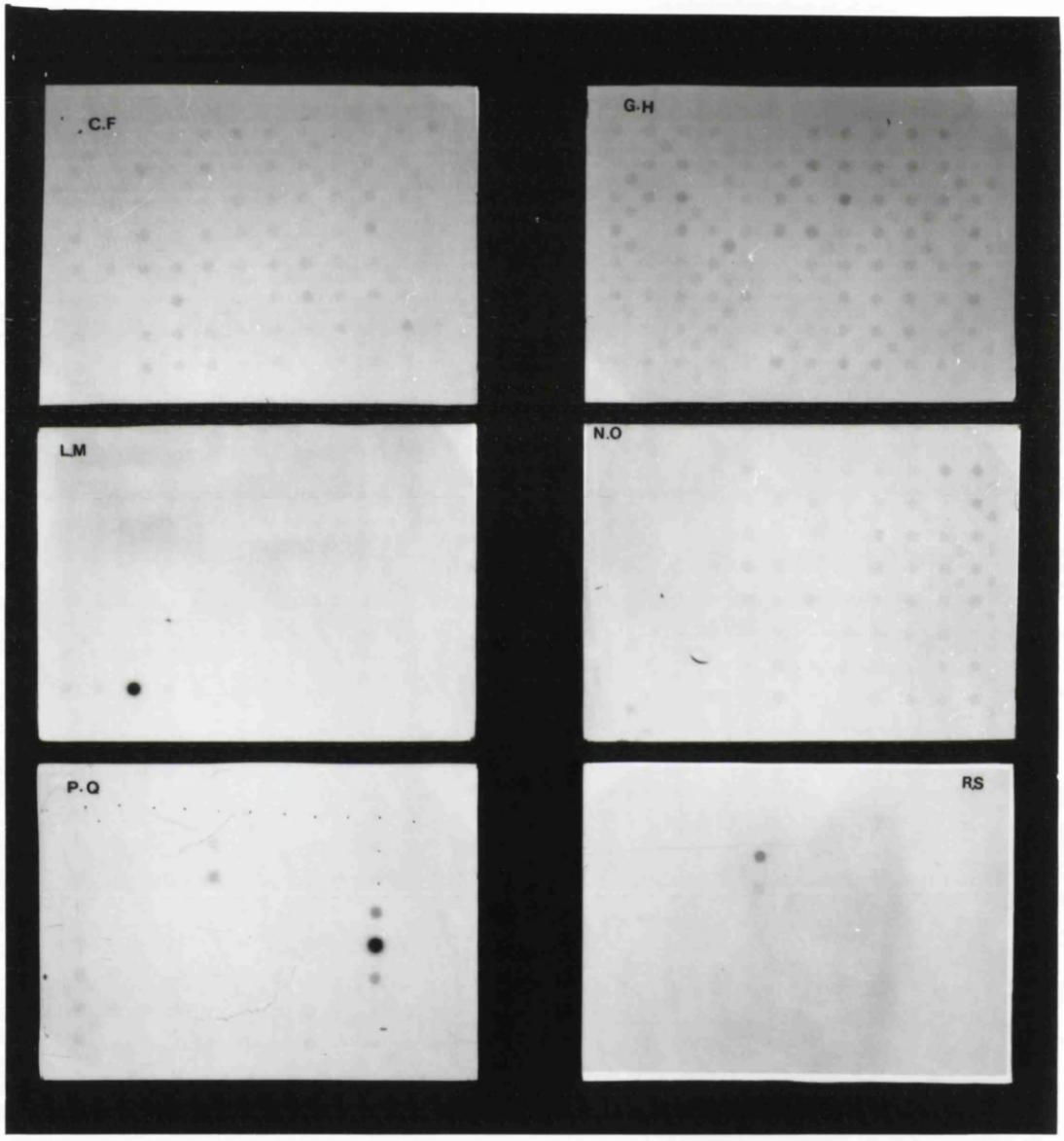


Figure 3.9 Hybridisation of the single copy probe GMGY15 to DNA from three clones on dot blot filters.

These filters (1152 clones) contain the three clones that hybridised to the probe (marked with arrows), and include examples of filters which do not contain any positive clones. In most cases of such hybridisation experiments, a background of weakly hybridising signals is observed (see filter GH). Positive clones are chosen somewhat arbitarily as those whose signal is noticeably higher than the background. The positive clone on filter PQ is flanked by weak signals. This is characteristic of the well to well contamination that sometimes occurs when handling clones in microtitre trays.

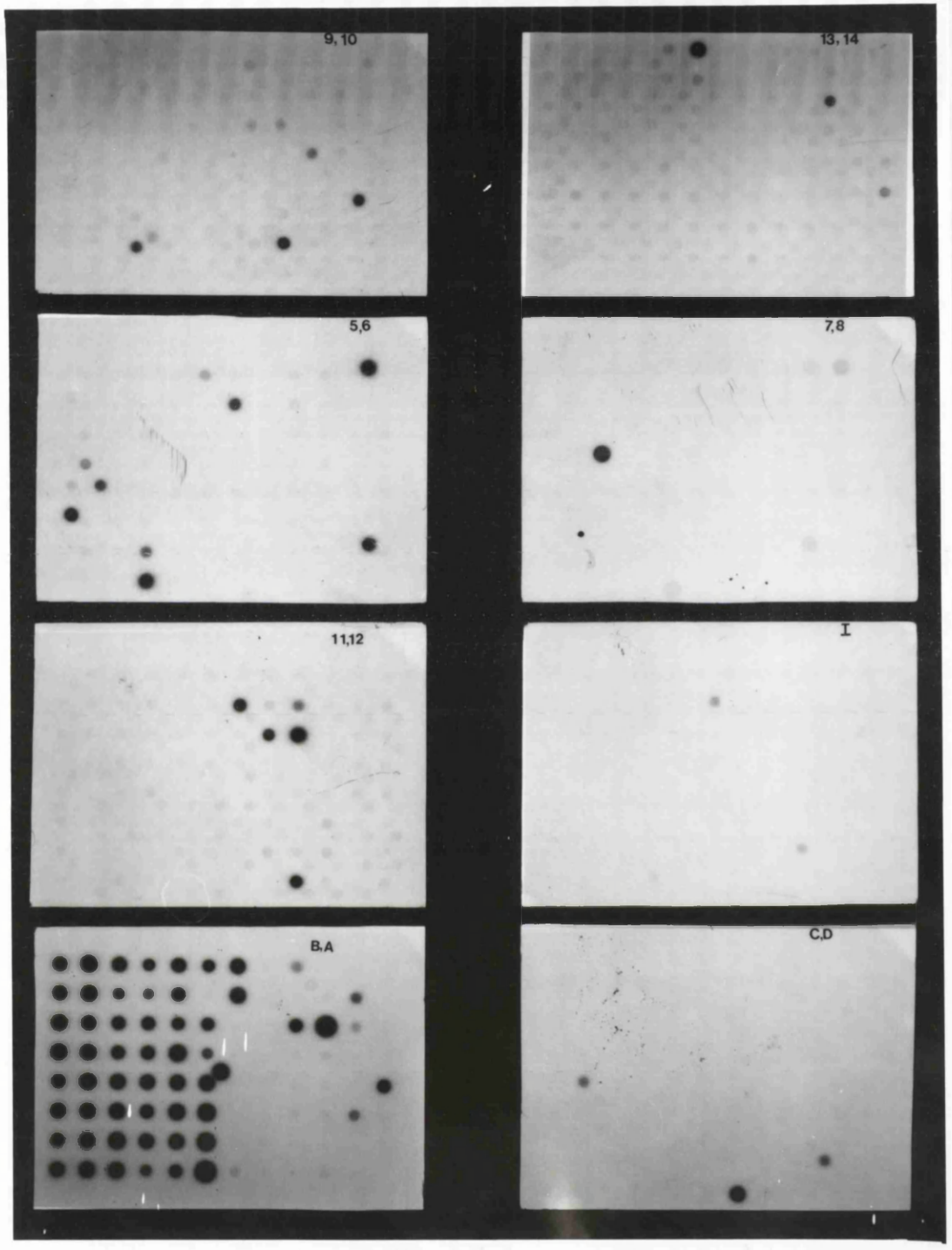


Figure 3.10 Hybridisation of the moderately repeated probe GMGY46 to DNA dot blots.

A series of cosmids from tray A have been contaminated and have identical fingerprints. They all hybridise to probe GMGY46. These clones have been excluded from the results. Washing stringency was 0.5xSSC/0.1% SDS.

3.1.2 *In situ* hybridisation

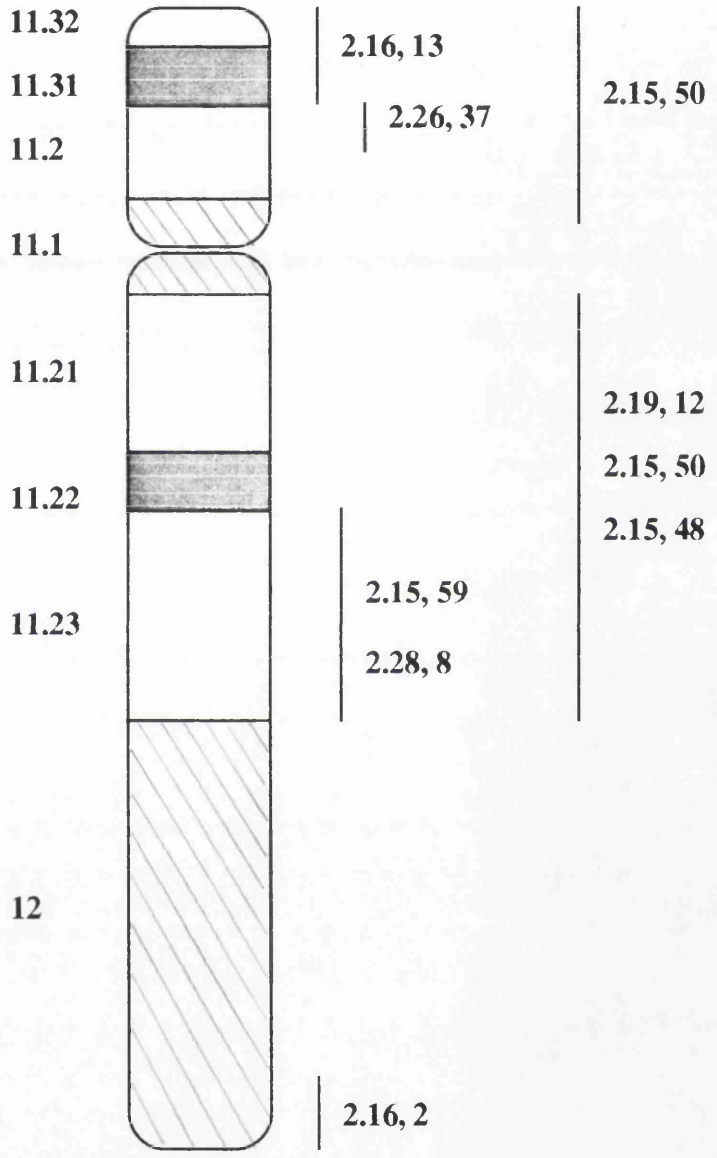
Several clones have been localised to chromosome regions by *in situ* hybridisation. This has been carried out by M. Leversha (Cambridge), D. Griffin (Galton Laboratory) and G. Prantera and M. Ferraro (Rome, Malaspina *et al.* (1990).

Clones were mapped for a number of different reasons. The two clones mapped by Malspina *et al.* formed part of the systematic search for Y chromosome polymorphism. Probe p2/2.9 (locus DYS149), a subclone of cosmid Lor2.16,2 recognises a strong Y-specific fragment and weak female fragments and is localised to Yq12 in the subterminal heterochromatic region. Probe p13/0.9 (DYS150), a subclone of cosmid Lor2.16,13 recognises four Y-specific sites at low stringency on the long and short arms but at high stringency is localised to distal Yp.

One cosmid localised by *in situ* in Cambridge hybridised to probe L24a which contains the acidic domain of the zinc finger gene *ZFY* (North *et al.* 1991). The probe recognises the cosmid Lor 2.26,37, already known to contain part of the *ZFY* sequence (Goodfellow personal communication based on hybridisation to cDNA probe pMF1), and maps to Yp as expected.

Several clones hybridised to probe CI 1-2 which is derived from a cosmid containing a long open reading frame and detects both Y chromosome and autosomal loci (Amanda O'Reilly personal communication) Most of these clones hybridise to a number of unrelated probes (including L24A) and it is possible that the hybridisation is non-specific, perhaps due to plasmid-like sequences within the clone. Some of these clones are located on the Y

Figure 3.11 Regional localisation of cosmid clones by in situ hybridisation



Chromosome 1 : 1p 35 -- 2.19, 27
-- 2.22, 2

1q 31/ 1q 32.1 -- 2.15, 59

Chromosome 12 : 12q 21.2 -- 2.18, 9

12q 14 -- 2.27, 6 CA repeat

chromosome but others are located on chromosomes 1 and 12 (see figure 3.11).

Furthermore, cosmid Lor2.27,6 which contains a polymorphic dinucleotide repeat (CA)n is located on chromosome 12q14 (Ulinowski *et al.* 1991, in press *Ann. Hum. Genet.* 55). See figure 3.12a

Obviously the somatic cell hybrid contains portions of human chromosomes 1 and 12 as well as the Y chromosome. It is difficult to assess at this stage what proportion of the clones in the library are derived from these contaminating chromosomes.

Figure 3.12a. Fluorescence *in situ* hybridisation of cosmid 2.27,6 to chromosome 12q 14.

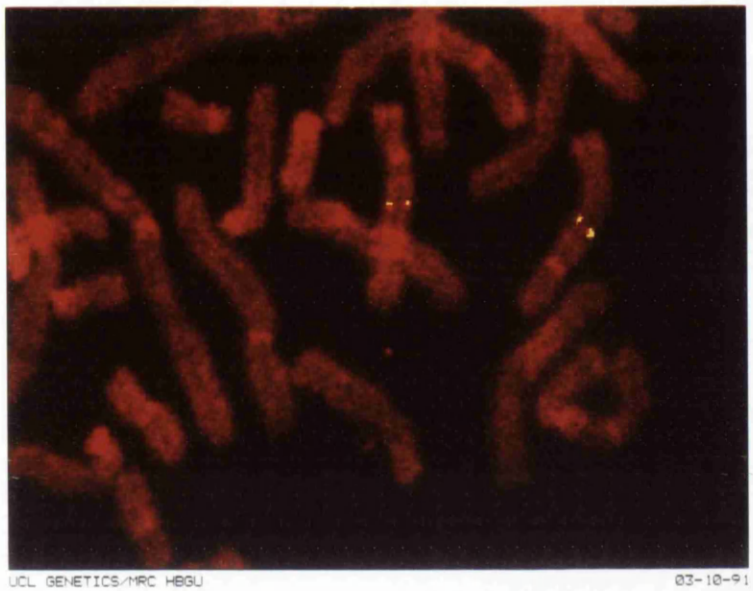


Figure 3.12b. Fluorescence *in situ* hybridisation of clone 2.28,8 to the Y chromosome long arm.

Both cosmids contain regions of $CA_{(n)}$ dinucleotide repeats. *In situ* hybridisation was carried out by D. Griffin.

3.1.3 A search for expressed sequences

I have used two different approaches in an attempt to identify cosmid clones in the library which may contain expressed sequences.

Screening the library with a probe derived from adult testis RNA.

Poly $(A)^+$ RNA prepared from adult testis was used as a template for cDNA synthesis by AMV reverse transcriptase with an oligo dT_{18} primer. From this, a radiolabelled second strand was synthesised in a random prime reaction. Since the probe is likely to contain repeated elements, these were removed by hybridisation to DNA coupled to cellulose as described in sections 2.8, 2.9 and 2.10.

The probe depleted of repetitive elements was hybridised to dot blots of the library in a solution containing 50% formamide at 42°C. The filters were washed in 2xSSC/0.1% SDS and were exposed to X-ray film for 3 nights at -70°C.

The resulting autoradiographs are shown in figure 3.13. Several duplicate signals were observed and were obviously not due to random hybridisation eventsor due to homology to highly repeated sequences. Most of these clones had not hybridised to any other sequences. However, one hybridises to the $(CA)_9$ oligonucleotide, one to probe GMGXY8 and two cosmids, which have sequential numbers and are probably the same clone, hybridise to GMGY20. It is perhaps surprising that the clone that is known to contain *ZFY* sequences did not hybridise to this probe, nor did any of the clones containing regions of homology to probe p19B3 which also detects an expressed sequence (Shortle 1990).

Figure 3.13 Hybridisation of a probe derived from adult testis cDNA to DNA dot blots.

Four sets of duplicate filters are shown. These are representative of the hybridisation patterns achieved with this probe. Each filter is labelled with the reference letter of the two microtitre trays that the DNA samples originated from, and Roman numerals which distinguish between the filters in a series of duplicates.

The probe was prepared, depleted of repeated sequences and hybridised to the filters as described in sections 2.8, 2.9 and 2.10. After washing in 2XSSC/0.1%SDS, the filters were exposed to X-Ray film for three nights.

In most cases, duplicate signals can be seen, although the intensity may vary due to uneven application of DNA to the filters. The large group of postive signals in the top right hand corner of filter **GHiv** which are not present on the duplicate filter (**GHiii**) probably arise from less stringent washing of this area. This non-uniform effect could be caused by washing too many filters together.

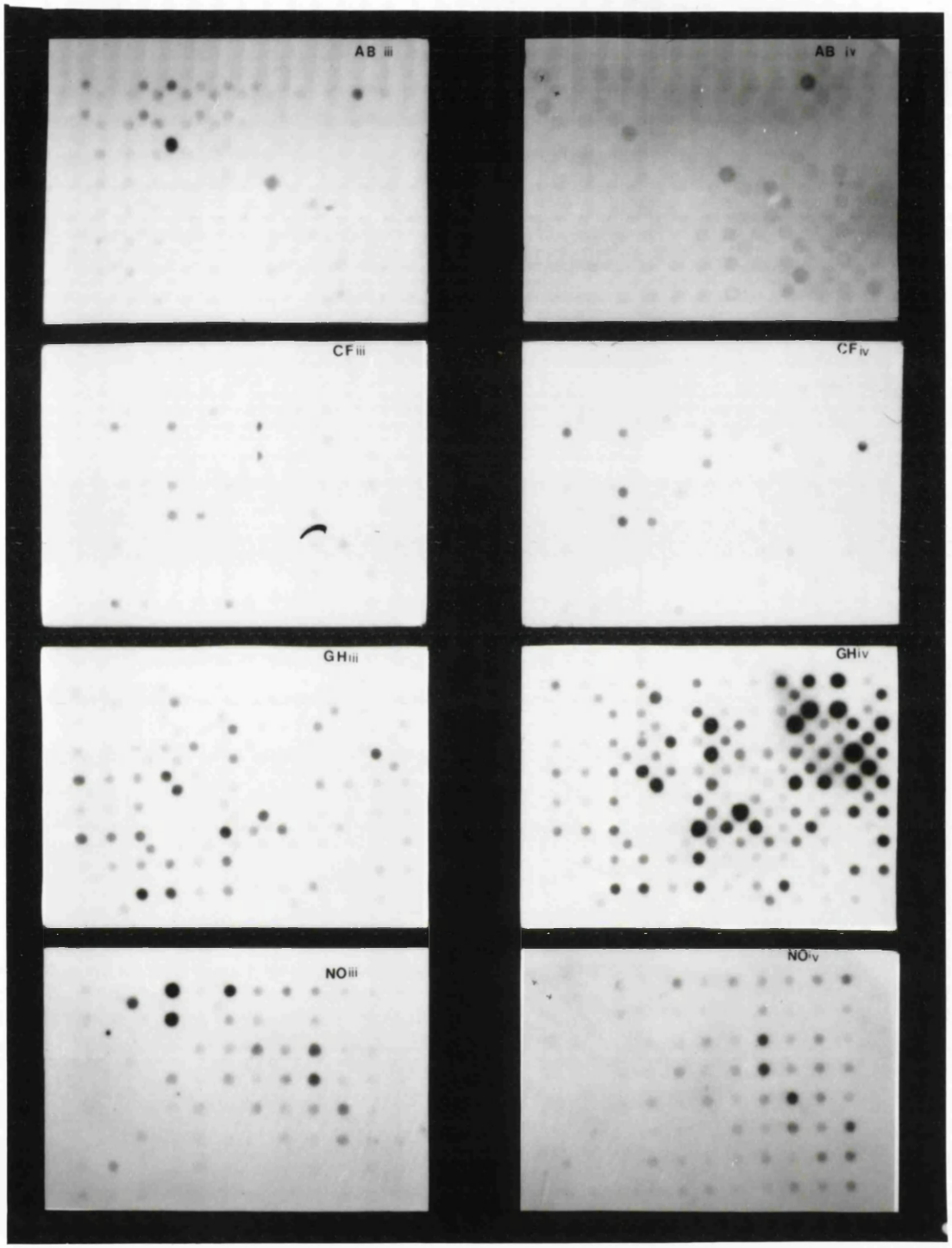


Figure 3.13 Hybridisation of a probe derived from adult testis cDNA to DNA dot blots.

See figure legend on facing page.

Hybridisation of PCR products from a foetal testis cDNA library to the Y chromosome library.

Insert sequences from a cDNA library derived from foetal testis RNA were obtained by PCR. Oligonucleotide primers situated adjacent to the *Eco*RI cloning site were used in a reaction described in section 2.11. Upon hybridisation to the dot blots, several faint signals were observed and one strong positive (see figure 3.14).

*Msp*I digestion

As a preliminary experiment to analyse the clones from both of these experiments further, I prepared digests of the clones using the restriction enzyme *Msp*I. Since regions of high unmethylated CpG content are known to be associated with the 5' end of expressed sequences (Bird 1986), it is likely that a clustering of sites for this enzyme (recognition site C'CGG) will be diagnostic of such regions.

The digested DNA fragments were labelled by end-filling using Klenow DNA polymerase in a reaction containing [$-^{32}$P]dCTP. These were separated on a 4% non-denaturing polyacrylamide gel run for 1 hour at 50 Watts and 45°C. Marker lanes were *Msp*I digested PBR322 labelled in the same way. The autoradiograph (figure 3.15) shows that most clones contain fragments corresponding to the digested vector with a few extra larger fragments of around 500bp upwards. One clone, however, (2.15 87) contains a large number of small fragments of 200 bp and less, indicative of a number of closely spaced *Msp*I sites. This clone was the most intensly positive clone when the foetal testis cDNA PCR product was hybridised to the library, and has not

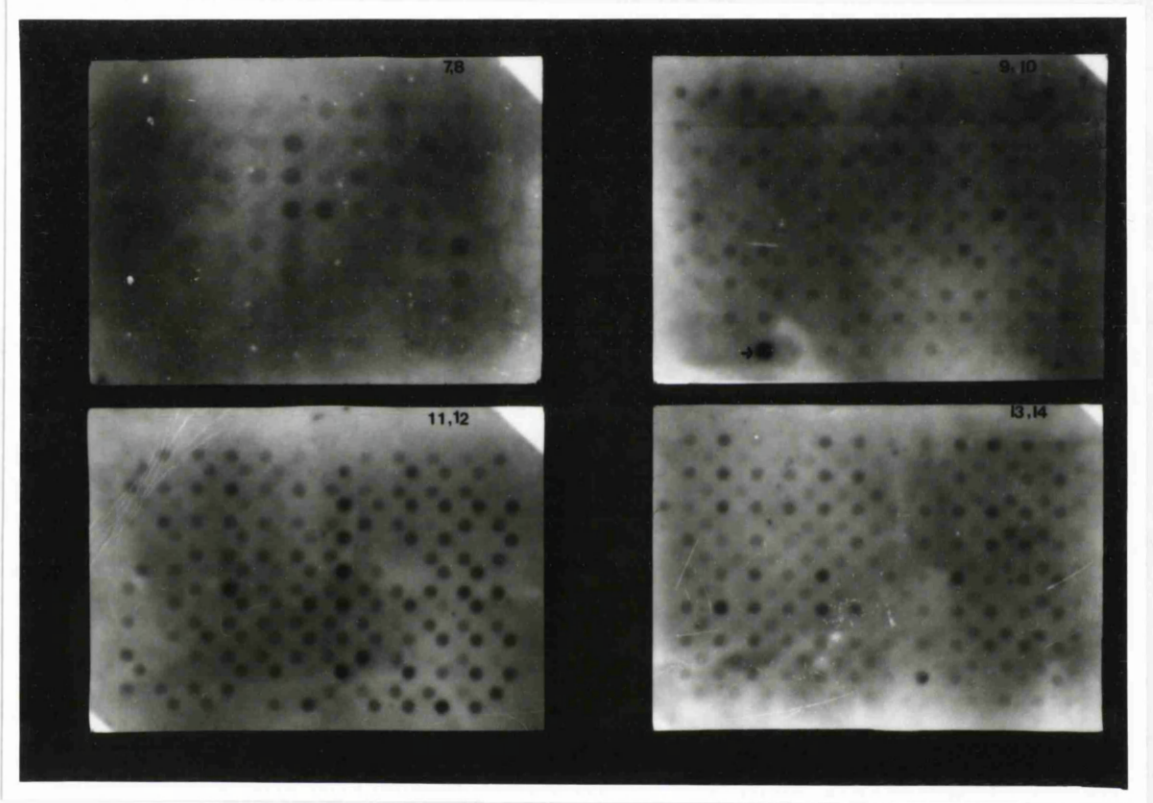


Figure 3.14 Hybridisation of labelled PCR products from a foetal testis cDNA library to the Y chromosome library.

Four filters are shown. These examples show that the probe results in a large amount of non-specific hybridisation. The most strongly hybridising signal (marked with an arrow) identifies the only clone deemed to be a true positive.

Probe preparation is described in section 2.11. The filters were hybridised to the probe at 65°C overnight and were washed for 30mins at 65°C in 0.5XSSC/0.1%SDS. Autoradiography was for three nights at -70°C using fuji film.

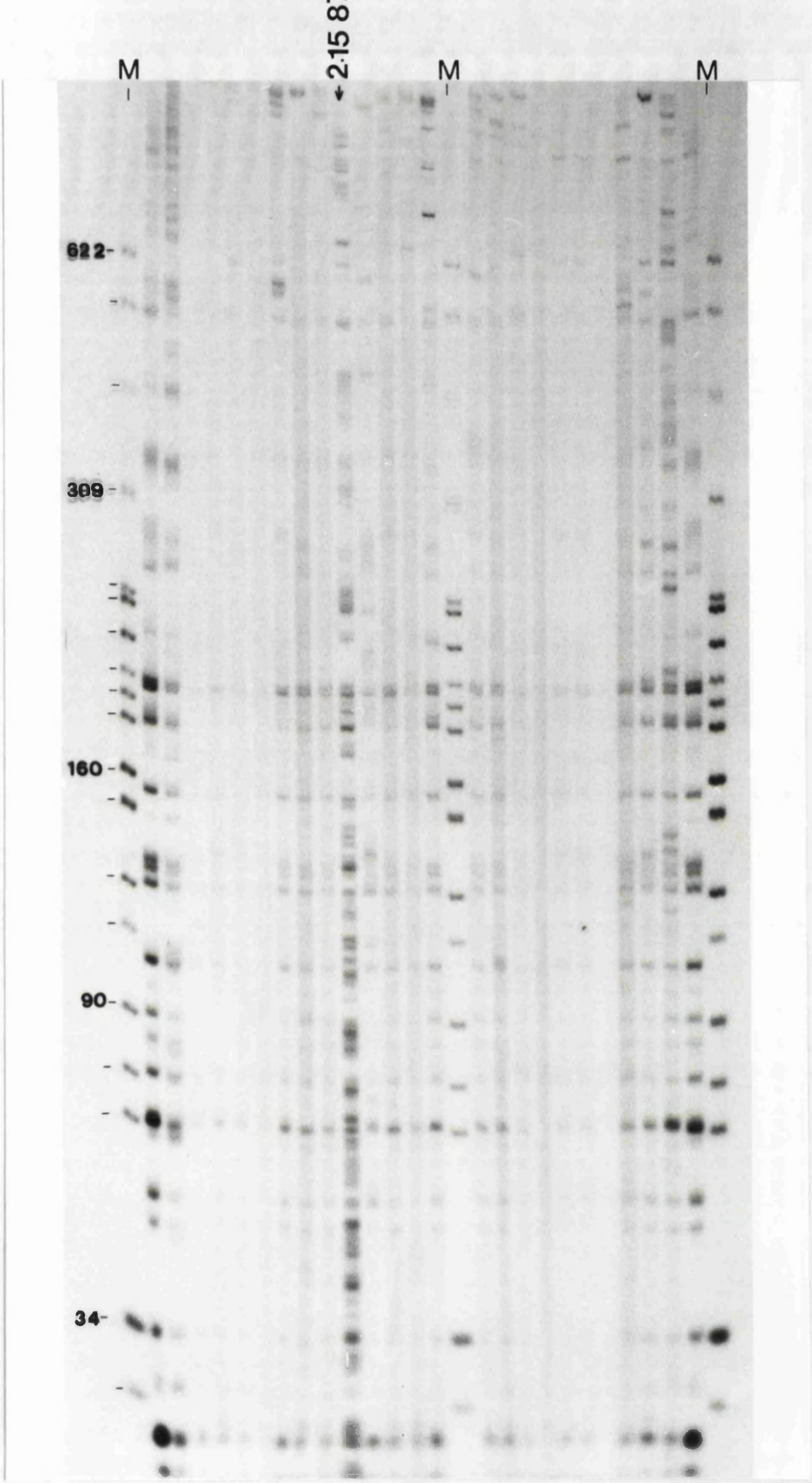


Figure 3.15 Radiolabelled *Msp*I-digested cosmid fragments separated by polyacrylamide gel electrophoresis.

Lanes are DNA from the cosmids identified as described in this section (3.1.3). The pattern of fragments obtained from cosmid 2.15,87 (marked by an arrow), suggests that it may contain a CpG island. Marker lanes (M) are end-labelled fragments of PBR322 digested with *MspI*.

hybridised to any other sequences tested to date.

I further examined the possibility that this clone may contain a 'CpG island' by digesting the DNA with other rare-cutting enzymes containing recognition sites rich in the nucleotides Cytosine and Guanine. First results (figure 3.16) indicate that there may be 6 sites for the enzyme *Bss*HII (recognition sequence G'CGCGC) of which 74% of sites appear to reside in islands, but no sites for the enzymes *Sal*I (G'TCGAC), *Sfi*I ($GGCCN_4$'NGGCC), *Eag*I (C'GGCCG) or *Sma*I (CCC'GGG).

If this cosmid does contain a sequence which is expressed in the foetal testis, it could be important from a developmental point of view.

There are a number of experiments which could be carried out to study this clone further. Initially it is important to confirm the chromosomal origin of the sequence given that the library does contain some clones derived from chromosomes 1 and 12 despite consisting predominantly of Y chromosome sequences. In addition to this the regional location may be determined by deletion mapping or *in situ* hybridisation.

