A cDNA selection approach to isolate Y-linked

genes expressed in testis

Eleni Makrinou

July 2000

Thesis submitted for the degree of Doctor of Philosophy
in the University of London

MRC Human Biochemical Genetics Unit
The Galton Laboratory
Department of Biology
University College London
University of London

ProQuest Number: U643916

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.
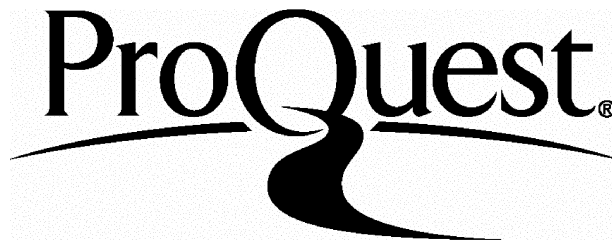
ProQuest U643916

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Statistical analysis has revealed that 2-7% of couples remain childless at the end of their reproductive life. In about half of these cases, the problem lies with the male partner. It is expected that genes expressed in the testis, and involved in spermatogenesis, will occur on the Y-chromosome. This view has been confirmed with the isolation of the Y-linked RBMY, TSPY and DAZ genes, deletions of which appear to be present in cases of oligo- and azoospermia.

In an effort to identify further Y-linked, testis-expressed genes, a cDNA selection library that was made by selecting testis cDNA with 1,000 Y-linked cosmid clones, was screened for Y-specific cDNAs. Screening with vector sequences and Y-linked repeat sequences, led to the elimination of more than 80% of the clones. Of the remaining 731 potential cDNA clones, sequencing revealed 79 clones with homology to several sequences in the database, including Y-linked and testis related sequences.

A number of sequences isolated from the cDNA selection library appeared to represent novel members of two gene families and they were further investigated. These studies revealed that the TTY2 gene, published by Lahn and Page (1997), is member of a large Y-linked multicopy gene family with an estimated number of 26 copies. Two members of this gene family, termed ml13d10 and ml22a3 TTY2-like genes, were further investigated, as representatives of the TTY2 gene family.

One more cDNA clone, termed 22d8, showed homology with several ESTs of unknown function and a 3.4kb cDNA that was recently released in the database (Feb. 2000) from a testis cDNA library. Investigation of this potential gene, revealed that either the whole, or part of it, is present in several chromosomal locations, including the Y chromosome and that it is ubiquitously expressed in many tissues, including testis.

Acknowledgements

Contents

6

Figures

CHAPTER FOUR

Isolation and investigation of TTY2-like genes

CHAPTER FIVE

Isolation and investigation of ml22d8/3.4 TT genes

Tables

CHAPTER FOUR

Isolation and investigation of TTY2-like genes

CHAPTER FIVE

Isolation and investigation of ml22d8/3.4 TT genes

Abbreviations

aa	amino acids
APS	ammonium persulphate
AZF	azoospermia factor
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
bp	base pair
cDNA	complementry deoxyribonucleic acid
CEPH	Centre d'Etude de Polymorphisme Humaine
Chr	chromosome
Ci	curie
cot	concentration x time
CpG	cytidine-phosphate guanosine
ddNTP	dideoxyribonucleoside triphosphate
DDRT	differential display reverse transcriptase
DEPC	diethylpyrocarbonate
dH_2O	distilled water (milli-RO plus)
ddH_2O	deionised and distilled water (milli-Q plus)
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
EST	expressed sequence tag
ETOH	ethanol
EMBL	European Molecular Biology Laboratory
FISH	fluorescence in situ hybridisation
FSH	follicle stimulating hormone
g	grams

Genbank	NIH genetics sequence database
HBGU	Human Biochemical Genetics Unit
HCl	hydrochloric acid
HGMP	Human Genome Mapping Project
HnRNP	heterogeneous ribobuncleoproteins
hnRNA	heterogeneous nuclear RNA
ICSI	intracytoplasmic sperm injection
kb	kilo base
kDa	kilo dalton
KCl	potassium chloride
l	litre
LH	luteininzing hormone
M	molar
Mb	mega bases
min	minute
ml	millilitre
mlcDNA	cDNA isolated by cDNA selection
mm	milli meter
MMuLV-RT	moloney murine leukaemia virus reverse transcriptase
MOPS	3-[N-morpholino] propanesulfonic acid
MRC	Medical Research Council
mRNA	messenger RNA
NaAC	sodium acetate
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	nanograms
°C	degrees Centigrade
ORF	open reading frame
PAC	P_1 artificial chromosome

PAR	pseudoautosomal region
PCR	polymerase chain reaction
PGM	phosphoglucomutase
pmoles	picomoles
RACE	rapid amplification of cDNA ends
RNA	ribonucleic Acid
RNAse	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal RNA
RT-PCR	reverse transcriptase polymerase chain reaction
SAP	shrimp alkaline phosphatase
SCO	sertoli cell only
SDS	sodium dodecyl sulphate
sec	second
SSC	salt sodium citrate
SSH	suppression subtractive hybridisation
SSPE	salt sodium phosphate EDTA
STS	sequence tag site
T ann	annealing temperature
TBE	Tris-borate/EDTA
TDF	testis determining factor
TE	Tris-EDTA
TEMED	$N_1N_1N'_1N'$-tetramethylethylenediamine
Tm	melting temperature
Tris	tris (hydroxymethyl) amino-methane
u	units
UCH	University College Hospital
UTR	untranslated region
uv	ultra violet
V	volts

v/v	volume per volume
W	watts
w/v	weight per volume
WHO	world health organisation
YAC	yeast artificial chromosome
Yco	Y-linked cosmid
μg	microgram
μl	microlitre
μm	micromolar

CHAPTER ONE

Introduction

1.1   Male infertility

1.1.1   Classification

Human fertility and reproduction vary considerably, influenced by religious background, geographic distribution, socioeconomic position and the structure of society. All these factors contribute to reproductive success, one of the strongest evolutionary forces. However, the success of human reproduction is also dependent on genetic factors, which affect our capacity to efficiently reproduce our species.

Analysis has revealed that of those couples, who try to have children, approximately 10% remain childless at the end of their reproductive life. In about half of the cases the problem lies with the male partner (Hull *et al.*, 1985). Male infertility is frequently associated with either a gross reduction in the number of sperm (oligozoospermia), or their complete absence in the ejaculate (azoospermia).

Over the past two decades, a number of scientists have reported a decline in the sperm quality and sperm count and a concomitant increase in reproductive problems of both animals and humans (Carlsen *et al.*, 1992; Irvine *et al.*, 1996; review, Irvine, 1997). In 1992 Niels Skakkebaek reported a decrease in human sperm count from 60×10^6/ml to 20×10^6/ml. This analysis was based on a total of 61 papers published between 1938 and 1990, which included data on 14,947 men. A separate study analysed the quality of sperm in a group of 577 men from Scotland, born between 1951 and 1973, and found a similar change and deterioration of sperm quality with time. However, more recent publications fail to confirm these findings and claim that there were methodological and geographic biases in the

earlier studies. The biases were ascribed to comparisons of data collected in different countries, at different times and to different methods of subject selection and laboratory methodology (Irvine, 1997; Fisch and Goluboff, 1996; Paulsen, 1996; Fisch et al., 1996).

According to the World Health Organization's (WHO) classification (1992-1999), the standard sperm count for normal males should be $\geq 20 \times 10^6$/ml, the sperm volume should be greater than 1.5ml, more than 70% of the spermatozoa should be motile and more than 60% of sperm should have normal morphology (quoted from Dawson and Whitfield, 1996; review on WHO report 1999; De Jonge and Barrat, 1999). If an individual has 20-60 million sperm/ml of semen he is defined as fertile; if there are less than 20 million motile sperm/ml the individual is considered infertile due to oligozoospermia; 20-40% of motility is designated as asthenozoospermia and more than 40% of abnormal morphology is termed teratozoospermia (Moosani et al., 1995).

1.1.2 Physical factors

Male infertility can arise as a consequence of physical blockage of the ductal system due to congenital epididymal defects, anatomical anomalies like varicocele, hypospadias and cryptoorchidism or congenital malposition of the ejaculatory ducts with an opening into the bladder or even into the ureter (Jequier, 1997). Physical blockage can arise as a secondary consequence of some other clinical conditions, for example contraction of a sexually transmitted disease, an autoimmune disease or other causes including infection of the epididymis. The above clinical conditions are generally described as epididymitis. Blockage can also be caused by physical injury due to an accident, or a lower abdominal surgery that can result in accidental vasectomy or blockage of the testis blood supply. In

addition, obstructive azoospermia can be caused by congenital bilateral absence of the vas deferens, due to mutations in the cystic fibrosis transmembrane conductance regulator gene (CFTR) (reviews, Jequier, 1993; Skakkebaek *et al.*, 1994; Meschede *et al*, 1995; De Kretser, 1997).

1.1.3 Environmental factors

Environmental factors are also important and include a number of lifestyle habits, such as diet, clothing, stress, use of tobacco and alcohol, as well as exposure to sperm toxic substances like pesticides, anabolic steroids which are widely used by athletes and body-builders, and numerous cytotoxic drugs that are used to treat cancer and autoimmune diseases (Vine *et al.*, 1994; Tiemessen *et al.*, 1996; Lloyd *et al.*, 1996). A hypothesis, which has attracted a good deal of attention, concerns exposure to environmental estrogens. Although estrogens are normally produced by both sexes, overexposure to exogenous synthetic estrogens, seem to be correlated with disorders of development of the male reproductive tract (Sharpe and Skakkebaek, 1993; Irvine., 1997; Saunders *et al.*, 1998).

1.1.4 Chromosome rearrangements

Genetic factors include a number of chromosome abnormalities, translocations, deletions and microdeletions of genes involved in the process of the control of spermatogenesis (Review Wieacker and Jakubiczka, 1997).

Chromosome abnormalities have been detected in sperm samples from azoospermic and oligozoospermic patients, with sex chromosome abnormalities being predominant (Van Assche *et al.*, 1996; Johnson, 1998). Examples of sex chromosome aneuploidies causing male infertility include Klinefelter syndrome (47, XXY), a disorder exhibiting female secondary

sexual characteristics and hypotrophic gonads and 47,XYY maleness, a disorder sometimes associated with male infertility due to abnormal pairing during meiosis (Speed *et al.*, 1991). Chromosomal abnormalities causing meiotic arrest or reduction in the chiasma formation during meiosis, are also thought to contribute to impairment of spermatogenesis (Pearson *et al.*, 1970; Vendrell *et al.*, 1999). Examples of sex chromosome rearrangements are X-autosome reciprocal translocations, which cause disruption of X-linked loci, and are associated with abnormal X-inactivation and arrest of germ cell maturation (Lifschytz and Lindsley, 1972). Rearranged Y chromosomes which are either acrocentric, ring, inverted or dicentric, also associated with spermatogenic impairment and meiotic arrest due to chromosomal instability (Maeda *et al.*, 1976; Chandley, 1979; Delobel *et al.*, 1998; review, Johnson, 1998). Finally 46,XX maleness is a heterogeneous disorder which can be caused by translocation of the testis determining factor (TDF) to the X chromosome, or by undetected 46,XX/47,XXY mosaicism, or by mutations in autosomal or X-linked, sex determining genes (see section 1.3) (review, Van Assche *et al.*, 1996; Wieacker and Jakubiczka, 1997).

In addition to numerical sex chromosome anomalies and rearrangements, there are other reports of both reciprocal and Robertsonian translocations and structural changes of autosomal chromosomes that can give rise to spermatogenic impairment. Examples include a Robertsonian translocation that occasionally is associated with male infertility and subfertility, t(13q14q) and a familial pericentric inversion on chromosome 1, which has been found to be associated with spermatogenic arrest at the level of primary spermatocytes (Meschede *et al.*, 1994).

In general, chromosome rearrangements can interfere with normal chromosome pairing and segregation during meiosis and thus result in unbalanced gametes. As a result, chromosomal abnormalities can cause a

number of disorders, where male infertility appears to be present as a secondary effect. There are numerous examples and some of these disorders are described in more detail in later sections. A very general review of these disorders would include the following; cystic fibrosis, often associated with congenital bilateral absence of the vas deferens; persistent Mullerian duct syndrome (OMIM 261550); Kallman syndrome (OMIM 308700); Kennedy's disease (OMIM 313200); infertile male syndrome caused by mild dysfunction of the androgen receptor gene (OMIM 308370); myotonic dystrophy (OMIM 160900); anomalies of imprinted genes and uniparental dysomy (review, Meschede and Horst, 1997; Skakkebaek *et al.*, 1994; Wieacker *et al.*, 1997; review, Johnson, 1998).

A significant proportion of cases of oligo- and azoospermia have no obvious cause. Over the last 20 years it has been shown that genetic defects make an important contribution towards these cases of idiopathic infertility. The World Health Organisation (WHO), has accumulated data from coordinated international research programs which suggest that genetic factors account for more than 10% of the entire spectrum of male infertility (Chandley, 1998). These genetic defects are restricted to spermatogenesis, and occur in otherwise healthy men. This indicates that any genes involved must either be expressed only in spermatogenesis, or be functionally required only during normal testis development (Elliot and Cooke, 1997).

Recently the introduction of assisted fertilization (e.g. intracytoplasmic sperm injection ICSI), has made it possible for some men with moderate or severe defects in spermatogenesis, to overcome their infertility and have children (Van Steirteghem *et al.*, 1993; Ron-El *et al.*, 1999). The ICSI technique is based on the microsurgical aspiration with the help of a needle, of epididymal or testicular spermatozoa, and the injection of a single spermatozoon into the oocyte. With this method any type of spermatozoa can be used, even those that have deficiencies in their motility

23

and morphology. With the development of microinjection of round spermatid nuclei into the oocyte it is no longer necessary that spermatogenesis should be completed (Ogura *et al.*, 1994).

However, this advance in technology, may bypass the normal control mechanisms for sperm selection and raises the possibility that male children born in this way will inherit genetic defects from their fathers and will themselves have reproductive problems (Kremer *et al.*, 1997; Kurinczuk and Bower, 1997; Chang *et al.*, 1999). Regardless of the disadvantageous impact on the dynamic of a population, ICSI has made it possible for males bearing chromosomal and gene abnormalities, to achieve biological paternity (Vogt *et al.*, 1992; Mulhall *et al.*, 1997; Page *et al.*, 1999). Recently, by using ICSI, a number of births of healthy normal infants have been achieved from fathers with non-mosaic Klinefelter's syndrome which are considered to be infertile due to primary testicular failure (Palermo *et al.*, 1998; Ron *et al.*, 1999). However, not all kinds of genetic abnormalities associated with defects of the male reproductive system can be overcome with ICSI. Studies have shown that males with deletions of the AZFa and AZFb region on the long arm of the Y chromosome (see section 1.4.6) have complete absence of spermatozoa at any state of maturation, and as a result ICSI cannot be applied (Brandell *et al.*, 1998).

1.2   The Y chromosome
1.2.1   Structure

In mammals, the Y chromosome is haploid, and is dispensable for normal development of the female. It is a small chromosome of approximately 59Mb, morphologically and genetically distinct from other chromosomes and comprising only 2-3% of the haploid genome.

The Y chromosome has two regions (PARs), which behave in a pseudoautosomal manner and are located at the tips of the long and short arms, close to the telomeres. There is in addition a dispersed non-recombining region, which accounts for 95% of the Y chromosome and represents the only haploid part of the human genome. This non-recombining region comprises blocks of X-Y homology and Y-specific regions, composed of both euchromatin and heterochromatin (Fig 1.1).

The euchromatic region (Ypter to Yq11) and the heterochromatic region (Yq12) are of approximately equal size. The heterochromatic region is largely composed of two repeat sequences DYZ1 and DYZ2 (section 1.2.2) and is assumed to be genetically inert (Goodfellow *et al.*, 1985). This view is in part based on the demonstration of length variations in the heterochromatin that have no apparent effect on the phenotype (Bishop, 1985). The euchromatic region also has numerous repeated sequences, but in addition, it contains genes associated with sex determination, control of spermatogenesis, the occurrence of gonadoblastoma, development of tooth enamel, control of height and Turner syndrome. Overall, the Y chromosome contains much more repetitive sequence than other chromosomes.

1.2.2   Repetitive sequences on the Y chromosome

The Y chromosome contains multiple families of repeated DNA sequences; some are similar to those found in autosomes, whereas others are Y-specific. Repetitive gene families on the Y chromosome include LINES (Long Interspersed Repeated Sequences), SINES (Short Interspersed Repeated Sequences), and tandemly arrayed repeated elements (minisatellites and microsatellites).

Examples of non Y-specific repeats are the *Alu* and *Kpn* families. *Alu* sequences comprise short, polymorphic, interspersed elements of
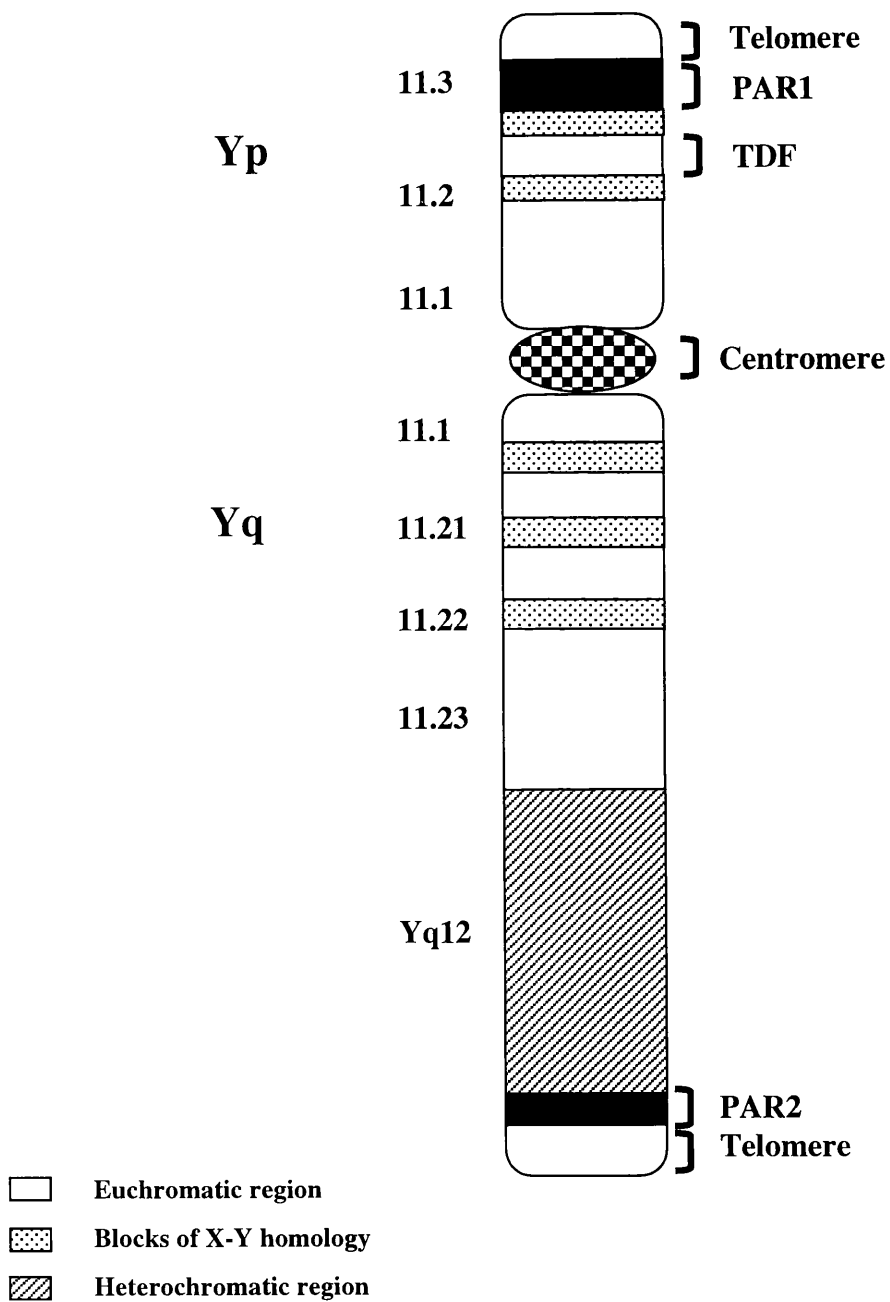
Figure 1.1 Diagram of recombining and non-recombining regions on the Y
chromosome. Numbers on the left hand side indicate the Yp and Yq regions

approximately 300bp in length, which occur most often in introns or 3'
untranslated regions. They are the most abundant interspersed repetitive
DNA sequence in all primate genomic DNA and are found in dense clusters
of up to two *Alu* elements per Kb. On autosomal chromosomes they make
up 5% of the genome, but on the Y chromosome occur at a lower frequency
and appear to have little homology with genomic consensus *Alu* sequences
(Deininger *et al.*, 1981; Wolfe *et al.*, 1984; Claverie *et al.*, 1994).

The *Kpn* family are long, dispersed sequences, each of around 5-6Kb
in length. Like *Alu*s they are located between single-copy DNA (Kunkel *et
al.*, 1982). They are often flanked by short repeats and some have a poly-
(A) tail at their 3' end (Grimaldi *et al.*, 1984). Southern blot analysis has
suggested that the distribution of *Kpn* elements on the Y chromosome is
equal to their distribution elsewhere in the genome, which is approximately
3-5%.

The human Y chromosome contains numerous Y- specific repetitive
DNA families in its heterochromatic region. The most abundant of them are
DYZ1 and DYZ2. They were originally described by Cooke (1976) as
3.4Kb and 2.1Kb male specific HaeII DNA fragments after restriction
enzyme digestion of total genomic DNA. Higher primates contain the DYZ1
and DYZ2 repeat units, but they both have autosomal locations and are not
considered to be Y-specific (Smith *et al.*, 1987).

The DYZ1 family constitutes a minimum of 25-30% of all Y DNA,
and is the major component of the heterochromatic portion of Yq (~50%)
(Bostock *et al.*, 1978). There are 4000-6000 copies on the Y chromosome,
arranged in tandem arrays. The 3.4Kb DYZ1 repeat unit consists of 713
pentanucleotides, of which 229 are a simple pentameric core (5'–TTCCA-
3') which is tandemly ordered. In between this main pentameric core, exist
other repeat units related to each other, but with small variations in their
core sequences (Nakahori *et al.*, 1986). DYZ1 repeat family is a mixture of

27

both male specific and non-specific segments, shared by both males and females (Bostock *et al.*, 1978; Nakagome *et al.*, 1991). The male specific segments are present only on the Y chromosome, whereas non-specific repetitive elements demonstrate homology with sequences at the heterochromatic regions of different chromosomes including 1, 9, 13, 14, 15, 16, 21 and 22. However, both the male specific and non-specific segments of this repeat have no homologous sequences on the X chromosome (Burk *et al.*, 1985).

DYZ2 comprises repeats of a Y-specific 2.1Kb HaeIII fragment and an autosomal 1.9Kb fragment embedded in a larger repeated element of around 2.4Kb. The 2.4Kb repeat unit is arranged on the Y chromosome as tandem clusters and has a 2.4Kb homologue on chromosome 14. The Y-linked 2.4Kb unit consists of a Y-specific fragment of 1.6Kb and an 800bp fragment homologous to part of the autosomal unit (Smith *et al.*, 1987). It is estimated that the DYZ2 family consists of around 2000 members, accounting for 10-20% of the Y chromosome. Sequence analysis of members of the DYZ2 family has revealed a complex structure involving both A-T and G-C rich subregions and an *Alu* repeat (Frommer *et al.*, 1984). However, it is not clear whether all DYZ2 fragments have exactly the same structure. Results from Southern blot analysis and FISH indicate that DYZ2 repeats extend through the entire heterochromatic region of Yq12 (Schmid *et al.*, 1990).

DYZ5, is an array of 20-40 tandemly repeated sequences located on the proximal Yp. It consists of a number of individual internal repeat units of 20Kb and the overall length varies between 540-800Kb (Tyler-Smith *et al.*, 1988). TSPY gene copies (see section 1.4.6) are considered to be part of these repetitive elements (Manz *et al.*, 1993). Comparative studies have shown that this repeat unit is conserved in great apes where, as in the case of the human Y, it is organised in one major block and confined to a small

region of the Y chromosome (short arm in chimpanzee and orangutan; long arm in gorilla) (Guttenbach *et al.*, 1992).

The Y chromosome also has a unique set of highly polymorphic, sex specific alphoid repeats, termed DYZ3, which are made up of EcoRI fragments of 340 and 680bp (Wolfe *et al.*, 1985). Alphoid repeats are tandemly clustered and concentrated around the centromere of chromosomes.

1.2.3   Pseudoautosomal regions

The Y chromosome pairs and recombines with the X chromosome in two regions, called pseudoautosomal regions (PARs), which are located at the tips of the long and the short arm of the X and Y chromosome (Fig 1.1). The PARs are required for correct sex chromosome segregation and completion of male meiosis. Failure either to pair or to recombine, results in meiotic arrest and is often associated with male sterility and numerous other disorders both in humans and mice (Hassold *et al.*, 1991; Mohandas *et al.*, 1992).