Further restriction mapping may identity sites for other rare-cutting enzymes and indicate whether these are clustered. Although this does not appear to be the case since three of the four insert-derived *Eco*RI fragments are digested by *Bss*HII. It is also necessary to determine whether these sites are unmethylated in the genome, since cloned DNA becomes subject to the methylation patterns of the bacterial host. A single *Bss*HII fragment from the cosmid should identify the same size fragment when it is used to probe male genomic

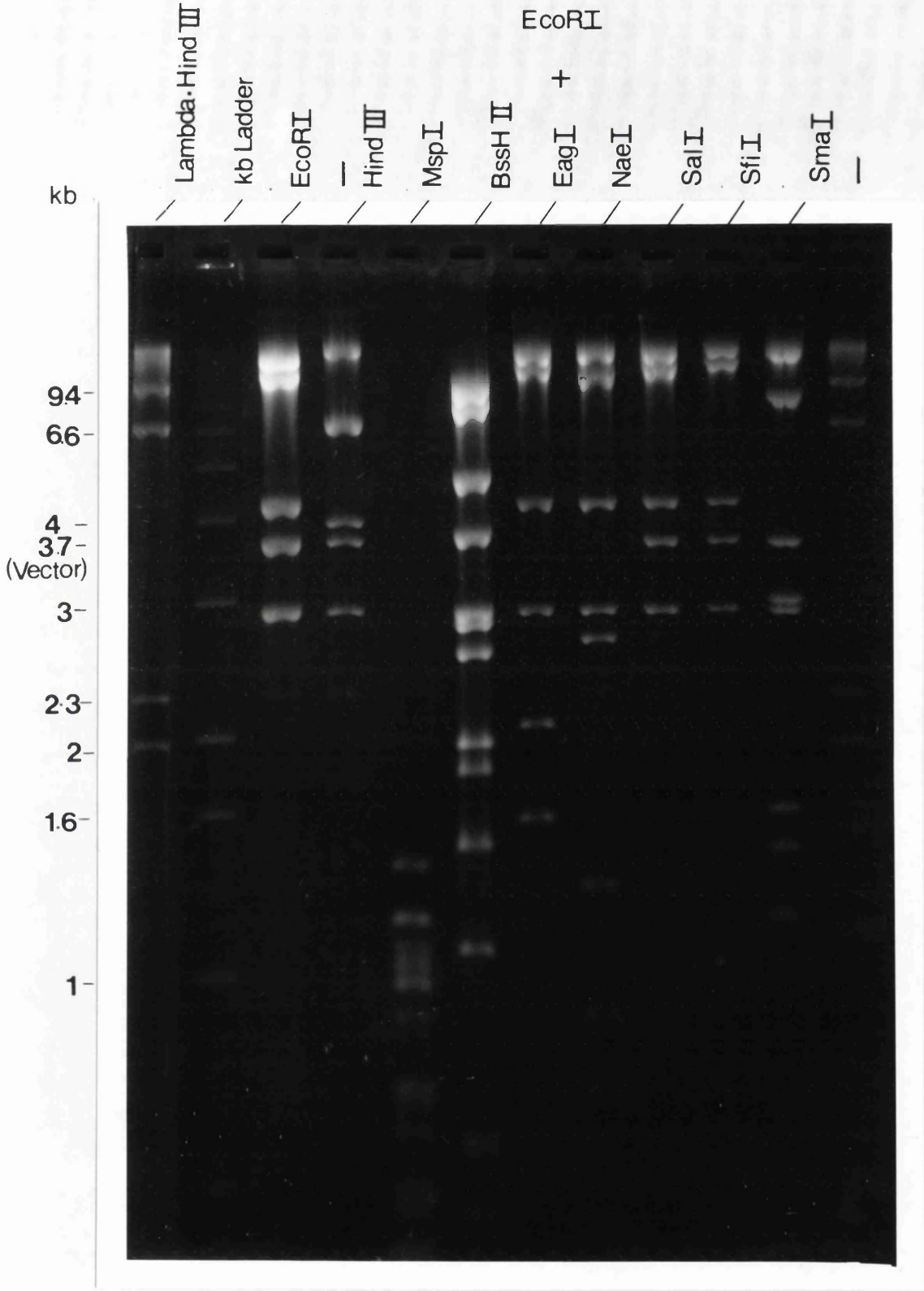


Figure 3.16 Cosmid 2.15,87 digested with *Eco*RI and double digests with 'rare-cutting' enzymes.

DNA digested with the same enzyme. A larger sized fragment would indicate that the genomic sites are methylated and do not represent a CpG island.

Also it would be appropriate to use probes derived from the cosmid (single copy internal fragments) in hybridisation experiments to filters of the original foetal testis cDNA library in order to identify corresponding cDNA clones. Similar probes could be hybridised to a Northern blot containing foetal and adult testis mRNA along with male and female control samples to study possible patterns of expression.

Due to a very limited amount of time that could be spent on these experiments they are far from conclusive. However, they provide a starting point which may lead to the identification of clones containing expressed sequences in the Y chromosome library.

3.2 DISCUSSION

I have examined the library for the presence of a number of different loci. These include ten anonymous single copy loci, seven low copy number repeat elements, and five gene loci along with the highly repetitive elements DYZ1, DYZ2 and DYZ3 and the dinucleotide repeat $(CA)_9$.

Repeated elements

The repetitive elements DYZ1 and DYZ2 are very under-represented in the library. Cooke (1976) estimates that the DYZ1 block should correspond to 40% of the total length of the chromosome (around 3,000-4,000 copies), in which case at least 40% of the cosmids should hybridise to this element. In fact only 11% of the clones screened contain the sequence DYZ1. The DYZ2 block may make up 25% of the chromosome (Smith 1987) but is only represented in the library in 5 clones. It is likely that these regions are rapidly deleted when cloned due to internal recombination events which occur even when the DNA is perpetuated in a recombination deficient host.

The alphoid elements, however, appear to be well represented. If the size of the alphoid DNA ranges from 300-900 kb (Oakey and Tyler-Smith 1991) then the largest region could be contained within around 22 cosmids of insert size 40Kb. The actual number of DYZ3-containing cosmids is 51, which probably corresponds to a three-fold chromosome coverage of the library. It is possible that the larger size of this repeat unit in comparison with that of DYZ1 and DYZ2 reduces the number of repeats per cosmid and therefore reduces the potential for recombination events.

Colony hybridisation of this kind can only give an indication of clones containing certain sequences. Spurious positive results appear

even when carrying out experiments in duplicate, and weak positives may be overlooked. 48 of the alphoid-containing clones have been analysed in more detail by Dr. Chris Tyler-Smith. On Southern blots of *Eco*RI digested DNA only 35 of these clones hybridised to the alphoid probe pY×1. However, probe pHY10 (DYZ1) hybridised to 6 clones not previously recognised implying that the number of DYZ1 containing clones is higner than detected in the initial hybridisation.

CA repeat oligonucleotide

The number of clones containing blocks or the dinucleotide CA is also lower than that expected of the genome as a whole, and, from the *in situ* data, it is apparent that at least one of these clones does not reside on the Y chromosome. Weber (1990) estimates that there are 35,000 copies of $(CA)_{12}$ in the haploid genome which, assuming random distribution, suggests that such elements should be found once in every 85kb or in one in two cosmids. The amount of $(CA)_n$-containing cosmids in the Y chromosome library appears to be around 10% of the expected value.

It is possible that repetitive blocks or this type are clustered in interval 6 of the Y chromosome. Vogt *et al.* (1991) have analysed a series of lambda and cosmid clones (some identified in this library by probe pYHSB65) which share homology to the fertility gene locus dhMiF1 of *Drosophila hydei*. A repeated sequence structure has been identified which is localised to interval 6 and contains blocks of $(CA)_n$ sequences. It is suggested that there is a similar sequence structure of Y chromosomal fertility genes in *Drosophila* and human and that the particular properties of stretches of DNA containing

such tandem repeats may have a functional significance on chromosome structure in this region.

Dinucleotide repeats of this kind are often polymorphic in length and may be a useful tool for studying populations (Weber 1990). One dinucleotide repeat $(CA)_{16}$ has been localised to the Y chromosome but does not show any polymorphism in males tested to date (Wolfe *et al.* unpublished, see figure 3.12b for *in situ* hybridisation).

Single locus DNA segments

On average, three clones share homology with each of the anonymous single-copy loci tested confirming that the library represents a three fold genome coverage. Of a total of 27 probes tested, two anonymous loci and two gene loci showed no hybridisation: GMGXY7 (DXYS33Y), GMGXY9 (DXYS35Y), pSGp1 (*MIC2*) and pAS1 (*ASSP6*). It is most likely coincidence that the two anonymous probes are derived from the same region of X:Y homology on Yp, but it is a possibility that this region is particularly under-represented in the library. The region around the pseudoautosomal boundary has been under-represented in other libraries (P. Goodfellow pers. comm.)

GMGY29 hybridises only to the group of cosmids that give positive signals with a large variety of probes. These clones have been shown to include regions of homology to pBR322 (C. Tyler-Smith pers. comm.) and are therefore reacting to contaminating vector sequence in the probe.

Probe pMA3.3Y is a single copy sequence, localised to proximal Yq, which also detects male specific sequences in apes (Burke *et al.*

1985). The fragment has been shown to be deleted in some azoospermic patients which, in keeping with its chromosomal location, suggest that this sequence may be functional in the control of spermatogenesis (K. Smith pers. comm.)

Low copy DNA segments

Probes which detect more than one locus hybridise, as expected, to a larger number of clones for example GMGY46, GMGY26 and p75/79. Probe GMGY8 hybridises to 67 cosmids even though it is assigned to only one locus. However, on Southern blots of male DNA this probe recognises a background smear (A. O'Reilly pers. comm.) and probably contains a moderately repeated element.

GMGY20 hybridises to 23 clones of which 6 also hybridise to p69/31. Five additional clones are detected only by p69/31. These two probes may be located within 40Kb of each other and therefore recognise common cosmids and also, cosmids covering adjacent regions. Alternatively, they may share a common locus and also recognise independent loci elsewhere on the chromosome.

Gene sequences

A single cosmid containing part of the *ZFY* gene has been identified by the two cDNA probes pMF1 and L24a. This clone contains the 3' region of the *ZFY* genomic sequence (North *et al.* 1991).

There are 28 clones with homology to the expressed sequence *TSPY*. A sequence localised to the Y chromosome that detects a Y-specific transcript has been identified independently by Arnemann *et al.* (1987) and Shortle (1990). The gene locus has been termed *TSPY* for Testis Specific Protein, Y chromosome, and is recognised by probes

pJA36B and pO19P3.

Probe pJA36B has been localised by deletion mapping (Arnemann *et al.*) to the median region of Yp, but *in situ* data obtained by Shortle using pO19p3, suggests that there are a number of loci spread along both Yp and Yq. Although the genomic sequence isolated by Arnemann *et al.* appears to contain part of an open reading frame, it has only 92% homology to the corresponding cDNA clone, does not contain an intron and is postulated to be a pseudogene. Clone pJA36B is comprised of two regions; a sequence that is repeated at least 10 times on the Y chromosome and has homology to a locus on Xq26-ter, and a single copy region on the Y chromosome containing the open reading frame.

It is likely that the large number of cosmids detected in this library using probe pO19P3 contain the repetitive element, but it is also possible that more than one pseudogene may exist along with the expressed sequence. It will be interesting to examine the similarity between these clones and assign them to separate loci, this may be possible using the fingerprinting technique. Also, the expressed locus may be identified by virtue of sequence identity to cDNA clones that I have retrieved from an adult testis cDNA library.

Hunter syndrome is an X-linked recessive gene disorder caused by deficiency of the enzyme alpha-L-iduronate sulfate sulfatase (IDS). A linkage study has localised the gene to the X chromosome long arm in the region Xq28-qter (Upadhayaya *et al.* 1985) and a cDNA clone has been isolated and mapped to Xq27. It has been suggested that homologous sequences may be identified on the Y chromosome (G. Camerino pers. comm.) and indeed, probe HFB2/E derived from the X-linked cDNA clone identifies 4 cosmids in this library.

Methods of presenting libraries for screening

I have screened the Y chromosome library in three different forms;

a) DNA dot blots — 10 filters 9 x 13cm

b) Colony arrays robotically positioned — 1 filter 11 x 7cm

c) Colony arrays positioned by hand — 1 filter 22 x 22cm

I found that the most spurious results were obtained using the robotically arrayed filters. This was unfortunate as the small filter size was very convenient for screening purposes. In several cases probes hybridised to a few or no cosmids on these filters but to several on the hand picked arrays. Also, the hybridisation was not consistent on duplicate filers and clones were difficult to identify due to the density of the array.

The colony arrays positioned by hand were very reliable and easy to identify. Although some problems were presented by uneven contraction of the filter whilst exposed to X-ray film at -70°C. These filters were obviously very tedious to prepare, it took some time to prepare the master filter which was then used to generate 20 replica filters.

The most convincing hybridisations were achieved with the dot blots. Although a lot of filters are required to represent the library and they are also time-consuming to prepare as the DNA must be isolated before transfer to the filters, the quality of results outweigh the effort involved.

Conclusion

A large amount of information may be obtained by screening the library with these regionally assigned probes. However, it should be noted that these are primary screenings and should be confirmed by secondary screening preferably of restriction enzyme-digested DNA from each clone.

Despite this, it would be valuable to continue screening with other loci in order to understand the relationship of clones to one another and as a means of localising contigs produced by the random fingerprinting technique.

These experiments were carried out in order to characterise the library in a broad sense. It appears that the cosmids correspond to a good representation of the Y chromosome, and are sufficient in number to provide a reasonable depth of coverage making a random fingerprint mapping approach feasible.

CHAPTER FOUR

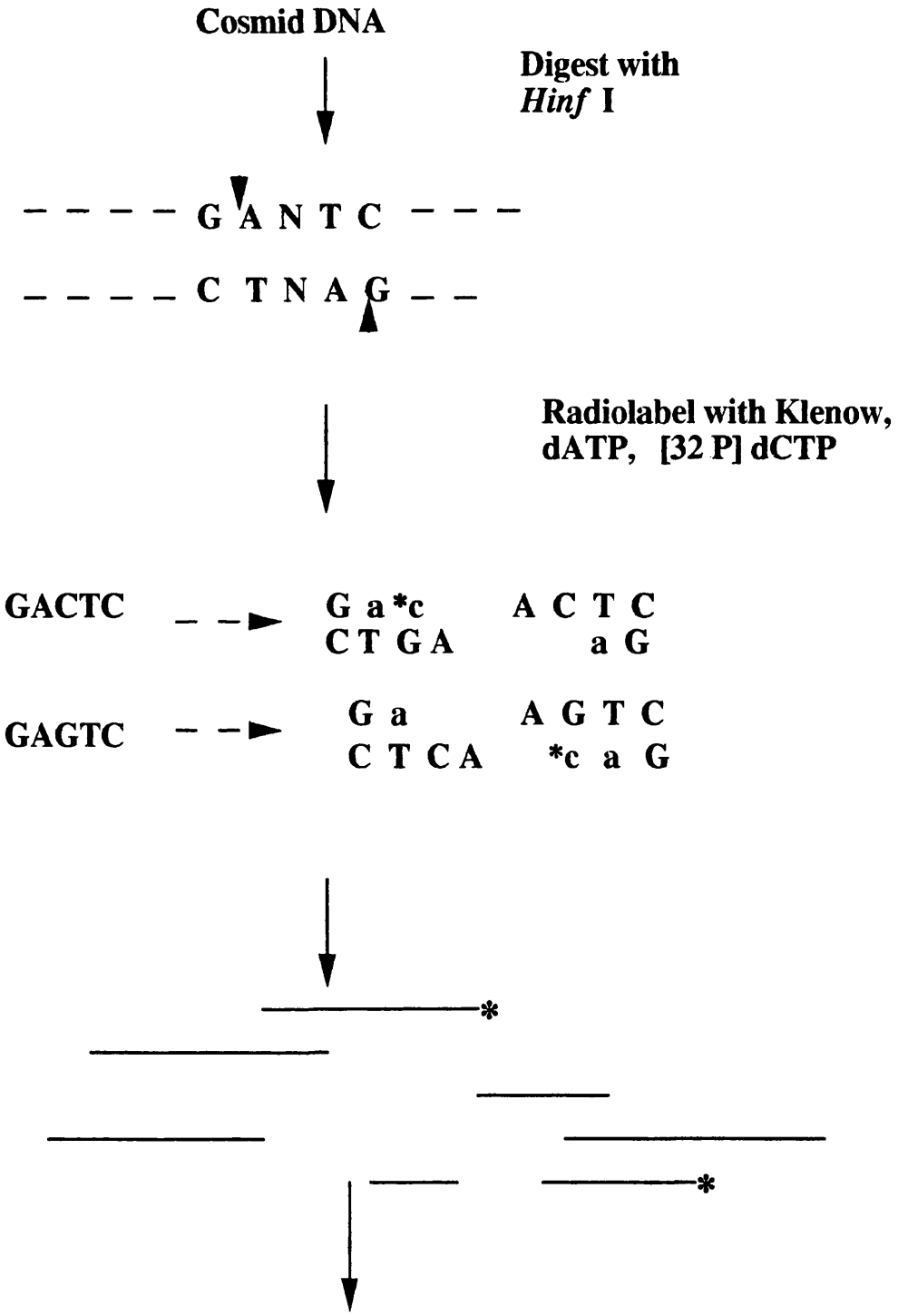
RESULTS PART 2

4.1 Introduction

Having prepared a cosmid library that appeared to be representative of the euchromatic portion of the human Y chromosome, I embarked upon the task of aligning these cosmids in the form of a contig map.

As discussed in part II of the introduction, there were a number of strategies that could be used. However, the most established at that time was the *Hin*dIII/*Sau*3AI double digest approach of Coulson *et al.* as applied to the genome of *C.elegans.* I spent several months attempting to use this method to analyse the Y chromosome cosmids but achieved limited success. I then turned to my own variation of a strategy developed by Knott *et al.* (1988) in the construction of the contig map of the *E.coli* genome. The results presented below are based on fingerprint analysis using this *Hin*fI digestion system. I have analysed the fingerprints of 723 cosmids and have identified 56 groups of overlapping clones. Of these, four groups do not contain linearly overlapping clones. Instead each group represents a different repeated element, which forms a false contig due to similarity between each cloned copy of the sequence.

Figure 4.1 Schematic of the fingerprinting strategy



4.2 Data generation and analysis

A brief description of the method (see sections 2.12, 2.13, 2.14)

Cosmid DNA was prepared in 96-well microtitre plates from small scale cultures. The use of semi-robotic devices such as repetitive dispensers and multichannel pipettes reduces some of the labour involved and results in a more uniform product. The DNA isolation method is an adaptation of an alkaline lysis procedure (Ish-Horowitz & Burke, 1981) but includes a purification stage using lithium chloride which precipitates a large amount of RNA and residual chromosomal DNA (Gibson & Sulston, 1987).

The DNA was digested with *Hinf*I in a medium salt buffer which was then diluted for the radio-labelling step. Cleaved fragments were 'end-filled' using Klenow DNA polymerase with added dATP and [α-^{32}P]dCTP. In this way one quarter of the fragments are labelled (see fig 4.1).

The terminated reaction was resolved by polyacrylamide gel electrophoresis. The 4% acrylamide, non-denaturing gels are bonded to the glass plate of the gel apparatus to prevent distortion upon drying. Electrophoresis was kept constant for each gel so that the samples were run to the same distance. This uniformity of gels allowed the computer programs to recognise standard lanes (end-filled *Hinf*I digests of Lambda DNA) and compare different clones between gels. Bands were visualised by autoradiography and the autoradiograph was digitised for analysis. Examples of fingerprint data produced by this method are shown in figs 4.2 and 4.3.

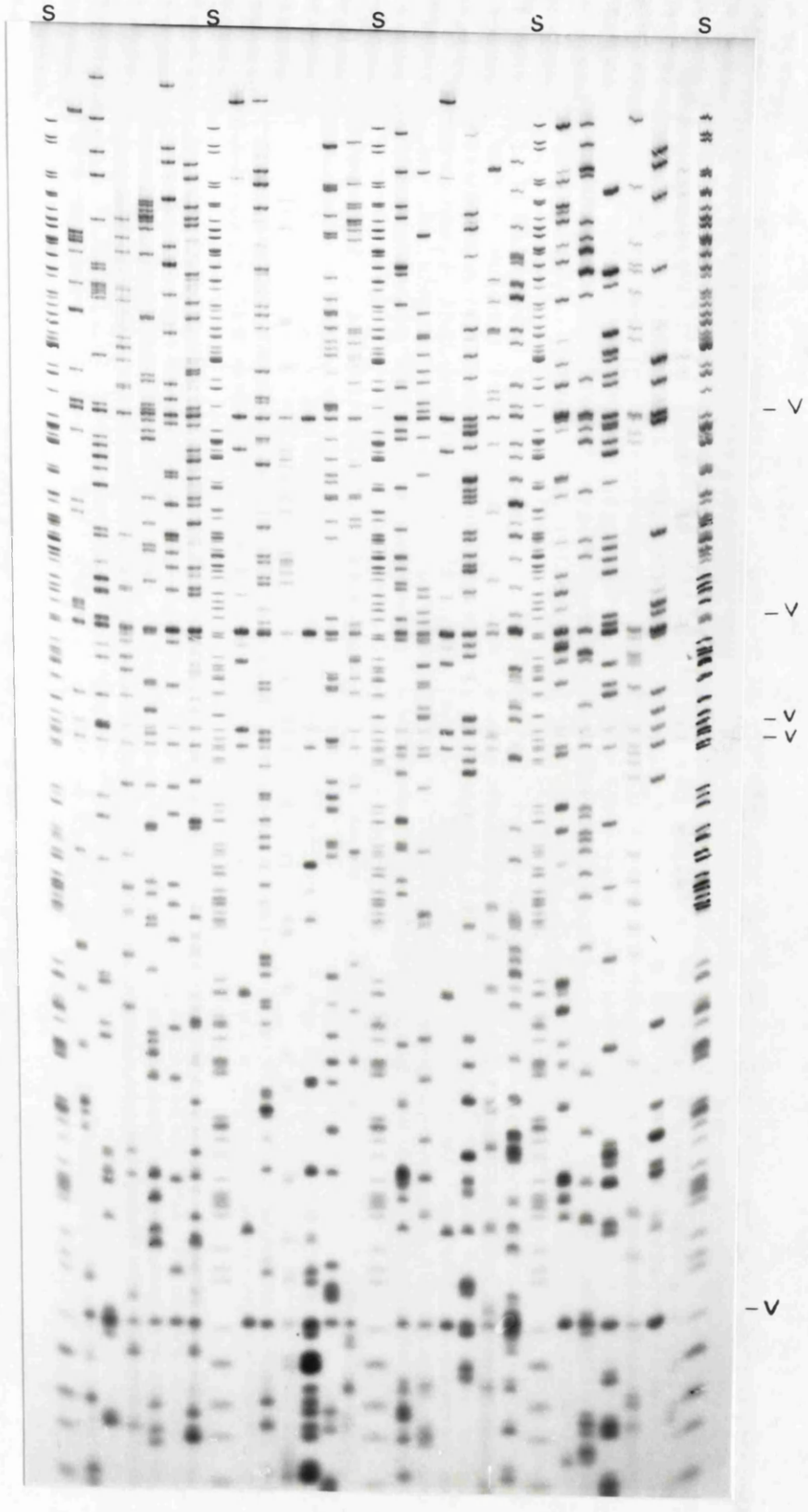


Figure 4.2 Example of a fingerprint gel: E946

Each lane contains a different, randomly selected, cosmid fingerprint. Standard lanes of end-filled *HinfI* digested lambda DNA are loaded every sixth lane (S). The bands derived from the cosmid vector are present in all lanes and are marked V.

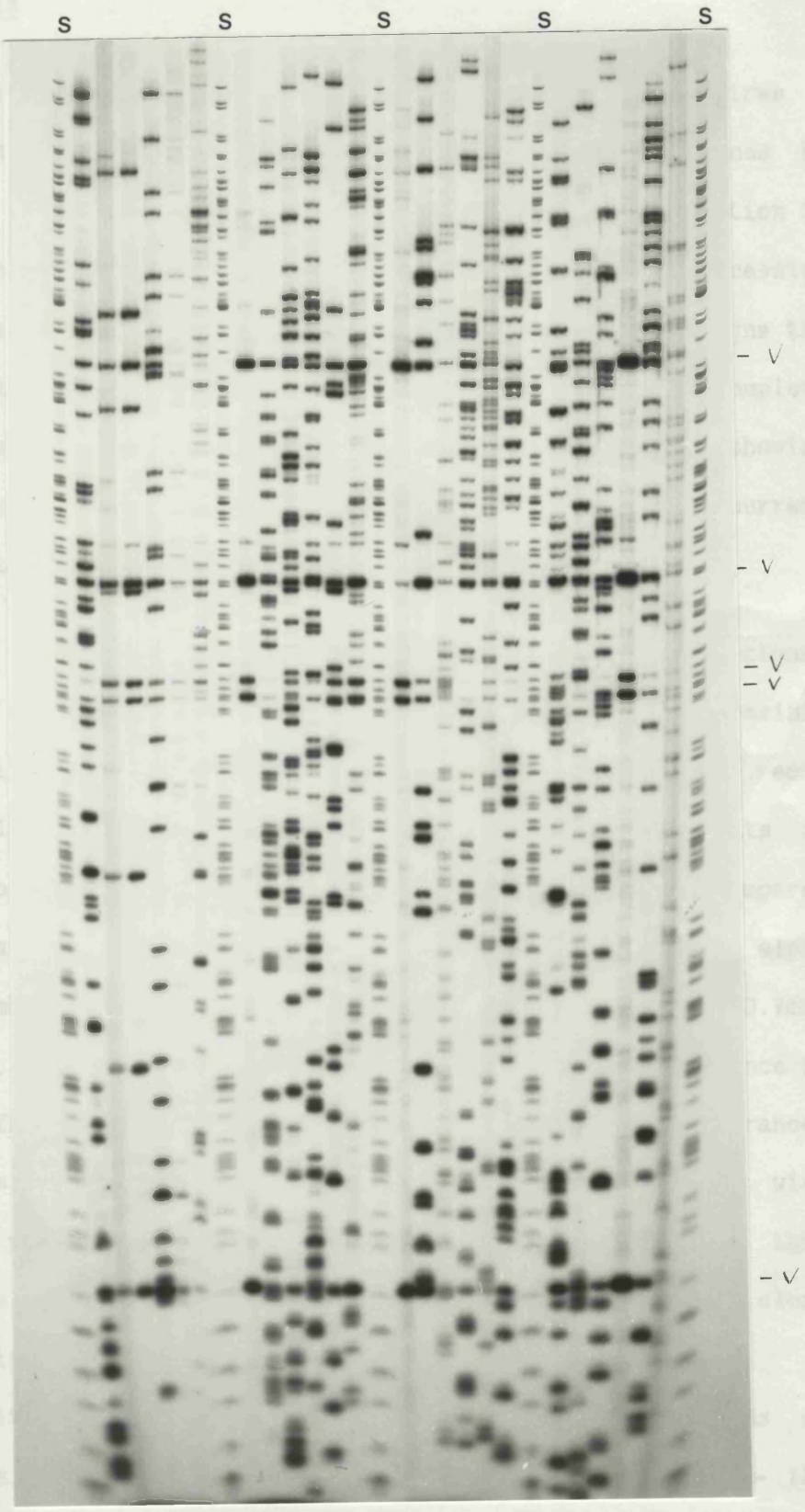


Figure 4.3 Example of a fingerprint gel: E949

Each lane contains a different, randomly selected, cosmid fingerprint. Standard lanes of end-filled *HinfI* digested lambda DNA are loaded every sixth lane (S). The bands derived from the cosmid vector are present in all lanes and are marked V.

Data analysis

The analysis of the data is semi-automatic but actually requires a large amount of manual interaction. The autoradiogram is scanned by an Amersham gel reader which digitises the image. The information is then transferred to a VAX 3100 workstation where an image processing package detects the bands, arranges them into lanes, and aligns the standards with marker bands (Sulston *et al.* 1988, 1989). The complete autoradiograph can be visualised on the workstation screen, showing the standard bands positioned over those of the current autoradiograph.

In order to enter data, the programs identify each band in a clone. Since marks on the film may be interpreted as bands, it is possible to edit out these, and vector fragments at this stage. For each clone a file is created which records the position of fragments in relation to the standard bands. This information is then compared with the data for all other clones in the database. For each clone pair the number of bands that agree within a preset tolerance (0.7mm) is recorded, and the probability of this event occurring by chance is calculated from the number of bands in each clone and the tolerance. All matches are ranked in terms of this probability and those with the lowest probability values are printed out. The nature of this printout is shown in figure 4.4 which shows the assignment of clone 797 to contig 39.

These results are used by further programs to construct diagrams of the contigs in terms of overlapping lines which correspond to the number of bands in each clone and the extent of the overlap. However, the computer often receives distorted information about the clones.

						A	B	C
960-797	(35b, 0)							
		15 matches	912-62	(34b,	39)	0.2E-04	0	20d
		15 matches	915-558	(36b,	39)	0.4E-04	0	12
		15 matches	9141-603	(57b,	0)	0.1E-02	0	12
		13 matches	939-253	(45b,	16)	0.4E-02	0	8
		13 matches	949-437	(46b,	0)	0.5E-02	0	8
	difmap:	9 matches	952-502	(41b,	0)	0.1E+00	0	12
	difmap:	10 matches	9531-648	(43b,	0)	0.6E-01	0	11
	difmap:	12 matches	9473-399	(44b,	0)	0.1E-01	0	11
945-334	(53b, 0)							
		12 matches	9521-505	(25b,	0)	0.2E-02	0	41d
		12 matches	9531-661	(28b,	0)	0.5E-02	0	11
		9 matches	956-679	(19b,	0)	0.7E-02	0	9
		13 matches	948-415	(33b,	0)	0.9E-02	0	12
		8 matches	923-137	(17b,	10)	0.1E-01	0	8
	difmap:	10 matches	903-535	(35b,	0)	0.2E+00	0	15
	difmap:	14 matches	9171-772	(46b,	0)	0.7E-01	0	18
	difmap:	7 matches	909-48	(18b,	0)	0.5E-01	0	9
	difmap:	8 matches	936-191	(21b,	0)	0.5E-01	0	10
945-335	(42b, 0)							
		23 matches	923-139	(41b,	24)	0.8E-08	0	19d
		21 matches	946-363	(44b,	24)	0.1E-05	0	15
		12 matches	9062-587	(28b,	0)	0.3E-03	0	7
		13 matches	940-286	(34b,	55)	0.2E-02	0	5
		15 matches	9531-643	(43b,	0)	0.2E-02	0	10
	difmap:	9 matches	948-426	(39b,	0)	0.2E+00	0	11
	difmap:	12 matches	951-492	(47b,	0)	0.7E-01	0	11
	difmap:	12 matches	9112-96	(42b,	38)	0.3E-01	20	10

Figure 4.4 Computer printout of overlap probabilities.

Each clone is described by gel number and clone number, and the number of bands is given along with the contig (if any) that the clone has been assigned to. The number of matches to other clones is displayed in descending order of significance, with each clone described in the same format. Columns A, B and C describe A) the probability of the overlap being coincidental, B) the shorter distance from the end of the matching clone to the end of its contig (in bands) and C) the number of bands not found in top matching clone (d) or number of these bands found in subsequent matching clones.