Pairing between the sex chromosomes was first observed in the rat by Koller and Darlington (1934) and in humans by Solari (1980). The term "pseudoautosomal" was suggested by Burgoyne (1982), based on the observations that the Yp region of homology with the X chromosome behaves like an autosomal segment during meiosis allowing recombination between the sex chromosomes. Proof of recombination was based on looking at the inheritance pattern of X and Y-linked allelic markers (Cooke *et al.*, 1985).

The pseudoautosomal regions are considered to be a remnant of an ancient large pseudoautosomal region through which the ancestors of the present sex chromosomes, proto-X and proto-Y, were thought to pair and

recombine during meiosis (Graves, 1998). A number of pseudoautosomal genes and repetitive sequences are conserved in other species, though not necessarily on the Y chromosome (Weber *et al.*, 1987).

PAR1, located in the distal portions of the short arms of the X (Xp22.3) and Y (Yp11.32) chromosomes, is about 2.6Mb and represents 3-5% of the human Y chromosome and 1.6% of the human X chromosome (Rappold, 1993). In humans, it is separated from the Y specific region by an insertional *Alu* element (a sequence of approximately 300bp with an *Alu* cleavage site at each end), followed by a 220bp stretch of reduced homology (78%) which forms the pseudoautosomal boundary (Ellis *et al.*, 1989). The distal boundary of the PAR is near the telomere, which in human germ cells is considerably longer than those in somatic cells (Cooke and Smith., 1986).

PAR1 contains, throughout its length, a number of repetitive sequences. Examples include repeats spanning part of PAR1 and the Yq homologous region (Fisher *et al.*, 1990), repeats specific to X and Y pseudoautosomal subtelomeric regions (STIR) (Rouyer *et al.*, 1990) and DXYZ2, a tandemly arranged repeat of 20-40 copies (Simmler *et al.*, 1985), which show variations in their copy numbers between individuals, and could reflect a possible length polymorphism of PAR1 as a whole (Page *et al.*, 1987). This repeat is also present in regions of meiotic crossing over on the sex chromosomes of both old and new world monkeys (Weber *et al.*, 1987; Schempp *et al.*, 1989).

It has also been suggested that PAR1 contains a high density of minisatellites, which could represent recombination hotspots involved in chromosomal pairing or play a role in the initiation of meiotic recombination (Cooke *et al.*, 1985).

At present 11 genes have been isolated from PAR1, all of which have been physically mapped such that their exact positions and order are known (Fig 1.2, Table 1.1). PAR1 genes escape X-inactivation, and most of them

are housekeeping genes ubiquitously expressed in many tissues (Rappold, 1993). They are involved in a number of processes such as cell adhesion, synthesis of melatonin and energy metabolism of the cell. In addition at least one of these genes has some role in Turner's syndrome (45,XO), which is thought to be caused by haploinsufficiency of a number of sex linked genes.

PAR2 is about 320Kb in length and flanks the telomeres of the long arms of the X and the Y chromosome, with a LINE element making the boundaries with the X and Y specific sequences (Freije *et al.*, 1992). Most of PAR2 has been sequenced (Speed and Chandley, 1990, Kvaloy *et al.*, 1994; Vogt *et al.*, 1997) and two genes have been assigned to this region; IL9R (interleukin 9 receptor) and SYBL1 (synaptobrevin-like 1) (Fig 1.2) (D'Esposito *et al.*, 1996). IL9R is a gene of around 17Kb, with an L1 element proximal to it, which is expressed from both sex chromosomes (Vermeesch *et al.*, 1997). A number of pseudogenes for IL9R have been identified and localised in chromosomes 9, 10, 16 and 18. It has been proposed that these pseudogenes may have arisen by duplication and translocations from the X and Y chromosomes. X-linked SYBL-1 is a 2.6Kb gene encoding a 220 aa protein of unknown function. It undergoes X-inactivation and its Y homologue is inactive (D' Esposito *et al.*, 1996). In addition, there are two autosomal homologues, SYB1 (OMIM 185880) and SYB2 (OMIM 185881), located on chromosomes 12 and 17 respectively, which encode integral membrane proteins.
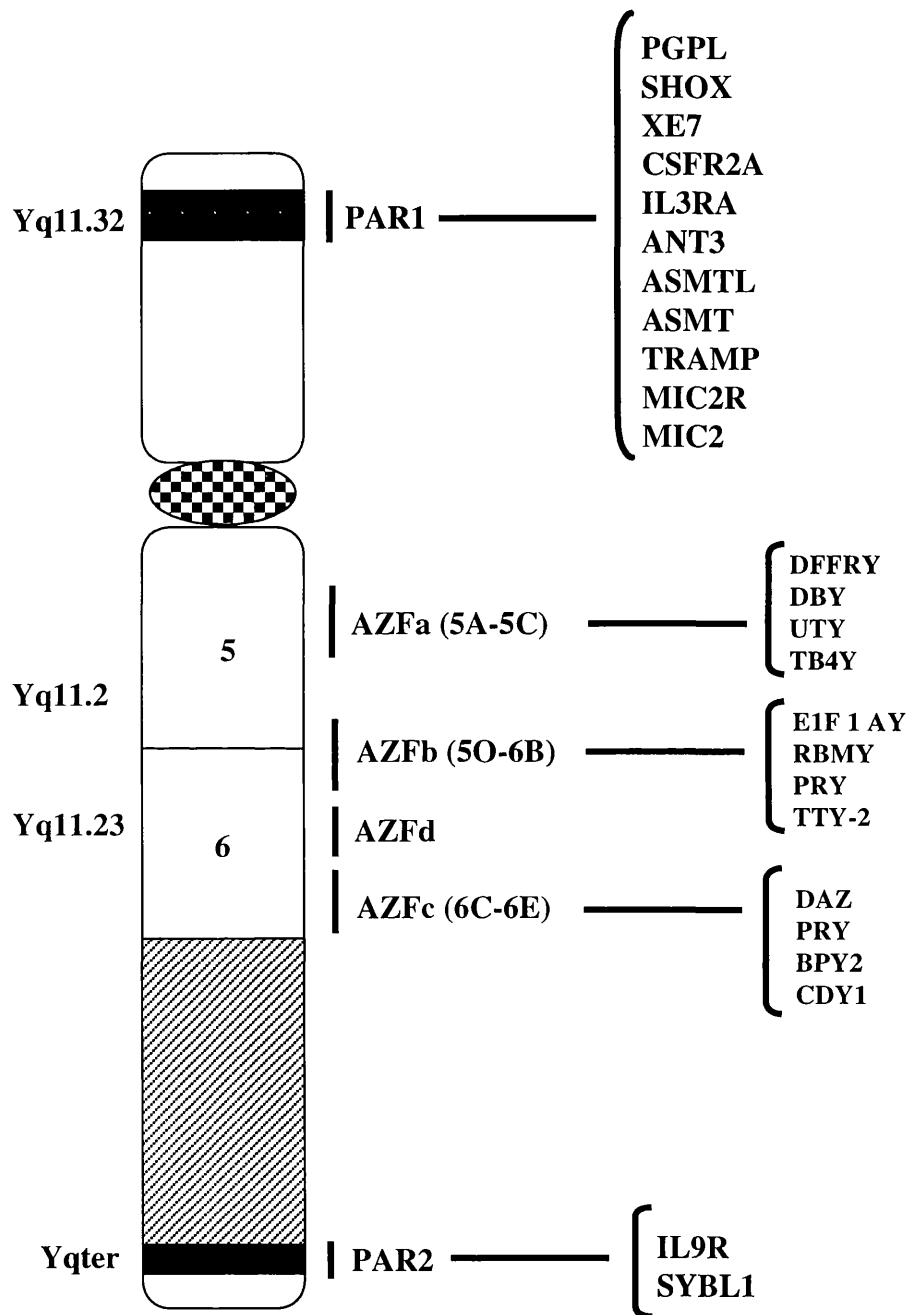
Figure 1.2 Diagram of Y-chromosome genes, located in the two PARs and the AZF regions. Genes in PAR1 are shown in order on Yp.

Table 1.1 Genes located in pseudoautosomal region 1 (PAR1)

Symbol	Name	Comments	References
MIC2	Monoclonal IC:Imperial Cancer Research 2	A cell surface antigen involved in cell adhesion processes	Buckle *et al.*, 1985
MIC2R	MIC2 Related	Associated with a CpG-rich island; no ORF; unknown function	Smith and Goodfellow, 1994
CSF2RA	Colony stimulating factor receptor α	Member of a cytokine receptors family; may be associated with type M2 acute myeloid leukaemia	Cough *et al.*, 1990
XE7	------	Two protein isoforms resulting from alternative splicing	Ellison *et al.*, 1992
SHOX	short stature homeobox containing gene	Two alternatively spliced isoforms; involved in idiopathic growth retardation/short stature in Turner syndrome	Henke *et al.*, 1991 Rao *et al.*, 1997
ASMT	Acetylserotonin-O-methyltransferase	Active in the synthesis of the hormone melatonin	Yi *et al.*, 1993
ASMTL	ASMT-Like	Resulted as duplication of the ASMT gene on the Y chr.	Ried *et al.*, 1998

Symbol	Name	Comments	References
PGPL	Pseudoautosomal GTP-binding protein-like	Encoding a 442 aa protein with GTP-binding protein domain motifs	Gianfrancesco *et al.*, 1998
TRAMP	Transposase	Encodes a protein of 694 aa, similar to transposases of the Ac family; may be ancient transposable element	Esposito *et al.*, 1999
ANT3	Adenine nucleotide translocase	Plays a role in energy metabolism	Slim *et al.*, 1993
IL3RA	Interleukin receptor subunit α	Cytokine receptor for interleukin-3	Kremer *et al.*, 1993

1.2.4 Non-recombining X-Y homologous regions

In addition to the 100% homologous pseudoautosomal regions, the X and Y chromosomes share several regions of lesser homology located outside the two PARs (Cooke *et al.*, 1984) (Figs 1.2 and 1.3). These X-Y homologous regions could be the result of duplications and translocations between the sex chromosomes that took place at different times in evolution, or they may be remnants of their ancient progenitors, the proto-X and proto-Y (Sargent *et al.*, 1996; Lahn and Page, 1999).

Information about the extent of X-Y homology emerged as the map of the Y chromosome was established. This process involved mapping deletions from patients with sex chromosome rearrangements (Affara *et al.*, 1986, 1987; Vollrath *et al.*, 1992) and assembling YAC contigs of cloned Y chromosome DNA (Foote *et al.*, 1992). This enabled the determination of the order and organisation of the X-Y homologous sequences on the Y chromosome. These maps, together with maps of the human pseudoautosomal regions (Brown, 1988; Kvaloy *et al.*, 1994), have indicated that of the 8-10Mb where the X-Y chromosomes share sequence, only around 3Mb correspond to the pseudoautosomal regions, whereas the rest comprise other regions of X-Y homology (Fig 1.3).

Comparative studies in different primate species give a view of the evolution of these X-Y homologous regions. These studies have demonstrated that some of the sequences within the human X-Y homologous blocks are strictly X-linked in some primate species. This finding suggests that the origin of the homologous sequences is the result of duplication and transfer of one copy from the X to the Y after the divergence of certain primate species (Lambson *et al.*, 1992). Other human X-Y regions appear to be X and Y-linked in some primates, and appear at similar chromosomal locations, although some have clearly been rearranged. For example a single contiguous segment of human Xq21, is homologous to two

Figure 1.3 X-Y homologous genes and their locations; hatched sections indicate the X-Y homologous regions. For the rest, the key is the same as for Fig 1.1

non-contiguous segments on Yp. Comparative studies show that this almost certainly arose as a single transposition event of 4Mb segment from the X to the Y chromosome associated with an inversion (Schwartz et al., 1998). This event occurred after the divergence of the human species from the chimpanzees.

A number of genes have been identified within the X-Y homologous regions (Fig 1.3, Table 1.2). These genes share similar features with those located within the PARs (Table1.1). They are usually found as single genes or small clusters including pseudogenes, they escape X-inactivation, and are ubiquitously expressed in many tissues. The degree of identity of the X-Y homologues at the nucleotide level varies between 80-95% (Lahn and Page, 1999). These genes are widely distributed throughout the Y chromosome, however, their X-linked homologues are mainly concentrated toward the distal end of the short arm with the exception of RPS4X that maps to Xq (Fig 1.3) (Vogt et al., 1997). The X-linked homologues of these genes are all functional, whereas in some cases their Y-linked homologues are pseudogenes (Table 1.2). This loss of function has occurred presumably as a result of continual rearrangements of the Y chromosome and accumulation of mutations due to lack of extensive recombination. In compensation for the degradation of the Y-linked genes, their X-homologues have acquired increased levels of expression. It is predicted that these genes will become subject to X-inactivation in order to balance and restore normal expression levels in both males and females (Jegalian and Page, 1998).

Some of the X-linked copies of these genes have been associated with genetic disease in man. For example mutations in the AMELX gene causes amelogenesis imperfecta a disorder of tooth enamel development. Although the Y-linked copy of this gene encodes a functional protein, its level of expression is only 10% of that of its X homologue. Nevertheless, despite this low level of expression of the Y-linked gene, there are detectable

Table 1.2 Genes with X and Y homologues

Symbol	Name of gene	Comments	X-Y homologues locations	References (for Y-linked homologues)
ZFY	Zing-finger-Y	Transcription factor with multiple zinc-finger motifs; X and Y copies functional	Xp22.2-p21.3 / Yp11.3 autosomal homologue 9p22-9pter	Page *et al.*, 1987 Affara *et al.*, 1989
AMELY	amelogenin	Development of tooth enamel; X and Y copies functional	Yp11.2 / Xp22.3-p22.1	Lau *et al.*, 1989 Salido *et al.*, 1992
STSP	Steroid sulfatase pseudogene	Y-linked non-processed pseudogene with functional X-linked copy	Xp22.32 / Yq11.21	Yen *et al.*, 1988
PRKY	Protein kinase Y-linked	protein kinase similar to catalytic subunit of cAMP dependent protein kinases; X and Y copies are functional; involved in ectopic recombination leading to XX male	Xp22.3 / Yp11.2 Pseudogene at Xq12-13 (PRKXP2) Pseudogene at 15q26 (PRKXP1)	Klink *et al.*, 1995 Schiebel *et al.*, 1997
DFFRY (also known as USP9Y)	*Drosophila* fat facets related Y-linked	two isoforms by alternative splicing; may regulate ubiquitin precursors; functional X copy, Y copy is non-processed pseudogene	Xp11.4 / Yq11.2	Jones *et al.*, 1996

Symbol	Name of gene	Comments	X-Y homologues locations	References (for Y-linked homologues)
SMCY	Selected mouse cDNA on Y	X and Y copies functional; encode human H-Y epitope HLA-B7	Xp11.1-p11.2 / Yq11.2	Agulnik *et al.*, 1994
XGPY	XG blood group system pseudogene Y-linked	Spans pseudoautosomal boundary of the X chr; multiple copies on Y, all lack an ORF. There is an intron containing pseudogene (XGPY) in Yq11.21	Xpter-p22.32 / Yq11.21	Ellis *et al.*, 1994 Weller *et al.*, 1995
DBY	Dead Box Y	Putative RNA helicase; functional X and Y copies	Xp11.3-p11.23 / Yq11	Lahn and Page, 1997
TB4Y	Thymosin β4 Y isoform	X homologue functions in actin sequestration; functional X and Y copies	Xq21.3-q22 / Yq11	Lahn and Page 1997
UTY	Ubiquitous TPR motif Y	Functional X and Y copies; Mouse homologue encodes an H-Y antigen; contains 10 tandem tetratricopeptide repeats (TPR);	Xp11.2 / Yq11	Lahn and Page, 1997
RPS4Y	Ribosomal protein S4 Y-linked	X and Y copies functional; encode ribosomal protein S4, implicated in Turner syndrome	Xq13 / Yp11.3	Fisher *et al.*, 1990

Symbol	Name of gene	Comments	X-Y homologues locations	References (for Y-linked homologues)
EIF1AY	Translation initiation factor 1A Y isoform	X homologue implicated in translation initiation; Y copy, unknown function	Xp11.2 / Yq11.2	Lahn and Page 1997
ARSE/D	Arylsulfatases E and D	Y-linked truncated pseudogenes with X-linked functional copies	Xp22.3 / Yq proximal to the centromere	Meroni *et al.*, 1996
KALP	Kallman syndrome gene	Y-linked pseudogene with a functional X-linked copy	Xp22.3 / Yq11.21	Del Castillo *et al.*, 1992

differences in the properties of human enamel proteins which can be distinguished according to the sex of the individual (Fincham *et al.*, 1991; Salido *et al.*, 1992). Mutations in the X-linked KAL1 gene are responsible for causing Kallman syndrome (see section 1.4.3), whereas the Y-linked homologue appears to be a pseudogene. Mutations of the X-linked ARSE and ARSD genes cause X-linked recessive chondroplasia punctata (CDPX1) and mutations in the X-linked STS gene cause X-linked ichthiosis and placental steroid sulfatase deficiency, while the Y-linked homologues are non-functional. Genes in the Xp11.2-p22.1 region which shares homology with genes located in the short and long arms of the Y chromosome have also been implicated in the determination of the short stature, ovarian failure, high-arched palate and autoimmune thyroid disease which are characteristic of Turner syndrome (45, XO) (Zinn *et al.*, 1998)

1.2.5   Non-recombining, Y-specific regions

There are a number of sequences, which are unique to the Y chromosome and appear to have evolved quite distinctly from genes on the X chromosome or other autosomes. These sequences are dispersed along the length of the Y chromosome, in both heterochromatic and euchromatic regions and within the X-Y homology regions.

To date, 12 genes have been identified, which do not have X-linked homologues. These include genes such as   DAZ and TSPY that exist as multiple copies and are specifically or predominantly expressed in testis (discussed in section 1.4.6). In addition, Lahn and Page (1997, 1999) after a large scale search of the non-recombining region identified 8 more potential genes with similar features (Table 1.3). The only example of single copy gene in the Y-specific region is SRY (discussed in section 1.3.2).

Table 1.3 Genes in the Y-specific, non-recombining region of the Y chromosome.

All exist as multiple copies and are expressed in testis

Symbol	Name	Comment	Location
CDY1	Chromodomain Y 1	Contains a chromodomain and a putative catalytic domain; intronless; arose by retroposition from CDYL gene (chr. 6); may modify DNA or chromosomal protein during spermatogenesis	Yq; interval 6F
CDY2	Chromodomain Y 2	Shares 98% identity at the aa level with CDY1	Yq; interval 5L
BPY1	Basic protein Y 1	Novel 125 aa protein; has an X homologue	Yq; interval 5G
BPY2	Basic protein Y 2	Novel 106 aa protein; same gene family as BPY1	Yq; interval 6E
XKRY	XK related Y	Novel protein with similarity to XK, a putative membrane transport protein	Yq; interval 5L XK; Xp21.2-p21.1
PRY	PTP-BL Related Y	Novel protein with similarity to the mouse PTP-BL putative protein tyrosine phosphatase	Yp; interval 4A Yq; intervals 6C and 6E
TTY-1	Testis-transcript 1	No ORF	Yp; intervals 3C and 4A
TTY-2	Testis-transcript 2	No ORF	Yp; interval 4A Yq; interval 6C

The multicopy nature of the Y-specific genes can be considered as a mechanism to maintain at least one functional copy. If only one copy were present there would be a high chance of loss of activity due to accumulation of mutations (Rice, 1994), a problem associated with the lack of extensive recombination. A similar explanation for the multiple copies of Y specific genes would be that some are located close to the heterochromatic region of the long arm of the Y chromosome (DAZ). If the heterochromatic region exerts a silencing effect on nearby genes, amplification may increase the distance of gene copies from the heterochromatin and sustain their function. Alternatively, the multiple copies of these genes may simply reflect the need for a large amount of mRNAs (Cooke *et al.*, 1998).

Exact functions for all these genes have not yet been determined. However, since the absence of the Y chromosome from individuals who are 45,XO (Turner syndrome), or the presence of extra Y copies (47,XYY maleness) do not cause conditions threatening to life, it is considered that these genes mainly control male sex determination, testicular differentiation, and regulation of spermatogenesis.

1.3 Male sex determination and the Y chromosome

1.3.1 Sex determination - description

Male and female human embryos are identical until around the 42nd day of gestation when they start differentiating. Prior to that stage, ovaries and testes are indistinguishable and therefore called bipotential or indifferent gonads. These bipotential gonads arise from the urogenital ridge, a structure containing cell precursors for the formation of the adrenal cortex, the gonads and the kidneys. The development of a male or female phenotype is the result of interactions of genetic, cellular and hormonal signals (Gubbay *et al.*, 1990; Carlson., 1994). When testes first develop, they are capable of

producing testosterone and anti-Mullerian hormone, although they lack any spermatogenic function. These hormonal products then control the subsequent steps of male sexual differentiation by causing the Mullerian ducts to degenerate and the Wolffian ducts to develop into the seminal vesicles, epidydimis and vas deferens. In the absence of testicular hormones, the Wolffian ducts regress and the Mullerian ducts form the female reproductive organs. So female differentiation is considered as the default pathway (Carlson , 1994).

A number of both X-linked (SOX3, DAX1) and autosomal genes (WT1, SF1, SOX9) are involved in sex determination. These will not be described in detail but their main features are summarised in Table 1.4 (review, Jimenez and Burgos, 1998). In addition to these genes, a single Y-linked gene, SRY, has been proposed as the major testis determining factor (TDF) (Section 1.3.2). Many of the genes involved in sex determination are transcription factors and it is thought they act alternatively to activate or suppress the level of expression of genes involved in pathways leading to the differentiation of testis or ovary. Numerous models discussing how these genes interact with each other to produce the male or female sex phenotypes have been proposed (Jimenez *et al.*, 1996; Schafer and Goodfellow, 1996; Graves, 1998; Edwards and Beard, 1999). Some of these models are very similar (Capel *et al.*, 1995; Graves, 1998) and a commonly proposed version is outlined in Fig 1.4, showing a number of genes that are expressed in different developmental stages and regulate the male and female sex differentiation. However, the detail of exactly how these genes and their products interact together to determine a sex phenotype still remain unclear.

The Y chromosome is thought to make an important contribution towards male sex determination through the action of gene(s) lying within the testis determining factor (TDF) (Mc Laren, 1988). This region of the Y (indicated in Fig 1.1) was identified by isolating fragments of the Y

Urogenital ridge (0-4 weeks gestation)

SF1
WT1

Bipotential gonad

SRY   DAX1
SOX9   SOX3

Testes   **Ovaries**

SF1
WT1   *testosterone*
SOX9

Wolffian ducts develop into:

•Internal genitalia
•Epididymis
•Vas deferens
•Seminal vesicles
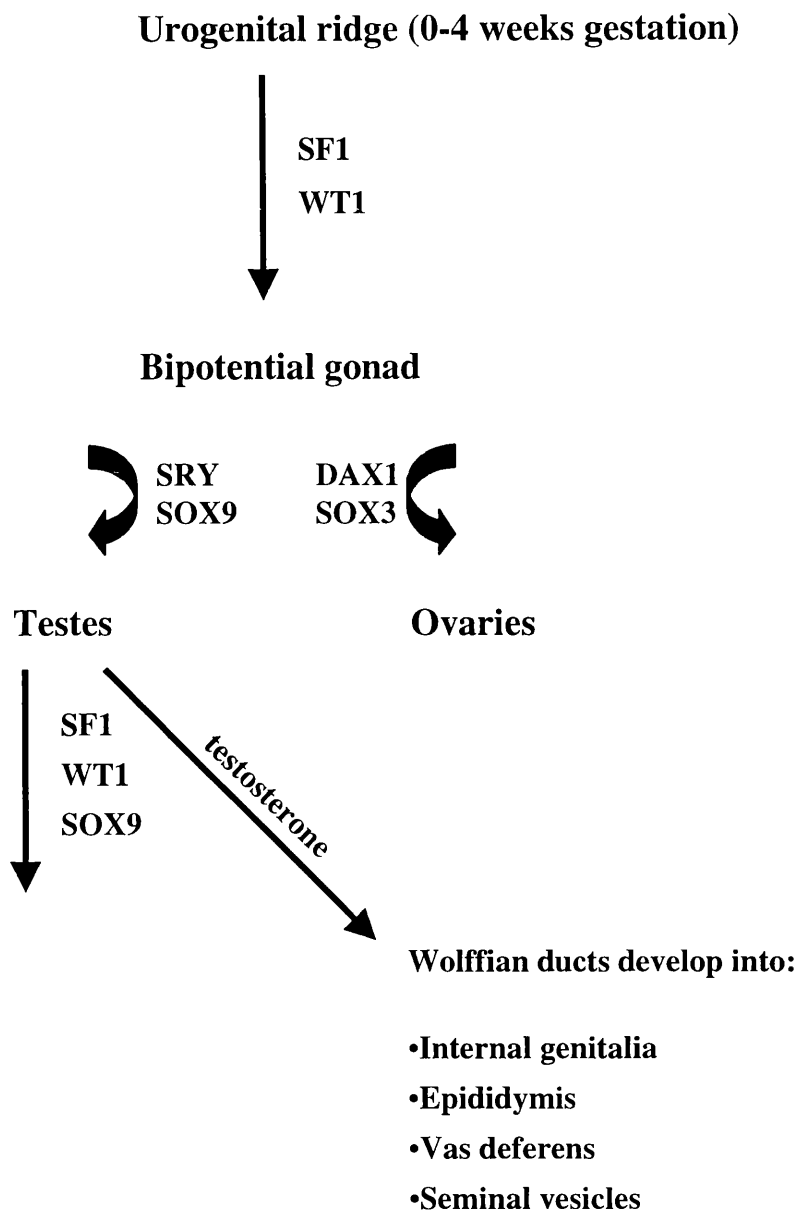
Figure 1.4 A schematic representation of the major stages of sex determination and gonadal development and the genes involved

45

Table 1.4 Autosomal and X-linked genes involved in sex determination and disorders of genitourinary development

Symbol	Name	Location	Comments	Disorders / Syndromes	OMIM No. for disorders	References
WT-1	Wilm's tumor suppressor gene 1	11p3	Zinc finger transcription factor; isoforms, generated by alternative splicing	**WAGR:** Cryptoorchidism, hypospadias **Denys-Drash:** streak gonads, sex reversal external and internal genitalia, pseudohermaphroditism **Frasier:** gonadal dysgenesis, pseudohermaphroditism	194070 194080 136680	Haber *et al.*, 1991 Kreidberg *et al.*, 1993
SF-1	Steroidogenic factor 1	9q33	Orphan nuclear receptor; regulates expression of P450, AMH and androgens	Complete XY sex reversal and adrenal failure; normal external genitalia and retention of the uterus	184757	Lala *et al.*, 1992 Luo *et al.*, 1994 Acherman *et al.*, 1999

46

Symbol	Name	Location	Comments	Disorders / Syndromes	OMIM No. for disorders	References
SOX9	SRY-box 9	17q24.3-q25.1	Transcriptional regulator; expressed in developing male gonad; stimmulates AMH	XY sex reversal	114290	Wright *et al.*, 1995 Haqq *et al.*, 1994
DAX-1	DSS-AHC X gene 1	Xp21.3-p21.2	Within DSS locus; DAX-1 is DSS; encodes a nuclear hormone receptor	Adrenal hypoplasia, congenital with hypogonadotropic hypogonadism	300200	Swain *et al.*, 1998
SOX3	SRY-box 3	Xq26-q27	Transcriptional regulator	Testicular failure	313430 (for gene)	Stevanovic *et al.*, 1993

chromosome present in XX males (De la Chapelle *et al.*, 1984). Thus far a single gene has been isolated from the TDF region and is called SRY.

1.3.2   SRY

SRY, is a small single copy gene located at Yp11.3, within a 35Kb region adjacent to the pseudoautosomal boundary. This gene was isolated by positional cloning (Sinclair *et al.*, 1990) and shown to transcribe an intronless transcript of around 1.1Kb (Lau *et al.*, 1993), with multiple transcription initiation sites and an open reading frame of 223 aa (Behlke *et al.*, 1993). The protein sequence contains a 79 aa region that shares a high degree of homology to a DNA binding motif called the HMG (High Mobility Group) box. SRY, which has a nucleolar localisation, and can induce bending in double stranded DNA, may act as a transcriptional regulator. Studies have shown that it binds with high affinity to the promoters of numerous genes, including P450 aromatase, and AMH (Anti-Mullerian inhibiting Hormone) (King and Weiss, 1993; Nasrin *et al.*, 1991).

Northern blot analysis and immunohistochemical staining, indicate testis specific expression, confined to foetal and adult male Sertoli and germ cells (Sinclair *et al.*, 1990). RT-PCR amplification has identified in addition, low expression of mRNA in a range of other tissues (Goodfellow *et al.*, 1993).

SRY belongs to a large family of genes, all of which have DNA binding domains related to that of the SRY HMG box, with which they share more that 60% homology at the aa level (Lovell-Badge, 1990; Goodfellow *et al.*, 1993). These genes are termed the SOX (SRY-box related) genes, and they seem to have diverse functions as regulators of development. Sequence comparison has shown that SOX3 is the X-linked homologue of SRY. SRY, SOX3 together with the autosomal SOX9, play

significant roles in sex determination (Foster *et al.*, 1994). Absence, mutation or deficiency of SRY, leads to 46,XY sex reversal. In contrast, overproduction of SRY has no effect on sex determination (e.g. 47,XYY males). Sequence analysis of XY females identified a number of mutations in the SRY gene (Cameron and Sinclair., 1997; Poulat *et al.*, 1998), all but one of which lies within the HMG box. In addition, 15% of individuals with gonadal dysgenesis also have mutations in SRY (Lovell-Badge, 1994; Tajima *et al.*, 1994).

In mice, *Sry* exists as an unusual non-polyadenylated circular transcript, which does not appear to be translated and is expressed by the germ cells in the adult testis and by somatic cells in the genital ridge at a time consistent with testis differentiation and development (Gubbay *et al.*, 1990; Hacker *et al.*, 1995). It contains a large CAG trinucleotide repeat region encoding a carboxy-terminal glutamine rich domain arranged as 20 blocks of 2-13 glutamines that act as a transcriptional trans-activator. Insertion of the *Sry* gene lacking this repeat unit in XX mice failed to induce testis formation, whereas insertion of the normal copy of the gene in XX mice resulted in testis development. This indicates that the CAG trinucleotide repeat is significant for testis formation and that *Sry* product may act via a fundamentally different biochemical mechanism in mice compared with other mammals (Koopman *et al.*, 1991; Bowles *et al.*, 1999).

Comparative mapping studies have identified SRY homologues on the Y chromosome of many other species, including mice and marsupials, but only the DNA binding HMG box is conserved, whereas outside this region, the sequence shows poor conservation with many amino acid changes, suggesting that the exact function of non-HMG box region varies from species to species (Goodfellow *et al.*, 1993; Tucker *et al.*, 1993). In addition, an SRY- related sequence, showing close homology with the mouse and human SOX3 gene, has been isolated from the marsupial X

chromosome. This suggests that SOX3 and SRY diverged in sequence and function from an ancestral autosome, at 130 million years ago (Graves *et al.*, 1998). In marsupials, SRY shows expression in many foetal and adult tissues, including testis and its role seems to be more similar to human SRY (Renfree *et al.*, 1995).

1.4 Spermatogenesis and the Y chromosome

1.4.1 Localisation of the Azoospermia factor (AZF)

AZF (Azoospermia Factor), is a general term used to describe a Y-chromosomal region, within which are located genes, which when deleted or mutated disrupt the normal process of spermatogenesis, and lead to reduction or loss of fertility in males.

Tiepolo and Zuffardi (1976) reported six cases of azoospermia associated with large deletions of the distal fluorescent heterochromatic part of the long arm of the Y and the non-fluorescent euchromatic region proximal to it. Subsequent studies of similar cases of infertility, confirmed this association and led to the proposal that a number of factors essential for spermatogenesis, exist on the long arm of the Y chromosome. This region was designated the azoospermia factor locus (AZF) and was mapped to interval 6 within the band Yq11.23 of the Y chromosome (Ma *et al.*, 1992). Vogt *et al* (1996) mapped this locus in more detail, by careful deletion analysis of the Y chromosomes of azoospermic men, all of which showed small Y-chromosome deletions. Results suggested that rather than a single AZF locus, there are three non-overlapping regions on the long arm of the Y chromosome, and these, have been termed AZFa, AZFb and AZFc (Figs 1.2 and 1.5).

Telomere

Pseudoautosomal region

X-Y homology region

Non recombining region

Centromere
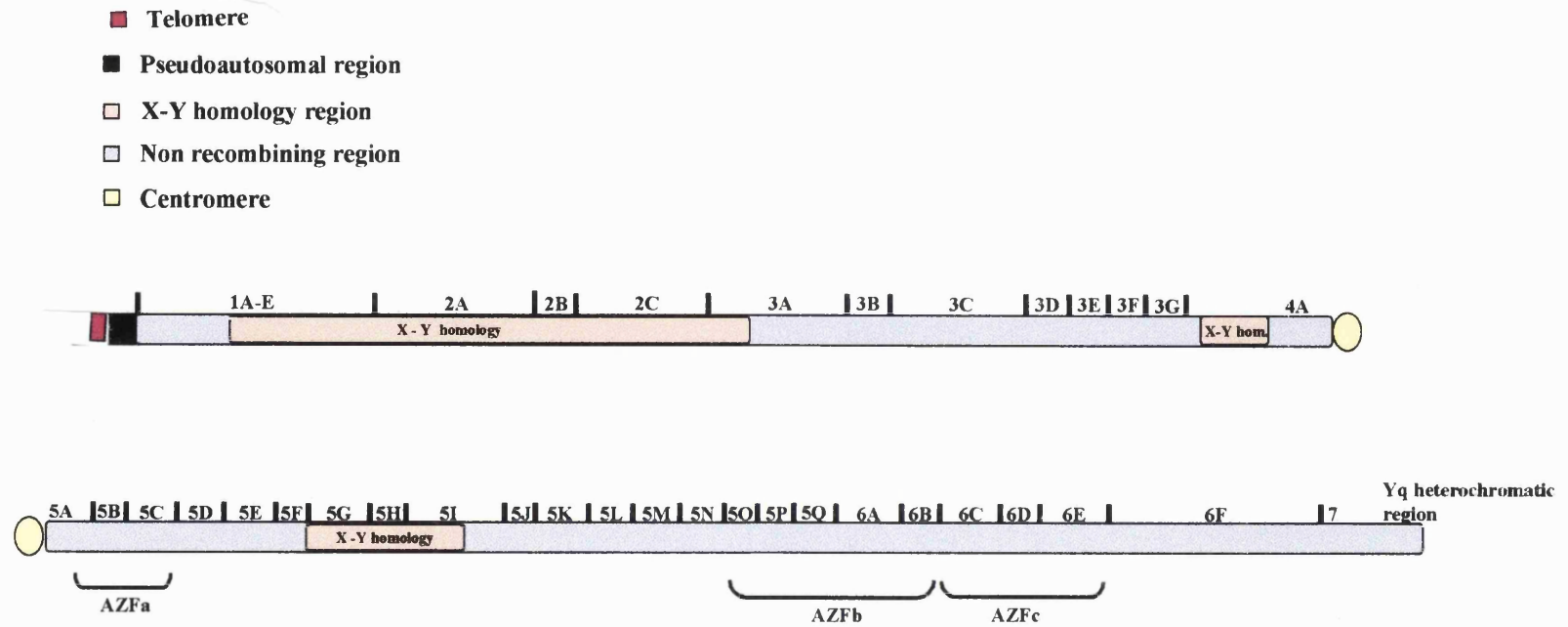
Figure 1.5 Diagram of chromosome Y mapping intervals according to Vollrath map (1992)

The AZFa region is located in proximal Yq11.2 within interval 5 and is estimated to be 1-3Mb in size. A number of genes, both Y-specific and X-Y homologous have been identified in this region. AZFb is located in Yq11.23 extending between intervals 5 and 6 and is around 1-3Mb in size, whereas AZFc (1.4Mb, formerly termed the AZF minimal region) is located to Yq11.23 proximal to the heterochromatic region (Fig 1.2) (Review Krausz and McElreavey, 1999). Detailed analysis of patient pathology indicated that each locus is important during different phases of germ cell development. For example testicular histology of a number of infertile or subfertile patients indicated that patients with deletions within the AZFa region predominantly suffered from SCO syndrome type I (Sertoli-Cell Only), with no germ cells present at any stage Therefore, disruption of normal spermatogenesis in these cases might occur before puberty. Patients with deletions within the AZFb region usually have pre-meiotic (spermatogonia and primary spermatocytes), but no post-meiotic germ cells, which suggests a disruption in spermatogenesis at puberty either before, or during meiosis. Finally patients with deletions within the AZFc region, appear to be either infertile or sub-fertile and usually possess a small number of germ cells in the seminiferous tubules. This suggests that AZFc gene products are involved in the maturation process of the postmeiotic germ cells (Vogt *et al.*, 1996). Recent studies based on more extensive deletion mapping, suggest the existence of a fourth region, AZFd between AZFb and AZFc (Kent-First *et al.*, 1999). To date, statistical data suggest that around 10% of men with primary azoospermia, have deletions or microdeletions within the AZF regions of the long arm of the Y chromosome (Nakahori *et al.*, 1996).

A number of candidate genes thought to be involved in spermatogenesis have been isolated from these regions by positional

cloning. Before discussing these in section 1.4.6, the process of spermatogenesis will be described.

1.4.2   Spermatogenesis - Description of stages

Spermatogenesis does not occur until puberty when sufficiently high levels of sex hormones, such as testosterone, luteinizing hormone (LH) and follicle-stimulating-hormone (FSH), are produced by the anterior pituitary gland.

Immature germ cells - spermatogonia, are located along the outer edge of the seminiferous tubules and are activated at puberty. The spermatogonia stop proliferating, differentiate into diploid primary spermatocytes and enter 1st meiotic prophase, where their homologous chromosomes pair and form crossovers. Secondary spermatocytes are produced, each containing 22 autosomes and either an X or Y chromosome. The result of meiotic division II is that each spermatocyte produces 2 spermatids, each with haploid number of chromosomes. The spermatids do not divide again, but undergo a series of morphological changes resulting in the production of mature spermatozoa, which escape from the lumen of the seminiferous tubules and pass into the epididymis. Here they are stored and undergo further maturation (Fig 1.6) (Alberts *et al.*, 1989).

The progeny of a single spermatogonium remain connected by cytoplasmic bridges throughout their differentiation into mature sperm. In this way, each haploid sperm, can be supplied with all the products of a complete diploid genome. Some sperm will receive an X chromosome while others will receive a Y. Since the X chromosome carries many essential genes not present on the Y chromosome, Y-containing sperm benefit from the cytoplasmic bridge by receiving the gene products that they are not able to produce (Alberts *et al.*, 1989).

53

MITOTIC STAGE

Migration into the testis early in embryogenesis

← **Primordial germ cell (diploid)**

← **Type A spermatogonia (diploid)**

← **Type B spermatogonia (diploid)**

MEIOTIC STAGE I

Prophase: Chromosome pairing
Condensation
Double strand breaks
Synapsis
Recombination
Chiasmata
Metaphase
Anaphase
Telophase

← **Primary spermatocyte (diploid)**

MEIOTIC STAGE II

Prophase
Metaphase
Anaphase
Telophase

← **Secondary spermatocyte (haploid)**

Germ cell maturation with loss of most of the cytoplasm

← **Round Spermatids (haploid)**

← **Elongated spermatids (haploid)**

Residual body

Phagocytosed by germ cells

← **Mature spermatozoon connected with cytoplasmic bridges (haploid)**

Figure 1.6 Diagram of spermatogenesis process

54

The transformation of spermatids into mature spermatozoa include a number of processes such as nuclear condensation; replacement of histones with the predominant sperm nuclear proteins, (protamines), acrosome formation, loss of most of the cytoplasm, development of a tail and arrangement of the mitochondria into the middle piece of the sperm which basically becomes the engine room to power the tail (review, Hecht, 1998). The duration of an entire spermatogenic cycle in humans is 74 days. During maturation in the epididymis, the sperm develop an increased capacity for motility and acquire the ability to penetrate the oocyte during fertilisation. It has been suggested that the epididymis secretes proteins that become part of the surface of the mature spermatozoon, and are important in post-ejaculatory function of the spermatozoa (reviewed in Cooper, 1990).

Terminally differentiated mature spermatozoa were thought to be transcriptionally inert (Kierszenbaum and Tres, 1975) however, several reports have demonstrated the presence of mRNAs as well as RNA and DNA polymerases (Kumar *et al.*, 1993; Miller *et al.*, 1999; Richter *et al.*, 1999). The mRNAs may play some role in the expression of proteins, which are synthesised in late stages of germ cell maturation (Schafer *et al.*, 1995; Sassone-Corsi, 1997). However, they may simply represent remnants of transcripts important to spermiogenesis, or untranslated, undegraded mRNAs (Miller, 1997; Kramer and Krawetz, 1997; Frayne *et al.*, 1999).

Frequently, a small number of abnormal spermatozoa can be observed in normal fertile individuals (Seuanez *et al.*, 1977), which are due to random alterations in the sperm heads and tails that have escaped any apoptotic mechanisms and proceeded to full maturation. However, in infertile patients with a high percentage of abnormal sperm ("teratozoospermia") (Chemes *et al.*, 1999), such abnormalities usually have a genetic origin (Lee *et al.*, 1996). Abnormalities can include sperm with more than one head or no head at all, round-headed spermatozoa with abnormal or non-existent acrosome

(Lalonde *et al.*, 1988; Baccetti *et al.*, 1991) and structurally abnormal flagella.

1.4.3 Hormonal control of spermatogenesis

As mentioned earlier, spermatogenesis does not occur until puberty, when levels of LH and FSH are high (De Kretser, 1993; review, Sassone-Corsi, 1997). LH stimulates the Leydig cells to produce testosterone, which rapidly enters the circulation. Testosterone diffuses into the seminiferous tubules, where it is converted into dihydrotestosterone. In the presence of combined FSH and testosterone stimulation, Sertoli cells secrete peptides and other components into the seminiferous tubules (Griswold, 1995). These stimulate the development of the male reproductive organs, cause the testes to descend and lead to the development and maintenance of secondary sexual characteristics.

A number of hormones secreted by both sexes play a role in spermatogenesis. For example, in males estrogens control the number of Sertoli cells by inhibiting their production throughout life and act directly on male germ cell development (Sharpe, 1993; Sharpe *et al.*, 1998). Synthesis of estrogens is catalysed by the enzyme aromatase cytochrome P450, deficiency of which causes spermatogenic impairment. (Robertson *et al*; 1999).

Another hormone secreted by both sexes is progesterone. In high concentrations progesterone inhibits the secretion of FSH and LH (De Kretser, 1993). At lower concentration in males, progesterone works together with the oocyte zona pellucida ZP3 protein, to induce the acrosome reaction (Osman *et al.*, 1989). In addition, 17OH-progesterone is a precursor of two other androgens, androstenedione and testosterone (Andersson and Moghrabi, 1997).

A number of male infertility disorders have been associated with hormone deficiency. Examples include Kallman's syndrome, "fertile eunuch" syndrome, isolated FSH deficiency and congenital hypogonadotropic syndrome. Kallman's syndrome (OMIM 14950), which is due to a deficiency of LH and FSH is uncommon (1 in 10,000), but is second only to Klinefelter's syndrome as a cause of hypogonadism. The "fertile eunuch" syndrome is associated with LH deficiency and low testosterone levels, with normal plasma FSH but few sperm in the semen. FSH deficiency, which is rare, is associated with a lack of response to gonadotropin-releasing hormone (GnRH) stimulation, and baseline levels of LH and testosterone. In this case the sperm count ranges from zero to a few (OMIM 228300). Congenital hypogonadotropic syndromes associated with secondary hypogonadism due to a deficiency of GnRH), include the Prader Willi syndrome (OMIM 176270), Laurence-Moon syndrome (OMIM 245800) and Bardet-Biedel syndrome (OMIM 209900).

1.4.4   Growth factors and spermatogenesis

Sertoli cells produce a number of growth factors of significance for spermatogenesis. One is the seminiferous growth factor (SGF), which stimulates somatic cell proliferation in the testis, during foetal and postnatal development (Feig *et al.*, 1980). Other examples include the transforming growth factors β (TGF-β), a large family of related polypeptides, which regulate a wide spectrum of cellular processes including spermatogenesis (review, Massague, 1998). Examples of TGF-β members involved in regulation of spermatogenesis include inhibin, activin (secreted by the Leydig cells), which can inhibit or activate the secretion of FSH from the pituitary (Bremner, 1989), and play a role in the regulation of testosterone synthesis in Leydig cells (Lin *et al.*, 1989). Bone morphogenetic protein 8B

(BMP8B) is another members of the TGF-β family, is expressed in male germ cells, suggesting a role in spermatogenesis. Studies of mice with targeted mutations in BMP8B gene indicated that this growth factor is associated with germ cell proliferation and survival (Zhao *et al.*, 1996).

1.4.5   Germ cell gene expression

Before discussing the Y-linked genes involved in spermatogenesis, a number of genes, which are important in this process and which are located on autosomes should be mentioned.

In general, genes expressed in the germ cells during spermatogenesis can be categorised into three classes: those genes where expression is altered during spermatogenesis, those that encode protein isoforms expressed specifically during spermatogenesis and those that are expressed either exclusively or mainly during spermatogenesis (Willson and Ashworth, 1987).

One example of a gene showing both altered levels of expression and differentially expressed isoforms is the transcriptional regulator CREM (cyclic AMP-responsive element modulator. In testis, the CREM gene appears to modulate its own expression by generating cell-specific isoforms, which regulate the cAMP signal transduction pathway, which in turn controls the hypothalamus-pituitary axis function (Stehle *et al.*, 1993). CREM α, β and γ isoforms are weakly expressed in premeiotic germ cells, where they act as spermatogenic inhibitors (Walker *et al.*, 1994), while the CREMτ isoform is generated by alternative splicing after the late pachytene spermatocyte stage (Foulkes *et al.*, 1993). CREMτ accumulates in large amounts and acts as a transcriptional activator; it has also been proposed that CREMτ may be incorporated into the structure of the sperm head, and plays

a role in post-fertilisation events (Foulkes *et al.*, 1992). Experiments in transgenic mice lacking the CREM gene have shown that deficiency can impair spermatogenesis and lead to germ cell apoptosis (Nantel *et al.*, 1996, Blendy *et al.*, 1996).

An example of a gene expressed specifically during spermatogenesis is phosphoglycerate kinase 2 (PGK-2). PGK-2 is an intronless autosomal gene, coding for a testis-specific isozyme of PGK whose expression is initiated with the onset of male germ cell meiosis, and is maintained at high levels in later stages of spermatocytes and in the round spermatids (McCarray and Thomas, 1987, McCarrey *et al.*, 1992).

Another example of this type is the heat shock proteins, which are the germ cell-specific isoforms of the HSP-70 gene family. HSP70-2 in both humans and mice and *Hsc70t* in mice, are synthesised during spermatogenesis. These protein chaperones are thought to be linked to mechanisms that inhibit apoptosis of spermatocytes (Dix *et al.*, 1996; Tsunekawa *et al.*, 1999).

An example of genes expressed exclusively during spermatogenesis are the protamines, which constitute the major class of sperm nuclear proteins. There are two protamine families PRM1 and PRM2, members of which are encoded by genes clustered to 16p13.3 (Viguie *et al.*, 1990). Protamines are located in the adluminal region of the seminiferous epithelium, and expressed postmeiotically in haploid round and elongating spermatids where they replace histones (Wykes *et al*; 1995). It has been suggested that they are associated with sperm nuclear shape and initiate condensation of the chromatin and packaging into the sperm head (protamine gene PRM1). The precise timing under which histone is replaced by protamine is crucial, since alteration in the timing of protamine synthesis results in arrest of spermatid differentiation (Lee *et al.*, 1995). Interestingly,

the *Hsc70t* chaperone is expressed at the round spermatid stage and may be involved with chromatin condensation.

The sperm adhesion molecule 1 (SPAM1) is another example of a spermatogenesis specific gene. The protein product is present on both sperm plasma and acrosomal membrane and is involved in sperm-egg adhesion by aiding the penetration of the egg's cumulus cells layer (Jones *et al.*, 1995).

1.4.6 Candidate AZF genes

In this section, genes that map to the azoospermia factor region of the Y chromosome will be considered.

RBMY   RBMY is a multicopy gene family, which is thought to be involved in spermatogenesis. RBMY1, the first of this family to be identified (Ma *et al.*, 1993) was isolated after the identification of a CpG island in a cosmid clone mapped to Yq11. The RBMY family consists of approximately 30 genes and pseudogenes spread over both arms of the Y chromosome, with functional copies being localised within the AZFb region (Prosser *et al.*, 1996; Elliot *et al.*, 1997). RBMY members code for proteins with a 90 aa RNA recognition motif (RRM) and copies of a tandemly repeated amino acid sequence of unknown function called the SRGY box (Ma *et al*; 1993; Prosser *et al.*, 1996). Studies have indicated the presence of at least six RBMY subfamilies termed RBMY1 to RBMY6. Of these, RBMY1 is the largest and is the only family to contain active members (Prosser *et al.*, 1996; Chai *et al.*, 1997). Six highly homologous RBMY1 genes have been identified (RBMY1A to RBMY1F), all of which differ by only one to seven bp, and each of which contains four SRGY repeats (Chai *et al.*, 1998).

The function of RBMY1 genes has not been elucidated. However, RBMY1 genes share around 67% homology with hnRNPG an autosomal gene located on chromosome 6. Both hnRNPG and RBMY genes are members of the hnRNP (heterogeneous ribonucleoproteins) protein family, which share a common RNA recognition motif (RRM) and are involved in pre-mRNA metabolism and splicing (Soulard *et al.*, 1993; Weighardt *et al.*, 1996). In contrast with RBMY, hnRNPG protein contains no SRGY repeat units (Elliot *et al.*, 1996). RBMY is expressed exclusively in male germ cells of foetal, prepubertal and adult testis. In adult male germ cells RBMY is expressed after the cells have entered meiosis, which indicates that RBMY is not absolutely essential for development up to the meiotic stages of spermatogenesis (Elliot *et al.*, 1997).