Clone 797 matches the two clones in contig 39 and is thus assigned to this contig.

This misleading information is due to imperfections in the autoradiographs.

In particular, at the threshold values used, the image analysis seems to fail to recognise faint bands. The printout of overlap probabilities can be used as a basis for manual assembly. In this case, predicted overlaps are examined more closely if the probability that the overlap is due to chance is less than 10^{-3}. For identical clones the probability may be as low as 10^{-26} and is usually in the range of 10^{-3} to 10^{-10} for conclusively overlapping clones. The fragments of each clone are compared on the original autoradiograph. It is helpful to mark corresponding bands with marker pen. If a third clone is compared and matches marked in different colours, then bands in common between all possible combinations of clones can be counted and used to deduce the relationship between the clones. A diagram showing the matching of four clones to each other is presented in figure 4.5. These clones form contig number 32.

This manual checking of results generated by the analysis programs has eliminated many false contigs which were produced by adjacent identical clones, has altered the proposed patterns of overlap, and has led to the recognition of those clones that contain repeated elements. The data has to be of a high standard before the computer routines will even begin to analyse it, but until it is possible to generate perfect data, with no missnapen bands, no partial digests, no extraneous marks and with every clone being completely independent, it will be necessary to continue with the manual analysis.

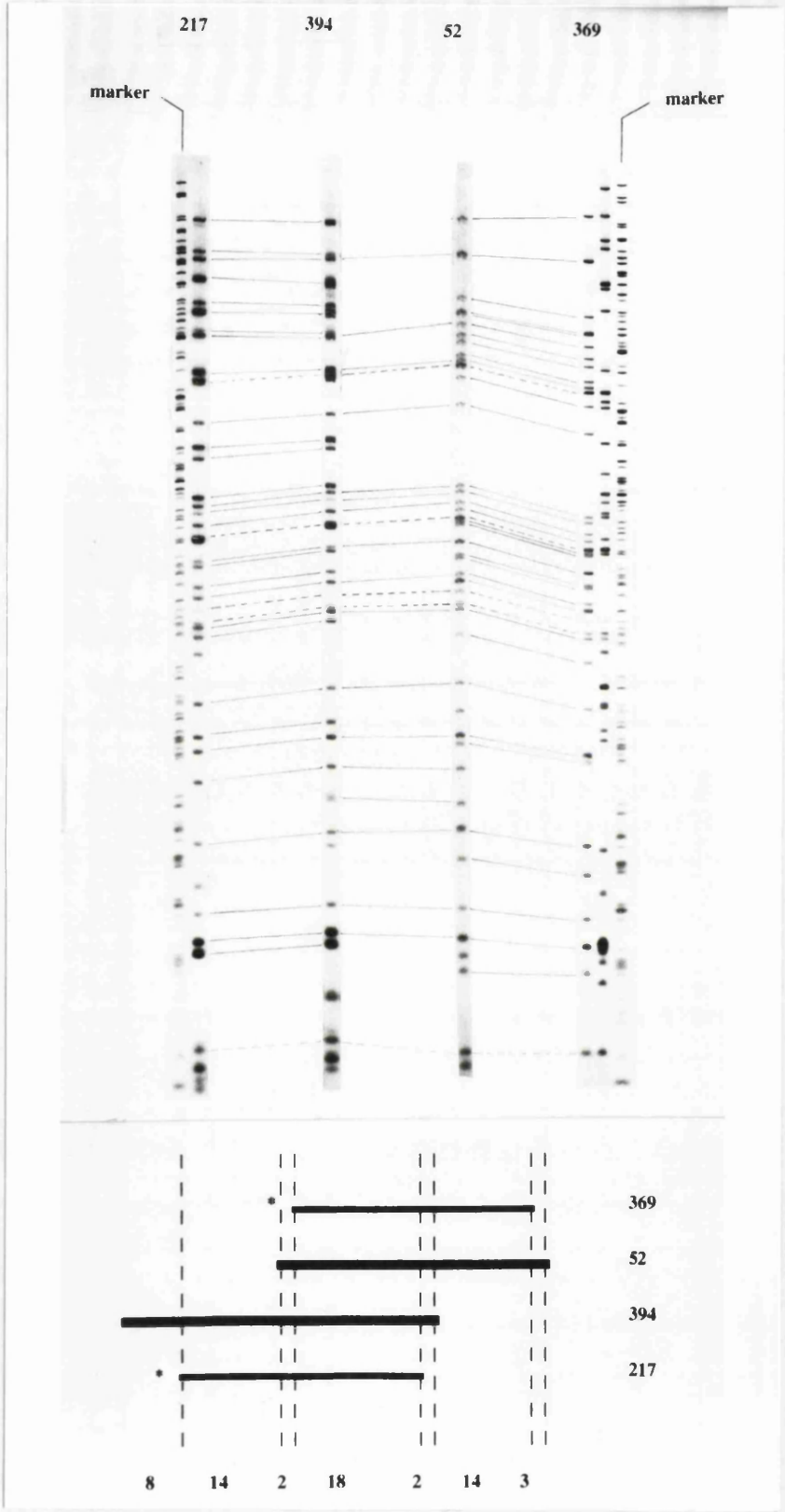


Figure 4.5 Manual comparison of clones in contig number 32

4.3 Descriptions of contigs

I will describe the nature of each contig produced in terms of the number of cosmids included, the extent of overlap between clones, and any corroborating or conflicting data produced by the hybridisation experiments.

NOTE: a) Each cosmid has two types of identification; the 'Lor' title refers to the specific library plating and the clone number within that plating. e.g. Lor 2.15 169. Each clone also has a separate number which refers to its identification in the database. To aid the retrieval of fingerprint data for manual checking, a gel number with the prefix 'e' is also given. e.g. Lor 2.15 169, database number 673, gel number e925.

b) The extent of overlap can only be expressed in terms of the number of bands shared between two clones as a fraction of the total number recorded. To suggest a length of contigs requires the assumption that all *Hin*fI sites are equally spaced given by 40kb/number of bands. In the diagrams of contigs, the figures below refer to the number of bands in each overlapping region.

Contig numbers have been generated by automatic analysis and some were eliminated on manual checking (tney sometimes consisted of two identical, adjacent clones), hence some numbers no longer exist (e.g. contig 2).

The contigs are drawn to the same scale so that the extents of genome coverage may be compared.

Contig 1

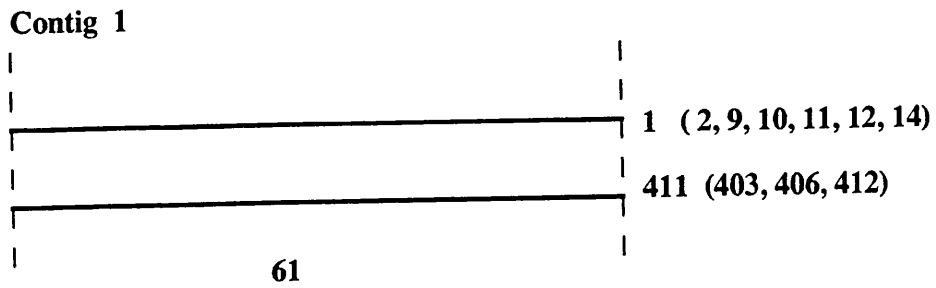
This contig consists of 11 clones. However, most of these are adjacent to one another on the gel, and are in sequence in terms of 'picking'. The fingerprints are identical and probably represent the same clone which has either been picked repeatedly or has contaminated other clones in subsequent manipulations.

Lor 2.15	169,	1	e925	
"	170,	2	"	
"	177,	9	"	
"	178,	10	"	
"	179,	11	"	
"	180,	12	"	[GMGY26]
"	182,	14	"	

The hybridisation of clone 12 to GMGY26 is spurious as none of the other identical clones show any hybridisation to this sequence. These clones cannot be counted as a contig. However, four further clones have also been identified as being very similar to this group;

Lor 1.1	42,	403	e948
"	38,	406	"
"	33,	411	"
"	32,	412	"

These clones are also adjacent and identical and represent a single isolate. Upon manual inspection of the autoradiograph, it is apparent that the second set of clones are identical to the first. However, as the canonical clone of this set was isolated and fingerprinted completely independently of the first, then these two clones do represent a contig which does not extend in either direction.

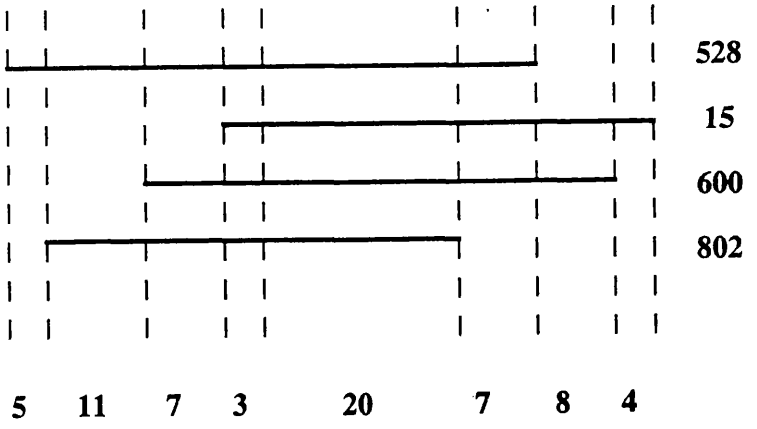


Contig 3

This is a true contig of four clones which has had clones added to it as more data was analysed. However, although the depth of coverage increased from three to four clones, the length has not increased as clones 600 and 802 are completely buried within 528 and 15. These two clones overlap by 46% of the combined number of bands and may cover 55-60kb of genomic DNA.

```
Lor 2.9 29,    528  e903
Lor 2.9 5,     600  e941 [same as 2.8 77, 616]
Lor 2.15 183,  15   e925 [same as 2.16 1, 16]
Lor 2.26 57,   802  e961
```

Contig 3



Contig 7

Contig 7 contains 12 clones. The fingerprint data for these clones is shown in figure 4.6 (figure prepared by V. Shortle). This figure excludes the clones 9,11 and 12 which are very similar to one another but share only 5 bands with the rest of the group. Clone 7 is highly deleted. This clone has 5 bands in common with 9,11 and 12 and the remaining 5 are in common with all other clones. It is possible that these fragments represent a common element which may be DYZ3 on the basis of hybridisation results.

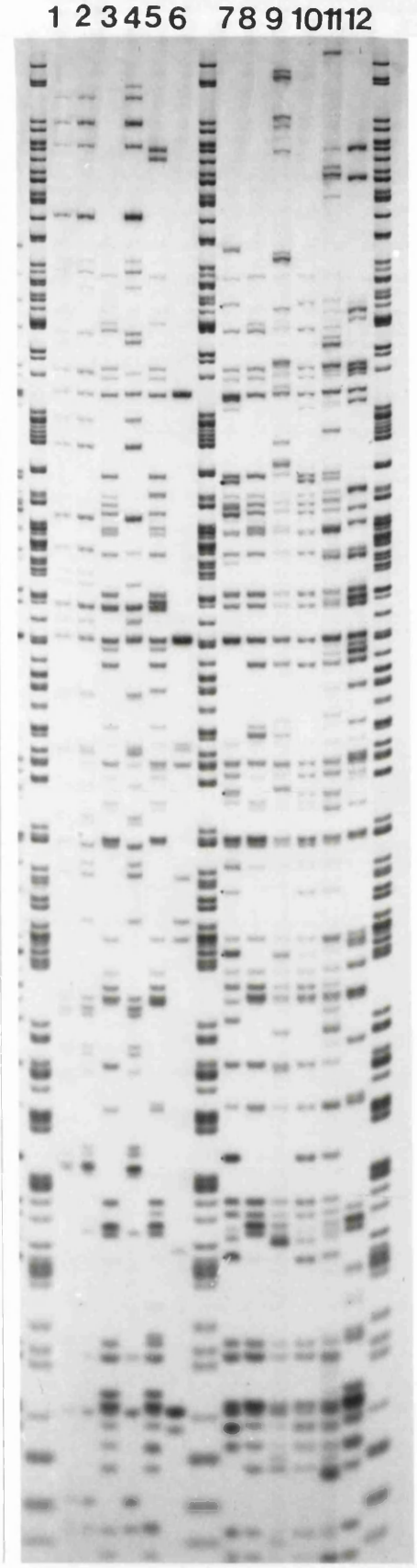
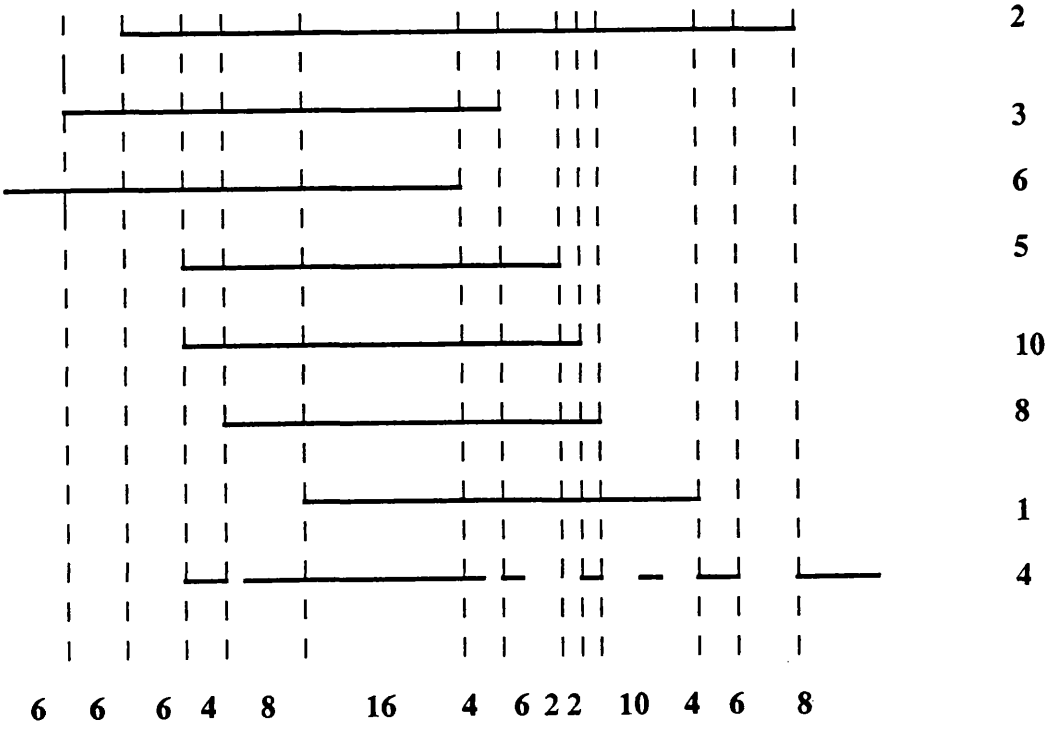


Figure 4.6 Fingerprint of twelve clones from contig 7.

It is impossible to construct a coherent formula for the overlaps between these clones, the degree of relatedness is shown below in the form of a contig diagram. The clones hybridise to a variety of DNA probes including DYZ3 and, predominantly, GMGY26. This type of data could be produced by a sequence which is repeated several times with a small amount of variation at each locus and probably consists of a number of smaller contigs. For example, clones 2,3 and 6 contain 5 fragments in common with each other but not any other members of the group. Unfortunately this observation does not seem to correlate with the hybridisation data.

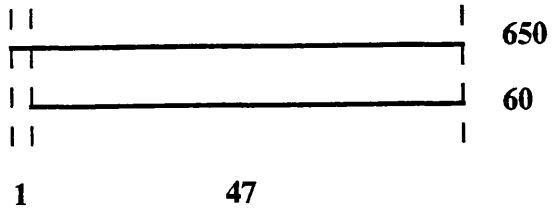
1) Lor 2.8 70, 753 [GMGXY8, GMGY26]
2) Lor 2.14 72, 90
3) Lor 2.14 86, 639 [GMGXY8, GMGY26]
4) Lor 2.15 172, 4 [$(CA)_9$, GMGY26]
5) Lor 2.18 70, 57 [DYZ3]
6) Lor 2.19 2, 55 [DYZ3, GMGY26]
7) Lor 2.19 8, 49 [DYZ1, DYZ3]
8) Lor 2.19 84, 205 [GMGY26]
9) Lor 2.21 55, 139
10) Lor 2.21 63, 138 [$(CA)_9$, GMGY26]
11) Lor 2.23 73, 335 [GMGXY8]
12) Lor 2.23 91, 363



Contig 8

This contig consists of two adjacent, identical clones and a highly similar independent clone.

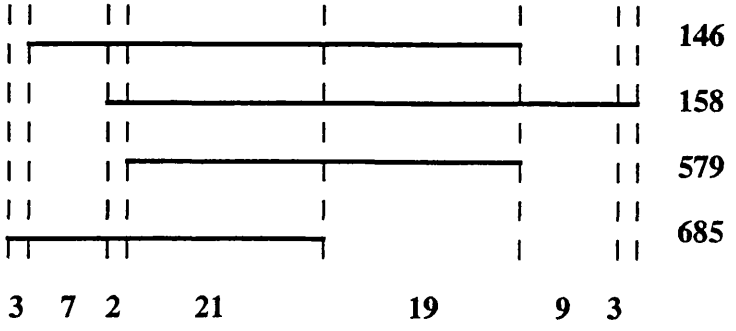
Lor 2.13 27, 60 e912
Lor 2.13 28, 61 e912
Lor 2.27 17 650 e9531



Contig 11

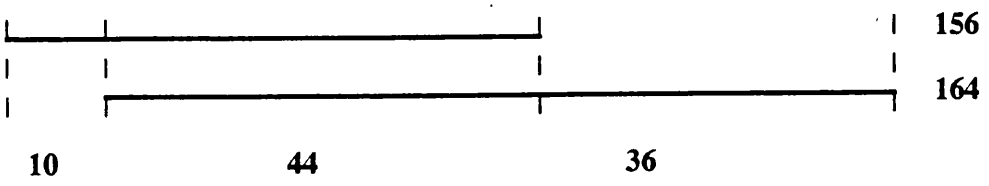
Of the four clones in this contig, two hybridise to the probe O19P3 which recognises the *TSPY* locus. It is possible that these clones represent one of the *TSPY* loci, although it is surprising that clone 579, buried within 146 and 158, did not hybridise to this probe under the original conditions. It is necessary to confirm this hybridisation result and that of GMGY26 with reference to these three clones.

Lor 2.21 33, 146 e923 [O19P3]
Lor 2.19 33, 158 e935 [O19p3, GMGY26]
Lor 2.15 45, 579 e9062
Lor 2.31 32, 685 e956



Contig 12

Contig 12 contains two cosmids which both show hybridisation to O19P3. However, the fingerprints are not related to those of the clones in contig 11. One of the cosmids (Lor 2.19 27, 156) has been localised by *in situ* hybridisation to chromosome 1 and hybridises to many other probes. Cosmid 156 is located 8 digits away from 164 and is therefore adjacent in the microtitre dish. This has been noted as a source of cross-contamination between clones and it is possible that 156 which appears to be a partial digest on the fingerprint may be a mixture of the two clones. Alternatively, but less likely, there could be a genuine *TSPY*-like locus on chromosome 1p35.

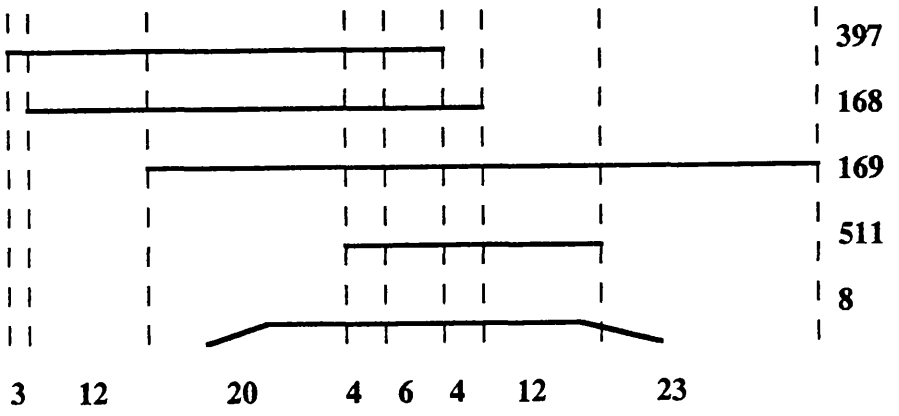


Contig 13

This contig consists of 5 clones. Although the clones 168 and 169 are adjacent, they are genuinely different but overlapping clones. The first four clones may be positioned into a contig but clone 8 cannot be included. It has 23 bands in common with 397 and 21 in common with 511, the remaining 18 bands cannot be fitted into the group which may also represent more than one locus.

Lor 1.1 2,	397	e9473 [GMGY39]
Lor 2.19 23,	168	e935
Lor 2.19 22,	169	e935
Lor 2.26 96,	511	e9521
Lor 2.15 176,	8	e925

Contig 13



Contig 14

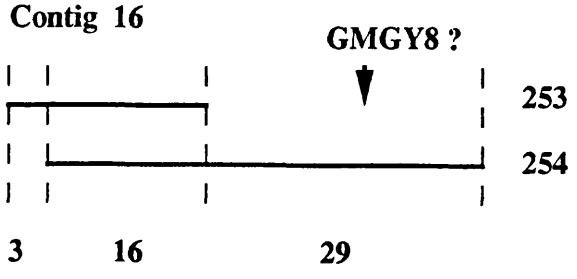
Contig 14 is another example of two cosmids that are independent isolates of the same region of DNA.

```
Lor 2.19 1,    211  e937
Lor 2.19 78,   56   e922
```

Contig 16

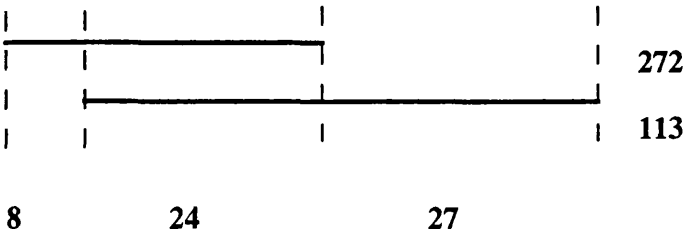
In this case two clones happened to be adjacent on the acrylamide gel but were not picked in succession, the overlap does not appear to be due to spillage between lanes of the gel.

```
Lor 2.18 1,      253  e939
Lor 2.16 92,     254  e939 [GMGY8]
```



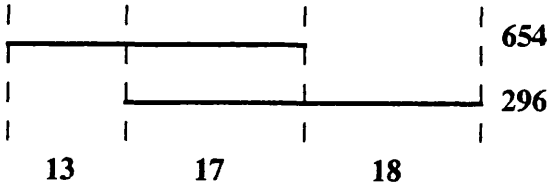
Contig 17

Lor 2.21 64, 113 e9211
Lor 2.15 162, 272 e940



Contig 18

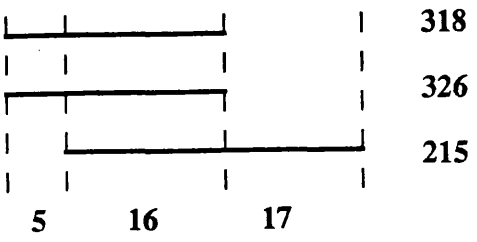
Lor 2.27 21, 654 e9531
Lor 2.15 136, 296 e941



Contig 20

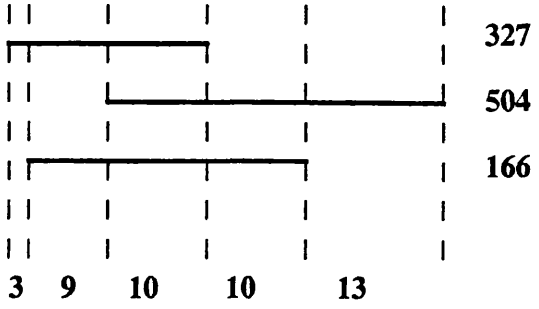
This contig includes two clones (318 and 326) that have the same fingerprint and are probably the result of contamination. However, a contig is formed by the inclusion of a further clone.

Lor 2.15 108, 318 e942
Lor 2.15 116, 326 e942
Lor 2.19 74, 215 e937



Contig 21

Lor 2.27 6, 504 e952
Lor 2.19 25, 166 e935
Lor 2.15 107, 327 e942



Contig 23

These two clones are independent isolates of an identical region.

Lor 2.85 151, 282 e940
Lor 2.23 97, 357 e946

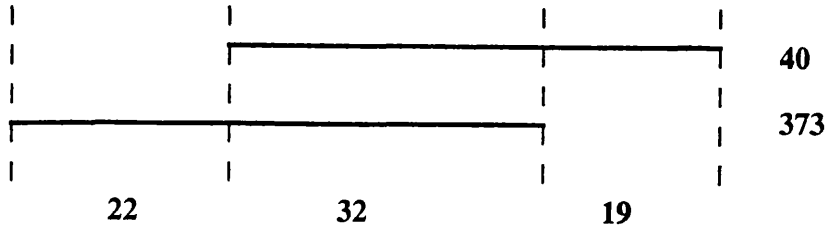
Contig 25

This contig has become incorporated with contig 32 (see contig 32).

Contig 26

Both of the clones in contig 26 hybridise to the probes 019P3, $(CA)_9$ and GMGY26. They have very similar fingerprints but do have unique bands. These cosmids do not share any significant homology with clones in contig 11 which also hybridise to 019P3.

Lor 2.16 34, 40 e909
Lor 2.23 81, 373 e946



Contig 30

Both of these clones hybridise to probe GMGY26. Clone 405 is completely contained within clone 98.

Lor 1.1 39,	405	e948	48 bands
Lor 2.14 64,	98	e9112	56 bands

Contig 31

It is not really correct to call this group of clones a contig. Twenty four clones are included but in fact they are all very deleted and share a characteristic pattern of very small fragments (see figure 4.7).

It is most likely that these represent cloned copies of the repeat element DYZ1.

Eight clones showed hybridisation to this probe in the initial screening, and a 3.5kb fragment is recognised on Southern blots of *Hae*III digested cosmid DNA using the probe pY3.4. (Data not shown). This is true for at least one cosmid that did not initially show hybridisation to the DYZ1 probe. The number of cosmids in this group represents 3.6% of the total number of fingerprints analysed. This is substantially less than the proportion (11%) of DYZ1-containing clones thought to exist in the library on the basis of hybridisation experiments. One explaination is that these clones are so deleted that they often appear to contain only vector bands and so may not have been entered into the database. The deleted nature of the clones is consistent with the accepted theory that tandemly repeated elements undergo substantial internal recombination events.

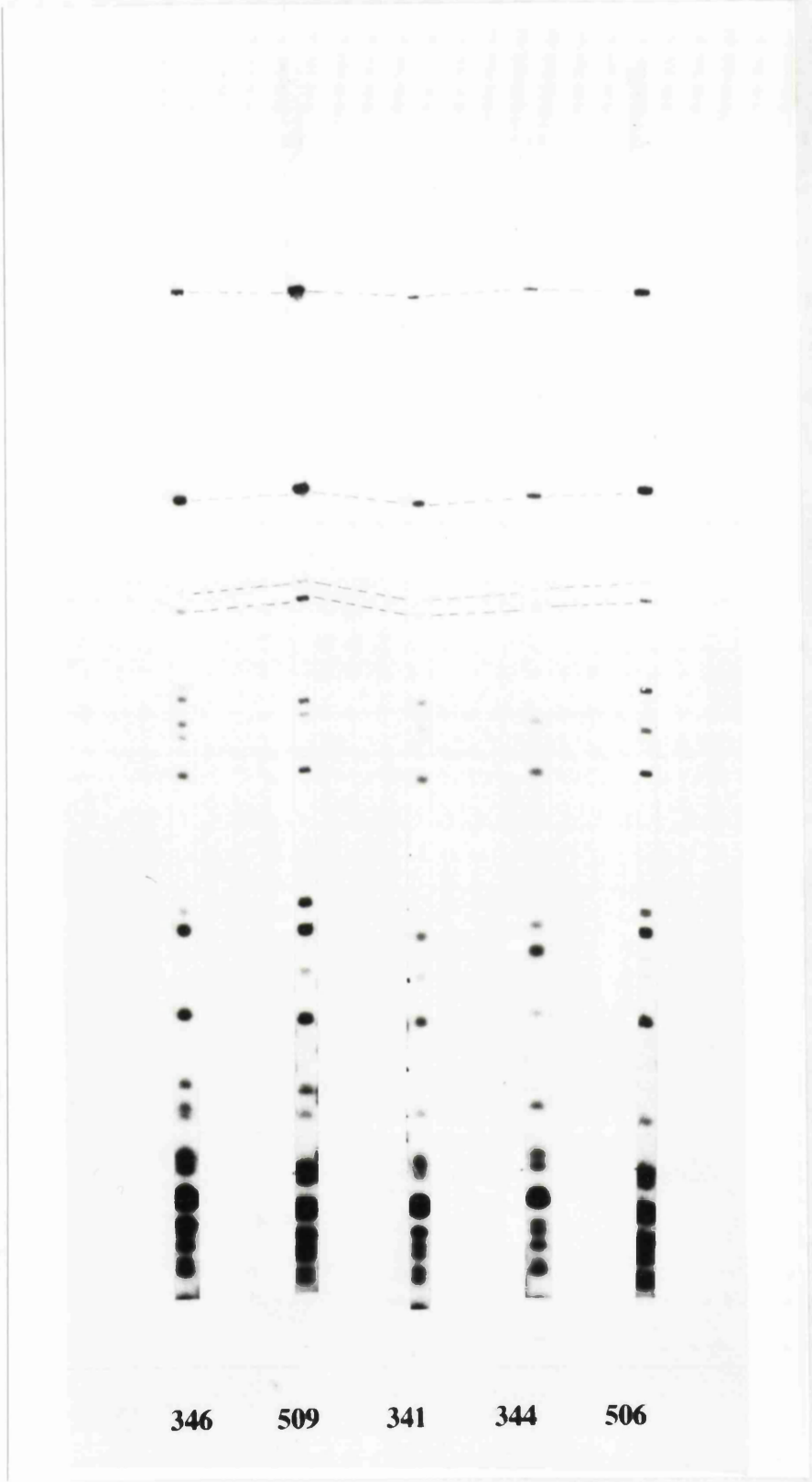


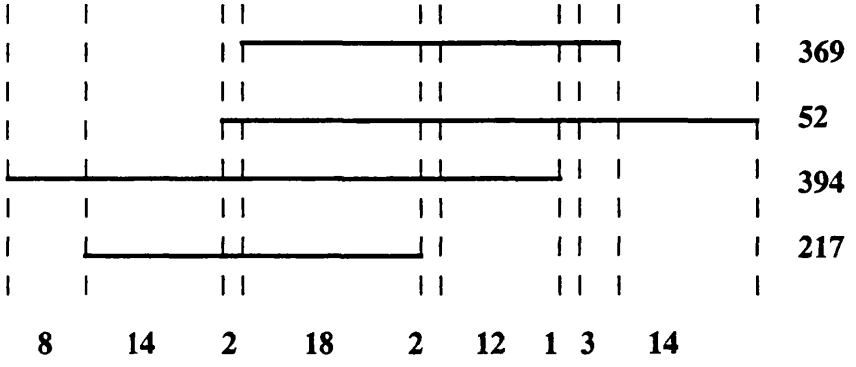
Figure 4.7 Fingerprints of clones containing the repeated element DYZ1.

Contig 32

Contig 32 includes four cosmids and represents the joining of two separate two-clone contigs into a single overlapping group (see figure 4.5).

Lor 2.23 85,	369	e946
Lor 2.19 5,	52	e922
Lor 2.19 72,	217	e937
Lor 2.3 8,	394	e9473

The two clones 369 and 52 were identified by the computer analysis as having highly similar fingerprints, in fact, 369 has 33 bands and is completely contained within the 39 bands of clone 52. Also, clone 217 only has bands that are present in clone 394. All four clones overlap in the central region but the contig is only extended in either direction by the two clones 394 and 52.



Contig 33

The five clones of this group have very few bands. They share a common feature of an intense band at 61bp and a less intense fragment at 120bp which are indicative of a satellite element (see fig 4.8). The clones most likely represent the same locus as they each seem to have two or less variant bands which could be explained by insert-vector junction fragments. The pseudoautosomal minisatellite element

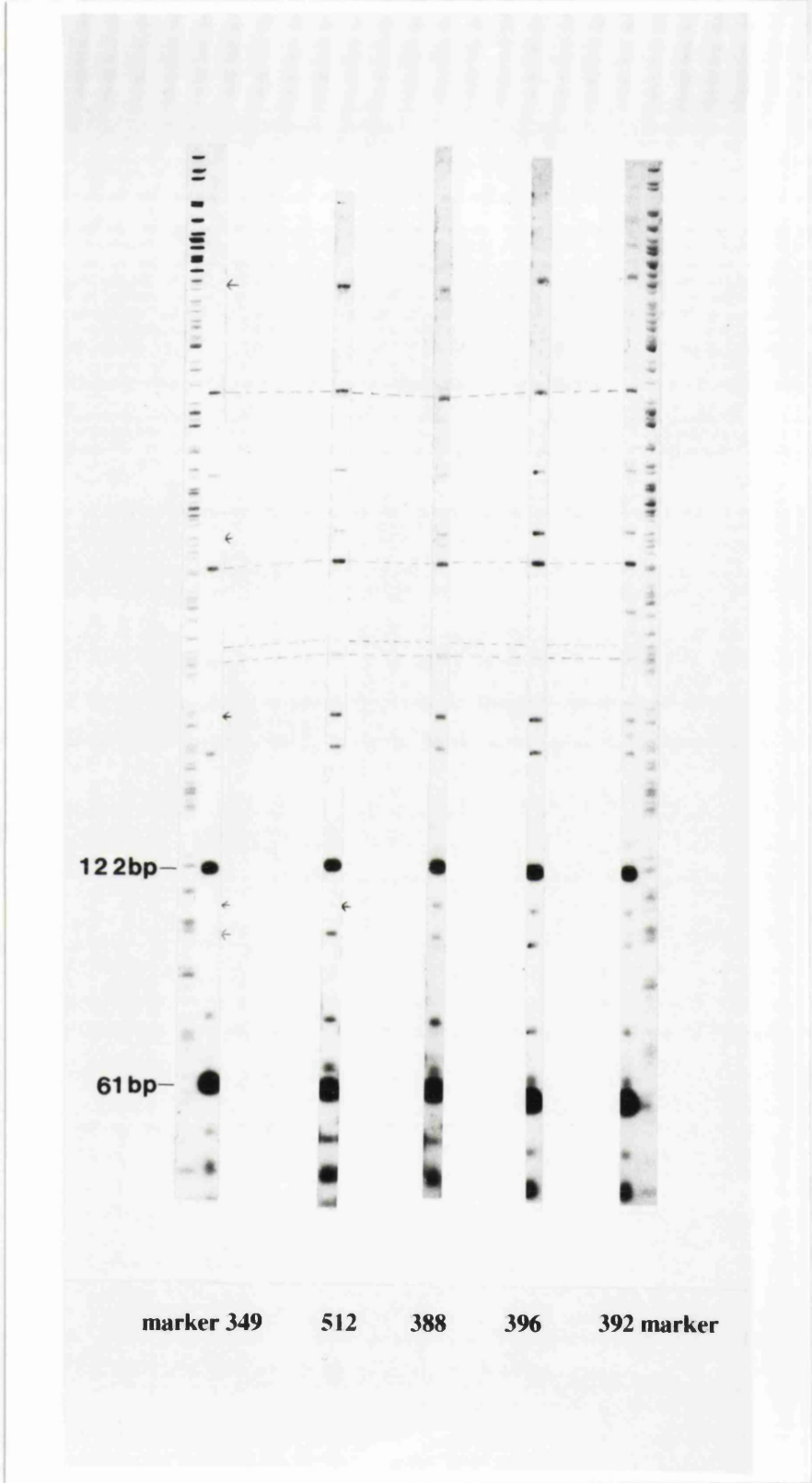


Figure 4.8 Fingerprints of clones containing a minisatellite element which may be DXYS20.

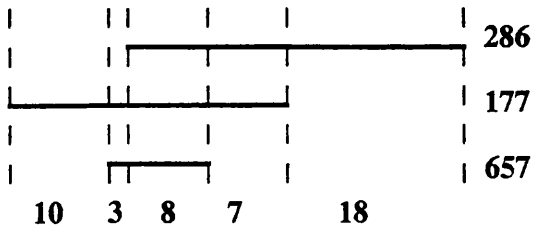
DXYS20 described by Page *et al.* (1987b) contains a *Hin*fI restriction site which would be labelled in this reaction (GACTC). Since the length of the DXYS20 repeat unit is 61bp, it appears that these clones are derived from this satellite region. This hypothesis could be confirmed by hybridisation experiments.

Four of the clones hybridise to the probe GMGY8. As stated previously, this probe may be contaminated with a repeated element, perhaps this is the minisatellite itself, or an element embedded within it. I am unaware of any evidence that GMGY8 and DXYS20 are derived from the same locus.

Contig 35

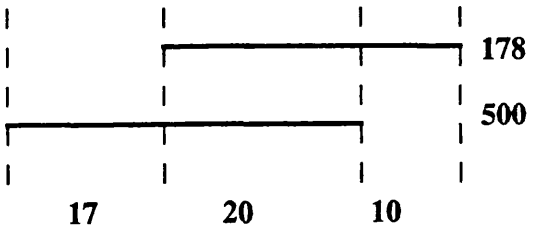
Lor 2.27 24, 657 e953 [p75/79]
Lor 2.19 64, 177 e936
Lor 2.15 147, 286 e940

Although clone 657 appears to contain sequences homologous to probe p75/79, the clones contained within it do not, and this hybridisation is probably spurious.



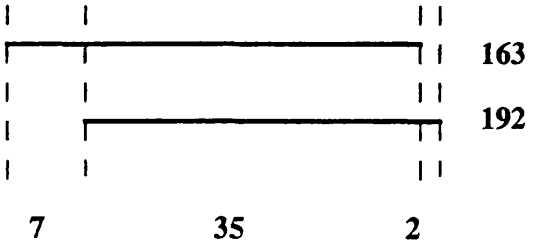
Contig 36

Lor 2.19 63, 178 e936
Lor 2.27 10, 500 e952



Contig 37

Lor 2.19 48, 192 e936
Lor 2.19 28, 163 e935



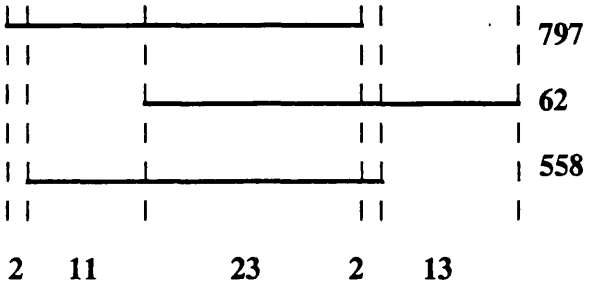
Contig 38

The eight clones in this group share a pattern of fragments but cannot be aligned into a contiguous sequence. It is probable that they represent a low copy repeated sequence but this has not been identified by any of the probes used in hybridisation experiments so far. The cosmids all contain a number of unique fragments and therefore do not result from the repeated cloning of one locus.

Contig 39

Lor 2.19 112, 558 e915
Lor 2.13 39, 62 e912
Lor 2.26 12, 797 e960

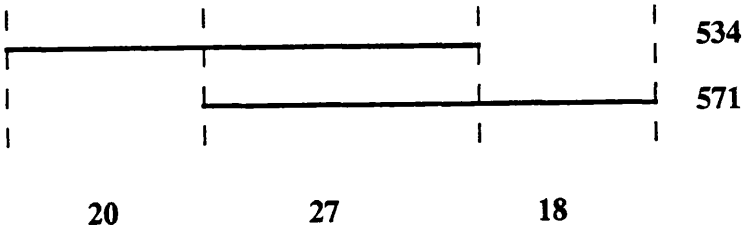
The printout of the probability analysis showing automatic assignment of clone 797 to this contig is shown in figure 4.4.



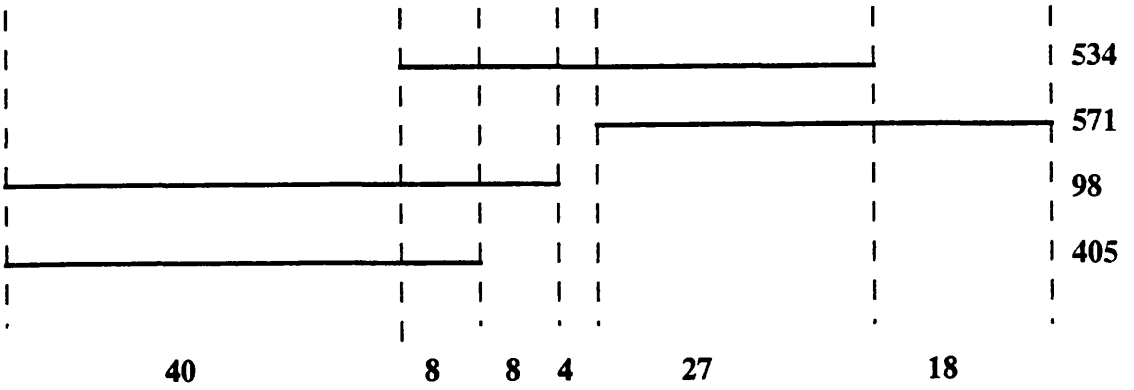
Contig 40

Lor 2.19 97, e915 571
Lor 2.9 21, e903 534

Both of these clones hybridise to probe GMGY26.



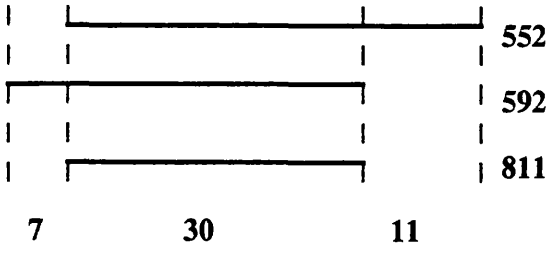
A similarity is also observed between clones 534 and 98. Clone 98 is already included in contig 30 and also hybridises to GMGY26. The probability that this overlap is due to chance is 10^{3}, and there are 16 fragments in common. A possible overlap between these two contigs may be represented thus;



No significant overlap is detected between 405 and 534, although 405 is entirely contained within 98, this places it at the extreme left of cosmid 98 in this diagram. The overlap between these two contigs would substantially extend the contig but needs to be confirmed by additional methods (see discussion). There is no significant similarity between clones in contigs 30 or 40 with those in contig 26 which hybridise to both pO19P3 and GMGY26.

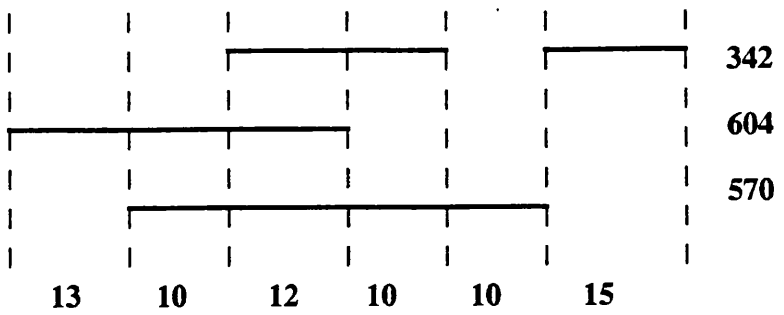
Contig 42

Lor 2.15 31, 592 e9062
Lor 2.21 22, 552 e919
Lor 2.26 47, 811 e961



Contig 43

Lor 2.9 1, 604 e9141
Lor 2.19 98, 570 e915



Clone 342 (2.23 63, e945) has high similarity to both 604 and 570 (with probability values of 10^{-5} and 10^{-6} respectively). However, it is difficult to rationalise all of the band positions within the contig. It is most likely that 342 is positioned as in the diagram above.

Contig 45

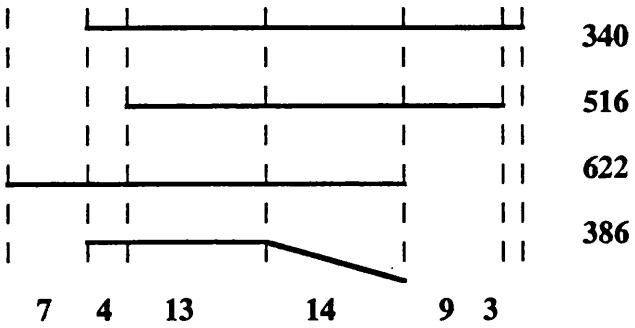
In this case, three cosmids have been picked simultaneously and are adjacent on the gel. Two (619 and 618, 2.8 74 and 2.8 75) are identical. They are similar, but not identical to the third clone 620 which contains, in addition to other bands, the pattern of fragments characteristic of the DYZ1 repeat element. It is possible that the third clone is a mixture of 619 or 618 with an independent cosmid containing the DYZ1 repeat element.

Contig 46

This contig also contains three clones which align well and a fourth which cannot be positioned accurately. Of course it is possible that such events may be due to coligation of human and mouse fragments in

the cloning process. One such event has been observed in a clone from this library (C. Tyler-Smith personal communication (Lor 2.19 108)).

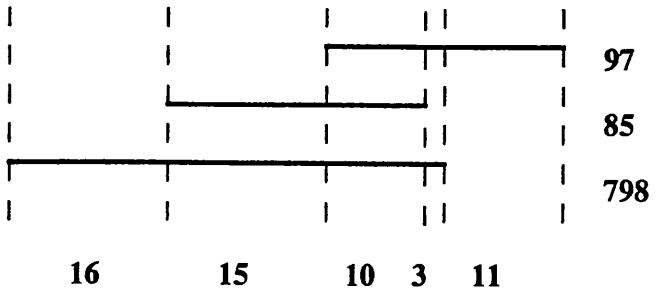
Lor 2.15 1, 622 e913
Lor 2.19 19, 172 e935
Lor 2.23 104, 386 e943
Lor 2.23 65, 340 e945



Contig A

Lor 2.26 61, 798 e961
Lor 2.14 65, 97 e9112
Lor 2.14 77, 85 e9112

Clone 799 (Lor 2.26 60, e961) is also included in this contig as it is identical to 2.26 61.

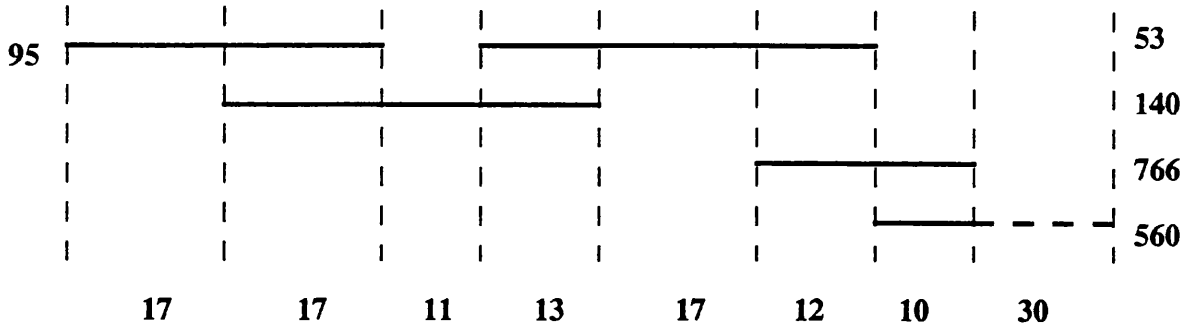


Contig B

Lor 2.14 67, 95 e9112
Lor 2.19 4, 53 e922 [GMGY26, $(CA)_6$]
Lor 2.21 49, 140 e923 [O19P3]

Clone 140 is recognised by the computer as having a substantial

overlap with both clones 95 and 53. Since no significant homology was detected between 95 and 53 alone, the relationship is probably like this;

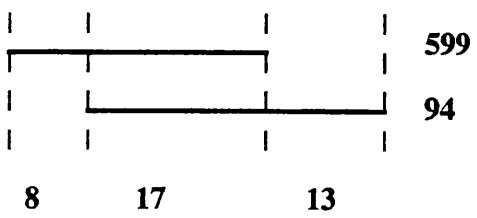


A further clone which may be included is 766 (Lor 2.8 57, e917) which only matches 53. Clone 766 also has 10 matches to clone 560 which doesn't have any matches to the other clones. (560 Lor 2.19 109, e915).

In this case I have used data generated by the computer to construct the contig which was too complex to construct manually. The structure of this contig is very different from those constructed manually and it is obvious that some oversimplification is inherent in the programs.

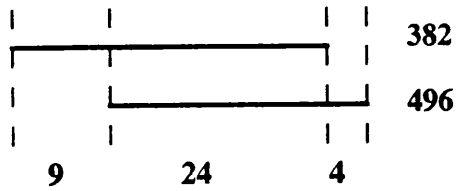
Contig C

Lor 2.9 6, 599 e914
Lor 2.14 68, 94 e9112



Contig D

Lor 2.23 101, 382 e9473
Lor 1.1 105, 496 e951



Contig E

The two clones in this contig are highly deleted with only 7 and 9 bands in addition to vector fragments. Of these there are 6 bands in common. The clones do not resemble the other deleted cosmids which contain the DYZ1 repeat and may have been deleted during growth.

Lor 2.26 77, 480 e951 [cY84]
Lor 1.1 79, 451 e950

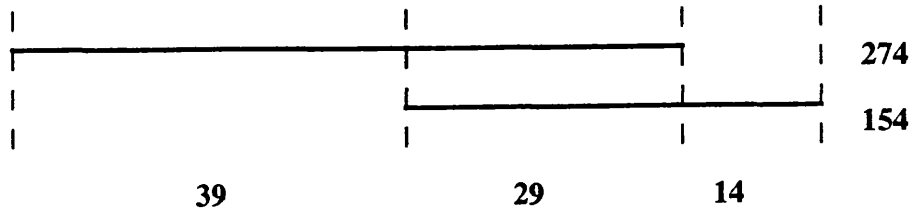
Contig F

This contig consists of two independent clones which have identical fingerprints. The identity of the clones is supported by the fact that they both hybridise to probe cY84 which detects the alphoid sequence DYZ3.

Lor 2.19 96, 572 e915 [cY84]
Lor 2.26 92, 515 e9521 [cY84]

Contig G

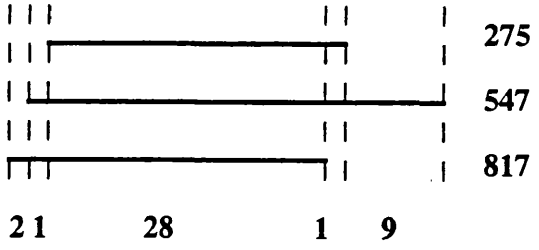
Lor 2.15 160, 274 e940
Lor 2.19 44, 154 e935



Contig I

Lor 2.15 159, 75 e940
Lor 2.21 29, 547 e919
Lor 2.26 14, 817 e961 [GMGY8]

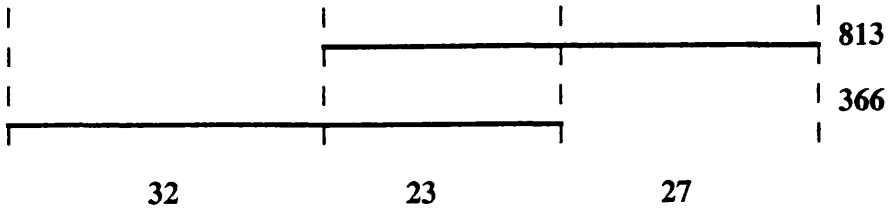
It does not seem appropriate that clone 817 has regions of homology to probe GMGY8 when the two clones to which it is so closely related do not.



Contig J

Lor 2.26 45, 813 e961 [p69/31]
Lor 2.23 88, 366 e946

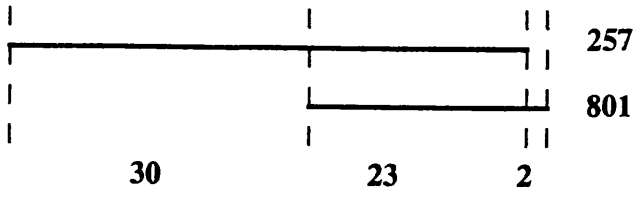
It is possible that the region of homology to probe 69/31 lies in the portion of cosmid 813 which is not in common with 366.



Contig K

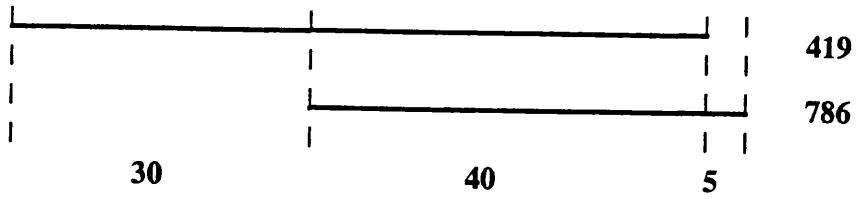
Lor 2.26 58, 801 e961
Lor 2.6 89, 257 e939

Clone 801 has two bands that are not contained within 257. It is probable that these are the boundaries between vector and insert DNA and so 801 may actually lie anywhere within the larger clone.



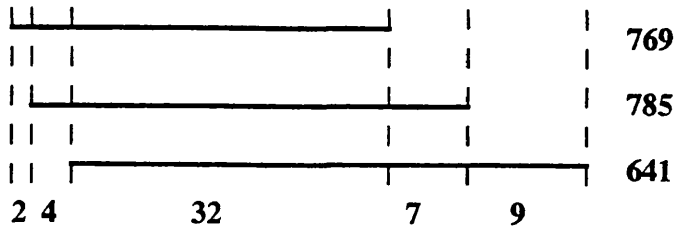
Contig L

Lor 2.26 23, 786 e960 [GMGY26]
Lor 1.1 21, 419 e948



Contig M

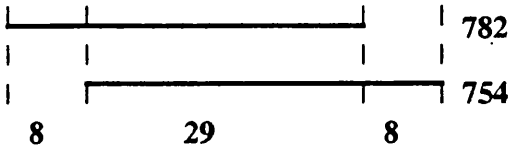
Lor 2.26 24, 785 e960
Lor 2.8 54, 769 e9171
Lor 2.14 81, 641 e913



Contig N

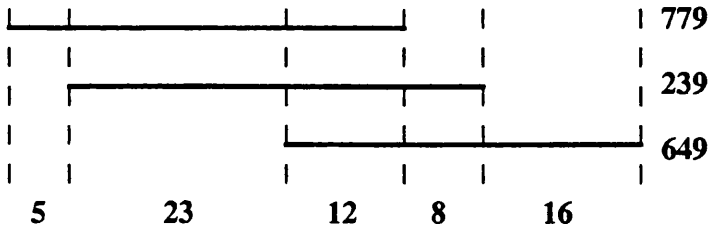
Lor 2.26 27, 782 e960 [O19P3, $(CA)_9$, cY84]
Lor 2.8 69, 754 e9171 [O19P3]

These clones seem to share a region with homology to one of the *TSPY* loci. The computer does not identify any similarity to other clones which include *TSPY* -homologous regions.



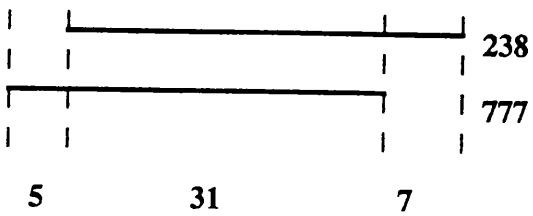
Contig O

Lor 2.26 31, 779 e960
Lor 2.15 78, 239 e938
Lor 2.27 16, 649 e9531



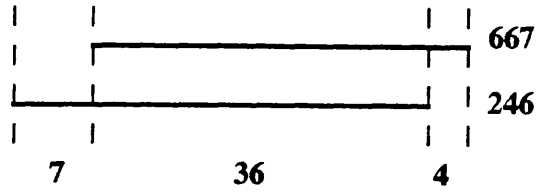
Contig P

Lor 2.26 33, 777 e960
Lor 2.15 79, 238 e938



Contig Q

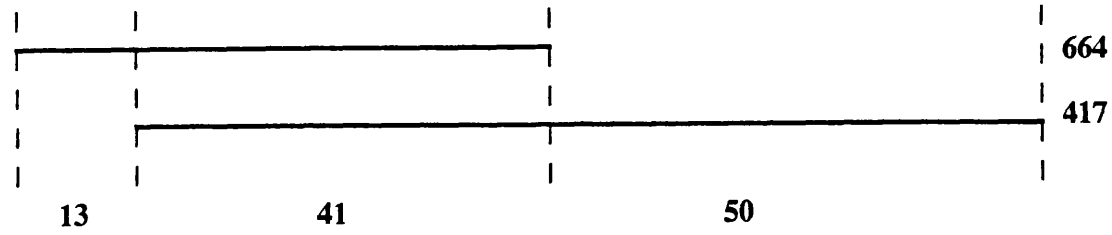
Lor 2.27 35, 667 e9531
Lor 2.19 91, 246 e938



Contig R

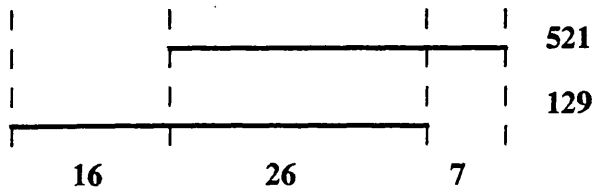
Lor 2.27 31, 664 e9531 [GMGY26]
Lor 1.1 25, 417 e948

Clone 417 has 47 bands in common with clone 419, which, although close on the gel, does seem to be independent. Clone 419 forms contig L with clone 786 which also hybridises to probe GMGY26. Clones 786 and 664 do have a number of bands in common, and it is possible that these two contigs may actually overlap.



Contig S

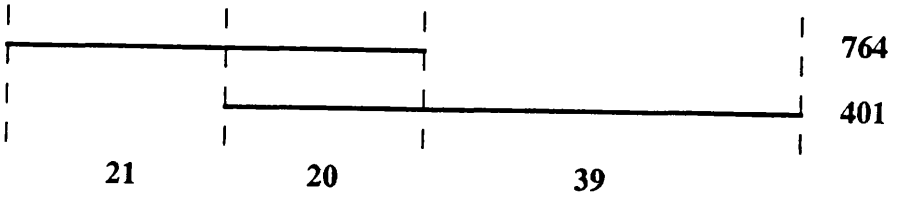
Lor 2.26 86, 521 e9521
Lor 2.14 9, 129 e902



Contig T

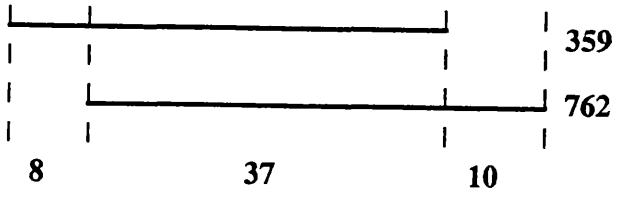
Lor 2.8 59, 764 e9171
Lor 1.1 6, 401 e9473 [GMGY8]

Although it is difficult to quantify the amount of overlap between these two clones manually due to differences in the running of the gels, there are definitely bands in common.



Contig U

Lor 2.23 95 359 e946
Lor 2.8 61, 762 e917

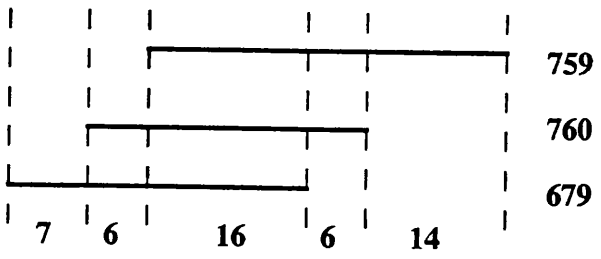


Contig V

Two adjacent clones, 759 and 760, share bands in common despite both having an average number of bands with some being unique to each clone. There are 22 common bands. Clone 760 also has 22 bands in common with a further clone 679 and 16 of these are shared with 759.

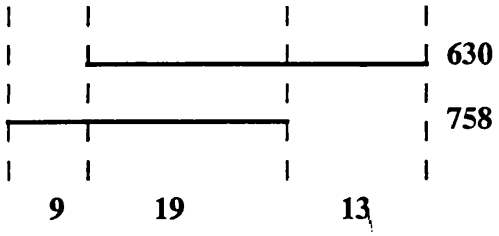
Lor 2.8 64, 759 e917
Lor 2.8 63, 760 e917
Lor 2.29 38 679 e956

Contig V



Contig W

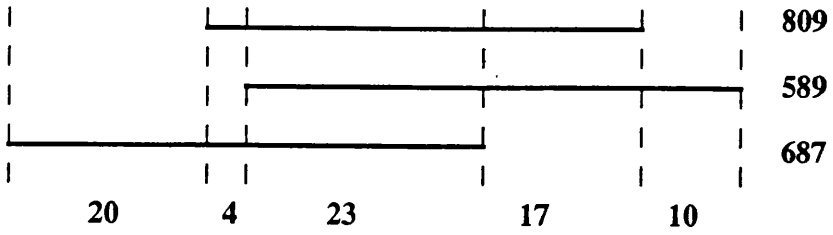
Lor 2.8 65, 603 e913 [O19P3, $(CA)_9$]
Lor 2.14 95, 758 e9171 [$(CA)_9$]



These clones may share a region containing a dinucleotide repeat element. There does not seem to be any significant similarity between these clones and the clones in contigs N or 26 which also hybridise to the O19P3 and $(CA)_9$ probes, either according to the overlap probabilities generated by the computer or by manual inspection of clones 630 and 40. These clones also both hybridise to the probe GMGY26 as do the clones of contig 26. It may be possible that the homologous region is too small to be detected by the fingerprint method but can be detected by hybridisation.