RBMY1 is evolutionarily conserved and homologues have been identified in many vertebrates, including marsupials (Ma *et al.*, 1993; Cooke *et al.*, 1996; Delbridge *et al.*, 1997). In mouse, an *Rbm* homologue with 66% similarity to human RBMY1 has been mapped on the short arm of the Y chromosome, between *Sry* and the centromere. *Rbm* seems to exist as a multicopy gene family, expressed only in the foetal and adult mouse germline, but unlike the human homologue, it contains only one SRGY motif (Elliot *et al.*, 1996; Laval *et al.*, 1995).

Recently, an X-linked homologue of RBMY designated RBMX has been found both in humans and marsupials. It encodes the widely expressed protein hnRNPG and is thought to have arisen by retroposition of the RBMY gene. Like other genes with counterparts on both the X and the Y chromosome (*Zfx/Zfy*, *Sox3/Sry*), RBMX is widely expressed in many tissues, in contrast to the testis-specific expression of RBMY1. In addition, FISH has detected other sequences homologous to RBMY on chromosomes 1, 4, 9 and 11 (Delbridge *et al.*, 1999).

61

DAZ The DAZ (deleted in azoospermia) gene, was isolated after detailed deletion mapping of the Y chromosomes of infertile males, followed by exon trapping to search for transcripts within a set of cosmids covering the common deleted region (Rejio *et al.*, 1995). DAZ is located in the AZFc region on the long arm of the Y chromosome (Vogt *et al.*, 1996) and encodes an RNA-binding protein with a single RRM (RNA Recognition Motif) and seven to ten tandemly arranged "DAZ specific"repeat units of 24 amino acids that appear to be polymorphic within the population (Reijo *et al.*, 1995; Yen *et al.*, 1997). Saxena *et al* (1996) realised that these repeat units are the same as found at the DYS1 locus, a highly polymorphic family of Yq specific sequences. Expression of DAZ is restricted to testis and exclusively to germ cells at early stages of spermatogenesis. It has been proposed that DAZ is involved in the maintenance of germ stem cell number and regulation of the first stages of spermatogenesis (Rejio *et al.*, 1995; Menke *et al.*, 1997).

Like RBMY1, DAZ is also a member of a multigene family. In addition to the multiple copies present on the Y chromosome, there is an autosomal homologue, termed DAZL1 (DAZ-Like) (Yen *et al.*, 1996; Saxena *et al.*, 1996; Rejio *et al.*, 1996; Cooke *et al.*, 1996) which maps to chromosome 3p24 in man and 17 in mouse. DAZL1 is a single copy gene expressed in testis and oocytes, that encodes a 3.3Kb transcript with only one "DAZ repeat" (Yen *et al.*, 1996; Seboun *et al.*, 1997; Nishi *et al.*, 1999).

DAZL1 has a 130bp segment at the 3' end of the coding region that is absent in the Y-linked DAZ copies (Yen *et al.*, 1996; Cooke *et al.*, 1996). Recently, Slee *et al* (1999), has tested the capacity of human DAZ gene, to complement the sterile phenotype of the *Dazl* knockout male mouse (*Dazl*$^{-/-}$), which is characterised by severe germ cell depletion and meiotic failure. Although introduction of the human transgene failed to completely restore

fertility, histological examination revealed a partial and variable rescue of the mutant phenotype.

Autosomal DAZ homologues have been identified in marsupials, old and new world monkeys and various other mammals (Seboun et al., 1997; Delbridge et al., 1997). Y-linked DAZ copies have been identified only in old world monkeys, apes and humans, which suggests that the Y chromosome has only recently acquired a copy of the DAZ gene probably within the last 30-50 million years ago (Seboun et al., 1997; Vogt et al., 1997; Agulnik et al., 1998). A second homologue of DAZ, boule, has been described in Drosophila. Loss of function of boule results in azoospermia due to blockage of meiotic divisions and limited spermatid differentiation (Eberhart et al., 1996).

Agulnik et al (1998) have shown that the intron and exon sequences of the Y-linked DAZ copies are acquiring base pair changes at an equal rate. This implies that the exons are subject to a neutral genetic drift and an absence of any functional selective pressures on these genes. However some, but not all, studies, have shown deletions within the DAZ cluster in 5-15% of infertile and subfertile patients (Rejio et al., 1995; Vereb et al., 1997; Ferlin et al., 1999).

TSPY   In addition to RBMY and DAZ, there is another multicopy gene family on the human Y chromosome thought to be a candidate AZF gene, the Testis-specific protein Y-encoded gene (TSPY) (Arnemann et al., 1987). The gene family consists of around 20 to 40 copies some of which are functional and some pseudogenes. The sequence of members of this family are highly conserved; functional genes share homology as high as 98-99% with each other and up to 90% with their pseudogenes. Members of this gene family are widespread along both arms of the Y chromosome,

where they are arranged in clusters. On Yp, there are two clusters, TSPYA and TSPYB, within interval 3; two further clusters on Yq are located in intervals 4 and 5 (intervals seen in Fig 1.5) (Arnemann *et al.*, 1991; Vogt *et al.*, 1997). TSPY transcription units are around 2.8Kb and are organised as components of the 20Kb DYZ5 repeat units (Manz *et al.*, 1993).

The TSPY genes, with minor differences to each other, are composed of six exons, five introns, and a promoter region of undefined length and sequence (Schnieders *et al.*, 1996). TSPY products have been found as splice variants, but whether these are generated by alternative splicing of the same transcript or derive from different transcripts is still unclear. The 1.3Kb TSPY transcript encodes a 33KD protein, homologous to the proto-oncogene SET (myeloid leukemia associated) and the nucleosome assembly protein 1 (NAP 1) (Schnieders *et al.*, 1996).

Expression of TSPY is restricted to testis in both embryos and adults primarily to the cytoplasm of a subset of spermatogonial cells and around the basal lamina of the seminiferous tubules (Arnemann *et al.*, 1987; Zhang *et al.*, 1992; Schnieders *et al.*, 1996). With these data in mind, Schnieders *et al* (1996), suggested that TSPY might regulate the normal proliferation of spermatogonia, and their entry into meiotic differentiation. TSPY is expressed aberrantly in tumors with germ cell origins and in epithelial cells of prostate cancers and may play a role in the genesis of these tumors. The case for TSPY as an important gene in gonadoblastoma is reviewed by Lau (1999). It is relevant that the "gonadoblastoma" critical region has been located to the short arm of the Y chromosome, where TSPY copies are located (Tsuchiya *et al.*, 1995).

TSPY homologous sequences are conserved in great apes (Schempp *et al.*, 1995), and mammals (Vogel *et al.*, 1997). In mouse, *Tspy* has been identified as a Y-linked, single copy gene. However, it produces only low levels of aberrantly spliced transcripts, and is thought to be a pseudogene

(Mazeyrat and Mitchell, 1998). In rat, there are two TSPY genes located to Yp, one is functional and expressed in testis and the other is truncated and possibly non-functional (Mazeyrat and Mitchell, 1998; Dechend *et al.*, 1998). Human and cow represent the most phylogenetically distant species in which functional TSPY sequences were found. In the cow, the TSPY cluster is located on Yp and comprises 50-200 copies. It is not yet clear whether TSPY is present on the marsupial Y chromosome (Delbridge *et al.*, 1997).

Recently, searches of EST databases have identified a number of novel genes, with significant homology to the TSPY-SET-NAP1 family. These have been designated TSPY-Like (TSPYL) and occur in mouse and man. One TSPYL gene is located to human chromosome 6 and mouse chromosome 10 and is ubiquitously expressed (Vogel *et al.*, 1998). Both the human and mouse *Tspyl* homologues are intronless, indicating that TSPYL has arisen by an ancient retroposition event.

1.5   cDNA direct selection technique

1.5.1   Methods used to identify expressed sequences

The fast progress of the human genome mapping project has emphasised the need to be able to identify expressed genes especially those related to clinical disorders. Numerous techniques for gene identification are at present available. The following section explains briefly their principles, advantages and disadvantages.

A method for isolation of expressed sequences is the identification of CpG islands. CpG islands are short stretches of hypomethylated GC-rich DNA sequence that contain rare cutting restriction enzyme sites and lie on the 5' ends of many genes (Lindsay; 1987). CpG islands have been identified either by restriction digest (Ma *et al.*, 1993), or

by a methyl-CpG binding domain column which has the ability to bind and retain methylated sequences according to their degree of methylation and separate them from the hypomethylated GC-rich regions (Bird, 1986; Cross *et al.*, 1994). Nevertheless, around 60% of tissue specific genes do not contain CpG islands and as a result will not be identified by this procedure.

Another method is exon trapping. This involves the cloning of genomic fragments into an artificial minigene that contains a viral origin of replication and promoter site, all located within a plasmid vector. The vector is then transfected into a eukaryotic cell line, where the insert DNA is transcribed into RNA, which is spliced resulting in the removal of intron sequence. The transcribed, spliced RNA is used for cDNA synthesis by RT-PCR amplification. The cDNA products are then PCR amplified with primers corresponding to the flanking sequence, a procedure called exon amplification. The PCR products are subcloned, usually into the pAMP vector and in this way exons present in the genomic segment are isolated (Duyk *et al.*, 1990; Buckler *et al.*, 1991). This method is considered to be efficient at finding exons, but gives false positives due to the presence of cryptic intron/exon splice sites in non-coding genomic DNA sequence, or may miss genes if the cloned fragment lacks exon/intron junctions and are not spliced. In addition, some splice events are temporarily regulated, or are tissue-specific and may not occur in the packaging cell lines.

Differential display RT-PCR (DDRT), is a method that has been developed for the identification of genes that show spatial or temporal specificity of expression (Liang and Pardee, 1992; Zhu and Liang, 1997). The technique involves amplification of partial single stranded cDNA sequences from subsets of mRNAs by reverse transcription and PCR, using an anchored poly-(A) primer and an arbitrary 10-mer primer. Different types of cDNA subsets can be compared, for example cDNAs derived from cells at different stages of induction or differentiation, or cells from neoplastic

and normal tissue. RT-PCR products are electrophoresed side by side and compared. cDNAs specific for the induced state, developmental stage or neoplasia can then be identified and isolated. Reproducibility is a problem associated with DDRT. Since this is a low stringency PCR based method, there may be variable results between experiments, arising because primers may anneal randomly to non-specific cDNA sequences and generate false positives/negatives (personal communication, Moddaressi, 1998).

Subtractive hybridisation is a similar technique and is based on hybridisation of expressed sequences from one cDNA population (tester) to excess of cDNA from another (driver). Unhybridised cDNAs, which represent differentially expressed, tissue specific sequences (Hara *et al.*, 1991) can thus be isolated. However, this method requires a large amount of mRNA and is not particularly effective in identifying low abundance transcripts. A modified version of this technique is suppression subtractive hybridisation (SSH). In this method, differentially expressed cDNA fragments are selectively amplified and non-target DNA amplification is simultaneously suppressed by adding to the hybridisation mix LINE repeat elements (Diatchenko *et al.*, 1996). These procedures select cDNAs by their cell type specificity of expression but it is difficult to see how they could be adapted to provide chromosome specificity.

In addition to these techniques a number of screening strategies have been developed in order to identify potential genes of known chromosomal location. One example is screening of hnRNA somatic cell hybrid cDNA libraries, usually prepared from hybrids that contain only a single human chromosome. The heterogeneous RNA in these hybrids is made up of ubiquitously expressed and tissue specific gene transcripts at various stages of maturation. Human cDNAs can be selected from this pool (or cDNA library) by PCR using *Alu* primers or by hybridisation to total human DNA and further characterised. However, the expression of many genes in

somatic cell hybrids is not the same as in human tissues and many tissue or cell-type specific genes will be overlooked by this method (Liu *et al.*, 1989; Corbo *et al.*, 1990).

Other laboratories have successfully screened human cDNA libraries with genomic sequence probes that are known to be conserved between species (Page *et al.*, 1987), or with chromosome specific YACs and cosmids in order to identify cDNAs whose genes lie in specific chromosomal regions (Hanks *et al*; 1987). Screening with human genomic clones is technically demanding and variable as some genomic clones, particularly YACs, give a very high background due to their complexity. An alternative version of this method which has been used with some success is to screen chromosome specific genomic libraries with pooled cDNAs made from specific tissues (Hochgeschwender *et al.*, 1989; Ferrari *et al.*, 1996).

In addition to these screening techniques, random sequencing of clones from tissue specific cDNA libraries can generate a large number of tissue specific expressed sequence tags (ESTs) (Okubo *et al.*, 1991; Boguski *et al.*, 1993); ESTs can then be rapidly mapped to a chromosome by PCR amplification of somatic cell hybrids (Polymeropoulos *et al.*, 1993; Jones *et al.*, 1997). ESTs represent exons of genes and can be used to retrieve entire genes by screening genomic libraries. Computational analysis of genome DNA sequences using programs like GRAIL or NIX (HGMP), can quickly identify exon intron boundaries.

1.5.2   Direct selection technique

Direct selection, is a fairly new technique, developed in 1991 by two independent groups for rapid identification of cDNAs (Lovett *et al.*, 1991; Parimoo *et al.*, 1991). This procedure, outlined in Fig 1.7, enables the identification of genes, which are expressed in selected tissues or cell type

and which map to a specific chromosome(s) or smaller regions of chromosomes.

In brief, cDNAs are prepared from either a single tissue/cell type or a mixture of different tissues. Prior to selection, cDNAs are either cloned, or attached by ligation to PCR primers. cDNAs can then be amplified en masse by PCR to provide relatively large quantities as starting material for the direct selection procedure. If cDNAs come from more than one tissues they may be attached to different linkers and separated at a later stage.

Highly repetitive elements present in the cDNAs are suppressed by hybridisation with either CotI DNA or total genomic DNA, to a low $Cot_{1/2}$ value. However, low copy repeats like Y specific DYZ5 (section 1.2.2) may not be efficiently blocked and can be present in the resulting cDNA selection library. The levels of contamination from low copy repeat units can vary according to the chromosomal region and the tissue of interest, but in general have been found to represent around 10% of the selection library (Lovett, 1994; Tassone *et al.*, 1995).

cDNAs are selected by hybridisation to cloned genomic DNA from a specific chromosomal region. When the technique was first developed, the genomic target cDNA clones were immobilised onto filters. However, experience has shown that the hybridisation steps are more efficient when the reaction is carried out in solution (Lovett, 1994; Del Mastro *et al.*, 1995; Del Mastro and Lovett, 1997). The target genomic DNA is tagged with biotin 16-UTP by nick translation, to allow capture of genomic DNA and associated cDNAs by streptavidin-coated magnetic beads, after liquid hybridisation. Flow-sorted chromosomes (Rouquier *et al.*, 1995), chromosome or chromosome region specific YACs or cosmids (Guimera *et al.*, 1997; Lahn and Page, 1997; Simmons *et al.*, 1997) are the most commonly used genomic target.

Tissue derived mRNA made into double stranded cDNAs

PCR amplified
pool of cDNAs

Biotinylated, chromosome specific genomic DNAs

block repeats
with Cot-I

Cot-I

Cot-I

Mix and hybridise

cDNA/genomic target products captured by
using streptavidin coated paramagnetic beads

Streptavidin
paramagnetic
bead

•Beads are washed and captured primary cDNAs are eluted
•Re-PCR amplified for second round of selection
•Secondary selected cDNAs cloned and further analysed

Figure 1.7 Diagram of the cDNA direct selection technique

Hybridisation is carried out at high stringency at 65^0 C, to an intermediate $Cot_{1/2}$ of between 100 and 200 (mole x sec/l). After hybridisation, streptavidin-coated paramagnetic beads are used to isolate the biotinylated genomic DNA and their associated cDNAs (Tagle et al., 1993). The magnetic beads are washed to remove any unbound single stranded sequences, and then the cDNAs are eluted from the beads. These cDNAs represent the primary selected cDNAs and are PCR amplified for a second time and passed through a further round of selection, exactly as above (Fig 1.7). Experiments have shown that two rounds of direct selection are usually sufficient to yield an efficient enrichment of up to 100,000-fold; this enrichment is not increased substantially by additional rounds of selection. The secondary selected cDNAs are then cloned into a plasmid vector and can be further analysed (Lovett, 1994; Del Mastro and Lovett, 1997).

The direct selection method appears to have many advantages over the techniques described in the early part of this section. This method rapidly identifies coding sequences, which map to large genomic regions in a way, which does not depend on number and size of introns or cryptic splice sites. The method does not depend upon the presence of CpG islands or transcription in hybrid cell lines (Parimoo et al., 1991). Since this method allows great flexibility regarding both tissue and genomic region, it can be used to identify genes that are active at different stages of development and body locations (Matsubara and Okubo, 1993). The cDNA fragments are collected as ESTs and can be used for the isolation of whole cDNAs and genes (Parimoo et al., 1991).

cDNA direct selection was used by Osbourne-Lawrence et al (1995), for isolation of expressed sequences within a 1Mb region of human chromosome 17q21 that flanks BRCA1 gene. In this case, a direct comparison of efficiency of direct selection with exon trapping was made (Brody et al., 1995). Results indicated similar efficiency in identification of

expressed sequences between direct selection and exon trapping (>90% of sequences were covered by both techniques). In addition, direct selection enabled the identification of two extra low copy transcripts. Since the direct selection method relies on PCR amplification, only small amounts of RNA or cDNA as a starting material are required. However, the selected PCR amplified products consist only of short fragments. As a result, full-length cDNA sequence needs to be obtained by further screening of full-length cDNA libraries or by 5' RACE.

Since 1991, a number of laboratories have used cDNA direct selection with various modifications of the original method to successfully identify and characterise novel genes. A modification of the technique was used to generate selected sublibraries, which contained expressed sequences conserved between human and mouse or pig (Sedlacek et al., 1993). In this experiment the genomic template comprised cosmids covering ~80Kb from human chromosome Xq28. Mouse and pig cDNAs were used as hybridisation probe and allowed the identification of novel cDNA sequences that share homology in humans and mouse and/or pig. This procedure has the potential to identify interesting sequences conserved between species, but is restricted to the isolation of strongly conserved sequences only, with numerous other less well conserved sequences going unidentified.

Tassone et al (1995) used 16 non-overlapping YACs, covering a total of an ~10Mb of human chromosome 21q to select 16 pools of cDNA (one from each YAC) from human foetal brain, whole human foetus, adult testis, thymus, liver and spleen. Analysis of 60 clones indicated that 19 sequenced cDNAs, matched to sequences in databases with a probability of $p<0.01$ and most of these detected mRNA transcripts by Northern blotting. The number of matched clones would be much higher if the experiment were carried out today. It is of interest to note that when the corresponding database sequences were analysed using the GRAIL program for exon

prediction, only 3 of the 19 sequences were predicted to contain easily recognised exons.

Del Mastro et al (1995) produced five cDNA selection libraries using as genomic target a human chromosome 5 cosmid library covering around 174Mb of DNA. As cDNA template five different sources cDNA pools were used; human placenta, foetal brain, thymus, activated T-cells and HeLa cells. 261 clones from the HeLa selected cDNAs, were analysed and more than 50% mapped back to chromosome 5 and represented ESTs, rRNAs and novel cDNAs.

Simmons et al (1995) used a cDNA selection library to identify candidate genes involved in cri-du-chat syndrome. In this case the genomic target DNA was composed of 30 cosmids corresponding to human chromosome 5p15.2 These cosmids were used to select cDNAs from a mixture of placenta, activated T cells and human cerebellum. Out of the 121 independent cDNAs selected, 21 were found to represent 5 separate novel cDNA sequences, emphasising the need to examine a large number of cDNA clones. Two more cDNA selected libraries were prepared from the same region of chromosome 5 by using two YACs spanning ~2Mb to select cDNAs from a foetal brain cDNA selection library (Del Mastro et al., 1995), and a normalised infant brain library (Simmons et al., 1997). In this case though, the selected cDNA libraries were generated by hybridising the purified YACs onto filters containing high-density arrays of the cDNA clones isolated from the previous cDNA selection experiment (Del Mastro et al., 1995). Sixteen cDNAs were identified from the libraries and arrays of these four clones mapped back to the cri-du-chat region and became candidate genes for cri-du-chat.

The project, which is the topic of this thesis, began in 1997 and during the course of that year two further reports appeared, indicating that the direct selection approach was a powerful tool in gene identification.

Guimera *et al* (1997) chose YACs from three regions of chromosome 21, the Down syndrome critical region, and used them to construct a cosmid library. The pooled cosmids were used to select 576 cDNAs from a human, total foetus, cDNA library. 45 cDNAs were sequenced, some of which showed homology with genes, ESTs and STSs, while others were characterised as novel.

Finally and relevant to the present project, Lahn and Page (1997) constructed a cDNA selection library using as genomic target 3600 cosmids containing flow-sorted Y chromosomes to select adult testis cDNA. 3600 clones were sequenced. Of these 308 were putative novel, Y-specific sequences. Of those, Lahn and Page (1997) more fully characterised 12 clones and obtained full-length cDNA for 10 of them. The expression pattern of these 12 clones was investigated by Northern analysis and all were expressed in testis. These data are described in more detail in chapter 3.

1.6 Research aims

When this project began (1997), there had been identified on the Y chromosome around 30 genes, including pseudogenes. To date, there have been 47 assigned Y-linked genes and pseudogenes. In comparison with the X chromosome, which is approximately 3 times larger and carries more than 300 genes, it might be expected that the Y chromosome would carry around 100 genes (Vogt *et al.*, 1997 3rd Chr Y workshop).

This thesis was based on a human "Y chromosome–testis specific" cDNA selection library, made in collaboration with M.Lovett and R.Del Mastro at the University of Texas. The technique was first applied in 1991 by both Lovett and Parimoo and the cDNA selection library described in my thesis was based on slight modifications of the first version, following a

standard protocol that Del Mastro and Lovett (1994) developed in their lab (section 1.5.2).

The ultimate objective of this project was the identification of novel Y-linked genes involved in the process of spermatogenesis, which may prove to be important in male fertility. Thus, the aims of this project were as follows:

1.   To evaluate the human "Y chromosome-testis specific" cDNA selection library.

2.   To identify Y-linked, testis-expressed cDNA clones.

3.   To further investigate some of these cDNAs. Characterisation of these clones included mapping studies, analysis of their expression pattern, exploration of their structure within the genome, identification of comparative homologous sequences.

4.   To investigate their potential involvement in the male infertility phenotype by PCR screening a panel of oligo- and azoospermic individuals.

CHAPTER TWO

Materials and Methods

2.1   MATERIALS

Standard reagents   Analar grade standard reagents were supplied by BDH/Merck, Fisons and Sigma

Enzymes   Restriction enzymes were supplied by Bethesda Research Laboratories (GIBCO-BRL) and New England Biolabs. MMuLV reverse transcriptase and Red Hot *Taq* polymerase were from Advanced Biotechnologies. All other enzymes were from Boehringer Mannheim.

Primers   The majority of primers were supplied by Oswel in solution. Some of the primers were purchased by Genosys or Pharmacia as lyophilised powder and were resuspended in 1ml of 1x TE. All primers were diluted providing a final concentration of 50pmoles/100μl of reaction.

Electrophoresis reagents   High melting point agarose was from Sigma. 40% (w/v) acrylamide, 2% (w/v) bisacrylamide solution was from Severn Biotech. 19:1 Sequagel (manual sequencing) and Sequagel XR (automatic sequencing) acrylamide were from National diagnostics. TEMED was from BDH.

Miscellaneous   Nick Sephadex G-50 columns, RNAse inhibitors, pd(N)$_6$ random hexamers and dNTPs were supplied by Pharmacia. Antibiotics were from Sigma. RNAzolB was from Biogenesis. Hybond N$^+$ membrane filters were supplied by Amersham Pharmacia Biotech. The same company also supplied [α-^{32}P]dCTP, [α-^{33}P]ddATP, [α-^{33}P]ddCTP, [α-^{33}P]ddGTP, [α-^{33}P]ddTTP, *T7* Sequenase v 2.0 and Thermo Sequenase radiolabelled

terminator cycle sequencing kits, *rediprime* DNA labelling kit and shrimp alkaline phosphatase. [α-^{35}S]dATP was supplied by DuPont. The dye terminator cycle sequencing ready reaction with AmpliTaq DNA polymerase, FS was supplied by Advanced Biosystems. 0.24-9.5Kb RNA and 1Kb DNA size ladders were supplied by GIBCO-BRL and 100bp DNA size ladder was supplied by Promega. Promega also supplied the Wizard miniprep DNA purification kit. QIAGEN supplied the QIAquick gel extraction, PCR purification, mini plasmid purification and maxi plasmid/cosmid purification kits. mRNA QuickPrep Micro purification kit was from Pharmacia and Micro-FastTrack mRNA isolation kit from *Invitrogen.*

Commonly used solutions and buffers   Glassware and solutions used in all experiments were sterilised either by autoclaving (15psi, 121°C for 20-25 minutes) or by filter sterilisation through a 0.22μm pore size "Acrodisk" filters (Gelman Sciences).

'Chloroform': refers to a 24:1 (v/v) mixture of chloroform and isoamyl alcohol.