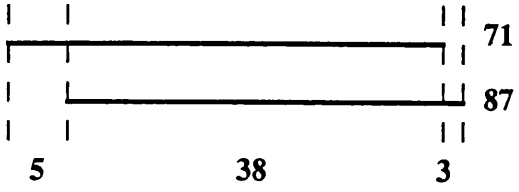
Contig X

Lor 2.26 49, 809 e961
Lor 2.15 34, 589 e9062
Lor 2.32 12, 687 e956



Contig Y

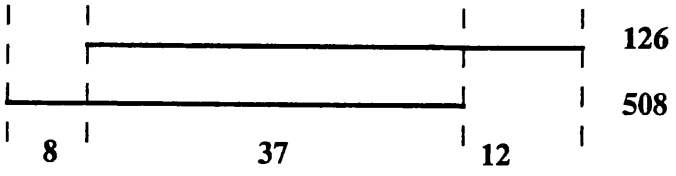
Lor 2.23 71, 426 e948
Lor 2.26 87, 520 e9521



Clone 520 has 16 matches to clone 558 which is located in contig 39. This is 39% of the total number of bands in 520. With this fraction of overlap I would not link the two contigs without confirming the homology by hybridisation (as discussed in the next chapter). However, the analysis program has detected an overlap between 558 and 426 with a probability of 10^{-2}, this confirms the likelihood of real similarities between the three clones.

Contig Z

Lor 2.14 12, 126 e902 [O10P3]
Lor 2.27 2, 508 e9521



The pMA3.3Y contig

The probe pMA3.3Y (Burke and Smith 1987, see section 3.2) detected 7 clones. These clones have been fingerprinted and their overlapping nature confirmed.

Other cosmids which one would expect to be overlapping, have been detected by means of the hybridisation experiments, for example, the clones detected by probes 69/31 and GMGY15. However, these clones have not yet been fingerprinted and do not feature in the contigs generated thus far.

4.4 Progress of the project

The goal of any project of this kind is obviously to produce a minimal set of overlapping clones which represent completely the chromosome or genome in question.

At the commencement of such a project comparison of cosmids to one another will yield almost no overlaps, but, as information in the database increases with the number of clones fingerprinted, contigs will be detected. The number and size of contigs produced increases with the number of clones up to a point where the contigs themselves begin to be joined together, causing a decrease in contig number. Eventually, the number of contigs is equivalent to the number of chromosomes in the genome.

Formulae for predicting the progress of projects of this nature have been developed by Lander and Waterman (1988). These formulae have been used to assess the progress of the Y chromosome project (see below).

A key factor in the rate at which contigs are detected is the amount of overlap between two clones, expressed as a fraction of clone length, which is required to detect that overlap. In the case of this project I have found that clones in contigs overlap by at least 50%. As the project proceeds it may be possible to confirm overlaps that are not so large by alternative methods, and this will increase the speed of contig assembly.

The predictions of Lander and Waterman

Assuming a perfectly representative genomic library with all inserts of equal size, the following symbols must be defined;

G - Haploid length of genome in base pairs 23×10^6

(based on Y chromosome as 1% of the human genome)

L - Length of clone insert in base pairs 4×10^4

N - Number of clones fingerprinted 1728 (total)

T - Amount of overlap in base pairs needed to detect overlap 2×10^4

$\theta = T/L$ 0.5

c - Redundancy of coverage $= LN/G$

$\alpha = 1 - \theta$

There are three levels of genome coverage i) Apparent islands which consist of one or more clones and are only apparent since some overlaps may go undetected, ii) Contigs consisting of two or more overlapping clones, and iii) oceans, which are gaps between islands. The following formulae can be used to assess the prevalence of each type of coverage and hence mapping progress;

A) The expected number of apparent islands is given by Ne^{-c}

B) The expected number of apparent islands consisting of j clones is $Ne^{-2c}(1-e^{-c})^{j-1}$.

C) The expected number of contigs (i.e. islands consisting of at least two clones is $Ne^{-c} - Ne^{-2c}$.

D) The expected number of clones in an apparent island is e^c

E) The expected length in base pairs of an apparent island is $L[((e^c - 1)/c) + (1-)]$.

Figure 4.9 shows the expected number of contigs with respect to the number of clones fingerprinted. According to the model, the number of contigs should begin to reduce after around 1,500 clones have been fingerprinted. This suggests that the number of clones available combined with this fingerprinting method are likely to be successful in representing the majority of the genome in terms of overlapping clones. Ideally, it would be appropriate to continue to fingerprint further clones up to a point where gaps are no longer being filled, when an alternative strategy should be assumed. From the same graph, the mean island length is expected to increase steadily as more clones are fingerprinted.

However, the actual number of contigs identified as the project has progressed has been substantially less than that predicted mathematically. After 200 clones have been fingerprinted, 12 rather than the predicted 27 contigs are observed, after 400 clones there are 26 rather than 83 contigs, and after 600 clones 43 rather than 145. These results are presented graphically in figure 4.10.

As would be expected from these results, since the number of contigs is reduced, so is the number of clones contained in contigs compared to the predicted values. The expected average number of clones per contig can be calculated by subtracting the number of islands containing two or more clones (contigs) from the number of apparent islands to give the number of singletons. This can then be subtracted from the number of clones fingerprinted to give the number of clones in contigs, which is divided by the number of contigs.

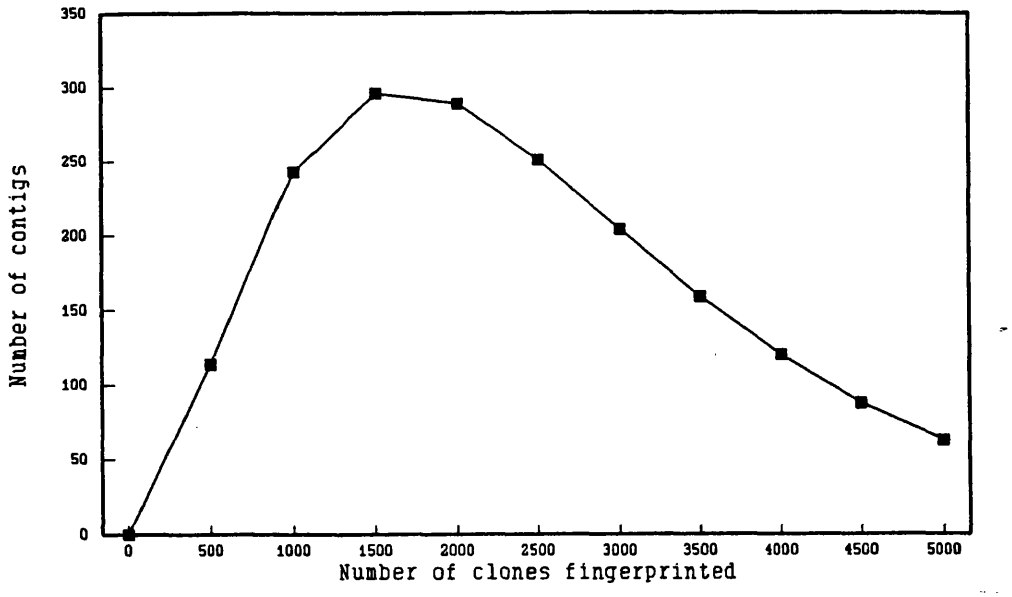


Figure 4.9 Expected progress of the mapping project using Lander and Waterman predictions. The number of contigs identified are plotted against the number of clones fingerprinted.

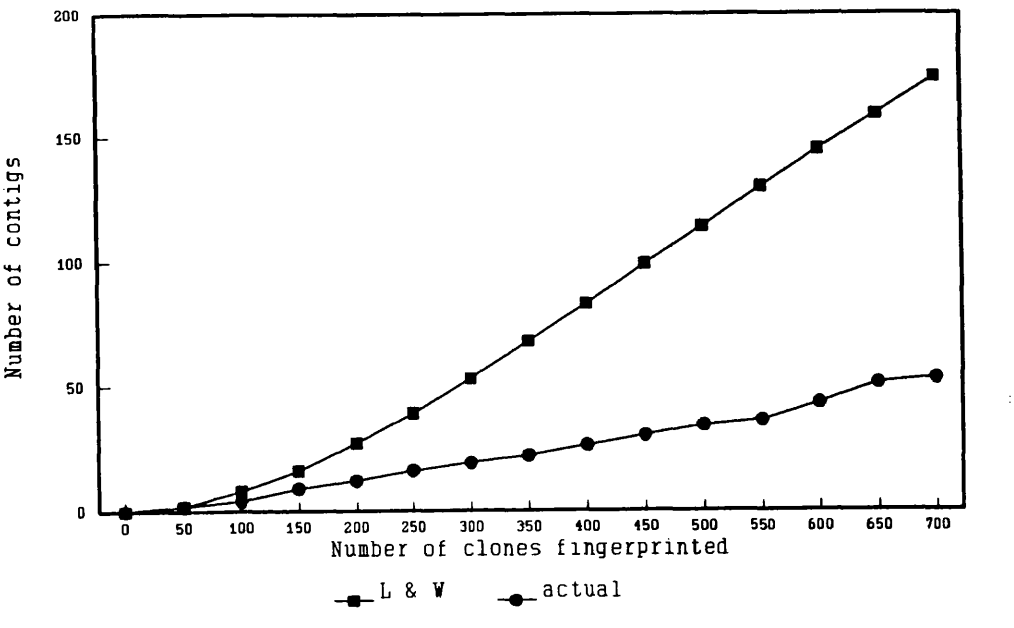


Figure 4.10 Actual progress of the project. The axes are as above.

After fingerprinting 723 clones;

	Expected	Actual
Number of contigs (not repeat groups)	180	51
Number of singletons	205	523
Clones in contigs	518	128
Average number of clones per contig	2.9	2.5

The expected number of contigs containing a certain number of clones can also be calculated.

Number of clones in contig.	Expected number of contigs.	Observed number of contigs
2	96	27
3	45	12
4	21	7
5	10	2

Obviously, the actual data gives a lower number of contigs but the amount of two-clone contigs compared to those containing 3,4 and 5 clones appears to be in proportion to the expected data.

4.5 Discussion

Projects of this kind often fall below predicted success rates. There are a number of factors which could be responsible for the reduced rate of progress in this case.

Most obviously, it is possible that the figure used for the genome size is an underestimate. In these calculations I have used the figure of 23Mb (1% of the human genome) for the size of the Y chromosome. Given the very recent data of Morton (1991) it may have been more appropriate to use a larger value (59Mb). This could certainly be a reason for the apparently very slow rate of contig production. However, as the heterochromatic region is not fully

represented in this library, and is therefore only partially included in this analysis, the size estimate used should probably lie somewhere between the two values of 59 and 23Mb. If the genome size were 59Mb, these formulae predict that 123 contigs should be observed after 723 clones have been fingerprinted, and that the number of contigs will only begin to reduce after around 5,000 clones have been fingerprinted. The observed number of contigs is still less than half of this predicted value.

The size of the cosmid inserts can range from 35-40kb and in some cases are highly deleted (particularly the repeat element-containing clones) thus 40kb is an optimum insert size and may be an overestimation. Such an overestimation is also likely to result in an artificially high predicted rate of progress.

Furthermore, it is likely that the library is not completely random, that is, some regions of the genome may be cloned disproportionately to others. This has been observed in the case of the *C. elegans* project in a 100 fold over-representation of ribosomal DNA. There is some evidence for non random cloning in the Y chromosome library in the under-representation of the repeat elements DYZ1 and DYZ2. The Y chromosome contigs seem to contain, on average, more clones than is expected given the small number of contigs. For example, if the genome size is 59Mb, then the 123 contigs are expected to contain an average of 2.3 clones per contig. The observed values of 51 contigs (excluding those consisting of repeated groups) with 2.5 clones per contig could be the result of non-random cloning, although the hybridisation data suggests a uniform coverage of the euchromatic regions.

Although the extent of detectable overlap between two clones is expected to be around 50% of their length, I have noticed that this is usually much greater. This suggests that overlapping clones are only being detected if the overlap is great. Indeed it is difficult to be convinced of an overlap unless two clones have a substantial number of bands in common. However, some clones with less substantial overlaps must be being overlooked. In some cases additional information such as an overlap with a third clone may provide verification of short overlaps, this has been the case in some of the contigs described above. As more clones are fingerprinted, this type of event will become more common, hopefully revealing a lot of hitherto undetected overlaps.

4.5.1 Additional methods for detecting overlaps

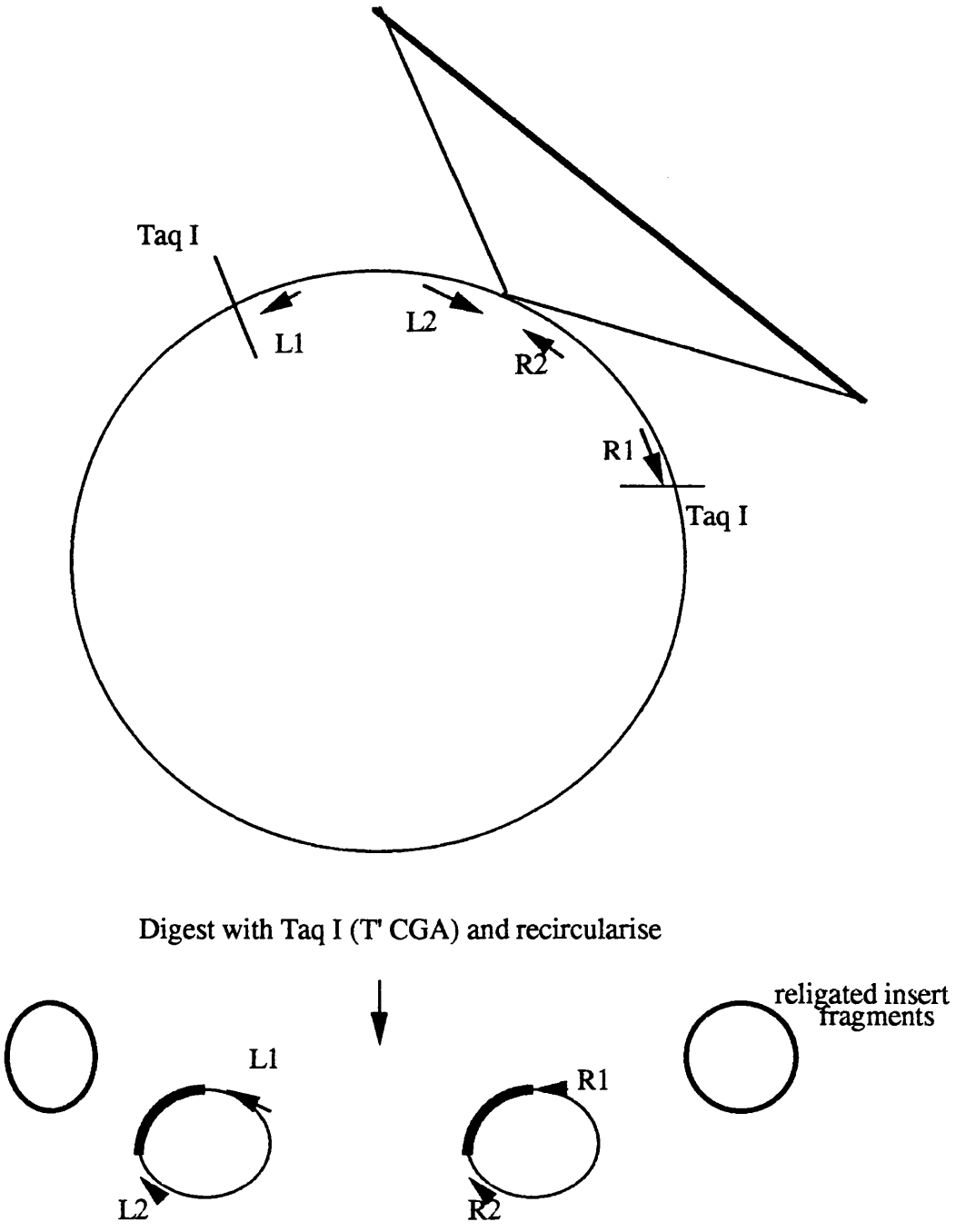
As a result of the problem described above, it is probable that alternative methods for detecting overlaps will be extremely valuable in this project. There are a variety of strategies that can be used to link islands into larger contigs. One way is to utilise existing clones but to reduce the percentage of homology that is required to detect an overlap.

This may be achieved by hybridisation of probes derived from the ends of contigs to the remainder of the library. Using the fingerprint data it is possible to determine which clones lie at the two ends of a contig. End-probes can then be prepared and should either hybridise back to clones in the contig or to novel clones depending on their position in the existing contig. The overlap detection fraction is reduced from 50% to less than 10% by hybridisation methods.

Of course it is necessary to use probes that are free of repetitive elements. Cosmid vectors such as those in the Lorist series include an RNA polymerase promoter region at either side of the cloning site (usually T7 or SP6). Thus radiolabelled transcripts can be produced from insert ends using the appropriate polymerase enzyme and a radionucleotide such as [α-^{32}P]UTP in the polymerisation reaction. By first cleaving the cosmid with a frequent-cutting restriction enzyme (e.g. *Rsa*I) the size of the transcript is reduced to around 200 bp and is unlikely to include repeated elements.

Alternatively, probes may be prepared using an inverse PCR method (Ochman *et al.* 1989). In this case, the cosmid is digested (with *Rsa*I or *Taq*I for example) and fragments are religated in a large volume to promote the occurrence of intramolecular ligation events. Oligonucleotide primers are prepared which allow polymerisation to occur when the cleaved vector molecule containing one end fragment is ligated to itself (see figure 4.11). Two different reactions may be carried out, each one specific for one end of the insert. This PCR amplified fragment may also be labelled by including a radionucleotide in the PCR reaction. The short length of this probe also reduces the likelihood of it containing repeated sequences. Repeats may also be removed by pre-association of the single-stranded probe with total human DNA before hybridisation to the library. In many of the projects described in the introduction, hybridisation methods have been used extensively towards the end of the project to confirm potential overlaps and detect clones to fill gaps.

Figure 4.11 The use of inverse PCR to produce end -probes from cosmid clones.



Oligonucleotide primers are shown as arrows labelled L and R to indicate amplification of specific insert ends.

In addition, contigs may be linked using large insert yeast artificial chromosome vectors. This provides a further source of cloned DNA, which in other projects has been shown to complement that cloned in cosmid or phage vectors. Selection of YACs would be by the hybridisation of cosmids probes to a YAC library which would preferably be specific for the Y chromosome or enriched for the Y chromosome.

It would also be possible to produce probes from individual YAC clones perhaps using *Alu* PCR (Nelson *et al.* 1989) in order to avoid repeated sequences. These could then be used to identify contigs contained within the region cloned in the YAC. In this way the YAC may include enough additional DNA to link two contigs positioned at a distance of more than 40kb which cannot be quickly linked using one or two cosmids.

A YAC contig map of the Y chromosome is being embarked upon by Page *et al.* (Cold Spring Harbor 1991 p218). 200 Y-specific probes anchored on a 30-interval deletion map are to be used to select YACs from a library derived from the DNA of a 49, XYYYY cell line.

Although it is likely that this map will be finished quickly due to the large insert size and the number of people working on the project, the cosmid map and the YAC map should remain complementary to one another since for further manipulation YAC clones are usually subcloned into cosmids. YAC clones provide long-range continuity which would be helpful in ordering cosmid contigs and correlating both maps to the deletion and pulsed-field maps already available. To this end it would be sensible to link the various maps by means of sequence tagged sites.

The technique of fluorescent *in situ* hybridisation (FISH) is also valuable. Single cosmids can be ordered along the chromosome by looking at the relative positions of probes labelled with different coloured fluorochromes on chromosome spreads and interphase nuclei (Trask *et al.* 1989). This technique will be valuable in identifying adjacent contigs which may be candidates for linkage using the hybridisation methods described above, and also in deducing the order of contigs so that the contig map can be related to the Y chromosome structure.

Particularly relevant to the Y chromosome project is the use of the comprehensive deletion map data. Information gained from the hybridisation of localised probes to the library, or by locating probes derived from contigs on the deletion map can be used to order contigs and identify those within a deletion interval which may be candidates for linkage.

CHAPTER 5

CONCLUSIONS.

I have attempted two different fingerprinting methods in order to produce a contig map of the Y chromosome. I found that the double enzyme strategy of Coulson *et al.* was not sufficiently robust in my hands to cope with variations in the quality of DNA extracted from each clone. This resulted in a large number of partial digestion products which would be impossible to analyse. This problem was also encountered by Siden-Kiamos *et al.* (1990) in their attempt to fingerprint cosmid clones from the *D. melanogaster* X chromosome. The single enzyme digestion with *Hin*fI was more successful and so far half of the cosmid library has been fingerprinted in this way.

There is one reason that the *Hin*fI digestion strategy may lead to spurious results. This is due to the large number of labelled fragments produced per clone. Although the average number of bands was around 30, the mean value is reduced due to the inclusion of clones with very few fragments, and with full length inserts, was usually in the region of 50-60 bands per cosmid. Siden-Kiamos gives an average of 60 fragments in their *Hin*fI procedures. For the *Hin*dIII/*Sau*3AI technique the average number of labelled fragments is only 23. An increased number of fragments will give more information about a clone up to a certain point. At this point, the chance that bands in different clones will be in the same position begins to rely more on chance than on the similarity between the two clones. In this project I have not observed any false overlaps where the computer routines have given a reasonably low probability rating that the

overlaps are due to chance (that is, anything less than 10^{-4}).

In terms of the contigs produced, it was at first disappointing to see that a large number of proposed contigs actually consisted of two identical clones from adjacent wells. However, this problem ceased after the first few trays and was due to inexperience in handling clones in microtitre trays rather than duplications in the library.

A number of misleading contigs have been generated. These appear to be based on homology between repetitive elements in the cosmids. Contig 7 is an example of this. Unlike the cosmids containing the DYZ1 repeat element which are highly deleted, cosmids in contig 7 contain an average number of bands. However, they cannot be aligned logically and there are too many clones to represent a single locus (unless it is preferentially cloned). This group of clones are most likely derived from a moderately repeated element that has undergone divergence at different loci. Thus common elements are retained, on which the contig is based, along with locus-specific fragments which are not shared between all clones.

Apart from these easily detectable groups of clones, the remaining contigs are more logical. Hopefully repeated elements will not create confusion later in the project as contigs are linked to one another.

With the mature mapping project of the *C.elegans* genome as a model, it is apparent that the Y chromosome map will not be completed by fingerprinting alone.

In a recent review of the *C.elegans* project, Coulson *et al.* (1991) describe the shortfalls of cosmid fingerprinting. These include

clones that have too few restriction fragments to be matched to others (or inconclusive overlaps), and the problem that clones do not always represent a random distribution in the genome. Although the first problem could be overcome to some extent by detecting overlaps by selective probing of the library, this approach was of no use where regions were simply not represented in the library.

At this point the use of YAC clones becomes invaluable. YACs provide long-range continuity and have been shown to support fragments which are more uniformly distributed in the genome than can be cloned in cosmids. However, due to the large insert size, it is not appropriate to fingerprint YACs in the same way as cosmids since too many fragments are produced to be interpretable. Coulson *et al.* undertook a hybridisation strategy whereby probes were produced from random large YACs and YACs preselected for their association with contig ends and were hybridised to a cosmid grid. This method was successful in placing a large number of YACs on the map. Most overlaps were supported by several lines of evidence (for example hybridisation to sequential cosmids in a contig) and the frequency of YAC co-ligation events was estimated at less than 5%. Using a grid of 958 YAC clones, it is now possible to locate any probe DNA to within a region of 100kb on the *C. elegans* physical map.

Although this method dramatically reduced the number of contigs from 700 to 200, the reciprocal probing strategy became ineffective when all cosmid end-probes had been used. In order to derive specific end-probes from YAC clones which protruded from cosmid contigs, a direct sequencing method primed with vector-specific oligonucleotides was used. Sequence data from the end of the insert provided a

sequence tagged site from which a PCR product could be generated and used to probe YAC grids. Overlaps detected in this way were usually based on a single bridging clone and may be due to chimaeric YACs or rare repeat sequences. However, agreement with the genetic map provides verification.

Some cosmid contigs could not be associated with YACs due to hybridisation problems; either due to the presence of repeat elements, or because they were cloned in a vector which had homology to the YAC vector. These were also sequenced directly from the vector to provide data for the synthesis of oligonucleotide pairs. PCR products were then used to probe the YAC grids. Hybridisations were checked by PCR of the hybridising YAC. These additional methods have reduced the contig number to 90 but the map is still not completely closed.

As stated by Evans (1991) the number of gaps in a contig-based genome mapping project is 'a function of the efficiency of the contig-building approach, as well as the representation and redundancy of the libraries and the sequence content of the region of interest.' Perhaps methods of cosmid mapping which reduce the amount of overlap required to detect an overlap such as the oligonucleotide hybridisation strategy of Craig *et al.* (1990) and Evans' own cosmid multiplex strategy (which may even be extrapolated to a three dimensional array), will increase the information derived from the initial cosmids to such an extent that the number of gaps will be substantially less than are produced by the 'first generation' approaches.

Evans suggests methods of map closure that are based on physical

location of contigs, that is, the use of radiation reduction somatic cell hybrids and fluorescence *in situ* hybridisation of cosmid clones to metaphase or interphase nuclei.

The Y chromosome project is at present in the first stage of the fingerprinting strategy. Many more cosmids must be added to the database before this method ceases to be productive in generating contigs. Since the progress does not follow predicted success rates, it is difficult to suggest how many cosmids must be fingerprinted. However, it is likely that at least the remainder of the library and probably 4 to 5,000 more should be analysed based on the results of the *C.elegans* project. Additional clones are to be made available in the form of a library prepared from the Y-only somatic cell hybrid 853 (provided by N. Affara) and also a library prepared from flow sorted Y chromosomes (provided by P. deJong). Hopefully clones from these different Y chromosome sources will compensate for the contaminating regions of chromosomes 1 and 12 in the 3E7 library.

In the *C. elegans* work, the use of YAC clones was highly advantageous and it would be desirable to begin to obtain Y chromosome YACs as soon as possible. YAC grids could be probed with cosmid end-probes alongside the cosmid grids so that overlapping cosmid contigs and corresponding YAC clones could be detected concurrently.
A combined Y chromosome deletion map is being generated by N. Affara, C. Tyler-Smith and J. Weissenbach. This well characterised map will be useful in correlating contigs to the physical map already available. There are still many DNA sequences which should be used to probe the cosmid clones and it would be appropriate to use any

sequence tagged site PCR products or end-probes to probe filters of patients with Y chromosome deletions. In this way contigs can be localised to deletion intervals and the scope of the deletion map increased. I am at present selecting cosmids from the library using probes which map within a microdeletion interval. This interval is the only detected portion of Y chromosomal DNA that is deleted in an azoospermic patient (A. Chandley pers. comm.). These cosmids will be used to create a contig spanning this region, within which it should be possible to search for potential gene sequences which may correspond to the *AZF* gene.

There are various ways that genomic material can be screened for coding sequences. Genomic clones can be used a) to probe blots of mRNA, b) to test for conservation on blots of DNA from a wide range of species or c) can be assayed for the presence of CpG islands. These approaches will be useful for analysing small numbers of cosmids, for example, those within a microdeletion interval. However, for screening whole libraries these approaches become impractical. Several more appropriate methods for screening large numbers of clones have recently been reported. Most involve preparing libraries in specialised vectors, for example, exon trapping, (Duyk *et al.* 1990). However, there is one method that is especially applicable in analysing large genomic libraries linked into contiguous segments. This is to screen with cDNA probes from different tissues and at different developmental times (Hochgeschwender *et al.* 1989). I have begun to develop this system in application to the Y chromosome library using adult and foetal testis cDNA, and further experiments will be important in the identification

of gene sequences on the Y chromosome.

It is obviously very difficult to close 'bottom up' maps of this kind. Due to large blocks of repetitive elements which cannot be rationalised by this fingerprinting method, the Y chromosome map can only consist of large chunks of contigs anchored to the euchromatic regions by virtue of deletion map and fluorescence *in situ* data.

In summary, the Y chromosome cosmid contig map is progressing in a very satisfactory way. With the involvement of YAC clones and in an environment in which many workers are producing new technology continuously, I believe that the project should be completed fairly rapidly. The map should provide a basis for the search for Y chromosome-encoded genes, studies of evolution and homology between the X and Y chromosomes, and, ultimately, the sequencing of the entire euchromatic chromosome regions.

REFERENCES

AFFARA, N.A., FERGUSON-SMITH, M.A., TOLMIE,J.L., KWOK, K., MITCHELL, M., JAMIESON, D., COOKE, A. & FLORENTIN, L. (1986). Variable transfer of Y-specific sequences in XX males. *Nucl. Acids Res.* **14**, 5375-5387.

AFFARA, N.A., FERGUSON-SMITH, M.A., MAGENIS, R.E., TOLMIE, J.L., BOYD,E., COOKE, A., JAMIESON, D., KWOK, K., MITCHELL, M. & SNADDEN, L. (1987). Mapping the testis determinants by an analysis of Y-specific sequences in males with apparent XX and XO karyotypes. *Nucl. Acids. Res.* 15 no 18, 7325-7347.

ANDERSSON, M., PAGE, D.C., PETTAY, D., SUBRT, I., TURLEAU, C., DE GROUCHY, J. & DE LA CHAPELLE, A. (1988). Y;autosome translocations and mosaicism in the aetiology of 45,X maleness: assignment of fertility factor to distal Yq11. *Hum. Genet.* **79**, 2-7.

ARNEMANN, J., EPPLEN, J.T., JAKUBICZKA, S., KAHLER, S., BERG, L.P., SCHRADER, U., SAUERMANN, U., COOKE, H.J., ENGEL, W. & SCHMIDTKE, J. (1987) A human Y-specific DNA sequence transcribed in human testis. (abstract)(9th international workshop on Human Gene Mapping)*Cytogenet. Cell Genet.* **46**, 572

ARNEMANN, J., EPPLEN, J.T., COOKE, H.J., SAUERMANN, U., ENGEL, W. & SCHMIDTKE, J. (1987) A human Y-chromosomal DNA sequence expressed in testicular tissue.

BALLABIO, A., PARENTI, G., CARROZZO, R., COPPA, G., FELICI, L., MIGLIORI, V., SILENGO, M., FRANCESCHINI, P. & ANDRIA, G. (1988). X/Y translocation in a family with X-linked ichthyosis,

chondrodysplasia punctata and mental retardation: DNA analysis reveals deletion of the steroid sulphatase gene and translocation of its Y pseudogene. *Clinical genetics* **34**, 31-37

BALLABIO, A. (1991) Contiguous deletion syndromes. *Current opinion in genetics and development*. **Vol 1, No 1**

BELLANNE-CHANTELOT, C., BARILLOT, E., LACROIX, B., LE PASLIER, D. & COHEN, D. (1991) A test case for physical mapping of human genome by repetitive sequence fingerprints: construction of a physical map of a 420kb YAC subcloned into cosmids. *Nuc. Acids Res.* **19**, no.3 505-510

BENGTSSON, B.O. & GOODFELLOW, P.N. (1987) The effect of recombination between the X and Y chromosomes of mammals. *Ann. Hum. Genet.* **51**, 57-64

BERNSTEIN,R., ROSENDORF, J., RAMSAY, M., PINTO, M.R. & PAGE, D.C. (1987) A unique dicentric X;Y translocation with Xq and Yq breakpoints: cytogenetic and molecular studies. *Am. J. Hum. Genet.* **41**, 145-156

BERTA, P., HAWKINS, J.R., SINCLAIR, A.H., TAYLOR, A., GRIFFITHS, B.L., GOODFELLOW, P.N. & FELLOUS, M. (1990) Genetic evidence equating *SRY* and the testis-determining factor. *Nature* , **348**, 448-450

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

BISHOP, C.E., GUELLEN, G., GELDWORTH, D., VOSS, R., FELLOUS, M. &

WEISSENBACH, J. (1983). Single-copy DNA sequences specific for the human Y chromosome. *Nature* 303, 831-832

BROWN, W.R.A. (1988) A physical map of the human pseudoautosomal region. *EMBO journal* 7,8 2377-2385

BUCKLE, V., MONDELLO, C., DARLING, S., CRAIG, I.W. & GOODFELLOW, P.N. (1985) Homologous expressed genes in the human sex chromosome pairing region. *Nature* **317**, 739-741

BUCKLE, V.J., BOYD, Y., FRASER, N., GOODFELLOW, P.N., GOODFELLOW, P.J., WOLFE, J. & CRAIG, I.W. (1987). Localisation of Y chromosome sequences in normal and 'XX' males. *J. med. Genet.* **24**, 197-203.

BULL, J.J., HILLS, D.M., O'STEEN, S., (1988) Mammalian *ZFY* sequences exist in reptiles regardless of sex determining mechanism. *Science,* **242**, 567-569

BURGOYNE, P.S. (1987). The role of the mammalian Y chromosome in spermatogenesis. *Development 101 supplement* 133-141.

BURGOYNE, P.S., LEVY, E.R. & McLAREN, A.(1986) Spermatogenic failure in male mice lacking H-Y antigen. *Nature,* 320, 170-172.

BURK, R.D., MA, P. & SMITH, K.D. (1985) Characterisation and evolution of a single-copy sequence from the human Y chromosome. *Molec. Cell. Biol.* 5, 576-581

CARLE, G.F., FRANK, M. & OLSON, M.V. (1986) Electrophoretic separation of large DNA molecules by periodic inversion of the electric field. *Science* **232**, 65-68

CARRANO, A.V., LAMERDIN, J., ASHWORTH, L., WATKINS, B., BRANSCOMB,

E., SLEZAK, T., RAFF, M., DE JONG, P.J., KEITH, D., McBRIDE, L., MEISTER, S. & KRONICK, M. (1989) A high resolution, fluorescence-based semiautomatic method for DNA fingerprinting. *Genomics* **4**, 129-136

CATTENACH, B.M., POLLARD, C.E. & HAWKES, S.G. (1971) Sex reversed mice: XX and XO males. *Cytogenetics* **10**, 318-337.

CHANDLEY, A.C., GOETZ, P., HARGREAVE, T.B. & JOSEPH, A.M. (1984). On the nature and extent of XY pairing at meiotic prophase in man. *Cytogenet Cell Genet* **38**, 241-247

CHAPMAN, V.M., KEITZ, B., DISTECHE, C.M., LAU, E.C. & SNEAD, M.L. (1991). Linkage of amelogenin (*Amel*) to the distal portion of the mouse X chromosome. *Genomics* **10**, 23-28

DE LA CHAPELLE, A. (1987). The Y-chromosomal and autosomal testis-determining genes. *Development 101 supplement* 33-38.

COOKE, H. (1976) Repeated sequence specific to human males. *Nature, Lond.* **262**, 182-186

COOKE, H.J., BROWN, W.A.R. & RAPPOLD, G.A. (1984) Closely related sequences on the human X and Y chromosomes outside the pairing region. *Nature* **311**, 259-261

COOKE, H.J., BROWN, W.R.A. & RAPPOLD, G.A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* 317, 687-692

COULSON, A. SULSTON, J., BRENNER, S. & KARN, J. (1986) Towards 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. & KOHARA, J. (1988) Genome linking with yeast artificial chromosomes. ***Nature*** **335**, 184-186

COULSON, A., KOZONO, Y., LUTTERBACH, B., SHOWKEEN, R., SULSTON, J. & WATERSTON, R. (1991) YACs and the *C.elegans* genome. Bioessays **13**, 8 413-417.

CRAIG, A.G., NIZETIC, D., HOHEISEL, J., ZEHETNER, G. & LEHRACH, H. (1990) Ordering of cosmid clones covering the Herpes Simplex virus type 1 (HSV-1) genome: a test case for fingerprinting by hybridisation. *Nuc. Acids Res.* 18, NO.9 2653-2660

CROSS, S.H. & LITTLE, P.F.R. (1986) A cosmid vector for systematic chromosome walking. *Gene* **49**, 9-22

CROW, T.J., (1989) Pseudoautosomal locus for the cerebral dominance gene. *Lancet* II 339-340.

DAIGER, SP., ILDIN, R.S. & SU,T-S. Sequences on the human Y chromosome homologous to the autosomal gene for arginosuccinate synthetase. *Nature*, **298**, 682-4

DARLING, S.M., GOODFELLOW, P.J., PYM, B., BANTING, G.S., PRITCHARD, C. & GOODFELLOW, P.N. (1986) Molecular genetics of *MIC2* : a gene shared by the human X and Y chromosomes. *Cold Spring Harbor Symp. Quant. Biol.* **51**, 205-212

DINGWALL, C. & LASKEY, R.A., (1986) Protein import into the cell nucleus *Ann. Rev. Cell. Biol* **2**, 367-390

DISTECHE, C.M., CASANOVA, M., SAAL, H., FRIEDMAN, C., SYBERT, V., GRAHAM, J., THULINE, H., PAGE, D.C. & FELLOUS, M. (1986). Small deletions of the short arm of the Y chromosome in 46, XY females. *Proc. Natl. Acad Sci. U.S.A.* **83**, 7841-7844

DUYK, G.M., KIM, S., MYERS, R.M. & COX, D.R. (1990) Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci.* 87, 8995-8999

EICHER, E.M. & WASHBURN, L.L. (1986). Genetic control of primary sex-determination in mice. *A. Rev. Genet.* **20**, 249-263.

ELLIS, N. & GOODFELLOW, P.N. (1989) The mammalian pseudoautosomal region. *Trends in genetics* **5**, 12 406-410

ELLIS, N.A., GOODFELLOW, P.J., PYM, B., SMITH, M., PALMER, M., FRISCHAUF, A-M. & GOODFELLOW, P.N. (1989) The pseudoautosomal boundary in man is defined by an *Alu* repeat sequence inserted on the Y chromosome. *Nature* **327**, 81-84

EVANS, G.A. & LEWIS, K.A. (1989) Physical mapping of complex genomes by cosmid multiplex analysis. *Proc. Natl. Acad. Sci.* **86**, 5030-5034

EVANS, E.P., BURTENSHAW, M.D. & CATTENACH, B.M. (1982). Meiotic crossing over between the X and Y chromosomes of male mice carrying the sex-reversing (*Sxr*) factor. *Nature* **300**, 443-445.

FAIRALL, L., RHODES, D. KLUG, A., (1986) Mapping the sites of protection on a 5S RNA gene by the *Xenopus* transcription factor IIIA: a model for the interaction. *J. Mol Biol,* **192**, 577-591

FERGUSON, M.W.J. & JOANEN, T., (1982) Temperature of egg incubation determines sex in *Alligator mississippiensis. Nature,* **296**, 850-853.

FERGUSON-SMITH, M.A., NORTH, M.A., AFFARA, N.A. & BRIGGS, H. (1990) The secret of sex *The Lancet* 336, 809-810

FERGUSON-SMITH, M.A. (1966) X-Y Chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome. The Lancet 2, 475-476

FISHER, E.M.C., BEER-ROMERO, P., BROWN, L.G., RIDLEY, A., McNEIL, J.A., BENTLEY LAWRENCE, J., WILLARD, H.F., BIEBER, F.R. & PAGE, D.C. (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* , 63, 1205-1218

FISHER, E.M.C., ALITALO, T., LUOH, S-W., DE LA CHAPELLE, A. & PAGE, D.C. (1990) Human sex-chromosome-specific repeats within a region of pseudoautosomal/Yq homology. *Genomics* 7, 625-628

FORD, C.E., JONES, K.W., POLANI, P.E., de ALMEIDA, J.C. & BRIGGS, J.H. (1959) A sex chromosome anomaly in a case of gonadal dysgenesis (Turners syndrome). *Lancet* i, 711-3

FRASER, N., BALLABIO, A., ZOLLO, M., PERSICO, G. & CRAIG, I. (1987) Identification of incomplete coding sequences for steroid sulfatase on the human Y chromosome: evidence for an ancestral pseudoautosomal gene? *Development* 101, supplement, 127-132

FROMMER. M., PROSSER, J. & VINCENT, P.C. (1984) Human satellite I sequences include a male specific 2.47kb tandem repeat unit containing one Alu family member per repeat. *Nuc. Acids Res.* 12, 2887-2900

GARZA, D., AJIOKA, J.W., BURKE, D.T. & HARTL, D.L. (1989) Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* 246, 641-646

GIBSON, T.J. & SULSTON, J. (1987) Preparation of large numbers of plasmid DNA samples in microtitre plates by the alkaline lysis method *Gene Anal. Techn.* 4, 41-44

GILLARD, E.F., AFFARA, N.A., YATES, J.R.W., GOUDIE, D.R., LAMBERT, J., AITKEN, D.A. & FERGUSON-SMITH, M.A. (1987) Deletion of a DNA sequence in eight of nine families with X-linked ichthyosis steroid sulphatase deficiency (1987) *Nuc. Acids Res.* **15**, No 10 3985-3977

GOODFELLOW, P. (1983) Expression of the 12E7 antigen is controlled independently by genes on the human X and Y chromosomes. *Different.* **23**, 535-539

GOODFELLOW, P.J., DARLING, S., BANTING, G., PYM, B., MONDELLO, C. & GOODFELLOW, P.N. (1987a) Pseudoautosomal genes in man. *Development* **101** supplement, 119-125

GOODFELLOW, P.J., PRITCHARD, C., TIPPET, P. & GOODFELLOW, P.N.(1987b) Recombination between the X and Y chromosomes at the *YG* locus: implications for the relationship between *MIC2*, *XG* and *YG*. *Ann. Hum. Genet.* 51, 161-167

GOUGH, N.M., GEARING, D.P., NICOLA, N., BAKER, E., PRITCHARD, M., CALLEN, D.F. & SUTHERLAND, G. (1990) Localisation of the human GM-CSF receptor to the X-Y pseudoautosomal region. *Nature* **345**, 734-736

GUBBAY, J., COLLIGNON, J., KOOPMAN, P., CAPEL, B., ECONOMOU, A., MUNSTERBERG, A., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature,* 346, 245-250

GUELLAN, G., CASANOVA, M., BISHOP, C., GELDWORTH, D., ANDRE, G., FELLOUS, M. & WEISSENBACH, J. (1984). Human XX males with Y single-copy DNA fragments. *Nature,* 307, 172-173.

HARRISON-LAVOIE, K.J., JOHN, R.M., PORTEOUS, D.J. & LITTLE, P.F.R. (1989) A cosmid clone map derived from a small region of human chromosome 11. *Genomics* 5, 501-509

HAZOUT, S. & LUCOTTE, G. (1986) Vers une genealogie du chromosome Y *Ann Genet* **29**, 246-252

HOCHGESCWENDER, U., SUTCLIFFE, J.G. & BRENNAN, M.B. (1989) Construction and screening of a genomic library specific for mouse human chromosome 16. *Proc Natl. Acad. Sci.* **86**, 8482-8486

HOHEISEL, J.D. & LEHRACH, H. (1991) (abstract) Genome mapping and sequencing meeting, Cold Spring Harbor.

ISH-HOROWICZ, D. & BURKE, J.F. (198) Rapid and efficient cosmid cloning. *Nucl. Acids. Res.* **9**, 2989-2998

JACOBS, P.A. & STRONG, J.A. (1959) A case of human intersexuality having possible XYY sex determining mechanism. *Nature* **183**, 302-3

JAGER, R.J., ANVRET, M., HALL, K. & SCHERER, G. (1990) A human XY female with a frame shift mutation in the candidate testis-determining gene *SRY* . *Nature* , 348, 452-453

JAKUBICZKA, S., ARNEMANN, J., COOKE, H.J., KRAWCZAK, M. SCMIDTKE, J. (1989) A search for restriction fragment length polymorphism on the human Y chromosome. *Hum Genet* **84**, 86-88

JANTZEN, HM., ADMAN, A., BELL, S.P. & TIJAN, R. (1990) Nucleolar transcription factor hUBF contains a DNA binding motif with homology to HMG proteins. *Nature,* **344**, 830-836

JOST, A., VIGIER, B., PREPIN, J., PERCHELLET, J. (1973) Studies on sex differentiation in mammals. *Rec. Prog. Horm. Res.* **29**, 1-34

KELLY, M., BURKE, J., SMITH, M., KLAR, A. & BEACH, D. (1988) Four

mating-type genes control sexual differentiation in the fission yeast. *EMBO J.* **7.** 1537-1547

KEITGES, E., RIVEST, M., SINISCALCO, M. & GARTLER, S.M. (1985). X-linkage of steroid sulphatase in the mouse is evidence for a functional Y-linked allele. *Nature,* **315,** 226-227

KOHARA, Y., AKIYAMA, K. & ISONO, K. (1987) The physical map of the whole *E. coli* chromosome: Application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50,** 495-508

KONGUSWAN, K, YU, Q., VINCENT, A., FRISARDI, M.C., ROSBASH, M., LENGYL, J.A. & MERRIAM, M. (1985) A *Drosophila minute* mutation encodes a ribosomal protein. *Nature* **317,** 555-558

KOOPMAN, P., GUBBAY, J., COLLIGNON, J. & LOVELL-BADGE, R. (1989) *ZFY* Gene expression patterns are not compatible with a primary role in mouse sex determination. *NATURE,* **342,** 940-941

KOOPMAN, P., MüNSTERBERG, A., CAPEL, B., VIVIAN, N. & LOVELL-BADGE, R. (1990) Expression of a candidate sex-determining gene during mouse testis differention. *Nature* , **348,** 450-452

KOOPMAN, P., ASHWORTH, A., LOVELL-BADGE, R., (1991) The *ZFY* gene family in humans and mice. *Trends in genetics,* **7,** vol 4, 132-136

KOOPMAN, P., GUBBAY, J., VIVIAN, N., GOODFELLOW, P. & LOVELL-BADGE, R. (1991) Male development of chromosomally female mice transgenic for *Sry. Nature* 351, 117-121

KNOTT, V., REES, D.J.G., CHENG, Z. & BROWNLEE, G.G. (1988) Randomly picked cosmid clones overlap the *pyrB* and *oriC* gap in the physical

map of the *E. coli* chromosome. *Nuc. Acids Res.* **16**, 2601-2612

KNOTT, V. BLAKE, D.J. & BROWNLEE, G.G. (1989) Completion of the detailed restriction map of the *E. coli* genome by the isolation of overlapping cosmid clones. *Nuc. Acids Res.* **17**, no. 15, 5901-5912

KWOK, K., MITCHELL, M.J., AFFARA, N.A., BOYD, E., TOLMIE, J., MAGENIS, R.E. & FERGUSON-SMITH, M.A. (1987) A deletion map of the long arm of the human Y chromosome and the detection of possible paracentric inversions. *Cytogenet. Cell Genet.* **46**, 642

LANDER, E.S. & WATERMAN, M.S. (1988) Genomic mapping by fingerprinting random clones: A mathematical analysis *Genomics* **2**, 231-239

LAU, E.C., MOHANDAS, T., SHAPIRO, L.J., SLAVKIN, H. & SNEAD, M. (1989). Human and mouse amelogenin gene loci are on the sex chromosomes. *Genomics* **4**, 162-168

LEVY, E.R. & BURGOYNE, P.S. (1986). The rate of XO germ cells in the testes of XO/XY and XO/XY/XYY mouse mosaics: evidence for a spermatogenesis gene on the mouse Y chromosome. *cytogenet. Cell Genet.* **42**, 208-213.

LINK, A. & OLSON, M. (1991) Physical map of the *Saccharomyces cerevisiae* genome at 110-kilobase resolution. *Genetics* **127**, 681-698

LITTLE, P.F.R. (1987) Choice and use of cosmid vectors. *In* "Gene Cloning" (D.M. Glover ed.), Vol III pp. 19-42 IRL press Oxford.

LOVELL-BADGE, R. & ROBERTSON, E. (1990) XY female mice resulting from a heritable mutation in the primary testis determining gene. *Development*, **109**, 647-653

MAGENIS, R.E., TOCHEN, M.L., HOLAHAN, K.P., CAREY, T., ALLEN, L. & BROWN, M.G.(1984) Turner syndrome resulting from partial deletion of Y chromosome short arm: Localisation of male determinants. *J. Paediatrics* **105**, 916-919.

MALASPINA, P., PERSICHIETTI, F., NOVELLETTO, A., IODICE, C., TERRENATO, L., WOLFE, J., FERRARO, M. & PRANTERA, G. (1990) The human Y chromosome shows a low level of DNA polymorphism. *Ann. Hum. Genet.* **54**, 297-305

MANIATIS, T., FRITSCH, E.F. & SAMBROOK, J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory. CSH, NY.

MARCUS, M., TANTRAVAHI, R. DEV, V., MILLER, D. & MILLER, O. (1976) Human-mouse cell hybrid with human multiple Y chromosomes. *Nature* **262**, 63-65

MARDON, G., MOSHER, R., DISTECHE, C.M., NISHIOKA, Y., McLAREN, A., PAGE, D.C. (1989) Duplication, deletion, and polymorphism in the sex-determining region of the mouse Y chromosome. *Science*, **243**, 78-79

MARDON, G. & PAGE, D.C. (1989) The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers. *Cell*, 56, 765-770

McKAY, R.D.G., HERITAGE, J., BOBROW, M. & COOKE, H.J. (1978) Endonuclease analysis of Y chromosome DNA. *Cytogenet. Cell Genet.* **22**, 357-358

McLAREN, A., SIMPSON, E., TOMONORI, K., CHANDLER, P. & HOGG, H. (1984). Male sexual differentiation in mice lacking H-Y antigen. *Nature,* **312**, 552-555.

McLAREN, A. (1988). Sex determination in mammals. *Trends in genetics,*

4 *no* 6, 153-157.

MILLER, J., McLACHLAN, A.D., & KLUG, A.,(1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4**, 1609-1604

MITCHELL, M.J. & BISHOP, C.E. (1991) The cloning of Sby, a candidate for Hya and Spy, from the Sxr region of the mouse. HGM11 abstract.

MONDELLO, C., ROPERS, H-H., CRAIG, I.W., TOLLEY, E. & GOODFELLOW, P.N. (1987) Physical mapping of genes and sequences at the end of the human X chromosome short arm. *Ann. Hum. Genet.* 51, 137-143

MORTON, N.E., (1991) Parameters of the numan genome *Proc. Natl. Acad. Sci.* 88, 7474-7476

MOORE, G.E., IVENS, A., CHAMBERS, J., FARRALL, M., WILLIAMSON, R., PAGE, D.C., BJORNSSON, A., ARNASON, A. & JENSSON, O. (1987). Linkage of an X-chromosome cleft palate gene. *Nature* 326, 91-92

MOYZIS, R.K., TORNEY, D.C., MEYNE, J., BUCKINGHAM, J.M., WU, J-R., BURKS, C., SIROTKIN, K.M. & GOAD, W.B. (1989) The distribution of interspersed repetitive DNA sequences in the human genome. *Genomics* **4**, 273-289

MÜLLER, U. (1987). Mapping of testis-determining locus on Yp by molecular genetic analysis of XX males and XY females. *Development 101 supplement* 51-58.

MÜLLER, U. & LALANDE, M. (1990) A physical map of the Y chromsome short arm. *Genomics,* 517-523

MURRAY, N.E., BRAMMER, W.J. & MURRAY, K. (1977) Lamboid phages that

simplify the recovery of *in vitro* recombinants. *Mol. Gen. Genet.* **150**

NAGAMINE, C.M., CHAN, K., KOZAK, C.A., LAU, Y.F. (1989) Chromosome mapping and expression of a putative testis-determining gene in mouse. *Science*, **243**, 80-83

NAGAMINE, C.M., CHAN, K., HAKE, L.E. & LAU, Y.C. (1990) The two candidate testis-determining genes (*Zfy-1* and *Zfy-2)* are differentially expressed in fetal and adult mouse tissues. *Genes & Development*, **4**, 63-74

NAKAGOME, Y., NAGAFUCHI, S., SEKI, S., NAKAHORI, Y., TAMURA, T., YAMADA, M. & IWAYA, M. (1991) A repeating unit of the DYZ1 family on the human Y chromosome consists of segments with partial male-specificity. *Cytogenet. Cell Genet.* 56, 74-77

NAKAHORI,Y., MITANI,K., YAMADA,M. & NAKAGOME,Y. (1986) A human Y-chromosome specific repeated DNA family (DYZ1) consists of a tandem array of pentanucleotides. *Nucl. Acids Res.* **14**, 7569-7580

NAKAHORI, Y., TAKENAKA, O. & NAKAGOME, Y. (1991a) A human X-Y homologous region encodes "Amelogenin" *Genomics* **9**, 264-269

NAKAHORI, Y., TAMURA,T. NAGAFUCHI, K., FUJIEDA, K., MINOWADA, S., FUKUTANI, K., FUSE, H., HAYASHI, K., KUROKI, Y., FUKUSHIMA, Y., AGEMATSU, K., KUNO, T., KANEKO, S., YAMADA, K., KITAGAWA, T., NONOMURA, M., FUKUDA, S., KUSANO, M., ONIGATA, S., HIBI, I. & NAKAGOME, Y. (1991b) Molecular cloning and mapping of 10 new probes on the human Y chromosome. *Genomics* **9**, 765-769

NELSON, D.L., LEDBETTER, S.A., CORBO, L., VICTORIA, M.F., RAMIREZ-SOLIS, R., WEBSTER, T.D., LEDBETTER, D.H. & CASKEY, C.T. (1989) *Alu* polymerase chain reaction: A method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **86**, 6686-6690

NGO, K.Y., VERNAUGD, G., JOHNSSON, C., LUCOTTE, G. & WEISSENBACH, J. (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. *Am J Hum Genet* **38**, 407-418

NORTH, M.A.J. (1991) Molecular genetic studies on the sex determining region of the Y chromosome. PhD thesis. University of Cambridge.

NORTH, M., SARGENT, C., O'BRIEN, J., TAYLOR, K., WOLFE, J., AFFARA, N.A. & FERGUSON-SMITH, M.A. (1991) comparison of ZFY and ZFX gene structure and analysis of alternative 3' untranslated regions of ZFY. *Nuc. Acids. Res.* **19**, 2579-2586

OAKEY, R. & TYLER-SMITH, C. (1990) Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. *Genomics* **7**, 325-330

OCHMAN, H. AJIOKA, J., GARZA, D. & HARTL, D. (1989) Inverse polymerase chain reaction. in'PCR Technology. Principles and Applications for DNA amplification'. Ed Erlich, Pge 105 Macmillan Press.

OLSON, M.V., DUTCHIK, J., GRAHAM, M., BRODEUR, G., HELMS, C., FRANK, M., MacCOLLIN, M., SCHEINMAN, R. & FRANK, T. (1986) Random-clone strategy for genomic restriction mapping in yeast. *Proc. Natl. Acad. Sci. USA* **83**, 7826-7830

OOSTHUIZEN, C.J.J., HERBERT, J.S., VERMAAK, L.K. BRUSNICKY, J., FRIKE,J., DU PLESSIS, L. & RETIEF, A.E. (1990) Deletion mapping of 39 random isolated Y-chromosome DNA fragments. *Hum. Genet.* 85, 205-210

ORITA, M., SUZUKI, Y., SEKIYA, T. & HAYASHI, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the Polymerase Chain Reaction. *Genomics* 5, 874-879

PAGE, D.C., HARPER, M.E., LOVE, J. & BOTSTEIN, D. (1984) Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* 311, 119-123

PAGE, D.C., BROWN, L.G. & DE LA CHAPELLE, A. (1987a) Exchange of terminal portions of X and Y chromosomal short arms in human XX males. *Nature* 328, 437-440

PAGE, D.C., BIEKER, K., BROWN, L.G., HINTON, S., LEPPERT, M., LALOUEL, J-M., LATHROP, M., NYSTROM-LAHTI, M., DE LA CHAPELLE, A. & WHITE, R. (1987b) Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. *Genomics* 1, 243-256

PAGE, D.C., MOSHER, R., SIMPSON, E.M., FISHER, E.M.C., MARDON, G., POLLACK, J., McGILLVRAY, B., De La CHAPELLE, A. & BROWN, L.G. (1987c) The sex-determining region of the Human Y Chromosome encodes a finger protein. *cell*, 51, 1091-1104

PAGE, D. (1987d) Hypothesis: a Y-chromosome gene causes gonadoblastoma in dysgenetic gonads. Development 101 Suppl. 151-155

PAGE, D.C., FISHER, E.M.C., McGILLVRAY, B. AND BROWN, L.G. (1990) Additional deletion in sex-determining region of Y chromosome resolves paradox ot X, t(Y;22) female. *Nature,* **346,** 279-281

PALMER, M.S., SINCLAIR, A.H., BERTA, P., ELLIS, N.A., GOODFELLOW, P.N., ABBAS, N.E. & FELLOUS, M. (1989) Genetic evidence that *ZFY* is not the testis-determining factor. *NATURE,* **342,** 937-939

PALMER, M.S., BERTA, P., SINCLAIR, A.H., PYM, B. & GOODFELLOW, P.N. (1990) Comparison of human *ZFY* and *ZFX* transcripts. *Proc. Natl. Acad. Sca. USA* , **87,** 1681-1685

PERRET, X., BROUGHTON, W. & BRENNER, S. (1991) Canonical ordered cosmid library of the symbiotic plasmid of *Rhizobium* species NGR234. *Proc. Natl. Acad. Sci. USA* 88, 1923-1927

PETIT, C., LEVILLIERS, J. & WEISSENBACH, J. (1988) Physical mapping of the human pseudoautosomal region; comparison with genetic linkage map. *EMBO journal* 7,8 2369-2376

PONDER, B. (1988) Gene losses in human tumours *Nature* 355, 400-402

PTASHNE, M., (1988) How eukaryotic transcriptional activators work. *Nature,* 335, 683-389

RAPPOLD, G.A. & LEHRACH, H. (1988) a long range restriction map of the pseudoautosomal region by partial digest PFGE analysis from the telomere. *Nuc. Acids Res.* 16, 5361-5377

RUDD, K., MILLER, W., WERNER, C., OSTELL, J., TOLSTOSHEV, C. & SATTERFIELD, S. (1991) Mapping sequenced *E.coli* genes by computer: software, strategies and examples. *Nuc. Acids Res.* **19,** no.3 637-647

ROUYER, F., SIMMLER, M-C., JOHNSSON, C., VERGNAUD, G. COOKE, H.J. & WEISSENBACH, J. (1986) A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* 319,

291-295

SAIKI, R.K., GELFAND, D.H., STOFFEL, S., SHARF, S.J., HIGUCHI, R., HORN, G.T., MULLIS, K.B. & ERLICH, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239 487-491

SCHEMPP, W., WEBER, B. & MULLER, G. (1989) Mammalian sex-chromosome evolution:a conserved homeologous segment on the X and Y chromosomes in primates. *Cytogenet Cell Genet* 50, 201-205.

SCHERER, G., SCHEMPP, W., BACCICHETTI, C., LENZINI, E., BRICARELLI, F.D., CARBONE, L.D.L. & WOLF, U. (1989) Duplication of an Xp segment that includes the *ZFX* locus causes sex inversion in man. *Hum Genet*, 81, 291-294

SCHLESSINGER, D. (1990) Yeast artiticial chromosomes: tools for mapping and analysis of complex genomes. *Trends in genetics* 6, no.8 248-258

SCHMID, M., GUTTENBACH, M., NANDA, I., STUDER, R. & EPPLEN, J.T. (1990) Organisation of DYZ2 repetitive DNA on the human Y chromosome. *Genomics* 6, 212-218

SCHWARTZ, D.C. & CANTOR, C.R. (1984) Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* 37, 67-75

SHAPIRO, L.J., YEN, P., POMERANTZ, D., MARTIN, E., ROLEWIC, L. & MOHANDAS, T. (1989). Molecular studies of deletions at the human steroid sulfatase locus. *Proc. Natl. Acad. Sci.* 86, 8477-8481

SHERTON, C.C. & WOOL, I.G. (1972) Determination of the number of proteins in liver ribosomes and ribosomal subunits by polyacrylamide gel electrophoresis. *J. Biol. Chem.* **247**, 4460-4467

SHORTLE, V.P. (1990) CG rich islands on the human Y chromosome: An approach to cloning Y-linked genes. PhD thesis. University of London

SIDEN-KIAMOS, I., SAUNDERS, R., SPANOS, L., MAJERUS, T., TREANEAR, J., SAVAKIS, C., LOUIS, C., GLOVER, D.M., ASHBURNER, M. & KAFTOS, F.C. (1990) Towards a physical map of the *Drosophila melanogaster* genome: mapping of cosmid clones within defined genomic divisions. *Nuc. Acids Res.* **18**, no.21 6261-6270

SIMMLER, M-C., ROUYER, F., VERNAUGD, G., NYSTROM-LAHTI, M., NGO, K.Y., DE LA CHAPELLE, A. & WEISSENBACH, J. (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317**, 692-697.

SIMPSON, E., CHANDLER, P., McLAREN, A., GOULMY, E., DISTECHE C.M., PAGE, D.C. & FERGUSON-SMITH, M.A. (1987). Mapping the H-Y gene. *Development 101 supplement* 157-161.

SINCLAIR, A.H., FOSTER, J.W., SPENCER, J.A., PAGE, D.C., PALMER, M.S., GOODFELLOW, P.N. & GRAVES, J.A.M. (1988) Sequences homologous to *ZFY*, a candidate human sex-determining gene, are autosomal in marsupials. *Nature,* **336**, 780-783

SINCLAIR, A.H., BERTA, P., PALMER, M.S., HAWKINS, J.R., GRIFFITHS, B.L., SMITH, M.J., FOSTER, J.W., FRISCHAUF, A., LOVELL-BADGE, R. & GODFELLOW, P.N. (1990) A gene from the human sex-determining region

encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240-244

SINGH, L. & JONES, K.W. (1982). Sex reversal in the mouse (*Mus musculus*) is caused by a recurrent nonreciprocal crossover involving the X and an aberrant Y chromosome. *Cell* **28**, 205-216.

SMITH, K.D., YOUNG, K. CONOVER TALBOT, C. & SCMECKPEPER, B.J. (1987) Repeated DNA of the human Y chromosome. *Development 101 supplement* 77-92

SOUTHERN, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517

STALLINGS, R., TORNEY, D., HILDEBRAND, C.E., LONGMIRE, J.L., DEAVEN, L., JETT, J., DOGGETT, N. & MOYZIS, K. (1990) Physical mapping of human chromosomes by repetitive sequence fingerprinting. *Proc. Natl. Acad. Sci.* **87**, 6218-6222

SULSTON, J., MALLETT, F., STADEN, R., DURBIN, R., HORSNELL, T. & COULSON, A. (1988) Software for genome mapping by fingerprinting techniques. *CABIOS* **4**, No 1, 125-132

SULSTON, J., MALLETT, F., DURBIN, R., HORSNELL, T.(1989) Image analysis of restriction enzyme fingerprint autoradiograms. *CABIOS* **5**, No 2, 101-106

SUTCLIFFE, M.J. & BURGOYNE, P.S. (1989) Analysis of the testes of H-Y negative XOSxr^{+} mice suggests that the spermatogenesis gene (*Spy*) acts during differentiation of the A spermatogonia. *Development*

107, 373-380

TIEPOLO, L. & ZUFFARDI, O. (1976). Localisation of factors controlling spermatogesis in the non-flourescent portion of the human Y chromosome long arm. *Human Genet.* **34**, 119-124.