'Phenol': refers to phenol equilibrated with TE, pH 7.5

'Polyacrylamide': refers to a 19:1 mixture of acrylamide and bisacrylamide for manual gels

Denaturing solution: 1.5M NaCl, 0.5M NaOH

Neutralising solution: 1.5M NaCl, 0.5M Tris, pH 7.5, 1mM EDTA

DEPC-treated water: 0.1% Diethyl pyrocarbonate added to ddH$_2$O, incubated for 2 hours o/n at 37°C and then autoclaved

Sequencing plate's bonding solution: 3ml 95% ethanol, 5μl γ-methocyloxylpropyltrimethoxy silane and 50μl of 10% acetic acid

100x Denhart's: 2% Ficoll, 2% polyvinylpyrrolidone, 2% BSA

10x MOPS: 0.2M MOPS, 50mM sodium acetate, 10mM EDTA, pH7.0

20x SSC: 3M NaCl, 300mM sodium citrate, pH 7.0 with citric acid

10x TBE: 890mM Tris-HCl, 890mM boric acid, 20Mm EDTA

1x TE: 10mM Tris-HCl and 1mM EDTA, pH 7.5

Loading buffer: 30% glycerol in TE plus bromophenol blue

TKM1: 0.24gr Tris-HCl, 0.41gr $MgCl_2$, 0.15gr KCl, 0.15gr EDTA, 200ml dH_2O

TKM2: 0.24g Tris-HCl, 0.41g $MgCl_2$, 0.15g KCl, 0.15g EDTA, 4.68g NaCl, 200ml dH_2O

Microbiology media and bacterial strains

Tryptone, Yeast extract and Bacto agar were from Difco.

L-broth (per litre): 10g Tryptone, 5g Yeast extract, 5g NaCl

L-agar (per litre): as L-Broth plus 14g Bacto agar

2x TY broth (per litre): 10g Tryptone, 10g Yeast extract, 5g NaCl

2x TY agar (per litre): as 2x TY Broth plus 14g Bacto agar

DH5α competent cells (host strain for pUC19, GIBCO-BRL)

Ampicillin: 100mg/ml in ddH_2O, filter sterilised through a 0.2μm filter. Stored at 4°C for a short term period used at a working concentration of 100μg/ml

Kanamycin: 25mg/ml in ddH_2O, filter sterilised through a 0.2μm filter. Stored at 4°C for a short term period used at a working concentration of 25μg/ml

DNA samples   Control samples included the Centre d'Etudes de Polymorphisme Humaine (CEPH) families and a small panel of male and female DNA samples from individuals in the MRC HBGU.

DNA from the mouse cell hybrids 3E7 (chr Y), HORL9X (chr X), HORLI (chr 15) hamster cell hybrid 853 (chr Y) and their parents IR3E (mouse) and

Wg11 (hamster) were kindly provided by Prof. S. Povey (MRC HBGU ,UCL).

DNA from YACs containing inserts that form a contig along the Y chromosome were kindly provided by N. Affara.

Human and mouse tissues   The healthy adult human testis tissue was dissected from a testicular cancer patient. Human foetal tissues were obtained from the MRC embryonic tissue bank (Hammersmith Hospital) and were provided by Dr. L. Wong. The sex of the tissues was not determined and their menstrual age was determined by either hand or foot measurements.

Mouse testis and kidney were obtained from a male adult mouse, supplied by UCL biological services.

The human Y chromosome/testis specific cDNA selection library

The selection procedure was carried out by Dr. Michael Lovett and Dr. Richard Del Mastro using first-strand cDNA prepared by J. Cameron from adult human testis RNA and 480 cosmids from each of the two Y chromosome cosmid libraries (n=960). 4,608 cDNAs were selected and amplified in the plasmid pAMP10. The clones were arrayed in 48 96-well microtitre plates in 100μl of 2x TY agar medium containing 100μg/ml ampicillin.

Y-cosmids used for the construction of the cDNA selection library

The Y chromosome cosmids used for the construction of this cDNA selection library originated from two separate genomic libraries. The first was constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, CA 94550 U.S.A (LLOYNCO3 'M') under the auspices of the National Laboratory Gene Library project sponsored by the U.S Department of Energy. The library was donated to our group by Dr. P.J De

Jong. This library was prepared by flow sorting Y chromosomes from the somatic cell hybrid J640-51 and then partially digesting these with the enzyme *Mbo*I. The digests were size fractionated and ligated into the cosmid vector Lawrist 16. The library comprises ~13,000 clones of which 82% are human, 13% are hamster and 5% are non-recombinant.

The second library was prepared by Taylor et al., (1996). For this library, DNA from the somatic cell hybrid 3E7 was digested with the enzyme *Sau3*A1, size fractionated and ligated into the cosmid vector Lorist B. This library comprises 1728 clones and is thought to contain <10% contamination with human chromosomes 1 and 12, caused by the presence of small fragments of these two chromosomes in the genome of 3E7.

2.2   METHODS

2.2.1   DNA analysis

DNA extraction methods

Mini-preparation of DNA from plasmids   cDNA clones were grown individually overnight in a shaking incubator at 37°C in 2mls of L-broth containing 100μg/ml ampicillin. Plasmid DNA was extracted using either the Wizard™ (Promega) or the ABI PRISM™ (Applied Biosystems) miniprep kits according to the manufacturers instructions.

Mini-preparation of DNA from cosmids   Cosmid clones were grown individually overnight in a shaking incubator at 37°C in 2mls of 2x TY broth containing 25mg/ml kanamycin. 200μl of the overnight culture and were then centrifuged for 5 minutes at 13,000rpm in a microfuge. The supernatant

was discarded, whereas the pellet was resuspended in 100μl of 1x TE. The suspension was boiled for 10 minutes at 95°C and centrifuged again at 1,500rpm for 4 minutes. The supernatant was transferred into a new 0.5ml Eppendorf tube and kept at 4°C for further use, whereas the pellet was discarded. For a standard PCR reaction of 50μl volume, 5μl of the extracted DNA was used.

Maxi-preparation of DNA from cosmids   500ml of 2x TY plus 25mg/ml kanamycin were divided into 2x 250ml, placed into two 1 litre conical flasks and each was inoculated with 100μl of cosmid overnight culture. The suspension was then incubated overnight in a shaking incubator at 37°C. DNA was extracted using the QIAGEN Plasmid Purification Midi and Maxi kit, following the manufacturer's instructions for DNA purification of low-copy cosmids.

DNA extraction from blood   DNA was isolated from human blood collected in tubes containing 15% EDTA by the method of Lahrini and Nurnberger (1991). 5ml of whole blood were placed in 20ml polypropylene tubes together with equal volume of a low salt solution (TKM1) and 125μl of Nonident-40 (NP40) to lyse the blood cells. The contents of the 20ml tube were mixed well by inversion and centrifuged at 6,000g for 10mins in a bench centrifuge (Centaur centrifuge). The supernatant was discarded and the pellet was washed in 5ml of TKM1 solution. Another 125μl of NP40 were added and the suspension was again centrifuged at 6,000g for 10mins. The above washing steps were repeated in total three times, or until the pellet became white i.e. all the red blood cells had been washed away.

After the final wash, the pellet was gently resuspended in 50μl of TKM1. 800μl of high salt solution TKM2 and 50μl of 10%(w/v) Sodium Dodecyl

Sulphate (SDS) were added to lyse the white blood cells. After mixing thoroughly by inverting the tube several times, the suspension was incubated at 55°C, with occasional agitation, for a minimum of 30mins, or until it became clear. The suspension was then removed from the 55°C water bath and centrifuged at 10,000g for 5mins. The pellet was discarded, whereas the supernatant was transferred to a new polypropylene 20ml tube, to which 2 volumes of ice cold ethanol (100%) were added. The tube was then gently inverted several times, until all the DNA had precipitated. The DNA strands were then transferred to a microcentrifuge 1.5ml Eppendorf tube containing 1ml of ice cold 70% ethanol, with the help of a sterile inoculation loop. This step was followed by another centrifugation at 10,000g for 5mins to pellet the DNA.

The supernatant from the above step was carefully discarded and the DNA pellet was air dried for approximately 15-30mins. DNA was then dissolved in an appropriate volume of ddH$_2$O, or 1x Tris EDTA and stored at 4°C until further use.

DNA extraction from tissues   DNA was isolated from several human tissues using a modified protocol from Maniatis et al (H. Modarressi, personal communication). Approximately 1µg of tissue was placed in a 5ml plastic bijou with 1ml of lysis buffer, and homogenised using the mechanical homogeniser. The cells were centrifuged at 7,000rpm for 5 minutes in a microfuge. The supernatant was discarded and the pellet was resuspended in 1ml of lysis buffer, mixed well by repeated inversions and centrifuged again at 7,000rpm for 5 minutes. This step was repeated on average three timers, or until the pellet became white. After the final wash, the pellet was dissolved in 100µl of 50mM NaOH and the suspension was boiled at 99°C for 30 minutes. 20ml of 1M Tris-HCL (pH 7.5) were then added in the suspension and mixed well. The suspension was centrifuged at 7,000rpm for 3-5 minutes in a

microfuge, pellet was discarded and supernatant was transferred to another 1.5ml Eppendorf tube and kept at 4°C until further use.

DNA modifications

DNA precipitation   DNA precipitations were routinely performed by the addition of 1/10 volume of 3M sodium acetate (NaAC) of pH 5.5 and 2 volumes of 100% ethanol. The mixture was placed overnight at –20°C. The precipitated DNA was recovered by centrifugation at 13,000rpm in an Eppendorf centrifuge. DNA pellets were washed in 70% (v/v) ethanol, air dried and resuspended either in 1x TE buffer, or ddH$_2$O.

Restriction enzyme digestion   Digests were performed using the incubation buffers provided with the enzymes and according to the manufacturers recommended conditions. Incubations took place in a 37°C water bath, with the exception of *Sma*I, which gives optimum restriction of DNA at room temperature.

Dephosphorylation   The terminal phosphate of the linearised pUC19 plasmid vector DNA was removed in order to prevent self-ligation in subcloning procedures. 1μg of vector DNA was linearised by standard restriction enzyme digestion, in a total volume of 20μl and phosphatased with 0.5 units of shrimp alkaline phosphatase (SAP), for 1 hour at 37°C. The enzyme was heat inactivated for 15 minutes at 65°C. When double digests of pUC19 were performed, the vector DNA was not phosphatased, but simply digested, gel purified and ligated to the DNA fragment to be subcloned.

Ligation   Ligation reactions included 20-50ng of linearised, phosphatased plasmid vector and a 3-fold molar excess of insert DNA. The ligation mixture also included 1x ligase buffer (10x buffer is 660mM Tris-HCl, 50mM MgCl$_2$, 10mM DTT, 10mM ATP, pH7.5) and 1 unit T4 DNA ligase, in a total volume of 10μl. Ligations were incubated at 15°C overnight.

PCR amplification of genomic DNA

A standard PCR reaction combined the following materials; 50-200ng of genomic DNA, 1.5 units of Taq polymerase with 1x Buffer (15mM $MgCl_2$, 1.25ml of 200mM $(NH_4)_2SO_4$, 750mM Tris-HCl, pH 8.8, 0.1% Tween), 0.2mM of dNTPs, 25pmoles of forward and reverse primers and sterile distilled water. The final volume of reaction was 25μl, 30μl or 50μl. Each reaction was covered with 30μl of paraffin oil to prevent evaporation during thermal cycling. After the initial denaturation step (95°C for 5 minutes), 30-35 cycles were performed using conditions established for each primer pair, on a Hybaid thermocycler machine as follows: denaturation at 95°C for 30 seconds, annealing for 30 seconds and extension at 72°C for 30 seconds. The annealing temperatures used were T_m-5°C. The melting temperature T_m is the temperature at which the proportion of annealed and dissociated DNA is 50:50 and is calculated from the primer sequence using the formula:

$$T_m = 69.3 + 0.41(\%G+C \text{ content of primer}) - (650/\text{length of primer})$$

Annealing temperatures were usually modified (either increased or decreased) after the initial PCR, to achieve the best conditions for specific PCR amplification.

For the purification of PCR products, the QIAquick PCR purification kit was used, according to the manufacturer's instructions (QIAGEN).

For PCR amplification the following primers were used:

Table 2.1

Primer name	Sequence	T ann	PCR product (bp)
pAMP10.F	5' TAAGCTTGGATCCTCTAGAGCG 3'	51°C	
pAMP10.R	5' GAATTCCCGGGTCGACTACTAC 3'		
PGM1.F	5' GAAAAATCAAAGCCATTGGTGGG 3'	56°C	420
PGM1.R	5' GGACCGAGTTCTTCACAGAGGAT 3'		
AMLXY.F	5' CTGATGGTTGGCCTCAAGCCT 3'		Chr X: 432
AMLXY.R	5' ATGAGGAAACCAGGGTTCCA 3'		Chr Y: 252
13d10.F (non-specific)	5' TCACCACAGATAGCCACTGAGAC 3'	50°C	299
13d10.R (non-specific)	5' ATCAGGTCCATGGGATTGGAATG 3'		
13d10.2F	5' CAGACTGTGAGTTGGTTCTG 3'	54°C	233
13d10.2R	5' TATGTGAGAGAGACCCTGTG 3'		
22a3.F (non-specific)	5' AGGATGGACCCTAATTTGCAC 3'	49°C	152
22a3.R (non-specific)	5' CACTGAGAAAGACAGATGACAC 3'		
22a3.2F	5' CCTATCTGAGCAGGTACTTTAC 3'	56°C	178
22a3.2R	5' GTGTCATCTGTCTTTCTCAGTG 3'		
22d8.F	5' GATTTAGCCATACTGCTCGG 3'	55°C	264
22d8.R	5' GTATCTGTCTTCCCTGCCAC 3'		

Agarose gel electrophoresis

DNA fragments were size separated through 0.8-2.0% agarose gels, prepared and run in 1x TBE buffer; ethidium bromide was included at a concentration of 100ng/ml. Bromophenol blue loading buffer was added to the DNA samples before loading and the gels were electrophoresed in 1x TBE. Gels were routinely run at approximately 10Vcm-1. DNA fragments were

visualised by ultraviolet (UV) transillumination. DNA fragment size was checked by comparison with 1Kb or 100bp ladder molecular weight markers.

DNA fragments required for sequencing, subcloning or ^{32}P labelling reactions were excised from the agarose gels using the QIAquick gel extraction kit, according to the manufacturer's instructions (QIAGEN).

DNA sequencing

Radioactive sequencing Sequencing was carried out using the dideoxynucleotide chain termination method (Sanger et al., 1977) . Most radioactive sequencing was performed using the Thermosequenase radiolabelled terminator cycle sequencing kit (Amersham Pharmacia Biotech) with ^{33}P labelled Redivue^{TM} terminators. However, at the beginning of this project, sequencing was carried out using the T7 Sequenase version 2.0 DNA polymerase (Amersham Pharmacia Biotech) with [α-^{35}S]dATP. Here, I will describe both techniques.

Radioactive sequencing using T7 Sequenase DNA polymerase and [α-^{35}S]dATP Sequencing was carried out using the Version 2.0 T7 Sequenase kit (Amersham). The first stage of the sequencing reaction is the annealing of the primer to the sequence; this modified method minimises the re-annealing of the template DNA, which can be a problem for short double-stranded PCR products. This was done by a rapid heat-denaturation of the double stranded product and then immediate annealing of the primer. 0.5-1pmoles of DNA, 0.5pmole of primer, 2μl of 5X T7 Sequenase reaction buffer (200mM Tris-HCL pH 7.5, 100mM MgCl$_2$, 250mM NaCl) and H$_2$O were mixed and boiled for 10 minutes. This mixture was then briefly centrifuged and left to cool for approximately 1 minute on ice.

For the labelling reaction, 1μl of 0.1M DTT, 2ml of labelling mix

(diluted 1:5), 0.5μl [α-^{35}S]dATP (1000 Ci/mmol) and 2μl (3.25 units) diluted

T7 Sequenase enzyme (1:8 dilution) were combined. 5.5μl of the labelling

mix was added to the primer-annealed template and incubated at room

temperature for 15-45 seconds. For the termination reaction, Four 3.5μl

aliquots of the labelled mixture were transferred to microtitre wells containing

2.5μl of each ddNTP mix. The mixture was then incubated at 37°C for 2

minutes. The reaction was stopped by adding 4μl of stop mix (95%

formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol

FF).

Radioactive sequencing using Thermo Sequenase DNA polymerase and
[α-^{33}P] ddNTP terminators Sequencing was carried out using

the Thermo Sequenase radiolabelled terminator cycle sequencing kit

(Amersham). 8 units of Thermosequenase DNA polymerase (2μl) were mixed

with 2μl of 10x reaction buffer (260mM Tris-HCl, pH 9.5, 65mM MgCl$_2$),

2.5pmol of primer and 50-500ng of DNA, made up to a final volume of 20μl

with H$_2$O. Four 4.5μl aliquots were mixed with 2.5μl of A,C,G and T

termination mixes (each termination mix is 7.5μM dATP, dCTP, dGTP, dTTP

with 0.075μM (0.225μCi) of the appropriate terminator [α-^{33}P]ddNTP

(1500Ci/nmol)). Each reaction was covered with 30μl of paraffin oil and

placed on a Hybaid Omnigene thermocycler machine using the following

conditions: 40 cycles at 95°C for 30 seconds, annealing for 30 seconds and

extension at 72°C for1 minute. At the end of the sequencing reaction, 4μl of

stop mix (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05%

xylene cyanol FF) were added and reactions were kept at 4°C prior to

electrophoresis.

Electrophoresis of sequencing products   Electrophoresis was

carried out using Biorad sequencing apparatus (21 x 40cm in size, 0.4mm gel

thickness). A denaturing polyacrylamide gel (6% 19:1

acrylamide/bisacrylamide) containing 7M urea, was polymerised with 1.7%

(v/v) TEMED (60μl) and 25% (w/v) ammonium persulphate (80μl) in a total

of 60ml gel mix. The gel was pre-run at 50W for 30 minutes in 1x TBE to

warm the gel to 40-50°C. Samples were denatured at 80-95°C for 5 minutes

before loading and routinely 4-5μl of the sample were loaded using a 12-well

shark's tooth comb. The sequencing gel was electrophoresed at constant 50W

(45-50°C) in 1x TBE buffer.

After disassembling the sequencing apparatus, the gel remained

attached to the front glass plate. The plate was soaked for 20 minutes in 10%

acetic acid and then 20 minutes in tap water. The gel was then dried to the

glass plate in an 80°C oven for 45-60 minutes and then exposed to X-ray film

for 12-48 hours at room temperature, in a light-tight black bag. After use, the

gel was removed from the glass plate by soaking in tap water.

Fluorescent sequencing   DNA was sequenced using BigDye^{TM} and

Di-Rhodamine^{TM} fluorescently labelled terminators, provided with the

Advanced Biosystems dye terminator cycle sequencing ready reaction mix kit.

8μl of Terminator Ready reaction Mix were mixed with 3.2-4pmol of primer

and 200-500ng of DNA, made up to 20μl with ddH$_2$O and overlaid with 30μl

of mineral oil. Cycling was performed on a Perkin Elmer DNA thermal cycler

480 using the following conditions: 25 cycles at 96°C for 30 seconds, 50°C for

15 seconds and 60°C for 4 minutes. After cycling and removal of the mineral

oil, the sequencing products were precipitated by using 10% 3M NaAc and 2x

the volume of 95% EtOH.

Electrophoresis of sequencing products   The following steps

were performed by either Ms Katie Morrison or Ms Wendy Pratt in the

HBGU. The electrophoresis gel consists of 18g urea, 5.2ml 40%(v/v) 19:1

acrylamide SequagelXR (National Diagnostics), 27.5ml H2O and 0.5g

Amberlite MB-1A mixed bed resin. 250µl of 10%(w/v) ammonium

persulphate solution and 35µl of TEMED polymerised the gel mix. Just prior

to loading, the pellet was solubilised in 5µl loading buffer (5 volumes

formamide, 1 volume 25mM EDTA (pH 8.0) containing 50µg/µl blue

dextran) denatured for 2 minutes at 95°C and placed on ice.

1.5µl of the sample were electrophoresed on the ABI 373XL automated

DNA sequencer (Advanced Biosystems) at 30W for 12 hours. Data was

collected using 373XL Collection software and analysed using Sequencing

Analysis version 3.0 and Sequence Navigator version 1.0.1 software.

Primers used for sequencing   In some cases, primers described

for PCR amplification were also used for sequencing. However, in others,

primers were designed specifically for sequencing and are listed below.

Table 2.2

Primer name	Sequence	Tann
M13.F (commercial)	5'CCCAGTCACGACGTTGTAAAACG 3'	60°C
M13.R (commercial)	5 AGCGGATAACAATTTCACACAGG 3'	60°C
M13for1	5' CCTTCTACCTCAGAGTC 3'	50°C
M13rev1	5' AGCCAGACTGGTATCTG 3'	50°C
M13rev2	5' GATGAAGTTGAGAGAGGG 3'	50°C
M13rev3	5' ACTCACTTCCTGGAATGC 3'	50°C
M13rev4	5' ACAGTCGCCCAGTCATTG 3'	50°C

2.2.2   RNA analysis

For RNA work, water was treated overnight with 0.1%

diethylpyrocarbonate (DEPC) at 37°C before autoclaving. All solutions used

for RNA work were prepared with DEPC-treated water unless stated

otherwise and with chemicals from sealed containers. Cuvettes used for

spectrophotometry were filled with 1:1 HCl:Methanol for 1 hour and then

rinsed thoroughly in DEPC-H$_2$O before use.

Isolation of RNA

RNA was extracted from foetal and adult mouse and human tissues

using RNAzol B, a solution containing guanidium thiocynate and phenol,

based on a modification (by Biogenesis) of the method described by

Chomczynski and Sacchi (1987). In brief, frozen tissue was homogenised in

RNAzol B in a mechanic homogeniser (2ml per 100mg of tissue) and RNA

extracted with chloroform (1 volume homogenate to 0.1 volume chloroform).

The aqueous phase was collected by centrifugation and precipitated with

isopropanol (1 volume of isopropanol per volume of aqueous phase) at 4°C for

a minimum of 3 hours and washed with 75% ethanol. After air or vacuum

drying, the RNA pellet was resuspended in DEPC-treated sterile distilled

water. Typical concentrations of RNA extracted from tissues using this

method are 1-5µg/l..

Isolation of mRNA

Polyadenylated RNA from small amount of fetal and adult mouse and

human tissues were isolated using the QuickPrep Micro mRNA Purification

(Pharmacia) and Micro-FastTrack™ (*Invitrogen*) kits according to the

manufacturer's instructions. Homogenisation of tissues took place in a motor

driven homogeniser and after inhibition of RNAses in a solution containing a

high concentration of guanidinium thiocyanate (GTC), the mRNA was

captured to oligo (dT) cellulose molecules, which allow efficient hydrogen-bonding between them and poly(A) tracts. Each purification allowed the isolation of 4-5µg of mRNA.

RNA gel agarose electrophoresis

RNA samples were electrophoresed in 1% agarose midigels (14 x 11cm, 100ml), made up in 1x MOPS and 20% (v/v) formaldehyde in a fume hood. The gel was electrophoresed in 1x MOPS buffer for 1.5 hours at 100V. RNA was visualised by ethidium bromide staining (at a concentration of 100ngml^{-1}) that was added in the samples and uv transillumination.

Preparation of cDNA by reverse transcription and RT-PCR

All procedures were carried out on ice. First strand cDNA was prepared from RNA using MMuLV reverse transcriptase enzyme. Approximately 5µg of total RNA were mixed with 7µl of 5x reverse transcriptase buffer (250mM Tris-HCl pH 8.3, 375mM KCl, 15mM $MgCl_2$ and 50mM DTT), 1µl of 1000pmoles random hexamer primers, 2µl of 20mM dNTPs, 1µl of RNAse inhibitor, 3.5µl of 0.1M DTT and DEPC-treated H_2O to make up a total volume of 33µl. The mix was heated at 65°C for 10 minutes, placed on ice and 2ml (400units) of reverse transcriptase added. The reaction was incubated at 37°C for 90 minutes. Reactions with DEPC-treated H_2O instead of reverse transcriptase were also performed and used as control to check at the PCR stage, for the presence of genomic DNA contamination. Single stranded cDNA was then stored at –20°C until further use.

cDNA was amplified by RT-PCR, using 2.5µl of single stranded cDNA, 25pmoles of both forward and reverse primers, 5mM of dNTPs, 1x enzyme buffer ((1.5mM $MgCl_2$, 1.25ml of 200mM $(NH_4)_2SO_4$, 750mM Tris-HCl, pH 8.8, 0.1% Tween) and 1.5 units of Taq polymerase, in a final volume

of 50μl made up with distilled water. Each reaction was covered with 30μl of parafin oil to prevent evaporation during thermal cycling. After an initial denaturation step, 30-35 cycles were performed, using conditions established for each primer as described in Table 2.1.

2.2.3   Filter hybridisation

Preparing filter lifts of the cDNA selection library

Hybond N+ nylon membranes were cut to be the size as a 96-well microtitre plate and labelled using a pencil, with the plate number. Membranes were then placed on L-agar plates containing ampicillin (100mg/ml). Plasmid clones were spotted onto the filters from the thawed microtitre tray glycerol stocks, using a 96-pin gridding tool. The agar plates were wrapped in Cling film and incubated overnight at 37°C such that the colonies had grown to 0.1-0.2mm in diameter.

After incubation, filters were transferred from the L-agar plates onto a tray with Whatman 3MM paper soaked in denaturing solution, were they were left for 5 minutes (DNA side up). Filters were then transferred onto a tray containing neutralising solution, where they were floated for another 5 minutes. Finally the filters were immersed in 2x SSC for 5 minutes, bacterial debris was removed very gently and allowed to dry for 30 minutes. The filters were then baked at 80°C for 2 hours.