TRASK, B., PINKEL, D. & VAN DER ENGH, G. (1989) The proximity of DNA sequences in interphase cell nuclei is correlated to genomic distance and permits ordering of cosmids spanning 250 kilobase pairs. *Genomics* **5**, 710-717

TYLER-SMITH, C. (1987) Structure of repeated sequences in the centromeric region of the human Y chromosome. *Development 101 supplement*, 93-100

ULINOWSKI, Z., TAYLOR, K., GRIFFIN, D., DELHANTY, J. & WOLFE, J. (1991) A highly polymorphic locus on human chromosome 12q14. *Ann. Hum. Genet.* In press.

UPHADAYA, M., SAFARAZI, M., BAMFORTH, J.S., THOMAS, N.S.T., OBERLE, I., YOUNG, I, & HARPER, P.S. (1986) Localisation of the gene for Hunter syndrome on the long arm of the X chromosome. *Hum. Genet.* **74**, 391-398

VERGNAUD, G., PAGE, D.C., SIMMLER, M-C., BROWN, L., ROUYER, F., NOEL, B., BOTSTEIN, D., DE LA CHAPELLE, A. & WEISSENBACH, J. (1986). A deletion map of the human Y chromosome based on DNA hybridisation. *Am. J. Hum. Genet.* 38, 109-124.

WADA, M., LITTLE, R.D., ABIDI, F., PORTA, G., LABELLA, T., COOPER, T., DELLA VALLE, G., D'URSO, M. & SCHLESSINGER, D. (1990) Human Xq24-Xq28: Approaches to mapping with yeast artificial chromosomes.

Am. j. Hum. Genet. **46,** 95-106

WEBER, J. (1990) Informativeness of human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ polymorphisms. *Genomics* 7, 524-530

WELSHONS, W. & RUSSELL, R. (1959) The Y chromosome as the bearer of male determining factors in the mouse. *Proc. Natl. Acad. Sci* 45 560-566

WOLFE, J., ERICKSON, R.P., RIGBY, P.W.J. & GOODFELLOW, P.N. (1984). Regional localisation of 3 Y-derived sequences on the human X and Y chromosomes. *Ann. Hum. Genet.* 48, 253-259

WOLFE, J., DARLING, S.M., ERICKSON, R.P., CRAIG, I.W., BUCKLE, V.J., RIGBY, P.W.J., WILLARD, H.F. & GOODFELLOW, P.N. (1985) Isolation and characterisation of an Alphoid centromeric repeat family from the human Y chromosome. *J.Mol. Biol.* **182,** 477-485

WONG, P., MACDONALD, I. & TENNISWOOD, M. (1989) Use of the polymerase chain reaction for the differential cross screening of libraries cloned into phage-lambda-based vectors. *Gene* **85,** 59-65

WRIGHT, J.M., & DIXON, G.H., (1988) Induction by torsional stress of an altered DNA conformation 5' upstream of the gene for a high mobility group protein from trout and specific bimding to flanking sequences by the gene product HMG-T. *Biochemistry,* **27,** 576-581

WYMAN, A.R. & WERTMAN, K.F. (1987) *Methods Enzymol.* **152,** 173-180

YEN, P.H., ALLEN, A., MARSH, B., MOHANDAS, T., WANG, N., TAGGART, R.T. & SHAPIRO, L. (1987). Cloning and expression of Steroid Sulfatase cDNA and the frequent occurrence of deletions in STS

deficiency: Implications for X-Y interchange. *Cell* **49**, 443-454

YOKOBATA, K., TRENCHAK, B. & DE JONG, P.J. (1991) Rescue of unstable cosmids by *in vitro* packaging. *Nuc. Acids Res.* **19**, 2 403-404

APPENDIX

Each clone in the libary is listed, with the following pieces of information;

a) **VAX number**: this is the number given to a clone when it has been fingerprinted and entered into the database.

b) **Contig number**: if the clone is contained within a contig, the contig number or letter is listed in this column.

c) and d) Cosmids which hybridise to the probes DYZ1 or DYZ3 are indicated with the letter Y.

e) **(CA)n**: Clones hybridising to the oligonucleotide $(CA)_9$ are described with the letter Z or D, refering to hybridisation on the hand-picked or dot blot filters respectively.

f) **Probe**: Positive hybridisation to other probes is recorded in this column.

g) ***in situ*** : some clones have been localised by *in situ* hybridisation, their location is recorded here.

h) **cDNA**: cosmids which hybridise to the adult or foetal testis cDNA probes described in section 3.1.3 are indicated.

This data will be very valuable in the processes of constructing contigs, confirming potential overlaps and in the positioning of contigs on the deletion map.

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.5 A1								
LOR 2.5 A2								
LOR 2.5 A3								
LOR 2.5 A4					D			
LOR 2.5 A5								
LOR 2.5 A6								
LOR 2.5 A7								
LOR 2.5 A8								
LOR 2.5 A9								
LOR 2.5 A10								
LOR 2.5 A11								
LOR 2.5 A12					D			
LOR 2.5 A13								
LOR 2.5 A14								
LOR 2.5 A15								
LOR 2.5 A16								
LOR 2.5 A17								
LOR 2.5 A18								
LOR 2.5 A19								
LOR 2.5 A20								
LOR 2.5 A21								
LOR 2.5 A22								
LOR 2.5 A23								
LOR 2.5 A24								
LOR 2.5 A25								
LOR 2.5 A26								
LOR 2.5 A27								
LOR 2.5 A28								
LOR 2.5 A29								
LOR 2.5 B1								
LOR 2.5 B2								
LOR 2.5 B3								
LOR 2.5 B4								
LOR 2.5 B5								
LOR 2.5 B6								
LOR 2.5 B7								
LOR 2.5 B8								
LOR 2.5 B9								
LOR 2.5 B10								
LOR 2.5 B11								
LOR 2.5 B12								
LOR 2.5 B13								
LOR 2.5 B14								
LOR 2.5 B15								
LOR 2.5 B16								
LOR 2.5 B17								
LOR 2.5 B18								
LOR 2.5 B19								
LOR 2.5 B20	47							
LOR 2.5 B21								
LOR 2.5 C1								
LOR 2.5 C2								
LOR 2.5 C3								
LOR 2.5 C4								
LOR 2.5 C5								
LOR 2.5 C6								
LOR 2.5 C7								
LOR 2.5 C8					DZ			
LOR 2.6 1								
LOR 2.6 2								
LOR 2.6 3					D			
LOR 2.6 4	26				Z		pMA3.3Y	
LOR 2.6 5								
LOR 2.6 6					Z			
LOR 2.6 7							GMGY39	
LOR 2.6 8								
LOR 2.6 9	46						GMGY39 & pO19p3	
LOR 2.6 10								
LOR 2.6 11								
LOR 2.6 12								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.6 13								
LOR 2.6 14								
LOR 2.6 15								
LOR 2.6 16								
LOR 2.6 17	48					DZ	GMGY46 & pO19p3	
LOR 2.6 18						Z	HFB2/E	
LOR 2.6 19						D		
LOR 2.6 20								
LOR 2.6 21								
LOR 2.6 22								
LOR 2.6 23								
LOR 2.6 24								
LOR 2.6 25								
LOR 2.6 26								
LOR 2.6 27								
LOR 2.6 28								
LOR 2.6 29								
LOR 2.6 30								
LOR 2.6 31				Y				
LOR 2.6 32								
LOR 2.6 33								
LOR 2.6 34								
LOR 2.6 35								
LOR 2.6 36								
LOR 2.6 37								
LOR 2.6 38								
LOR 2.6 39				Y		Z		
LOR 2.6 40								
LOR 2.6 41								
LOR 2.6 42								
LOR 2.6 43								
LOR 2.6 44								
LOR 2.6 45								
LOR 2.6 46								
LOR 2.6 47								
LOR 2.6 48								
LOR 2.6 49								
LOR 2.6 50								
LOR 2.6 51								
LOR 2.6 52								
LOR 2.6 53								
LOR 2.6 54								
LOR 2.6 55								
LOR 2.7 A1								
LOR 2.7 A2								
LOR 2.7 A3								
LOR 2.7 A4								
LOR 2.7 A5								
LOR 2.7 A6								
LOR 2.7 A7								
LOR 2.7 B1								
LOR 2.7 B2								
LOR 2.7 B3								
LOR 2.7 B4								
LOR 2.7 B5								
LOR 2.7 B6								
LOR 2.7 B7								
LOR 2.7 B8								
LOR 2.7 B9								
LOR 2.7 B10								
LOR 2.7 B11								
LOR 2.7 C1								
LOR 2.7 C2								
LOR 2.7 C3								
LOR 2.7 C4				Y				
LOR 2.7 C5								
LOR 2.7 C6								
LOR 2.7 C7								
LOR 2.7 C8								
LOR 2.7 C9								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.7 C10								
LOR 2.7 C11								
LOR 2.7 C12								
LOR 2.7 C13								
LOR 2.7 C14								
LOR 2.8 1					Z			
LOR 2.8 2								
LOR 2.8 3						p69/31		
LOR 2.8 4								
LOR 2.8 5								
LOR 2.8 6								
LOR 2.8 7								
LOR 2.8 8								
LOR 2.8 9					Z			
LOR 2.8 10								
LOR 2.8 11								
LOR 2.8 12								
LOR 2.8 13								
LOR 2.8 14								
LOR 2.8 15								
LOR 2.8 16								
LOR 2.8 17					Z			
LOR 2.8 18								
LOR 2.8 19					Z			
LOR 2.8 20					Z			
LOR 2.8 21								
LOR 2.8 22								
LOR 2.8 23								
LOR 2.8 24								
LOR 2.8 25								
LOR 2.8 26				Y				
LOR 2.8 27								
LOR 2.8 28								
LOR 2.8 29								
LOR 2.8 30								
LOR 2.8 31				Y				
LOR 2.8 32						p116\21		
LOR 2.8 33								
LOR 2.8 34								
LOR 2.8 35								
LOR 2.8 36				Y				
LOR 2.8 37					Z			
LOR 2.8 38					Z			
LOR 2.8 39								
LOR 2.8 40								
LOR 2.8 41								
LOR 2.8 42			Y		Z			
LOR 2.8 43								
LOR 2.8 44			Y					
LOR 2.8 45			Y					
LOR 2.8 46								
LOR 2.8 47								
LOR 2.8 48			Y					
LOR 2.8 49	774		Y					
LOR 2.8 50	773							
LOR 2.8 51	772							
LOR 2.8 52	771							
LOR 2.8 53	770							
LOR 2.8 54								
LOR 2.8 55	768							
LOR 2.8 56	767			Y				
LOR 2.8 57	766	b						
LOR 2.8 58	765							
LOR 2.8 59	764	t						
LOR 2.8 60	763							
LOR 2.8 61	762	u						
LOR 2.8 62	761							
LOR 2.8 63	760	v						
LOR 2.8 64	759	v						
LOR 2.8 65	758	w				GMGY26		

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.8 66	757							
LOR 2.8 67	756							
LOR 2.8 68	755			Y				
LOR 2.8 69	754	n				GMGY46 & GMGY14		
LOR 2.8 70	753	7				GMGXY8 & GMGY26		
LOR 2.8 71	752		Y					
LOR 2.8 72	751	31	Y					
LOR 2.8 73	620	45	Y					
LOR 2.8 74	619	45						
LOR 2.8 75	618	45						
LOR 2.8 76	617							
LOR 2.8 77	616	3						
LOR 2.8 78	615		Y					
LOR 2.8 79	614	31	Y					
LOR 2.8 80	613							
LOR 2.8 81	612							
LOR 2.8 82	611							
LOR 2.8 83	610					p75/79		
LOR 2.8 84	609		Y					
LOR 2.8 85	608					GMGXY8		
LOR 2.8 86	607		Y					
LOR 2.8 87	606		Y			GMGY46		
LOR 2.8 88	605	31	Y					
LOR 2.9 1	604	43	Y					
LOR 2.9 2	603							
LOR 2.9 3	602							
LOR 2.9 4	601							
LOR 2.9 5	600	3						
LOR 2.9 6	599	c						
LOR 2.9 7	598							
LOR 2.9 8	597							
LOR 2.9 9	544					GMGXY6		
LOR 2.9 10	543							
LOR 2.9 11	542							
LOR 2.9 12	541							
LOR 2.9 13	540							
LOR 2.9 14	539							
LOR 2.9 15	538							
LOR 2.9 16						GMGY14		
LOR 2.9 17	537							
LOR 2.9 18								
LOR 2.9 19	536	38						
LOR 2.9 20	535							
LOR 2.9 21	534	40				GMGY26		
LOR 2.9 22	533							
LOR 2.9 23	532					GMGY39		
LOR 2.9 24								
LOR 2.9 25								
LOR 2.9 26	531				Z	p116\21		
LOR 2.9 27	530							
LOR 2.9 28	529							
LOR 2.9 29	528	3						
LOR 2.9 30	527							
LOR 2.9 31	526							
LOR 2.9 32						GMGXY8		
LOR 2.9 33	525							
LOR 2.9 34	524							
LOR 2.9 35	523							
LOR 2.9 36								
LOR 2.9 37								
LOR 2.9 38								
LOR 2.9 39								
LOR 2.9 40								
LOR 2.9 41								
LOR 2.9 42								
LOR 2.9 43						24R6		
LOR 2.9 44								
LOR 2.9 45								
LOR 2.9 46								
LOR 2.9 47						p116\21		

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.9 48								
LOR 2.9 49								
LOR 2.9 50								
LOR 2.9 51								
LOR 2.9 52								
LOR 2.9 53								
LOR 2.9 54								
LOR 2.9 55								
LOR 2.9 56								
LOR 2.9 57								
LOR 2.9 58						GMGY46		
LOR 2.9 59								
LOR 2.9 60								
LOR 2.9 61								
LOR 2.9 62								
LOR 2.9 63								
LOR 2.9 64								
LOR 2.9 65								
LOR 2.9 66				Y				
LOR 2.9 67								
LOR 2.9 68					Z			
LOR 2.9 69					Z			
LOR 2.9 70			Y					
LOR 2.9 71								
LOR 2.9 72								
LOR 2.9 73								
LOR 2.9 74								
LOR 2.10 1								
LOR 2.10 2	27					pMA3.3Y		
LOR 2.10 3								
LOR 2.10 4								
LOR 2.10 5								
LOR 2.10 6								
LOR 2.10 7								
LOR 2.10 8								
LOR 2.10 9								
LOR 2.10 10								
LOR 2.10 11								
LOR 2.10 12								
LOR 2.10 13								
LOR 2.10 14								
LOR 2.10 15								
LOR 2.10 16					Z			
LOR 2.10 17								
LOR 2.10 18								
LOR 2.10 19								
LOR 2.10 20								
LOR 2.10 21								
LOR 2.10 22								
LOR 2.10 23								
LOR 2.10 24			Y					
LOR 2.10 25			Y					
LOR 2.11 1								
LOR 2.11 2			Y					
LOR 2.11 3			Y			pY6HSB65		
LOR 2.11 4			Y					
LOR 2.11 5			Y					
LOR 2.11 6			Y					
LOR 2.11 7			Y					
LOR 2.11 8			Y					
LOR 2.11 9			Y	Y				
LOR 2.11 10			Y					
LOR 2.11 11			Y					
LOR 2.11 12			Y			p116\21		
LOR 2.11 13								
LOR 2.11 14								
LOR 2.11 15								
LOR 2.11 16								
LOR 2.11 17								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.11 18								
LOR 2.13 1								
LOR 2.13 2								
LOR 2.13 3					Z			
LOR 2.13 4								
LOR 2.13 5					Z			
LOR 2.13 6								
LOR 2.13 7			Y					
LOR 2.13 8								
LOR 2.13 9								
LOR 2.13 10						p116\21		
LOR 2.13 11								
LOR 2.13 12								
LOR 2.13 13								
LOR 2.13 14						p116\21		
LOR 2.13 15								
LOR 2.13 16								
LOR 2.13 17						GMGY46		
LOR 2.13 18						p116\21		
LOR 2.13 19								
LOR 2.13 20								
LOR 2.13 21								
LOR 2.13 22						p116\21		
LOR 2.13 23						GMGY46		
LOR 2.13 24			Y					
LOR 2.13 25								
LOR 2.13 26	59							
LOR 2.13 27	60	8						
LOR 2.13 28	61	8						
LOR 2.13 29	62	39						
LOR 2.13 30	63							
LOR 2.13 31	64							
LOR 2.13 32	65	31						
LOR 2.13 33	66		Y					
LOR 2.13 34	67		Y					
LOR 2.13 35	68		Y					
LOR 2.13 36	69		Y					
LOR 2.13 37			Y					
LOR 2.13 38			Y					
LOR 2.13 39								
LOR 2.13 40								
LOR 2.13 41						GMGY37		
LOR 2.13 42								
LOR 2.13 43								
LOR 2.13 44								
LOR 2.13 45								
LOR 2.13 46								
LOR 2.13 47								
LOR 2.13 48								
LOR 2.13 49								
LOR 2.13 50								
LOR 2.13 51								
LOR 2.13 52								
LOR 2.14 1								
LOR 2.14 2								
LOR 2.14 3								
LOR 2.14 4			Y					
LOR 2.14 5			Y					
LOR 2.14 6								
LOR 2.14 7								
LOR 2.14 8	130							
LOR 2.14 9	129	s						
LOR 2.14 10	128		Y		Z			
LOR 2.14 11	127		Y					
LOR 2.14 12	126	z	Y					
LOR 2.14 13	125							
LOR 2.14 14	859							
LOR 2.14 15	860							
LOR 2.14 16	858							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.14 17	857							
LOR 2.14 18	856							
LOR 2.14 19	855							
LOR 2.14 20	854							
LOR 2.14 21	853							
LOR 2.14 22	852							
LOR 2.14 23	851							
LOR 2.14 24	850							
LOR 2.14 25	849							
LOR 2.14 26								
LOR 2.14 27	848							
LOR 2.14 28	847		Y					
LOR 2.14 29								
LOR 2.14 30	846		Y					
LOR 2.14 31	845		Y					
LOR 2.14 32	844							
LOR 2.14 33								
LOR 2.14 34								
LOR 2.14 35								
LOR 2.14 36								
LOR 2.14 37								
LOR 2.14 38			Y					
LOR 2.14 39								
LOR 2.14 40			Y					
LOR 2.14 41								
LOR 2.14 42								
LOR 2.14 43								
LOR 2.14 44								
LOR 2.14 45								
LOR 2.14 46								
LOR 2.14 47								
LOR 2.14 48								
LOR 2.14 49								
LOR 2.14 50								
LOR 2.14 51							p116\21	
LOR 2.14 52								
LOR 2.14 53								
LOR 2.14 54								
LOR 2.14 55								
LOR 2.14 56								
LOR 2.14 57	105							
LOR 2.14 58	104							
LOR 2.14 59	103							
LOR 2.14 60	102						p75/79	
LOR 2.14 61	101							
LOR 2.14 62	100							
LOR 2.14 63	99							
LOR 2.14 64	98	30					GMGY26	
LOR 2.14 65	97	a						
LOR 2.14 66	96	38						
LOR 2.14 67	95	b			Z		GMGY26	
LOR 2.14 68	94	c						
LOR 2.14 69	93							
LOR 2.14 70	92		Y				24R6	
LOR 2.14 71	91							
LOR 2.14 72	90	7						
LOR 2.14 73	89							
LOR 2.14 74	88							
LOR 2.14 75	87						p116\21	
LOR 2.14 76	86							
LOR 2.14 77	85	a						
LOR 2.14 78	84							
LOR 2.14 79	83							
LOR 2.14 80	82							
LOR 2.14 81	644							
LOR 2.14 82	643							
LOR 2.14 83	642							
LOR 2.14 84	641	m						
LOR 2.14 85	640							
LOR 2.14 86	639	7					GMGXY8 & GMGY26	
LOR 2.14 87	638							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.14 88	637					GMGY46		
LOR 2.14 89	636							
LOR 2.14 90	635							
LOR 2.14 91	634							
LOR 2.14 92	633							
LOR 2.14 93	632							
LOR 2.14 94	631							
LOR 2.14 95	630	w		Z		GMGY26		
LOR 2.14 96	629							
LOR 2.14 97	628	31						
LOR 2.14 98	627							
LOR 2.14 99	626							
LOR 2.14 100	625							
LOR 2.14 101	624							
LOR 2.14 102	623					GMGY15		
LOR 2.15 1	622	46						
LOR 2.15 2	621					HFB2/E		
LOR 2.15 3								
LOR 2.15 4								
LOR 2.15 5								
LOR 2.15 6								
LOR 2.15 7								
LOR 2.15 8								
LOR 2.15 9								
LOR 2.15 10	667	31				HFB2/E & GMGY8		
LOR 2.15 11								
LOR 2.15 12	668							
LOR 2.15 13								
LOR 2.15 14	669							
LOR 2.15 15			Y					
LOR 2.15 16			Y					
LOR 2.15 17	769	m	Y					
LOR 2.15 18	671							
LOR 2.15 19	672	31				GMGY26		
LOR 2.15 20	673							
LOR 2.15 21	674					GMGXY8		
LOR 2.15 22	675							
LOR 2.15 23						24R6		
LOR 2.15 24								
LOR 2.15 25								
LOR 2.15 26	596							
LOR 2.15 27	595							
LOR 2.15 28	594							
LOR 2.15 29								
LOR 2.15 30	593							
LOR 2.15 31	592	42						
LOR 2.15 32	591							
LOR 2.15 33	590							
LOR 2.15 34	589	x	Y					
LOR 2.15 35								
LOR 2.15 36	588							
LOR 2.15 37	587		Y			pY6HSB65		
LOR 2.15 38	586							
LOR 2.15 39	585							
LOR 2.15 40	584							
LOR 2.15 41	583							
LOR 2.15 42	582							
LOR 2.15 43	581			Y	Z			
LOR 2.15 44	580							
LOR 2.15 45	579	11						
LOR 2.15 46	578							
LOR 2.15 47	577							
LOR 2.15 48	576					GMGXY8 & CI-1.2	Yq11	
LOR 2.15 49	575	31						
LOR 2.15 50						CI-1.2	Yp11,Yq11	
LOR 2.15 51						24R6		
LOR 2.15 52								
LOR 2.15 53								
LOR 2.15 54								
LOR 2.15 55								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.15 56					Z			
LOR 2.15 57			Y					
LOR 2.15 58			Y					
LOR 2.15 59						24R6 & CI-1.2 Yq11.23,1q31-32.1		
LOR 2.15 60								
LOR 2.15 61								
LOR 2.15 62								
LOR 2.15 63								
LOR 2.15 64								
LOR 2.15 65								
LOR 2.15 66								
LOR 2.15 67								
LOR 2.15 68								
LOR 2.15 69								
LOR 2.15 70								
LOR 2.15 71								
LOR 2.15 72								
LOR 2.15 73								
LOR 2.15 74	243							
LOR 2.15 75	242							
LOR 2.15 76	42					GMGY29		
LOR 2.15 76	241							
LOR 2.15 77	240					GMGY29		
LOR 2.15 78	239	o				GMGY29		
LOR 2.15 79	238	p						
LOR 2.15 80	237							
LOR 2.15 81	236							
LOR 2.15 82	235	38						
LOR 2.15 83	234							
LOR 2.15 84	233							
LOR 2.15 85	232							
LOR 2.15 86	231							
LOR 2.15 87	230				Z			foetal testis
LOR 2.15 88	229	38						
LOR 2.15 89								
LOR 2.15 90								
LOR 2.15 91								
LOR 2.15 92	227							
LOR 2.15 93	226				Z			
LOR 2.15 94								
LOR 2.15 95	225		Y					
LOR 2.15 96	224							
LOR 2.15 97	223							
LOR 2.15 98								
LOR 2.15 99	333							
LOR 2.15 100	332							
LOR 2.15 101								
LOR 2.15 102								
LOR 2.15 103	331							
LOR 2.15 104	330							
LOR 2.15 105	329							
LOR 2.15 106	328	22						
LOR 2.15 107	327	21						
LOR 2.15 108	326	20	Y					
LOR 2.15 109	325							
LOR 2.15 110	324							
LOR 2.15 111	323							
LOR 2.15 112	322							
LOR 2.15 113	321		Y					
LOR 2.15 114	320	22	Y					
LOR 2.15 115	319							
LOR 2.15 116	318	20						
LOR 2.15 117	317							
LOR 2.15 118	316							
LOR 2.15 119	315							
LOR 2.15 120	314							
LOR 2.15 121	313							
LOR 2.15 122	312							
LOR 2.15 123	311							
LOR 2.15 124	310				Z			
LOR 2.15 125	309							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.15 126	308							
LOR 2.15 127	307							
LOR 2.15 128	306	31						
LOR 2.15 129	305							
LOR 2.15 130	304							
LOR 2.15 131	303	31	Y					
LOR 2.15 132	302	31						
LOR 2.15 133	301							
LOR 2.15 134	299							
LOR 2.15 135	298							
LOR 2.15 136	296	18						
LOR 2.15 137	297	18						
LOR 2.15 138	295							
LOR 2.15 139	294							
LOR 2.15 140	293							
LOR 2.15 141					Z			
LOR 2.15 142			Y					
LOR 2.15 143								
LOR 2.15 144								
LOR 2.15 145								
LOR 2.15 146								
LOR 2.15 147	286	35						
LOR 2.15 148						GMGY46		
LOR 2.15 149	285							
LOR 2.15 150	284							
LOR 2.15 151	283	23						
LOR 2.15 152	282							
LOR 2.15 153	281							
LOR 2.15 154	280							
LOR 2.15 155	279		Y					
LOR 2.15 156	278					p75/79		
LOR 2.15 157	277							
LOR 2.15 158	276							
LOR 2.15 159	275	i						
LOR 2.15 160	274	g						
LOR 2.15 161	273							
LOR 2.15 162	272	17						
LOR 2.15 163	271		Y					
LOR 2.15 164	270	31	Y					
LOR 2.15 165	269		Y					
LOR 2.15 166	268	31	Y		Z			
LOR 2.15 167	267				Z			
LOR 2.15 168	266			Y				
LOR 2.15 169	1	1						
LOR 2.15 169	265	1						
LOR 2.15 170	2	1						
LOR 2.15 170	264	1						
LOR 2.15 171	3							
LOR 2.15 172	4	7			Z	GMGY26		
LOR 2.15 173	5							
LOR 2.15 174	6							
LOR 2.15 175	7	13						
LOR 2.15 176	8	13						
LOR 2.15 177	9	1						
LOR 2.15 178	10	1						
LOR 2.15 179	11	1						
LOR 2.15 180	12	1				GMGY26		
LOR 2.15 181	13		Y			p116\21		
LOR 2.15 182	14	1	Y					
LOR 2.15 183	15	3						
LOR 2.16 1	16	3		Y				
LOR 2.16 2	17							
LOR 2.16 3	18					p116\21		
LOR 2.16 4	19							
LOR 2.16 5	20							
LOR 2.16 6	21							
LOR 2.16 7	22					GMGY46 & GMGY39 & p019p3		
LOR 2.16 7	45							
LOR 2.16 8	23							
LOR 2.16 9	24							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.16 10								
LOR 2.16 11						GMGY39&p116\21		
LOR 2.16 12			Y			GMGY39&p116\21		
LOR 2.16 13						GMGY39&p116\21		
LOR 2.16 14								
LOR 2.16 15								
LOR 2.16 16								
LOR 2.16 17								
LOR 2.16 18								
LOR 2.16 19			Y					
LOR 2.16 20								
LOR 2.16 21								
LOR 2.16 22								
LOR 2.16 23			Y					
LOR 2.16 24								
LOR 2.16 25								
LOR 2.16 26								
LOR 2.16 27								
LOR 2.16 28								
LOR 2.16 29					Z			
LOR 2.16 30								
LOR 2.16 31								
LOR 2.16 32								
LOR 2.16 33								
LOR 2.16 34	40	26			Z	GMGY26		
LOR 2.16 35								
LOR 2.16 36					Z			
LOR 2.16 37								
LOR 2.16 38								
LOR 2.16 39								
LOR 2.16 40					Z			
LOR 2.16 41								
LOR 2.16 42								
LOR 2.16 43	70							
LOR 2.16 44	82				Z			
LOR 2.16 45								
LOR 2.16 46								
LOR 2.16 47						GMGY46		
LOR 2.16 48								
LOR 2.16 49								
LOR 2.16 50								
LOR 2.16 51								
LOR 2.16 52								
LOR 2.16 53								
LOR 2.16 54						PAX		
LOR 2.16 55								
LOR 2.16 56								
LOR 2.16 57			Y					
LOR 2.16 58								
LOR 2.16 59								
LOR 2.16 60								
LOR 2.16 61								
LOR 2.16 62								
LOR 2.16 63								
LOR 2.16 64					Z			
LOR 2.16 65								
LOR 2.16 66								
LOR 2.16 67								
LOR 2.16 68								
LOR 2.16 69								
LOR 2.16 70								
LOR 2.16 71	71	38						
LOR 2.16 72	72							
LOR 2.16 73	73							
LOR 2.16 74	74							
LOR 2.16 75								
LOR 2.16 76	75							
LOR 2.16 77	76					pY6HSB65		
LOR 2.16 78	77							
LOR 2.16 79	78							
LOR 2.16 80	79							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.16 81	80							
LOR 2.16 82	81							
LOR 2.16 83	263							
LOR 2.16 84	262							
LOR 2.16 85	261							
LOR 2.16 86	260							
LOR 2.16 87	259							
LOR 2.16 88	258							
LOR 2.16 89	257		k					
LOR 2.16 90	256							
LOR 2.16 91	255							
LOR 2.16 92	254	16				GMGY8		
LOR 2.17 1								
LOR 2.17 2								
LOR 2.17 3								
LOR 2.17 4								
LOR 2.17 5								
LOR 2.17 6						pY6HSB65		
LOR 2.17 7								
LOR 2.17 8								
LOR 2.17 9						p116\21		
LOR 2.17 10								
LOR 2.17 11						GMGY37		
LOR 2.17 12						p116\21		
LOR 2.17 13								
LOR 2.17 14								
LOR 2.17 15								
LOR 2.17 16								
LOR 2.17 17								
LOR 2.17 18								
LOR 2.17 19								
LOR 2.17 20								
LOR 2.17 21								
LOR 2.17 22								
LOR 2.18 1	253	16						
LOR 2.18 2						p116\21		
LOR 2.18 3								
LOR 2.18 4	252					p116\21		
LOR 2.18 5	251					GMGXY8&p116\21		
LOR 2.18 6	250							
LOR 2.18 7	249							foetal testis
LOR 2.18 8	248							foetal testis
LOR 2.18 9	247					CI-1.2	12q21.2	foetal testis
LOR 2.18 10								foetal testis
LOR 2.18 11								
LOR 2.18 12								foetal testis
LOR 2.18 13								
LOR 2.18 14								
LOR 2.18 15								
LOR 2.18 16								
LOR 2.18 17				Z				
LOR 2.18 18			Y					
LOR 2.18 19								
LOR 2.18 20								
LOR 2.18 21								
LOR 2.18 22						p75/79		
LOR 2.18 23								
LOR 2.18 24								
LOR 2.18 25								
LOR 2.18 26								
LOR 2.18 27				Z				
LOR 2.18 28								
LOR 2.18 29								
LOR 2.18 30	25					pMA3.3Y		
LOR 2.18 31								
LOR 2.18 32								
LOR 2.18 33								
LOR 2.18 34								
LOR 2.18 35								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.18 36								
LOR 2.18 37								
LOR 2.18 38								
LOR 2.18 39								
LOR 2.18 40								
LOR 2.18 41								
LOR 2.18 42						GMGY20		
LOR 2.18 43								
LOR 2.18 44								
LOR 2.18 45						p75/79		
LOR 2.18 46								
LOR 2.18 47						GMGY20 & p69/31		
LOR 2.18 48						GMGY20		
LOR 2.18 49								
LOR 2.18 50								
LOR 2.18 51								
LOR 2.18 52								
LOR 2.18 53								
LOR 2.18 54								
LOR 2.18 55								
LOR 2.18 56								
LOR 2.18 57								
LOR 2.18 58			Y					
LOR 2.18 59			Y					
LOR 2.18 60								
LOR 2.18 61								
LOR 2.18 62								
LOR 2.18 63								
LOR 2.18 64								
LOR 2.18 65	107			Y				
LOR 2.18 66				Y				
LOR 2.18 67	106			Y				
LOR 2.18 68								
LOR 2.18 69	58							
LOR 2.18 70	57	7		Y				
LOR 2.19 1	56	14		Y				
LOR 2.19 2	55	7		Y		GMGY26		
LOR 2.19 3	54					p75/79		
LOR 2.19 4	53	b						
LOR 2.19 5	52	32						
LOR 2.19 6	51							
LOR 2.19 7	50							
LOR 2.19 8	49	7		Y				
LOR 2.19 9						p116\21		
LOR 2.19 10				Y				
LOR 2.19 11	135							
LOR 2.19 12	134					CI-1.2	Yq11	foetal testis
LOR 2.19 13	133							
LOR 2.19 13	136							
LOR 2.19 14	132							
LOR 2.19 15	131							
LOR 2.19 16					Z			
LOR 2.19 17	174				Z			
LOR 2.19 18	173				Z			
LOR 2.19 19	172	46						
LOR 2.19 20	171							
LOR 2.19 21	170							
LOR 2.19 22	169	13						
LOR 2.19 23	168	13						
LOR 2.19 24	167							
LOR 2.19 25	166	21						
LOR 2.19 26	165							
LOR 2.19 27	41	12	Y			GMGY20	1p35	
LOR 2.19 27	164	12	Y					
LOR 2.19 28	163	37						
LOR 2.19 29	162							
LOR 2.19 30	161							
LOR 2.19 31	160							
LOR 2.19 32	159					GMGXY6		
LOR 2.19 33	158	11				GMGY46 & p019p3		