Southern blotting

A 0.8-1% agarose gel containing electrophoresed DNA was first placed in a shaking tray with denaturing solution for 30 minutes, rinsed in distilled water and then placed in a tray with neutralising solution for a further 30 minutes.

The blot was set up on a glass plate placed above a tray filled with 10x SSC; the gel was placed on a layer of Whatman 3MM paper covering the

glass and dipping into the tray filled with 10x SSC. The gel was blotted onto Hybond N+ membrane, pre-soaked in 10x SSC and covered with three layers of Whatman paper 3MM, all of which were cut to fit the size of the agarose gel. Air bubbles trapped in between the membrane and Whatman paper 3MM layers, were removed and transfer of the DNA to the membrane was achieved by capillary action, using a stack of paper towels and a weight on top of the assembly to drive the vertical buffer flow. The gel was left to blot overnight. The membrane was then removed, air dryed for 5-10 minutes and baked in an 80°C oven for 2 hour

Prehybridisation of DNA fixed onto filters

Membranes to be hybridised were placed into hybridisation bottles (Hybaid) with 50mls of hybridisation buffer and incubated for at least 2 hours in a temperature equivalent to the hybridisation temperature (57-65°C) in a rotating hybridisation oven (Hybaid).

Preparation of Sephadex G50 spin columns

Removal of unincorporated nucleotides from the radiolabelling mix was done either by preparation and use of Sephadex G-50 spin columns, or by using the commercial Sephadex G-50 Nick-columns (Pharmacia).

The plunger from a 1ml syringe was removed and the end of the syringe was pluged with siliconised cotton wool. The syringe was then filled with Sephadex G50 suspension using a 1ml Gilson pipette avoiding any air bubbles. The syringe was then placed into a conical 5ml plastic tube and centrifuged at 1,500rpm for 3 minutes. If necessary, the syringe was filled with Sephadex G50 again and re-centrifuged to a final level of 1ml. The Sephadex G50 column was equilibrated by adding 200ml of 1x TE and centrifuging for 1 minute at 1,500rpm. Columns were kept at 4°C until ready for use.

Preparation of ^{32}P-labelled probes

DNA used as probes were labelled by random priming (Feinberg and Vogelstein; 1984), using the Klenow fragment of DNA polymerase I to synthesise the second strand of DNA using denatured double-stranded template DNA and random oligonucleotide primers. DNA was labelled with [α-^{32}P]dCTP, using the *redi*prime DNA labelling system (Amersham Pharmacia Biotech). This system contains a dried, stable labelling mix of dATP, dGTP, dTTP, Klenow enzyme and random primers (9mers). In brief, approximately 25ng of DNA were diluted to a volume of 45μl in sterile distilled water, denatured by boiling for 5 minutes and added to the labelling mix, together with 2-4μl of [α-^{32}P]dCTP. After incubation at 37°C for 30-45 minutes, unincorporated [α-^{32}P]dCTP and primers were removed by passing the labelling reaction through a Sephadex G-50 column. The labelling reaction volume (50μl) was applied onto the filter matrix and eluted with 400μl 1X TE. Percentage incorporation was estimated by comparing the cpm retained on the G-50 Sephadex column, which represent unincorporated [α-^{32}P]dCTP, with the cpm in the eluate. Successful incorporation was generally 60-80%. The DNA probe was denatured by boiling for 5 minutes before addition to the hybridisation solution and filters. Hybridisation buffer was reduced from 50mls to 10mls during radioactive labelling hybridisation.

Post-hybridisation washes and signal detection

The membranes were primarily washed at low stringency (2X SSC) for 5 minutes at temperature which could vary from room temperature to 65°C and thereafter at increasingly higher stringency up to 0.1x SSC and 0.1% SDS (57°C for hybridisation to different species, 65°C for hybridisation to same species). The membranes were then briefly blotted on Whatman 3MM paper to remove any excess liquid, wrapped in Cling film and exposed to X-ray film

at -70^0C (rarely at room temperature) in a light-proof cassette with intensifying screens. Exposures ranged from 4 hours to 4 days. All autoradiography using ^{32}P, ^{33}P and ^{35}S isotopes was carried out by exposure of X-ray film (Kodak biomax MR), at room temperature for ^{33}P and ^{35}S and with an intensifying screen at -70°C for ^{32}P.

Removal of radioactive probe from the filters

Radioactive probes were removed from the membranes by boiling a large volume of 0.1x SSC, 1% SDS and place on the membranes for 10-15 minutes in a shaking tray. The above procedure was repeated three times, or until only a very low level of radioactivity could be detected using a hand-held Geiger counter. Membranes were then air-dried and stored at room temperature.

2.2.4   Fluorescent in situ hybridisation

FISH was carried out by Dr. Margaret Fox (MRC Human Biochemical Unit) on spreads of human, chimpanzee, pigmy chimpanzee and gorilla lymphocyte metaphase chromosomes, using cosmids 2e9 (ml13d10), 2f6 (ml22a3), DG11, GE10 and 2C4 (ml22d8) as probes.

Cultured lymphocytes were incubated with thymidine, to synchronise replication by blocking DNA synthesis. Cells were harvested, fixed and droplets placed onto cold slides. 400ng of DNA was labelled with biotin –14-dATP or digoxigenin by nick translation (Bionick kit, GIBCO-BRL). The labelled probe was resuspended in hybridisation mix, containing 50% formamide, 10% dextran sulphate, 2x SSPE, denatured and incubated with Cot1 DNA and herring sperm DNA for 24 hours at 37°C, prior to incubation with the spreads. Signal detection was achieved using fluorescein isothiocyanate-conjugated avidin. Preparations were mounted in anti-fade solution p-phenylenediamine dihydrochloride, to which fluorochromes

diamidinophenylindole and propidium iodide had been added for counter-staining and banding. Slides were examined under a Nikon Optiphot fluorescence microscope and images were captured using confocal laser microscopy.

2.2.5   Sequence analysis methods

BLAST search was performed using the Tokyo GenomeNet database (http://www.blast.genome.ad.jp/), the NCBI database (http://www.ncbi.nlm.nih.gov/BLAST/) and the HGMP blast search facilities (http://www.hgmp.mrc.ac.uk/). Part of the sequence analysis was performed using NIX (UK-HGMP), which is a blast search and gene identification program. All sequence comparisons were performed by programs supplied as part of the GCG v.10 suite (Genetics Computer Group, Winsconsin) available at HGMP.

The BESTFIT and GAP programs, also from the GCG suite, was used for most sequence alignments and identity statistics. By default, the gap creation penalty was set at 50 and the gap extension penalty set at 3.

The phylogenetic tree was constructed by using a series of programs (part of the GCG v.10 suite). These programs are described in section 4.7.

CLUSTALX was also part of the HGMP genome analysis facilities and provided an integrating system for performing multiple sequence alignments and analysing the results.

CHAPTER THREE

cDNA selection library

3.1 Analysis of the cDNA selection library clones

In order to identify Y-linked, testis expressed cDNAs, a "Y-chromosome, testis-specific" cDNA selection library was screened. The library was made as collaboration between my supervisor Dr. Kay Taylor and Dr. M. Lovett and Dr. R.Del Mastro (Texas University). 1000 Y-cosmid clones were used as a genomic target, to select human adult testis cDNAs. The Y-cosmid clones were derived from two sources; a flow sorted human Chr Y library (Lawrence Livermore) and a Chr Y specific library, prepared from the somatic cell hybrid 3E7, which contains only Chr Y as its human component (Taylor *et al.*, 1996) (M+M). By using the direct selection technique (summarised in Fig 1.7), 4,608 potential cDNA clones were selected and arrayed into 48 microtitre plates.

Although the eventual objective was to identify novel expressed sequences, as a starting point it was necessary to evaluate the cDNA selection library. This was necessary because, despite the fact that the cDNA selection procedure provides a considerable enrichment of cDNAs, experience in other labs has shown that this method is prone to contamination by a variety of non-cDNA sequences (Lovett *et al.*, 1991; Morgan *et al.*, 1992; Del Mastro *et al.*, 1995). The evaluation of the cDNA selection library is the subject of this chapter.

At the start of this project it was known from random sequencing of several selected clones that a proportion of them contain DNA derived from the Y-cosmid vectors Lorist B and Lawrist 16, while others contain fragments of a Y chromosome repeat (DYZ1). It was decided that a first step was to screen the library with cosmid vector DNA in order to identify and eliminate recombinant clones containing vector sequence. In all cases, filters

corresponding to plates 2 to 48 were screened; plate no. 1 was not available for analysis.

Since the cosmid vectors Lorist B and Lawrist 16 contain a significant proportion of sequence homologous to the plasmid vector pAMP10, screening with cosmid vector DNA as probe, failed to distinguish which recombinants contain vector sequence and which did not (Fig 3.1 A). Hence, it was necessary to identify and use as probe cosmid sequence with no homology to the cDNA cloning vector pAMP10. Therefore, the sequence obtained from 29 vector-containing recombinants was aligned. This alignment showed that the majority of the recombinants contained vector sequence from one of three areas of the cosmid sequence that correspond to different regions of the two cosmid vectors (Fig 3.2). One of these areas corresponded to sequence homologous to the *Tn5* (*E. Coli* transposon), which is present in the cosmid vector and contains the neomycin and kanamycin resistance genes. Insert DNA from 3 selected clones, ml1a10, ml1f1 and ml13c8, was chosen to represent the three regions of vector that appeared most frequently as contaminants. Non of these sequences are homologous with the pAMP10 vector. ^{32}P radiolabelled probes were amplified from these cDNAs using the primers pAMP10.F and pAMP10.R, designed very close to the ends of the vector cloning sites to avoid amplification of substantial vector sequence and were used to screen the cDNA selection library (Fig 3.1 B).

In order to eliminate recombinant clones containing DYZ1, the most common Y-linked repeat, the library was also screened with a 350bp PCR product, amplified from genomic DNA using SY160.F and SY160.R STS map pair primers that corresponded to the DYZ1 sequence. As a final screen and to ensure the elimination of clones containing *Alus* and *LINES*, plus other Y-linked repeat sequences, the library was also hybridised to labeled, sonicated genomic DNA from the OXEN cell line, which has the karyotype 49,XYYYY (Fig 3.1 C).

A

Plate ml10

B

Plate ml13

C

Plate ml24

Figure 3.1 Examples of gridded filters containing the cDNA selection library screened with A. cosmid vector genomic DNA; B. cosmid vector specific probes ml1a10, ml1f1 and ml13c8; C. a PCR product containing DYZ1, plus OXEN (49, XYYYY) genomic DNA

99

Figure 3.2 Alignment of cDNA clones ml1a10, ml1f1 and ml13c8 with three different

regions of the cosmid vectors Lawrist 16 and LoristB. Numbers indicate the position (in

bp) that corresponds to the vector and cDNA sequences.

```
Lawrist16    TCAATGCATTAAATGCTTATAACGCCGCATTGCTTGCAAAAATTCTCAAAGTTAGCGTTG    3371
LoristB      TCAATGCATTAAATGCTTATAACGCCGCATTGCTTACAAAAATTCTCAAAGTTAGCGTTG    3808
ml1a10       ------------------------ACGCGCATGCTACAA----TCTCAAG--TAGCGT--    1
                                     ***    *  ***   ******    ******


Lawrist16    AAGAATTTAGCCCTTCAATCGCCAGAGAAATCTACGAGATGTATGAAGCGGTTAGTATGC
LoristB      AAGAATTTAGCCCTTCAATCGCCAGAGAAATCTACGAGATGTATGAAGCGGTTAGTATGC
ml1a10       --GAGATTAGCC--TCAATCGCCAGAGAA-TCTACGAGATGTATGAAGCGGTTAGTATGC
               **   ******   ***************  ****************************


Lawrist16    AGCCGTCACTTAGAAGTGAGTATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGATGT
LoristB      AGCCGTCACTTAGAAGTGAGTATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGATGT
ml1a10       AGCCGTCACTTAGAAGTGAGTATGAGTACCCTGTTTTTTCTCATGTTCAGGCAGGGATGT
             ************************************************************


Lawrist16    TCTCACCTGAGCTTAGAACCTTTACCAAAGGTGATGCGGAGAGATGGGTAAGCACAACCA    3579
LoristB      TCTCACCTGAGCTTAGAACCTTTACCAAAGGTGATGCGGAGAGATGGGTAAGCACAACCA    3932
ml1a10       TCTCACCTAAGCT-G                                               177
             ******* ****



Lawrist16    CTCGCTTCCGGCAATACTCGTAAACCATATCAACCAGCTCGCTGACGTTTGGCAGTCCGG    1160
LoristB      CTCGCTTCCGGCAATACTCGTAAACCATATCAACCAGCTCGCTGACGTTTGGCAGTCCGG    1513
ml1f1        ----------CAATACNCGTAAACCATATCAACCAGCTCGCTN-CGAAAGGCAGTCCGG    1
                       ****** *********************** **    *********


Lawrist16    CGGTAACGGATGCTTCTTCCCGGCACCATGCAACAAACTGCCCGGGTGATGGCAGAAATG    1252
LoristB      CGGTAACGGATGCTTCTTCCCGGCACCATGCAACAAACTGCCCGGGTGATGGCAGAAATG    1605
ml1f1        CGGTAACGGATGCTTCTTCCCGGCTGAGTGAGA                             91
             ************************      **   *



Lawrist16    CGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTG    4151
LoristB      CGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTG    4501
ml13c8       ------------------------------GTCAAGACCGACGTGTCCGGTGCCCTG    1
                                           ***********  ***************


Lawrist16    AATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGC
LoristB      AATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGC
ml13c8       AATGAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGC
             ************************************************************


Lawrist16    GCAGCTGTGCTCG-ACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGT
LoristB      GCAGCTGTGCTCG-ACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGT
ml13c8       GCAN-TGTGCTCGTACGTTGTCACTGA-GCGGA---GGACTG-CTGCTAT--GGCGA-GT
             ***  *******  ************** ****    ****** ******* ***** **


Lawrist16    GCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC    4350
LoristB      GCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGC    4700
ml13c8       GC--GGGCAG------------------------------------------------    146
             **  ******
```

100

Typical autoradiographs of individual gridded filters are shown in Fig 3.1. It can be seen from these examples that a surprisingly large proportion of recombinants contain cosmid vector and Y chromosome repeat sequences. This point is illustrated by considering the results from screening a single microtitre plate, shown diagrammatically in Fig 3.3.

Screening of all 47 filters showed that overall, the library appear to comprise 55.6% recombinants containing vector inserts and 24.7% recombinants containing repeat sequences. In addition, a small proportion of the clones (3.5%) did not grow. As a result, only 16.2% (731 out of the 4,512 screened for contaminants), represent testis cDNAs. This procedure of screening the library for vector and repeat however did not entirely eliminate such sequences. A further substantial group was identified when individual clones were sequenced (see later in this section).

The library was next screened with mouse adult testis cDNA. Mouse cDNAs were chosen as a probe because the problems associated with detecting human specific repeats is removed and because such a screen will detect cDNAs which are conserved between human and mouse. Screening the library (47 plates) identified 41 strongly positive clones (Fig 3.4). 11 of these, had already been identified by previous screening experiments as containing vector (n=8), or repeat (n=3). Presumably, these plasmid clones contain more than one insert DNA, one of which corresponds to vector or repeat sequence and the other, to an expressed sequence conserved between mouse and man. Of the remaining 29 positive clones, 25 were sequenced from the pAMP10.F primer. The sequences were compared with those currently available in the Genbank, EMBL, dbEST and dbSTS databases and this showed that all, but 6 correspond to expressed sequences. More specifically, 16 corresponded to sequence homologous to the testis specific transcript TTY2 (Lahn and Page, 1997) (chapter 4). Three other clones shared overlapping sequences and were

101
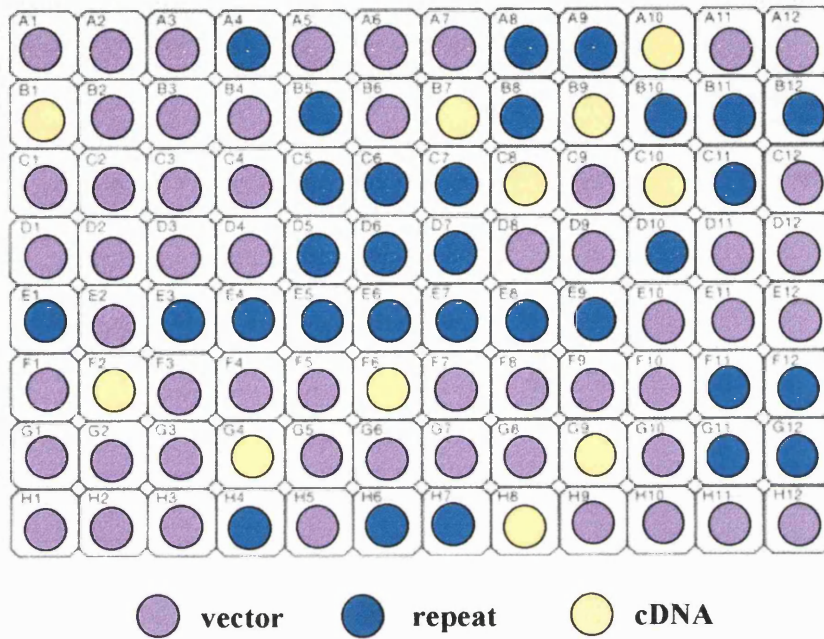
Figure 3.3 A typical example of microtitre plate indicating the large proporti on of clones containing vector or repeat sequences, identified by screening and sequencing.
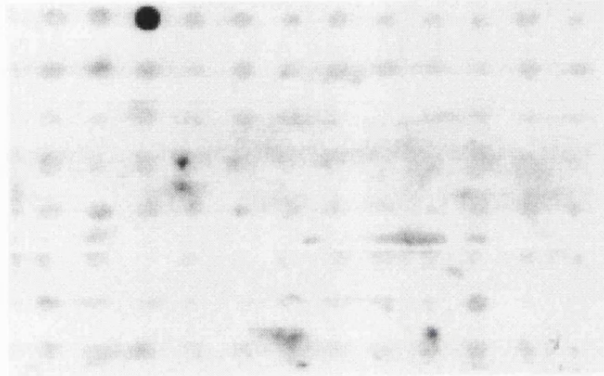
Plate ml24
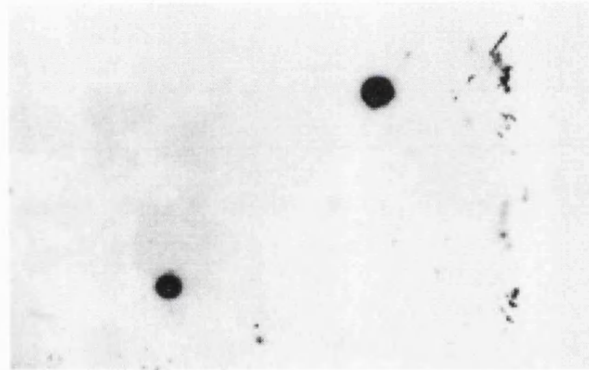
Plate ml11
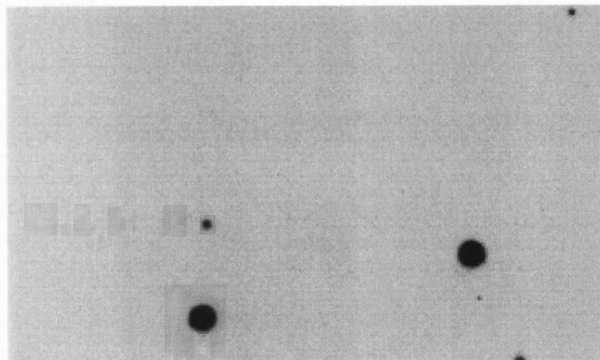
Plate ml16

Figure 3.4 Examples of three different microtitre plates from the cDNA selection library, screened with adult mouse testis cDNA

clearly part of a distinct cDNA (chapter 5). The six remaining clones contained only vector sequence. They were not identified in the earlier screening, as their sequence was not one of the 3 sequences represented in the probe. Thus, only two classes of cDNAs were identified, which were conserved between man and mouse.

Of the 731 clones identified as cDNAs by screening out vector and repeat sequences, 148 were partially or fully sequenced. A preliminary PCR amplification using primers flanking the cDNA inserts was carried out in order to estimate the size of the inserts. This showed that insert sizes range between 200-800bp (Fig 3.5). As each sequence was obtained, a database search using the BLAST program was carried out. Disappointingly, a large proportion of these putative cDNAs also proved to be vector insert or repeats (37.2% and 9.4% respectively). These were discarded and attention focused on the remaining 79 clones (53.4%).

Database comparisons revealed a wide variety of matches with expressed sequences and STSs. The data have been summarised for the sake of clarity, into 8 tables: 37 cDNA clones appeared to represent potential multicopy gene-families, members of which mapped to the Y-chromosome. Table 3.1 lists 19 clones (24.1%) that all showed 65-80% homology to a published testis specific Y-linked gene, TTY2 (Lahn and Page, 1997). In addition, some of them matched with 87-100% homology ESTs from a human adult testis cDNA library (Soares_testis_NHT). Two of these cDNA clones are described in more detail in chapter 4. Table 3.2 lists another 18 Y-linked clones (22.8%), which were related to each other and showed greater than 95% homology with several regions of genomic fragments AC005630 and AC006328 and numerous ESTs. Genomic fragment AC005630 contains sequence from chromosome 15, whereas AC006328 contains Y-linked sequence. Table 3.3 contains 7 cDNAs that matched with between 96 to100% to other Y-linked gene families, the RBMY and TSPY families (see chapter 1)

(i.e. is 1.3% and 7.6% of the sequenced cDNAs respectively). Table 3.4 contains 4 sequences that identified only Y-linked genomic clones, ESTs or STSs (5.0%). Table 3.5 shows two cDNA sequences that matched with >80% homology X-linked or autosomal sequences, including several ESTs derived from testis or germ cell libraries (2.5%). Table 3.6 lists 9 recombinant clones that show homology with non-Y linked genes already identified in the human genome (11.4%). Some of these may well be expressed in the testis and were presumably identified during the selection procedures. Table 3.7 lists 6 cDNA sequences that did not identify any significant match in the database and therefore represent novel sequences (7.6%). Table 3.8 lists 14 clones that contain repetitive elements and which match randomly to all sequences containing the same repeat (17.7%). In these cases, only the first two matches are shown.

A summary of the results from the sequencing of the 148 cDNAs is shown as a histogram in Fig 3.6 A. An overall summary of the evaluation of the cDNA selection library is shown in Fig 3.6 B with a histogram that combines data from both screening and sequencing. These results have indicated a high proportion of contamination from vector and repeat sequences. The proportion of recombinant clones that potentially contain expressed sequences is 16.2% and less than 1% from the library represent sequences that are conserved between mouse and human and are expressed in testis. However, database blast search with sequence available for 79 cDNA clones showed that a good proportion (64%) of them are Y-linked and/or are expressed in testis, since they show homology with ESTs from adult testis libraries. Despite the overwhelming contamination of the selected clones by vector and repeat containing recombinants, the cDNA selection technique has selected for Y-linked and/or testis expressed cDNA sequences.
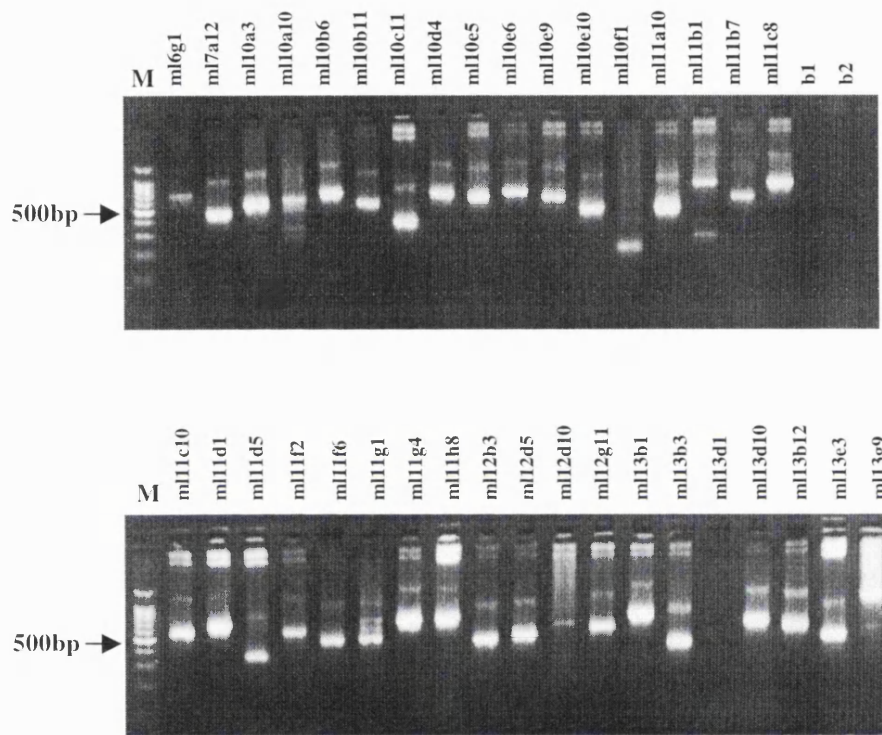
Figure 3.5 Examples of PCR amplification of cDNA selected clones
with pAMP10.F and pAMP10.R primers, carried out in order to
estimate the size of the insert; M: DNA size marker

A

55.6%

24.7%

16.2%

3.5%

Vector   Repeats   Blank   cDNAs

B

33.8%

19.6%

46.6%

Y-linked/testis   other cDNAs   Vector/repeat
expressed
cDNAs
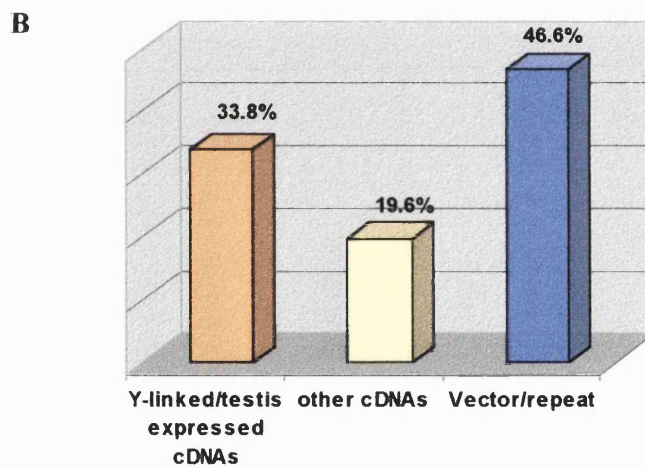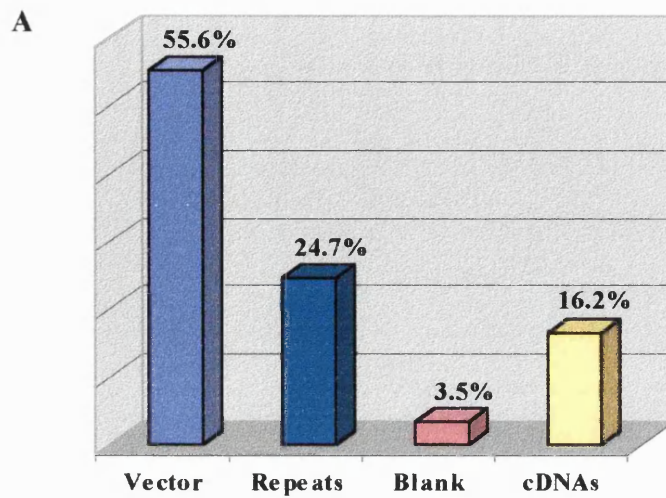
Figure 3.6 Charts showing: A the overall % of vector and repeat sequences present in the cDNA selection library after screening for vector/repeat contaminants and random sequencing of cDNA clones

B.an evaluation of 148 sequenced cDNA clones selected from A (16.2%); Y-linked and/or expressed cDNAs - see tables 3.1, 2, 3, 4 and 5; cDNAs in other categories - see tables 3.6, 7, 8

Table 3.1 cDNA clones with homology to TTY2. cDNA clones grouped under ml22a3, overlap to form a contig. The remainders are independent sequences. cDNA clones marked * were identified after screening the gridded cDNA selection library with mouse testis cDNA and sequencing. All others, were identified by random sequencing of clones from the cDNA selection library.

cDNA	size (bp)	sequenced (bp)	TTY-2 cDNA	
			% homology	region
ml13d10	399	399	72%	12-411
ml18g1	400	220	75%	1248-1468
ml22h10	------	301	77%	1162-1463
ml22a3	576	576	74%	1099-1670
ml18f2	400	337	73%	1104-1437
ml18f12	700	346	73%	1104-1446
ml19h4	600	228	70%	1481-1708
ml11b9	------	155	69%	1537-1692
ml11g4	700	143	65%	1077-1220
ml21a8	------	101	67%	1151-1700
ml20a4*	700	234	78%	1394-1628
ml24a2*	600	624	70.5%	1085-1703
ml24b8*	600	257	69%	1150-1402
ml35a3*	800	209	71%	1127-1332
ml44e5*	800	190	77%	1426-1629
ml44f8*	800	201	75%	1422-1623
ml48d6*	800	230	71%	1101-1396
ml44h5*	800	279	78%	1365-1642
ml6g1*	700	497	71%	1075-1568

Table 3.2 cDNA clones that showed >95% homology with several regions of the genomic clones AC005630 and AC006328 and numerous ESTs. Some cDNAs are indicated as overlapping with identical sequence, forming a small contig. These cDNAs also matched many ESTs with >95% homology, but only the first two matches are given

Name	Insert size (bp)	Sequenced (bp)	Genbank match: AC006328 and AC005630	dEST match
ml10d4	550	523		AW291677;
ml11f6	500	360		germ cell selected
ml7a12	450	338	Overlap with identical sequence	
ml22d8	450	450		AA195274; pooled human melanocyte
ml13d1	-----	374		pregnant uterus and
ml23a8	500	244		fetal heart
ml11a10	500	213	Overlap with identical sequence	AW162304; TAR1 repeat; fetal brain
ml11b1	300+700	134		
ml10b6	700	148		AL110434; testis
ml13g9	900	100	Overlap with identical sequence	------
ml19h6	-----	130		
ml16c1	450	277	Overlap with identical sequence	AW205403; germ cell selected
ml1g1	319	319		AI651385; pooled germ
ml18h8	500	222		cell tumors
ml22c6	389	389	Overlap with identical sequence	
ml24a3*	-----	376		------
ml35g3*	550	416		
ml10a3	600	128	------	------

Table 3.3 cDNA clones that show >96% homology with the Y-linked RBMY and TSPY genes. In addition, they showed homology with two genomic clones, which are bracketed.

cDNA	Insert size (bp)	Sequenced (bp)	Genbank match	% homology
ml10e10	500	156	RBMY2 (AC007320)	98% 98%
ml11c10	600	146		99% 99%
ml11h8	700	158		98% 98%
ml12b3	500	336	TSPY (AC006335)	96% 97%
ml16f10	700	104		100% 100%
ml17c5	450	128		97% 97%
ml22f4	500	94		100% 100%

Table 3.4 cDNA clones that show homology with Y-linked sequences, from either ESTs, STSs, or genomic clones

cDNA	Insert size (bp)	Sequenced (bp)	Genbank, dbEST, dbSTS match	% homology
ml11b7	600	437	AC007359 (chr Y),	82%
			AC006335 (chr Y),	83%
			GMGY26, DYS77 STS	88%
ml11c8	700	76	AC006335 (chr Y),	88%
			AC007967 (chr Y)	89%
ml13b1	750	206	AC006328 (chr Y),	92%
			AC006312 (chr 9),	91%
			AW016837; prostate selected	90%
			AW070214; mixture of fetal lung, testis and B-cell NCI_CGAP	91%
ml11d1	700	84	G42672 cosmid (Chr Y)	96%

Table 3.5 cDNA clones matching to testis/germ cells/prostate or related tumors mapped to X or have autosomal location

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml22a7	350	227	HS347H13; (chr. 22) HSU80040; nuclear aconitase mRNA, encoding mitochondrial protein	100% 100%	AI695810; clone similar to aconitase precursor; adult prostate AI205797; adult testis	98% 98%	G07039; WI-8900	
ml10e5	700	113	HUU91321; BAC clone (chr. 16)	89%	AI989709; Q69384 ENV mRNA; germ cell tumors	95%	------	

Table 3.6 cDNA clones matching to ESTs/genes of unknown, or non-testis origin and autosomal regions or regions not localised on a chromosome

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml12d10	750	534	HSINT2; int-2 proto-oncogene (11q13) AC002110; (9q34)	213bp with 81% 230bp with 81%	AI746920; mouse cDNA to alpha-actinin 2-like clone; Sugano mouse embryo	66bp 86%	G15776 CHLC.GGAT1D8. P9457	60bp 91%
ml13d12	600	114	AF026844; ribosomal protein L41 HSRPL41; human homologue of yeast ribosomal protein L41	100% 98%	AA515966; Ewing's sarcoma AI660373; Q16465 hypothetical protein; thymus	100% 100%	------	
ml13b3	400	374	AB027519; Spirometra erinacei mRNA, expressed in plerocercoid stage	100%	AF107109; laryngeal cancer	100%	------	

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml17h8	550	236	HSVDAC6; voltage-dependent anion channel gene HUMVDAC1X; voltage-dependent anion channel isoform 1 on chr. X	98% 98%	AW129305; clone similar to the outer mitochondrial membrane protein porin, renal cell tumor kidney	98%	------	
ml22e7	710	710	HUMGRAS09; human granulocyte-macrophage colony stimulating factor alpha subunit gene, exon 9	98%	T81944; fetal liver, spleen	98%	------	
ml22f2	------	155	AB019441; DNA for immunoglobulin heavy-chain variable region U66061; human germline T-cell receptor beta chain	87% 93%	AI375641; pooled human melanocyte, fetal lung and pregnant uterus	96%	------	

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml10b11	600	159	HS42A4; PAC 42A4 (6q26-q27) AC007103; (4p16)	97% 97%	AI940256; colon library AA381013; activated-T-cells	96% 95%	U26578; CEPH YAC, (chr X)	95%
ml10e6	800	101 only 42bp match	AC005034; (chr. 2) HUAC00230; (chr. 16)	100% 97%	AA599754; Gessler Willms tumor AA411438; ovary tumor	97% 97%	------	
ml17b7	500	216	U82696; (chr.Xq28) AC004553; (Xp22)	94% 94%	AI393931; B-cell, chronic lymphotic leukemia	99%	AL110090; (chr. 20)	88%

115

Table 3.7 cDNA clones with no match in the database to known expressed sequence

cDNA	size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	dbSTS match	% homology
ml24b8	600	258	AC006343; clone DJ0548K24	84%	------	microsatellite STS in rat	82%
ml10c11	350	119	AC004813; clone 277F10	51bp 90%	------	------	
ml10a10	600	201	------		------	------	
ml11d5	350	85	------		------	------	
ml16f3	300	113	------		------	------	
ml16f12	400	84	------		------	------	

116

Table 3.8 cDNA clones finding the matches shown in the table because of the presence of repetitive element. In each case only the first two matches are shown.

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml11f2	600	207	HSU59100; 1.4Kb centromeric tandem repeat (chr. 9,13,14, 21) HSY10752; centromeric NotI cluster	90% 88%	AA279932; MER22 repeat; tonsillar cells library, enriched for germinal center B cells AI792203; MERR22 repeat; breast library	99% 95%	------	
ml12d5	550	543	AC006502; clone hRPK.36_A_1	84%	AA715301; L1 repeat; germinal center B cell	91%	G59347 SHGC-110534	87%
ml12g11	650	286	HS1054C24; (chr. 20p12.1-13) HS21A2; clone 212A2 maps on chr. 22q12	94% 94%	AI889579; Alu and LTR1 repeats; endometrial adenocarcinoma AL039930;DKFZp434I1012 clone; testis	89% 89%	AF003745; (chr. 11) G63442 SHGC-141351	91% 91%
ml13e3	400	125	HSAC001164; subclone 2_e10 from PAC H92	90%	AA632513; Alu repeat; thyroid tumor AA775032; Alu repeat; adult lung	90% 90%	G10660	88%

117

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml10e9	700	148	AC004668; BAC clone, (7q22-q31) PAC 863K19 (chr. X)	96% 90%	AA682198; schizo brain AI697333; contains a PRT5 repetitive element; endometrial adenocarcinoma	96% 96%	------	
ml16d3	500	78	CA repeat (32bp); the rest shows no matches		------		CA repeat; rat STS	
ml18c10	400	149	AP000539; (chr. 22q11.2)	100bp with 89%	AI761237; CER repeat; kidney	97%	------	
ml21d5	600	244	AC003075; PAC clone, (chr7p21) HS225D2; PAC clone, (Xq11)	92% 91%	N57818; L1 repeat; fetal liver and spleen AI865741; L1 repeat; adult prostate	90% 87%	G27700; SHGC-32987	98%
ml24b8	600	258	AC006343; clone DJ0548K24	84%	------		microsatellite STS in rat	82%
ml27g10	400	176	AC004842; (chr. 7p22) HS416J7; (chr. 6p25) HSCOS10; (chr. 3) subtelomeric region	98% 98% 98%	AI024069; testis AI805123; Alu and MSR1 repeats; B-cell, chronic lymphotic leukemia	98% 97%	G10660; G63710; (chr. 3)	89% 87%

118

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml39e9	600	125	AC004987; clone DJ1173I20 HSCOS10; (chr. 3) subtelomeric region	100% 98%	------	------	------	------
ml29c10	------	231	AC005317; PAC clone, (chr. 15q26.1) HS102D24; (chr. 22q13.31-13.32); contains a novel mitosis-specific SMC1 LIKE protein gene, a novel gene, and exon 1 of the FBLN1 gene for Fibulin 1	100bp with 90% 110bp with 89%	AA644545; Alu and MER22 repeats; lung carcinoma AI524166; Alu repeat; B-cell, chronic lymphotic leukemia	97bp 90% 88%	AL117301; (chr. X)	106bp 87%

Name	Insert size (bp)	Sequenced (bp)	Genbank match	% homology	dbEST match	% homology	dbSTS match	% homology
ml19b3	550	415	HUMRCMYC3; rearranged plasma cell myeloma c-myc gene HSU59100; 1.4Kb centromeric tandem repeat (chr. 9,13,14 and 21)	89% 84%	AI632004; MER22 repeat; kidney AI684722: mixture of fetal lung, testis and B-cell	98% 98%	------	86%
ml10f1	200	118	GT repeat	60bp	------		------	

3.2 DISCUSSION

3.2.1 Evaluation of cDNA selection library starting material

A major resource for this project was the cDNA selection library, made in collaboration with M.Lovett and R.Del Mastro. From this library, several expressed sequences have been isolated and characterised (chapters 4 and 5).

There are at least 13 published examples where direct selection of cDNAs has been used successfully to isolate novel genes involving a variety of chromosomes and cells/tissues and their findings are summarised in Table 3.9. The library used in this project was constructed using the Y chromosome as a genomic target and adult testis as the cDNA source and it seems worth commenting on these two resources before evaluating the results of this cDNA selection experiment.

Comments on the Y chromosome as genomic target The Y chromosome differs from other chromosomes on the basis of its haploid, male specific status, its common ancestry and limited recombination with the X chromosome and the tendency of its genes to degenerate by accumulation of mutations during evolution (Charlesworth., 1978, Goodfellow *et al.*, 1985, Maxson., 1990). The persistence of this chromosome in the mammalian genome indicates that it is home to genes with important functions that during the course of evolution have made it indispensable for the survival of the species.

Compared to other chromosomes of similar size, like chromosomes 21 and 22, there are only a very few Y-linked transcribed genes. To date, only 47 genes have been assigned to the human Y chromosome, including at least 9 pseudogenes, whereas on chromosomes 21 and 22, 190 and 304 genes have been assigned respectively (including pseudogenes) (GDB database, last updated: 4/6/2000). Most of the Y-linked genes, have only poorly been

Table 3.9 List of projects that involve the identification of novel genes by using a cDNA selection approach

Author	Genomic DNA	CDNA library	Libraries produced	Total no. of clones	No. of clones analysed	Clones map back to genomic DNA	% repeats and rDNAs	% vector sequences	% ESTs, genes, STSs	% novel
Parimoo *et al.*, 1991	MHC I,II, β-globin cosmids MHC I, HLA-A YACs	spleen	1	ND	pilot study; search for MHC seq. only	Yes	ND	ND	ND	1
Lovett *et al.*, 1991	550Kb of human Chr 7 as YACs	foetal kidney	1	5000 non-arrayed	pilot study; search for EPO gene only (2%)	Yes	1%	Present but ND	1	ND
Sedlacek *et al.*, 1993	~300Kb of Chr Xq28 as cosmids	foetal brain	1	573	270	100%	26%	ND	24.4% but lots were redundant	75.6% but 7 novel in total

Author	Genomic DNA	cDNA	Libraries produced	Total no. of clones	No. of clones analysed	Clones map back to genomicDNA	% repeats and rDNAs	% vector sequences	% ESTs, genes, STSs	% novel
Cheng et al., 1994	Chr 21 as 72 cosmids	foetal brain	72 1/cosmid	200-400/cosmid	175 clones	61%	20-50%	ND	12%	88%
Rouquier et al., 1995	Flow-sorted Chr 17, 19	foetal brain	2 1/chr.	ND	ND	Yes, but ND	ND	ND	ND	ND
Del Mastro et al., 1995	24,768 arrayed chr. 5 cosmids covering~174Mb (commercial)	placenta, fetal brain, thymus, activated T-cells and Hela cells	5 1/cosmid	4608/library 23040 in total	261 / Hela cDNA	79.5%	21%	0.4%	24.2%	54.4%
Chiannikulchai et al., 1995	10-12Mb of Chr 15q as YACs	foetal and adult muscle cDNA	3 1/YAC	167	167	100%	ND	ND	36%, redundant; 5 genes in total	64% redundant; 12 genes in total
Tassone et al., 1995	10Mb of 21q as YACs	foetal brain, whole fetus, adult testis, thymus, liver and spleen	16 1/YAC	50-100 clones/100Kb of YAC; 5000-10000 in total	60	>90-95% and 77% for a chimaeric YAC	>5-10%	ND	19	40-50

123

Author	Genomic DNA	CDNA library	Libraries produced	Total no. of clones	No. of clones analysed	Clones map back to genomicDNA	% repeats and rDNAs	% vector sequences	% ESTs, genes, STSs	% novel
Simmons *et al.*, 1995	Chr 5p as 30 cosmids	placenta, activated T-cells and cerebellum	3 1/cDNA	2500	121	61%	44%	85% from cerebellum;	0%	7.4%
Simmons *et al.*, 1997	~2Mb of Chr 5p as YACs	chr. 5 specific foetal brain cDNA selection library; foetal brain	2 1/cDNA	4608	56 34	30% 59%	57% 21%	ND ND	4.0% 79%	39% 0%
Touchman *et al.*, 1997	9,600 flow sorted Chr 7 cosmids	foetal brain, placenta, HeLa cells and thymus	4 libraries, with a mixture of cDNA	4 x 4608	2172	96% of 196 ESTs checked	10%	7.3%	82.5%	5%
Guimera *et al.*, 1997	3 regions of Chr 21 as 502 cosmids	foetal brain	1	576	120	100%	20.8%	ND	44%	56%

124

Author	Genomic DNA	CDNA library	Libraries produced	Total no. of clones	No. of clones analysed	Clones map back to genomicDNA	% repeats and rDNAs	% vector sequences	% ESTs, genes, STSs	% novel
Lahn and Page., 1997	30Mb of Chr Y as 3,600 cosmids	adult testis	1	3,600 sequences generated	3,600	24.6%	10.6%	13.8%	59%	8.5%
Makrinou et al., 2000, unpublished	980 Chr Y cosmids	adult testis	1	4608	4512	1% of the sequenced	24.7%	55.6%	16.2%	0.44%

characterised and those that have been most extensively investigated are those like RBMY, TSPY, DAZ and SRY that have a sex determining function, or a role in spermatogenesis (see sections 1.3.2 and 1.4.6). It was not clear at the start of this project what proportion of Y-linked genes had been identified and what remained to be discovered. Even if the number waiting to be identified were low, it seemed likely that any functional genes present on the Y, would be too important to be ignored. A number of other problems were also foreseeable; a significant number of Y-linked gene copies exist as expressed pseudogenes, which cannot be distinguished from functional genes by the selection process. In addition, a relatively high proportion of functional genes will contain repeated sequences and it seemed likely that these might make the interpretation of data more complex.

Comments on testis as a source of cDNA   Testis is a transcriptionally very active tissue, responsible for fulfilling two important functions; the production of sperm, and of the male sex hormones. UniGene, an experimental system that automatically positions GenBank and EST database sequences, into a non-redundant set of gene-oriented clusters and covers almost 50% of the expressed sequences in the genome, derived from a plethora of tissues, has listed a large number of sequences expressed in testis or epididymis.

The testis comprises a very large number of specialised cell types, like the sertoli cells and leydig cells and undergoes two waves of differentiation, one during embryonic development and another marking sexual maturity. Interactions between different cell types in the testis play an important role in the control and maintenance of testicular functions, as well as the growth and differentiation of testis itself. Examples of gene products that contribute, like in many other tissues, in the testis structure are laminin, collagen type I and IV, fibronectin, and proteoglycans (De Kretser, 1993). In addition, hormones,

like estrogens (Sharpe, 1993) and numerous growth factors, like the seminiferous growth factor (SGF) (Feig *et al.*, 1980), or transforming growth factor β (TGF-β) (Review Massague, 1998), play an important role in the regulation of specialised testis cells, like the Leydig and Sertoli cells.

A large number of diverse functions are required to coordinate spermatogenesis; this requires genes that are able to control functions like cell differentiation, specification, survival and proliferation. Some genes have already been identified and include the protamines (Wykes *et al.*, 1995, Lee *et al.*, 1995), the phosphoglycerate kinases 1 and 2 (PGK-1 and 2) (McCarray *et al.*, 1992) and the cyclic AMP-responsive element modulator gene (CREM) (Nantel *et al.*, 1996). Many more remain to be discovered and characterised. Since the Y chromosome is important to the normal control of testis formation and the process of spermatogenesis, it is expected that important genes that control these processes will be located on the Y chromosome. For the purposes of this study the testis was viewed as a male specialised tissue and an excellent source for the isolation of novel, Y-linked male specific genes.

3.2.2   Evaluation of the cDNA selection library

Contamination with non-specific sequences   The evaluation of the Y-linked, testis-expressed cDNA selection library by hybridisation with a variety of probes and by sequencing, revealed that only a small proportion of selected clones contain expressed cDNAs as inserts (731 recombinants, 16.2% of the library). The remaining recombinants contained either vector sequences or genomic Y Chr repeat sequences (>83%). The expected level of such contamination was approximately 10% (Del Mastro and Lovett, 1997) and thus the proportion of contaminants in this library is unusually high.

Fundamental factors that affect the outcome of the selection process are the quality of the starting materials. More specifically, the quality of the

cDNA source is of great importance. A cDNA source that has good length distribution and is low in ribosomal RNA has the potential to give excellent results. In addition, the quality and purity of the genomic template, is also critical. The technique relies on capturing the biotinylated genomic target and its hybridised cDNAs. Thus, it is essential to obtain good incorporation of biotin within the genomic target and experience has shown that cosmids give high levels of enrichment and low background. In addition, the binding capacity of the streptavidin-coated magnetic beads should be adequate, in order to capture the biotinylated genomic fragments with their hybridised cDNAs. Finally, the relative amount of cDNA to genomic target is an important parameter in a selection experiment. In general, the genomic target should be in excess in relation to the cDNA amount. This ensures that the lower abundance cDNAs are efficiently selected.

The starting material used for the construction of the cDNA selection library was made under the supervision of Dr. K. Taylor and is described in Dr. J. Cameron's thesis. The total RNA was prepared using the LiCl/urea extraction method and its concentration was estimated using the spectrophotometer, to be 540µg. The estimated amount of polyadenylated RNA was 10µg and was isolated using Message Affinity Paper. Following the isolation of mRNA, first strand cDNA was synthesised using oligo-dT primers. The yield of single stranded cDNA used in the construction of the selection library was 0.57µg. As a genomic target 480 cosmids from the Lawrence Livermore library (plates 1-5) and 480 Y-cosmids from the Taylor *et al* library (plates D, E, F, G and H) were used. These cosmids provided a coverage of ~51Mb DNA, which was equivalent to approximately 1 Y chromosome. The Lawrence Livermore library was prepared by flow sorting Y chromosomes from the somatic cell hybrid J640-51 and of the clones produced, 82% are estimated to be human, 13% are hamster and 5% are non-recombinant. The Taylor *et al* library was prepared using DNA from the

mouse somatic cell hybrid 3E7 and is thought to contain <10% contamination of human Chrs 1 and 12. Consequently, it would be acceptable to assume that the starting material used for the construction of the cDNA selection library described in this thesis was checked sufficiently and was of good quality.

It is clear that vector, repetitive and ribosomal RNA contamination appears to be for most cDNA selection libraries, a common problem, with figures for proportion of clones containing repeats and ribosomal RNAs, varying from >5% to 60% (Table 3.9). These are fewer estimates of the extent of contamination by vector fragments. Interestingly, in a similar experiment conducted by Del Mastro *et al* (1995) using Chr 5 cosmids and 4 different human tissues, although vector contaminants were not abundantly present (0.4%), the level of contamination from repeat elements was in a similar proportion (21%) with the library described in this thesis (24%).

It is a widely held opinion that there are only a very low number of Y-linked testis expressed sequences. This feature, in combination with the high level of repetitive elements on the Y chromosome, could account for the relatively low proportion of clones, which were genuine cDNAs. However, the Y-chromosome, testis-specific cDNA selection library constructed by Lahn and Page (1997), was found to have an overall level of non-cDNA sequences of >11%, close to that suggested as acceptable and which corresponds to the contamination level suggested by Lovett and Del Mastro (1997). Why do the present results and those of Lahn and Page (1997) differ? A partial explanation may be found in modified experimental methodologies in the construction of the two libraries. An example is that Lahn and Page used a genomic target that provided a 5-fold coverage of only the Y euchromatic region, whereas the cosmids used for the construction of the described selection library, provided only a 1-fold coverage of the Y chromosome. In addition, these cosmids were randomly selected and could include both the euchromatic and the highly repetitive Y heterochromatic

region. Another difference in the two protocols is that Lahn and Page carried out four rounds of cDNA selection, followed by two rounds of subtraction with human Cot-1 DNA. In contrast, in the method that Lovett and Del Mastro followed, the repeats present in the cDNA were blocked with Cot-1 before the genomic target / cDNA hybridisation and also the cDNA selection process was carried out only twice.