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.19 34	157							
LOR 2.19 35	156	12						
LOR 2.19 36	155							
LOR 2.19 37	153					p75/79		
LOR 2.19 38	44					p019p3		
LOR 2.19 38	152							
LOR 2.19 39	151					GMGY14		
LOR 2.19 40	198					GMGY14 & p75/79		
LOR 2.19 41	197							
LOR 2.19 42	196					GMGY46		
LOR 2.19 43								
LOR 2.19 44	154	g						
LOR 2.19 45	195							
LOR 2.19 46	194				Z			foetal testis
LOR 2.19 47	193	38						
LOR 2.19 48	192	37						
LOR 2.19 49	191							
LOR 2.19 50	190					GMGY46&p116\21		
LOR 2.19 51								
LOR 2.19 52	189			Y				
LOR 2.19 53	188							
LOR 2.19 54	187							
LOR 2.19 55	186					GMGY46		
LOR 2.19 56	185				Z			
LOR 2.19 57	184				Z			
LOR 2.19 58	183					pY6HSB65		
LOR 2.19 59	182							
LOR 2.19 60	181				Z	GMGY20		
LOR 2.19 61	180							
LOR 2.19 62	179							
LOR 2.19 63	178	36						
LOR 2.19 64	177	35						
LOR 2.19 65	176							
LOR 2.19 66	175							
LOR 2.19 67	222							
LOR 2.19 68	221							
LOR 2.19 69	220							
LOR 2.19 70	219							
LOR 2.19 71	218							
LOR 2.19 72	217	32						
LOR 2.19 73	216							
LOR 2.19 74	215	20						
LOR 2.19 75	214					p116\21		
LOR 2.19 76	213							
LOR 2.19 77	212							
LOR 2.19 78	211	14						
LOR 2.19 79	210							
LOR 2.19 80	209							
LOR 2.19 81	208							
LOR 2.19 82	207							
LOR 2.19 83	206							
LOR 2.19 84	205	7				GMGY26		
LOR 2.19 85	204							
LOR 2.19 86	203							
LOR 2.19 87	202				Z			
LOR 2.19 88	201							
LOR 2.19 89	200							
LOR 2.19 90	199							
LOR 2.19 91	246	q						
LOR 2.19 92	245							
LOR 2.19 93	244							
LOR 2.19 94	574		Y					
LOR 2.19 95	573							
LOR 2.19 96	572	f		Y				
LOR 2.19 97	571	40				GMGY26		
LOR 2.19 98	570	43						
LOR 2.19 99	569							
LOR 2.19 100								
LOR 2.19 101	568							
LOR 2.19 102	567							
LOR 2.19 103	566							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.19 104	565		Y					
LOR 2.19 105	564	31	Y					
LOR 2.19 106	563							
LOR 2.19 107	562		Y					
LOR 2.19 108	561	31	Y	Y				
LOR 2.19 109								
LOR 2.19 110	560	b						
LOR 2.19 111	559							
LOR 2.19 112	558	39						
LOR 2.19 113	557							
LOR 2.19 114	886							
LOR 2.19 115	885							
LOR 2.19 116	884					GMGY20		
LOR 2.19 117	883							
LOR 2.19 118	882							
LOR 2.19 119	881							
LOR 2.19 120	880					p69/31		
LOR 2.19 121	879			Y				
LOR 2.19 122	878					GMGY15		
LOR 2.19 123	877							
LOR 2.21 1	876							
LOR 2.21 2	875							
LOR 2.21 3	874							
LOR 2.21 4	873							
LOR 2.21 5	872							
LOR 2.21 6								
LOR 2.21 7	871							
LOR 2.21 8	870							
LOR 2.21 9	869							
LOR 2.21 10	868					GMGY37		
LOR 2.21 11	867							
LOR 2.21 12								
LOR 2.21 13								
LOR 2.21 14								
LOR 2.21 15	556							
LOR 2.21 16	555							
LOR 2.21 17								
LOR 2.21 18	554							
LOR 2.21 18	147							
LOR 2.21 19					Z			
LOR 2.21 20					Z			
LOR 2.21 21	553							
LOR 2.21 22	552	42						
LOR 2.21 23								
LOR 2.21 24								
LOR 2.21 25	551							
LOR 2.21 26	550							
LOR 2.21 26	150							
LOR 2.21 27	549							
LOR 2.21 27	149							
LOR 2.21 28	548							
LOR 2.21 28	148							
LOR 2.21 29	547	i						
LOR 2.21 30								
LOR 2.21 31	546							
LOR 2.21 32								
LOR 2.21 33	545	11				p019p3		
LOR 2.21 33	146	11				p019p3		
LOR 2.21 34								
LOR 2.21 35								
LOR 2.21 36								
LOR 2.21 37								
LOR 2.21 38								
LOR 2.21 39	145			Y				
LOR 2.21 40	144			Y				
LOR 2.21 41	143			Y				
LOR 2.21 42	142							
LOR 2.21 43	141							
LOR 2.21 44								
LOR 2.21 45								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.21 46								
LOR 2.21 47								
LOR 2.21 48								
LOR 2.21 49	140	b						
LOR 2.21 50								
LOR 2.21 51					Z			
LOR 2.21 52								
LOR 2.21 53	124							
LOR 2.21 54	123			Y				
LOR 2.21 55	139	7						
LOR 2.21 56	121							
LOR 2.21 57	120				Z			
LOR 2.21 58	119			Y				
LOR 2.21 59	118							
LOR 2.21 60	117					p75/79		
LOR 2.21 61	116							
LOR 2.21 62	115				Z			
LOR 2.21 63	138	7			Z	GMGY26		
LOR 2.21 64	113	17						
LOR 2.21 65	112							
LOR 2.21 66	111							
LOR 2.21 67	110							
LOR 2.22 1	109							
LOR 2.22 2	108					GMGY29	1p35	
LOR 2.22 2	38						1p35	
LOR 2.22 3					Z			
LOR 2.22 4						24R6		
LOR 2.22 5								
LOR 2.22 6								
LOR 2.22 7								
LOR 2.22 8								
LOR 2.22 9			Y	Y				
LOR 2.22 10				Y				
LOR 2.22 11			Y					
LOR 2.22 12				Y				
LOR 2.22 13								
LOR 2.22 14								
LOR 2.22 15								
LOR 2.22 16								
LOR 2.22 17								
LOR 2.22 18	34					GMGY20 & p69/31		
LOR 2.22 19								
LOR 2.22 20								
LOR 2.22 21						24R6		
LOR 2.22 22								
LOR 2.22 23								
LOR 2.22 24								
LOR 2.22 25								
LOR 2.22 26								
LOR 2.22 27			Y					
LOR 2.22 28				Y				
LOR 2.22 29				Y		GMGY46		
LOR 2.22 30				Y				
LOR 2.22 31				Y				
LOR 2.22 32								
LOR 2.22 33								
LOR 2.22 34								
LOR 2.22 35								
LOR 2.22 36						24R6		
LOR 2.22 37								
LOR 2.22 38								
LOR 2.22 39								
LOR 2.22 40								
LOR 2.22 41								
LOR 2.22 42								
LOR 2.22 43								
LOR 2.22 44								
LOR 2.22 45								
LOR 2.22 46								
LOR 2.22 47						p75/79		

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.22 48			Y					
LOR 2.22 49			Y					
LOR 2.23 1			Y					
LOR 2.23 2			Y		Z			
LOR 2.23 3								
LOR 2.23 4								
LOR 2.23 5								
LOR 2.23 6								
LOR 2.23 7			Y					
LOR 2.23 8								
LOR 2.23 9			Y					
LOR 2.23 10			Y					
LOR 2.23 11			Y					
LOR 2.23 12								
LOR 2.23 13								
LOR 2.23 14								
LOR 2.23 15								
LOR 2.23 16								
LOR 2.23 17								
LOR 2.23 18								
LOR 2.23 19								
LOR 2.23 20								
LOR 2.23 21								
LOR 2.23 22								
LOR 2.23 23								
LOR 2.23 24					Z			
LOR 2.23 25			Y			p116\21		
LOR 2.23 26								
LOR 2.23 27								
LOR 2.23 28								
LOR 2.23 29						p75/79		
LOR 2.23 30								
LOR 2.23 31								
LOR 2.23 32			Y					
LOR 2.23 33			Y					
LOR 2.23 34			Y					
LOR 2.23 35			Y					
LOR 2.23 36								
LOR 2.23 37								
LOR 2.23 38								
LOR 2.23 39								
LOR 2.23 40								
LOR 2.23 41								
LOR 2.23 42								
LOR 2.23 43								
LOR 2.23 44								
LOR 2.23 45								
LOR 2.23 46								
LOR 2.23 47								
LOR 2.23 48								
LOR 2.23 49								
LOR 2.23 50								
LOR 2.23 51	354							
LOR 2.23 52	353							
LOR 2.23 53	352							
LOR 2.23 54	351							
LOR 2.23 55	350							
LOR 2.23 56	349	33				GMGY8		
LOR 2.23 57	348							
LOR 2.23 58	347	38						
LOR 2.23 59	346	31						
LOR 2.23 60	345							
LOR 2.23 61	344	31						
LOR 2.23 62	343					PAX		
LOR 2.23 63	342	43						
LOR 2.23 64	341	31						
LOR 2.23 65	340	46						
LOR 2.23 66	750							
LOR 2.23 67	339			Y				
LOR 2.23 68	338			Y				

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.23 69	337							
LOR 2.23 70	336							
LOR 2.23 71	426	y				p116\21		
LOR 2.23 72	425							
LOR 2.23 73	335	7				GMGXY8		
LOR 2.23 74	334			Z				
LOR 2.23 75	379							
LOR 2.23 76	378							
LOR 2.23 77	377							
LOR 2.23 78	376					GMGXY8		
LOR 2.23 79	375					GMGXY8		
LOR 2.23 80	374							
LOR 2.23 81	373	26		Z		GMGY26		
LOR 2.23 82	372							
LOR 2.23 83	371							
LOR 2.23 84	370							
LOR 2.23 85	369	32						
LOR 2.23 86	368							
LOR 2.23 87	367							
LOR 2.23 88	366	j						
LOR 2.23 89	365	31						
LOR 2.23 90	364							
LOR 2.23 91	363	7						
LOR 2.23 92	362					GMGY46		
LOR 2.23 93	361							
LOR 2.23 94	360							
LOR 2.23 95	359	u						
LOR 2.23 96	358							
LOR 2.23 97	357	23						
LOR 2.23 98	356							
LOR 2.23 99	380							
LOR 2.23 100	381							
LOR 2.23 101	382	d						
LOR 2.23 102	384							
LOR 2.23 103	385	31						
LOR 2.23 104	386	46						
LOR 1.1 1	396	33				GMGY8		
LOR 1.1 2	397	13		D		GMGY39		
LOR 1.1 3	398							
LOR 1.1 4	399			D		GMGY15		
LOR 1.1 5	400							
LOR 1.1 6	401	t				GMGY8		
LOR 1.1 7	424							
LOR 1.1 8	423					p75/79		
LOR 1.1 9	402							
LOR 1.1 10	422							
LOR 1.1 11								
LOR 1.1 12								
LOR 1.1 13								
LOR 1.1 14								
LOR 1.1 15	421							
LOR 1.1 16								
LOR 1.1 17								
LOR 1.1 18				D				
LOR 1.1 19								
LOR 1.1 20	420							
LOR 1.1 21	419	l						
LOR 1.1 22								
LOR 1.1 23								
LOR 1.1 24	418							
LOR 1.1 25	417	r						
LOR 1.1 26	416							
LOR 1.1 27								
LOR 1.1 28	415							
LOR 1.1 29	414							
LOR 1.1 30	413							
LOR 1.1 31								
LOR 1.1 32	412	1						
LOR 1.1 33	411	1						
LOR 1.1 34	410							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 1.1 35	409							
LOR 1.1 36	408							
LOR 1.1 37	407				D			
LOR 1.1 38	406	1						
LOR 1.1 39	405	30				GMGY26		
LOR 1.1 40								
LOR 1.1 41	404							
LOR 1.1 42	403	1						
LOR 1.1 43	426							
LOR 1.1 44								
LOR 1.1 45	427							
LOR 1.1 46	428							
LOR 1.1 47								
LOR 1.1 48								
LOR 1.1 49	429							
LOR 1.1 50	430							
LOR 1.1 51	431							
LOR 1.1 52	432							
LOR 1.1 53	433				D			
LOR 1.1 54	434							
LOR 1.1 55	435				D			
LOR 1.1 56	436							
LOR 1.1 57	437							
LOR 1.1 58	438							
LOR 1.1 59	439							
LOR 1.1 60								
LOR 1.1 61								
LOR 1.1 62					DZ			
LOR 1.1 63	440							
LOR 1.1 64	441				D			
LOR 1.1 65	442				D			
LOR 1.1 66								
LOR 1.1 67	443				D			
LOR 1.1 68	444							
LOR 1.1 69								
LOR 1.1 70	445							
LOR 1.1 71					D			
LOR 1.1 72	446							
LOR 1.1 73	447							
LOR 1.1 74								
LOR 1.1 75	448							
LOR 1.1 76	449			Y				
LOR 1.1 77								
LOR 1.1 78	450							
LOR 1.1 79	451	e						
LOR 1.1 80	452							
LOR 1.1 81	453							
LOR 1.1 82	454							
LOR 1.1 83								
LOR 1.1 84	455							
LOR 1.1 85	456							
LOR 1.1 86	457							
LOR 1.1 87	458				D			
LOR 1.1 88	459							
LOR 1.1 89	460							
LOR 1.1 90	461							
LOR 1.1 91	462				D			
LOR 1.1 93	463				D			
LOR 1.1 94	464				D			
LOR 1.1 95	465			Y				
LOR 1.1 96	466							
LOR 1.1 97	498							
LOR 1.1 97	467							
LOR 1.1 98	468							
LOR 1.1 99	469							
LOR 1.1 100	470							
LOR 1.1 101	471				D			
LOR 1.1 102	472							
LOR 1.1 103	473				D			
LOR 1.1 104	497							
LOR 1.1 105	496	d						

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 1.1 106								
LOR 1.1 107	495			D				
LOR 1.1 108	494							
LOR 1.1 109	493							
LOR 1.1 110	492							
LOR 1.1 111	491							
LOR 1.1 112	490							
LOR 1.1 113	489			D				
LOR 1.1 114	488			D				
LOR 1.1 115	487							
LOR 1.1 116	486							
LOR 2.1 1				D				
LOR 2.1 2								
LOR 2.1 3								
LOR 2.1 4	485							
LOR 2.1 5								
LOR 2.1 6								
LOR 2.1 7								
LOR 2.1 8				DZ				
LOR 2.1 9								
LOR 2.1 10								
LOR 2.1 11	484							
LOR 2.1 12								
LOR 2.1 13				Z				
LOR 2.1 14								
LOR 2.1 15								
LOR 2.1 16								
LOR 2.1 17								
LOR 2.1 18								
LOR 2.1 19				DZ				
LOR 2.1 20				DZ				
LOR 2.1 21								
LOR 2.1 22						HFB2/E		
LOR 2.1 23								
LOR 2.1 24								
LOR 2.1 25								
LOR 2.1 26								
LOR 2.1 27				Z				
LOR 2.1 28								
LOR 2.1 29								
LOR 2.1 30				Z				
LOR 2.1 31								
LOR 2.1 32								
LOR 2.1 33				DZ				
LOR 2.1 34				D				
LOR 2.1 35								
LOR 2.1 36								
LOR 2.1 37								
LOR 2.1 38								
LOR 2.1 39								
LOR 2.1 40								
LOR 2.1 41								
LOR 2.1 42								
LOR 2.1 43				Z				
LOR 2.1 44								
LOR 2.1 45								
LOR 2.1 46								
LOR 2.1 47				Z				
LOR 2.1 48								
LOR 2.1 49								
LOR 2.1 50								
LOR 2.1 51				D				
LOR 2.1 52								
LOR 2.1 53								
LOR 2.1 54								
LOR 2.1 55								
LOR 2.1 56								
LOR 2.1 57				D				
LOR 2.1 58				D				
LOR 2.1 59								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.1 60								
LOR 2.1 61								
LOR 2.1 62								
LOR 2.1 63								
LOR 2.1 64								
LOR 2.1 65								
LOR 2.1 66								
LOR 2.1 67								
LOR 2.1 68								
LOR 2.1 69								
LOR 2.1 70								
LOR 2.1 71								
LOR 2.1 72								
LOR 2.1 73				D				
LOR 2.1 74	39					GMGY29		
LOR 2.1 75								
LOR 2.1 76								
LOR 2.1 77								
LOR 2.1 78								
LOR 2.1 79								
LOR 2.1 80								
LOR 2.1 81								
LOR 2.1 82				Z				
LOR 2.1 83				Z				
LOR 2.1 84								
LOR 2.1 85								
LOR 2.1 86								
LOR 2.1 87								
LOR 2.1 88								
LOR 2.1 89								
LOR 2.1 90				Z				
LOR 2.1 91								
LOR 2.1 92								
LOR 2.1 93								
LOR 2.24 1								
LOR 2.24 2								
LOR 2.24 3								
LOR 2.24 4								
LOR 2.24 5								
LOR 2.24 6				Z				
LOR 2.24 7								
LOR 2.24 8								
LOR 2.24 9								
LOR 2.24 10								
LOR 2.24 11								
LOR 2.24 12								
LOR 2.24 13				D				
LOR 2.25 1						GMGY20 & p69/31		
LOR 2.25 2						GMGY17		
LOR 2.25 3								
LOR 2.25 4								
LOR 2.25 5								
LOR 2.25 6								
LOR 2.25 7								
LOR 2.25 8	35					GMGY20 & p69/31		
LOR 2.25 9								
LOR 2.25 10								
LOR 2.25 11								
LOR 2.25 12								
LOR 2.25 13								
LOR 2.25 14								
LOR 2.25 15								
LOR 2.25 16						GMGY20		
LOR 2.25 17								
LOR 2.25 18								
LOR 2.25 19								
LOR 2.25 20						p69/31		
LOR 2.25 21				Z				
LOR 2.25 22								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.25 23								
LOR 2.25 24					D			
LOR 2.25 25								
LOR 2.25 26								
LOR 2.25 27								
LOR 2.25 28								
LOR 2.25 29								
LOR 2.25 30								
LOR 2.25 31								
LOR 2.25 32					D			
LOR 2.25 33								
LOR 2.25 34								
LOR 2.25 35								
LOR 2.25 36								
LOR 2.25 37								
LOR 2.25 38								
LOR 2.26 1								
LOR 2.26 2								
LOR 2.26 3								
LOR 2.26 4								
LOR 2.26 5								
LOR 2.26 6								
LOR 2.26 7								
LOR 2.26 8								
LOR 2.26 9								
LOR 2.26 10								
LOR 2.26 11								
LOR 2.26 12	797	39				p75/79		
LOR 2.26 13	796							
LOR 2.26 14	795	23						
LOR 2.26 15	794							
LOR 2.26 16	793			Y				
LOR 2.26 17	792							
LOR 2.26 18	791	38		Y	D			
LOR 2.26 19	790				D			
LOR 2.26 20	789							
LOR 2.26 21	788							
LOR 2.26 22	787				D			
LOR 2.26 23	786	l				GMGY26		
LOR 2.26 24	785	m			D			
LOR 2.26 25	784							
LOR 2.26 26	783							
LOR 2.26 27	782	n		Y	Z			
LOR 2.26 28	781					GMGY20		
LOR 2.26 29	780							
LOR 2.26 30								
LOR 2.26 31	779	o		Y				
LOR 2.26 32	778				DZ			
LOR 2.26 33	777	p						
LOR 2.26 34	776							
LOR 2.26 35	775							
LOR 2.26 36	821					GMGY20		
LOR 2.26 37	37					L24a		
LOR 2.26 38	820							
LOR 2.26 39	819							
LOR 2.26 40	818							
LOR 2.26 41	817	i				GMGY8		
LOR 2.26 42	816							
LOR 2.26 43	815							
LOR 2.26 44	814				Z			
LOR 2.26 45	36					GMGY20 & p69/31		
LOR 2.26 45	813	j						
LOR 2.26 46	812							
LOR 2.26 47	811	42						
LOR 2.26 48	810			Y				
LOR 2.26 49	809	x				GMGY8		
LOR 2.26 50								
LOR 2.26 51	808							
LOR 2.26 52	807							
LOR 2.26 53	806							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.26 54	805							
LOR 2.26 55	804							
LOR 2.26 56	803							
LOR 2.26 57	802	3						
LOR 2.26 58	801	k						
LOR 2.26 59	800							
LOR 2.26 60	799	a						
LOR 2.26 61	798	a						
LOR 2.26 62								
LOR 2.26 63								
LOR 2.26 64								
LOR 2.26 65								
LOR 2.26 66								
LOR 2.26 67					D			
LOR 2.26 68								
LOR 2.26 69								
LOR 2.26 70								
LOR 2.26 71								
LOR 2.26 72								
LOR 2.26 73								
LOR 2.26 74	483			Y				
LOR 2.26 75	482							
LOR 2.26 76	481							
LOR 2.26 77	480	e		Y				
LOR 2.26 78	479							
LOR 2.26 79								
LOR 2.26 80	478				D			
LOR 2.26 81	477							
LOR 2.26 82	476							
LOR 2.26 83	475					GMGY39		
LOR 2.26 84	474					GMGY20 & p69/31		
LOR 2.26 85	522					GMGY20		
LOR 2.26 86	521	s						
LOR 2.26 87	520	y						
LOR 2.26 88	519							
LOR 2.26 89	518				D	GMGXY8		
LOR 2.26 90	517			Y				
LOR 2.26 91	516	46						
LOR 2.26 92	515	f		Y				
LOR 2.26 93	514							
LOR 2.26 94	513							
LOR 2.26 95	512	33				GMGY8		
LOR 2.26 96	511	13						
LOR 2.26 97	510				D			
LOR 2.27 1	509	31						
LOR 2.27 2	508	z						
LOR 2.27 3	507							
LOR 2.27 4	506	31						
LOR 2.27 5	505							
LOR 2.27 6	504	21			D		12q14	
LOR 2.27 7								
LOR 2.27 8								
LOR 2.27 9								
LOR 2.27 10	500	36						
LOR 2.27 11								
LOR 2.27 12	645							
LOR 2.27 13	646							
LOR 2.27 14	647							
LOR 2.27 15	648							
LOR 2.27 16	649	o						
LOR 2.27 17	650	8						
LOR 2.27 18	651							
LOR 2.27 19	652				DZ			
LOR 2.27 20	653							
LOR 2.27 21	654	18						
LOR 2.27 22	655							
LOR 2.27 23	656							
LOR 2.27 24	657	35				p75/79		
LOR 2.27 25	658							
LOR 2.27 26	659							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.27 27	660							
LOR 2.27 28	661							
LOR 2.27 29	662					p116\21		
LOR 2.27 30	663							
LOR 2.27 31	664	r				GMGY26		
LOR 2.27 32	665							
LOR 2.27 33								
LOR 2.27 34	666							
LOR 2.27 35	667	q						
LOR 2.27 36	668							
LOR 2.27 37								
LOR 2.27 38						PAX		
LOR 2.27 39								
LOR 2.28 1								
LOR 2.28 2								
LOR 2.28 3				Y				
LOR 2.28 4					Z			
LOR 2.28 5								
LOR 2.28 6								
LOR 2.28 7								
LOR 2.28 8					D			
LOR 2.28 9								
LOR 2.28 10								
LOR 2.28 11								
LOR 2.28 12					D			
LOR 2.29 1								
LOR 2.29 2								
LOR 2.29 3								
LOR 2.29 4								
LOR 2.29 5								
LOR 2.29 6								
LOR 2.29 7								
LOR 2.29 8	699							
LOR 2.29 9								
LOR 2.29 10								
LOR 2.29 11								
LOR 2.29 12								
LOR 2.29 13								
LOR 2.29 14								
LOR 2.29 15	33					p69/31		
LOR 2.29 16						GMGY20		
LOR 2.29 17								
LOR 2.29 18								
LOR 2.29 19	700							
LOR 2.29 20								
LOR 2.29 21								
LOR 2.29 22								
LOR 2.29 23					D			
LOR 2.29 24								
LOR 2.29 25					D			
LOR 2.29 26								
LOR 2.29 27								
LOR 2.29 28								
LOR 2.29 29	678							
LOR 2.29 30								
LOR 2.29 31								
LOR 2.29 32	695			Y	Z			
LOR 2.29 33	843							
LOR 2.29 34								
LOR 2.29 35								
LOR 2.29 36	833							
LOR 2.29 37								
LOR 2.29 38	679	v						
LOR 2.29 39								
LOR 2.29 40	690							
LOR 2.29 41								
LOR 2.29 42								
LOR 2.29 43								
LOR 2.29 44	696							

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.29 45	842							
LOR 2.29 46								
LOR 2.29 47								
LOR 2.29 48								
LOR 2.29 49								
LOR 2.29 50								
LOR 2.29 51	832							
LOR 2.29 52								
LOR 2.29 53								
LOR 2.29 54	680							
LOR 2.29 55	691			DZ				
LOR 2.29 56								
LOR 2.29 57	697							
LOR 2.29 58	841					GMGY39		
LOR 2.29 59	831							
LOR 2.29 60								
LOR 2.29 61								
LOR 2.29 62								
LOR 2.29 63	681							
LOR 2.29 64	692							
LOR 2.29 65								
LOR 2.29 66	698							
LOR 2.29 67	840							
LOR 2.31 1	830							
LOR 2.31 2								
LOR 2.31 3								
LOR 2.31 4	682							
LOR 2.31 5	693							
LOR 2.31 6								
LOR 2.31 7				D				
LOR 2.31 8								
LOR 2.31 9								
LOR 2.31 10								
LOR 2.31 11								
LOR 2.31 12	829			Z				
LOR 2.31 13								
LOR 2.31 14				D				
LOR 2.31 15	683			Z				
LOR 2.31 16	694							
LOR 2.31 17				D				
LOR 2.31 18								
LOR 2.31 19				DZ				
LOR 2.31 20	839							
LOR 2.31 21	828			D				
LOR 2.31 22				DZ				
LOR 2.31 23				D				
LOR 2.31 24	684							
LOR 2.31 25								
LOR 2.31 26								
LOR 2.31 27								
LOR 2.31 28	838			DZ				
LOR 2.31 29	827			D				
LOR 2.31 30				D				
LOR 2.31 31								
LOR 2.31 32	685	11						
LOR 2.31 33				Z		GMGY20		
LOR 2.31 34								
LOR 2.31 35				DZ				
LOR 2.31 36	837			D				
LOR 2.32 1	826							
LOR 2.32 2						p69/31		
LOR 2.32 3								
LOR 2.32 4	686							
LOR 2.32 5						GMGY20		
LOR 2.32 6						GMGY20		
LOR 2.32 7				DZ				
LOR 2.32 8	836			DZ				
LOR 2.32 9	825							
LOR 2.32 10								

COSMID #	VAX #	contig #	DYZ1	DYZ3	(CA)n	probe	in situ	cDNA
LOR 2.32 11								
LOR 2.32 12	687		x					
LOR 2.32 13								
LOR 2.32 14								
LOR 2.32 15								
LOR 2.32 16	835							
LOR 2.32 17	824							
LOR 2.32 18								
LOR 2.32 19								
LOR 2.32 20	688							
LOR 2.32 21								
LOR 2.32 22								
LOR 2.32 23								
LOR 2.32 24								
LOR 2.32 25	823							
LOR 2.32 26								
LOR 2.32 27								
LOR 2.32 28								
LOR 2.32 29	689		31					
LOR 2.32 30								
LOR 2.32 31								
LOR 2.32 32						GMGY30		
LOR 2.32 33	834							
LOR 2.32 34	822							
LOR 2.3 1	387							
LOR 2.3 2	388		33			GMGY8		
LOR 2.3 3	389				D			
LOR 2.3 4	390							
LOR 2.3 5	391				Z			
LOR 2.3 6	392		33					
LOR 2.3 7	393							
LOR 2.3 8	394		32		Z			
LOR 2.3 9	395							