Identified Y-linked sequences　　　　Despite the high frequency of clones not containing cDNAs, the described selection library was considered to be successful in the isolation of expressed genes. Of the 148 clones that were sequenced, 79 matched expressed sequences in the database and many of those were proved to be of Y origin (48, including redundant clones).

Random sequencing identified only two known Y-linked genes, TSPY and RBMY and these recombinants accounted for 8.9% of the sequenced cDNAs. Some known Y-linked genes, like DAZ and SRY, were not encountered amongst the 148 clones sequenced, but are expected to be amongst the remaining 583 clones, which were not sequenced. Lahn and Page (1997), who used 3,600 Y-linked cosmids as template, rather than the 980 used in this experiment, identified 16 known genes, both Y-specific and pseudoautosomal, which corresponded to 84.1% of the cDNA containing clones. The Lahn and Page genomic targets were selected to give them five-fold coverage of the 30Mb euchromatic region, which theoretically contained >95% of Y-linked STSs and would provide sufficient genomic material for capturing of most of the Y-linked genes. In contrast, the 980 cosmid clones used to construct the cDNA library evaluated in this thesis, was thought to provide only a one-fold coverage of the entire 50Mb Y chromosome.

Identified expressed sequences　　　　While the scale of success was smaller than for the cDNA selection experiment conducted by Lahn and Page

(1997), a large proportion (64%) of the expressed sequences, identified in the present UCL study, either mapped to the Y chromosome and/or were expressed in testis and its associated structures.

Interestingly, a number of the expressed sequences identified in the UCL testis cDNA selection library, appear to either be proto-oncogenes, for example ml12d10 which matched the int-2 proto-oncogene, or sequences expressed in tumorigenic tissues, like ml13d12 that is expressed in Ewing's sarcoma. Although the adult testis was obtained from a testis cancer patient (the cancerous parts were removed) and this would explain expression of genes related with tumors, proto-oncogene expression in testis is perhaps not surprising, given the high level of cell divisions, both mitotic and meiotic. Several proto-oncogenes, like int-1, c-*myc*, c-*fos* and c-*jun* are involved in cell proliferation mechanisms and as it has been demonstrated these genes appear to be present at several developmental stages of germ cells (Propst *et al.*, 1988, Kumar *et al.*, 1993).

Another interesting feature of the cDNAs isolated from the UCL selection library is that some match ESTs that contain repetitive elements. Table 3.8 lists 14 such cDNAs, which correspond to 17.7% of the sequenced clones. It has been proposed that testis is a tissue where, high levels of repetitive sequences, like MERs, LINES and SINES and retroposon like elements with long terminal repeats are transcribed (Lankenau *et al.*, 1994, Hendriksen *et al.*, 1997; Casau *et al.*, 1999). As a result, it is not unexpected that a proportion of cDNAs would fall into this category. However, although Lahn and Page found in their library repetitive sequences, they did not specify whether they were amongst the ones mentioned above, or other Y-specific repeats.

Some of the cDNA clones selected by Chr Y DNA, from the testis library are puzzling. One example is cDNA clone ml13b3 (Table 3.6), which shows 100% homology with a cDNA sequence obtained from Spirometra

erinacei (plerocercoid stage) and with an EST from a laryngeal cancer cDNA library. The connection between those tissues and Spirometra (tapeworm) is unlikely, unless the patient from whom the tissue samples were taken, were infected with this parasite. Even so, it is difficult to explain why the Y Chr template should have selected this cDNA.

3.2.3 Conclusion

If we assume that the starting material which was the testis cDNA and genomic DNA coming from the Y chromosome cosmids was of good quality, the high levels of contamination could be explained by postulating that there is only a very low number of Y-linked, testis expressed sequences. Nevertheless, it is clear that these genes although low in number have been evolutionarily conserved because they have retained a significant function and a cDNA selection approach can provide the raw material for isolation of those genes.

CHAPTER 4

TTY2-like genes

4.1 Identification of TTY2-like cDNAs

Of the recombinant clones identified as potential cDNAs, 19 demonstrated ~65-78% homology with the TTY2 testis specific transcript described by Lahn and Page (1997) (see chapter 3, Table 3.1). Lahn and Page isolated TTY2 (testis transcript-Y 2) using a cDNA selection approach, very similar to that described here. TTY2 was one of 12 novel genes expressed in testis that were identified in their study. The authors describe TTY2 as a 3.2Kb transcript with no significant ORF. They also comment that the TTY2 gene exists on the Y chromosome in multiple copies, as judged from the number of hybridising fragments on Southern blots and the occurrence of TTY2 related sequences at various positions on the Y chromosome. Multiple TTY2 copies were identified by PCR amplification of DNA from males carrying partial Y chromosome deletions (Vollrath et al., 1992) and by PCR amplification of YACs spanning the Y chromosome euchromatic region (Foote et al., 1992). In this way, it was demonstrated that copies of TTY2 map to Yp11.2 within the 4A subinterval, close to the centromere, and to Yq11.2 within subinterval 6C (AZFc region). It seemed likely that the cDNAs found in this study represent some of the members of the TTY2 family.

Of the 19 TTY2-like cDNA sequences identified here, 16 share 99-100% identity and probably represent parts of the same cDNA (chapter 3, Table. 3.1). For ease of reference this particular cDNA will be referred to as ml22a3, after the cDNA in the set that has been analysed most thoroughly. The remaining three TTY2 like cDNAs ml13d10, ml18g1 and ml22h10, showed distinct DNA sequences. ml13d10 aligns with ~72% similarity to the 5' end of the published TTY2 sequence, whereas ml22a3, ml18g1 and

ml22h10 align with around 74-77% homology to a region located approximately 1Kb further towards the 3' end (Fig 4.1). These results indicate there are a minimum of five distinct members of the TTY2 gene family. Of the new members, two were chosen as representatives for further study; ml13d10 and ml22a3. Since these two cDNAs lie at different positions along the TTY2 sequence, they could represent a single TTY2-like gene.

4.2   Identification of cosmids specific for ml13d10 and ml22a3

To investigate whether ml13d10 and ml22a3 are parts of two distinct genes, or are different exons of the same gene, a panel of Y-cosmids was screened to determine if each cDNA was present in a different cosmid. Cosmid inserts have a size of around 40kb, so it seemed most likely that they would contain sequence that corresponds to only one gene. In order to isolate cDNA specific cosmids, a panel of gridded Y-derived cosmids (n=480) were screened using ^{32}P labeled ml13d10 and ml22a3 as probes. Each probe identified the same 34 cosmids (Fig 4.2). It was realised that the primers used to prepare the two probes for library screening, amplified a region which showed substantial amount of sequence common to TTY2, ml13d10 and ml22a3 and that the probes would identify all TTY2-like cosmids (Figs 4.3 and 4.4). The relatively large number of cosmids containing TTY2 like sequences suggests that this is an extensive gene family. In order to isolate cosmids specific for ml13d10 and ml22a3, primers were relocated so that they had low 3' homology to the TTY2 sequence. These were primers 13d10.2F and 13d10.2R, (ml13d10) and 22a3.2F and 22a3.2R, (ml22a3) (Figs 4.3 and 4.4). PCR products amplified using these primers, identified a single cosmid corresponding to each cDNA (Yco2E9/ml13d10, Yco2F6/ml22a3) (Figs 4.5 B and C).

Restriction enzyme digestion of DNA from Yco2E9 and Yco2F6, both of which have the same cosmid vector (Lawrist 16), with six different
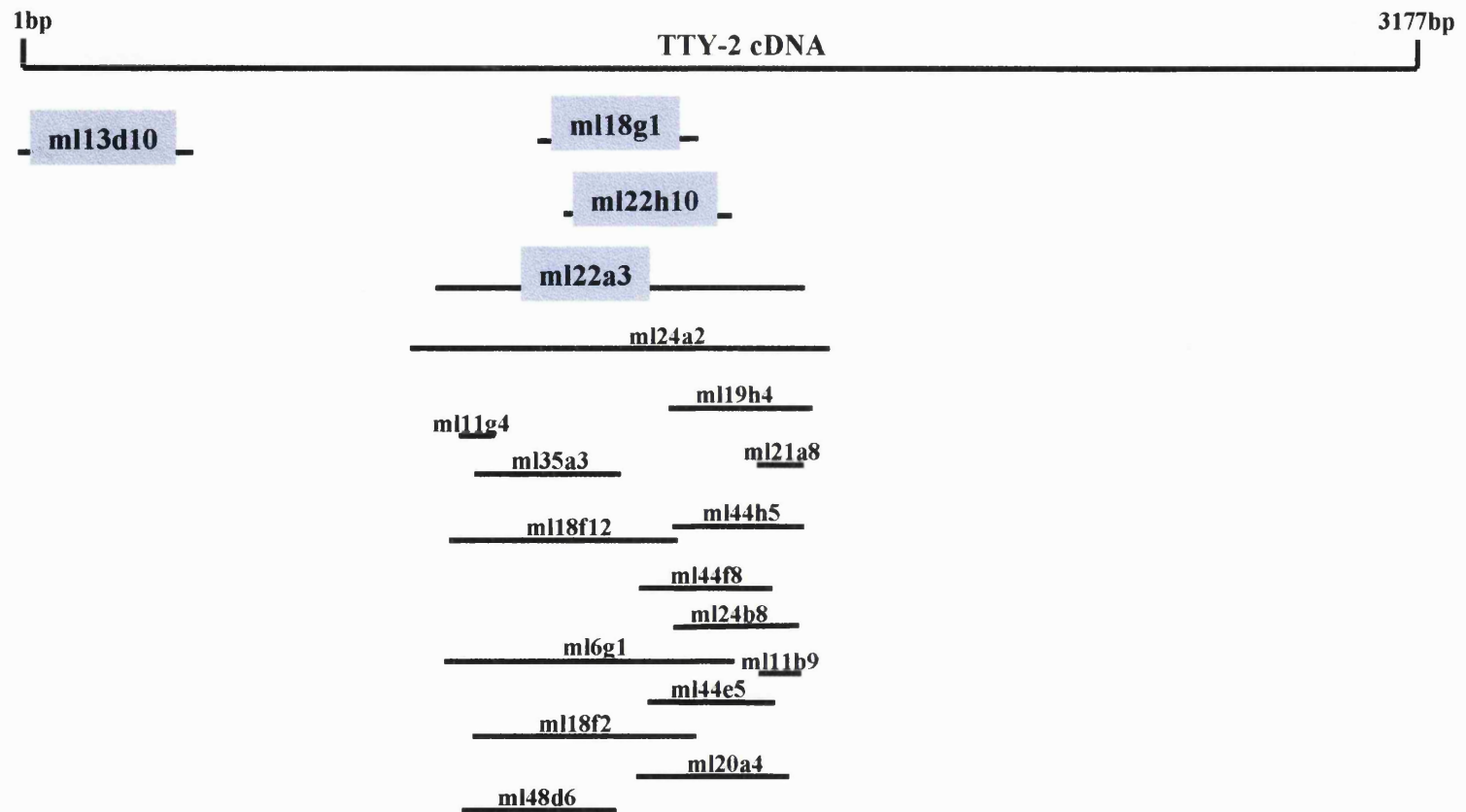
Figure 4.1 The position of the TTY-2 like cDNA clones derived by cDNA selection, relative to the published TTY-2 sequence. ml13d10, ml18g1, ml22h10 and ml22a3 (highlighted) represent distinct members of the TTY-2 gene family (72-77% homology). The remaining cDNAs all overlap and are identical to ml22a3.
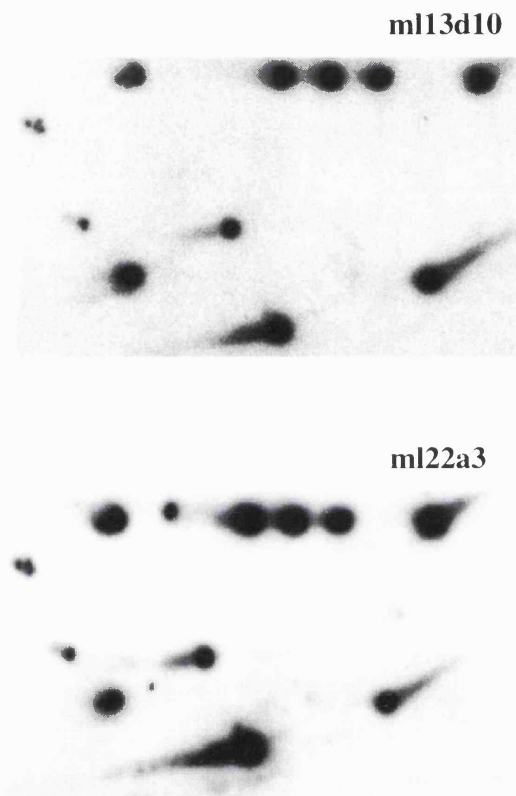
ml13d10

ml22a3

Figure 4.2 Screening a panel of 500 Y-derived cosmids with ml13d10 and ml22a3, identified as positives the same 34 cosmids. A typical example of a single nitrocellulose filter screened with probes for ml13d10 and ml22a3 on two separate occasions.

```
TTY-2      AGGCTTGCCATCACCACAGATGGCCTCTGAGACACTGTTTGAACCACATC 50
           ||||||||||| ||| |||||||| || | | ||| ||||
ml13d10    ........aatcaccacagatagccactgagaccctctctcaacttcatc 42
                   13d10.F(non-specific) →

TTY-2      TGCACCTGTGAGAGGCCAGTTTGAGGTATGAGAACACTGTTTCAATTTGG 100
           |||| |||||||||||||||||||||| ||||| ||| | |||||||||
ml13d10    tgcatgtgtgagaggccagtttgaggtgtgagagtacttctccaatttgg 92

TTY-2      ACTTGCCTTTGTCTTGGTTCCTGCTTTTCCCAGATGGCACCTACCCAACC 150
           |||||||||||||| || |||||||| | |    ||| | ||| ||||
ml13d10    acttgcctttgtcatgtttcctgcctatgtgagagagaccctgtgaaacc 142
                                  13d10.2F →

TTY-2      CAGGATGAATGAGTGCAGAGAGGTCAAGTGCCAGGCCATCTTTTGCTGAC 200
           |||| ||| |  |||| |||| || ||| ||||||||||||||||||
ml13d10    caggtagaagggaggcagtaaggttaatggcccagccatcttttgctgac 192

TTY-2      ACCCTTTTCTGGTATTTCAGGTATAAGTCCATCATCCAAAGACTGCTCAA 250
           ||| | ||||| | |||||||||| | || |||| ||||||||| || ||
ml13d10    ccccatctctggggtctcaggtatgattctatcacccaaagacacctaaa 242

TTY-2      CATCTCACCAGAATATATTTCAATCCTCATGGGGCATGATTCTTTCACAA 300
           ||  ||||||||| || ||| |||||| ||||||  ||||| || ||||
ml13d10    caaatcaccagactacattccaatccccatgggacctgattattgcacac 292
                             ← 13d10.R(non-specific)

TTY-2      AACCCCTTTCAGGAATGGAGTCAGAAGAGTAGTTTCCAGAGACAACCTCA 350
           | | ||||||| ||| |||||||||||| |||||    || | ||||
ml13d10    agcttctttcagcaat.gagtcagaagagcagtttttcagaaccaactca 341
                                         ← 13d10.2R

TTY-2      CAGTCTTGGAACGGCTCTGCCTCCCATGTGATCTGACCATGGAGATGGCA 400
           ||||||  || |   | ||||| | | || |   ||   |          |
ml13d10    cagtctgaaaaag...caacctccttttcatatatacacttaacatcaa 388

TTY-2      TATAAGGGCCCTAAGTTTGAGACTTTTAGGGTACTGCAATGCGTTATCAC 450
           | ||      |
ml13d10    aaaaaaaaagc........................................ 399
```

Figure 4.3 Comparison of TTY-2 and ml13d10 sequences using the GAP program. % similarity 71.93%.

13d10.F and 13D10.R are primers designed in regions of high 3' end homology between ml13d10 and the TTY-2 cDNA sequence; 13d10.2F and 13d10.2R are primers designed in regions of low 3' end homology with TTY-2.

```
TTY-2    GTGTGTGTAGGTGTCTGTGTGTGTGTATGTAAATGAATTCTGTGGATCAG 1100

ml22a3   ...................................................GC 2

                .         .         .         .         .
TTY-2    GAATCAGCAATGACTAGTTAAGCTGTCTGTGACCAGCCGGGTTCCCCATC 1150
         |||||  |  ||| |  | |    |     |    || |       |||  |||
ml22a3   GAATCCACTGTGAATTGGGAGGCATGCCAAGACCTGCCGGCATCCAAATC 52

                .         .         .         .         .
TTY-2    GTCTGCCCCTGCCAAAAAAACAGGTACTCTTCTACAAA.GAAGAGGAGAG 1199
           || ||||||| |||||| || ||  ||||||||||| ||| ||||||||
ml22a3   AACTCCCCCTG.CAAAAGAAAAGCCACTCTTCTAGAAAGGAAGAGGAACA 101

                .         .         .         .         .
TTY-2    CACCACACCCAAGAACAGACATCTCCCAGTGTTGCATTATAAAGCAGCCA 1249
         |||| |||| || |||||||||| || || ||   ||| |    |||||||
ml22a3   CACCCCACCAAAAAAACAGACATCTTCCTGTATTTATTGTCCTGCAGCCA 151

                .         .         .         .         .
TTY-2    ACCC....ACAGACACTAGCACTCTGGTCTGCATAGCCCCTTTAATTTAC 1295
           | |    | ||||||||||| || ||| |     |||||||||||
ml22a3   TCTCAGAGAGAGACACTAGCAATCCTGTCCACTGGACCCCTGGAATTTAC 201

                .         .         .         .         .
TTY-2    CTAGAATTCAGTTCCCAGCCAAGTAGGTGCTTCATGTCNTGAGGGTGCAA 1345
         ||  |||| | | || |  || |||| ||| |  ||:||||||  |
ml22a3   CTCAAATTGATTA<ins>CCTATCTGAGCAGGTACTTTAC</ins>ATCATGAGGGGACCC 251
                       <b>22a3.2F →</b>

                .         .         .         .         .
TTY-2    TCCTCCATCATCTTGAGATTTCATGCTGGTACAGAGAGTGTGA.CAGCAA 1394
         ||||||||||||||| |||||||| |||| ||| | |||||| | |||||
ml22a3   TCCTCCATCATCTTAGGATTTCATCCTGGGACATACAGTGTAAGAAGCAA 301

                .         .         .         .         .
TTY-2    TAAGGTCAGATAGGGGTGAGTATACAACCTGGTGAAGGGTGGATGGGGTC 1444
         ||  |||||||| |||||||| |||||| |||||||||  |||||||||
ml22a3   TACAGTCAGAT.GGGGTGAGGATACAATCTGGTGAGAGATGGATTGGGTC 350

                .         .         .         .         .
TTY-2    CCGTACCTTCACCAGCAAAAAGGGTGAAAATAGATGACACAGAATGTGCT 1494
         |    |  |||||||| |||| |  |     || | |||||||||||| |||||
ml22a3   CTACAACTTCACCTGCAA<ins>CACTGAGAAAGACAGATGACAC</ins>AGAAGGTGCT 400
               <b>22a3.F(non-specific) →        ← 22a3.2R</b>

                .         .         .         .         .
TTY-2    TCCAACTCCATCCCCACA.TTCCCATAATTGCAAAATCAGTCAACAACAT 1543
         ||||||| |||||  ||||  ||||  |||||||||| || ||||| | |  ||
ml22a3   TCCAACACCATTCGCACATTTCCTTTAATTGCACAAGCAGTCCATACAAT 450

                .         .         .         .         .
TTY-2    GGCCTGGTGTTTAGGTGGGAGTACTCCAACCTGCAGGAAGAATTTGGAGT 1593
         ||||  | |||| ||||||||||||||||||||  ||   || |||||| ||
ml22a3   GGCCCGATGTTGAGGTGGGAGTACTCCAACACGCCAAAAACATTTGGT<ins>GT</ins> 500

                .         .         .         .         .
TTY-2    GCAAATT.GTGGCCAATCTGGAAAACTCCTGGTTTGAGGGTTTTAATACC 1642
         ||||||| |  | |||  ||||  ||||||||| |||| || ||| |||||
ml22a3   <ins>GCAAATTAGGGTCCATCCT</ins>GGCTAACTCCTGATTTGTAGGCTTTCATACC 550
              <b>← 22a3.R(non-specific)</b>


                .         .         .         .         .
TTY-2    TGTAGTCAAATGGAAGTGGAATAGATTGATGCTGGGTGGGTTGTGGCCTC 1692
         |||||||| || |||| ||  || ||
ml22a3   CAGAGTCAAACGGGAGTGCAACGGAGTG...................... 578
```

Figure 4.4 Comparison of TTY-2 and ml22a3 sequences using the GAP program. % similarity 73.86%.

22a3.F and 22a3.R are primers designed in regions of high 3' end homology between ml22a3 and the TTY-2; 22a3.2F and 22a3.2R are primers designed in regions of low 3' end homology with TTY-2.
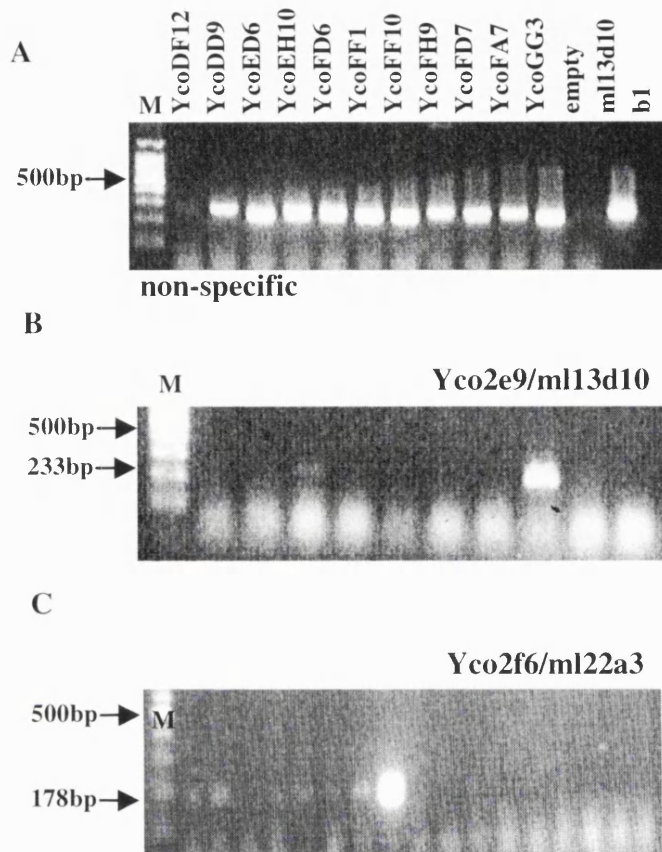
Figure 4.5 PCR amplification of DNA from 34 Y-cosmids that contain
TTY2-like sequence, using as primers A.Example of PCR amplification
with primers13d10.F and 13d10.R (non-specific) that share high homology
at their 3' ends with the published TTY2 sequence; B.13d10.2F and 13d10.2R
(ml13d10 specific), with low homology at their 3'
ends with the TTY2; C.22a3.2F and 22a3.2R (ml22a3 specific) with low
homology at their 3' ends with TTY2; M:DNA size marker

restriction enzymes showed that the two cosmids have different restriction patterns, which do not overlap, confirming that ml13d10 and ml22a3 are not part of the same gene. Southern blot analysis of the digests using ^{32}P labeled cDNA as probe gave a single hybridizing band for each enzyme digest (Fig 4.6). A 3.8Kb HindIII band from Yco2E9 (ml13d10) and the 2.6Kb XbaI band from Yco2F6 (ml22a3) were isolated and subcloned into the pUC19 plasmid vector. The subcloned fragments were partially sequenced with primers, located in vector sequence at each end of the cloning site (M13 primers) and with primers specific for ml13d10 or ml22a3 cDNAs. In this way, 713bp of the 2.6Kb XbaI/2F6 cosmid fragment and 2405bp of the 3.8Kb HindIII/2E9 cosmid fragment were obtained (Figs 4.7 and 4.8). In each case the sequenced portion contained the respective cDNA sequence as one continuous stretch of sequence and this was identical to that of the corresponding cDNAs, confirming the specificity of the cosmids.

4.3 Mapping studies

Localisation to the Y chromosome using somatic cell hybrids   In order to confirm that the genes encoding the TTY2-like ml13d10 and ml22a3 are located on the Y chromosome, PCR amplification using specific primers (13d10.2F and 13d10.2R/ml13d10, 22a3.2R and 22a3.2F/ml22a3), was carried out on DNA from two somatic cell hybrids, 3E7 and 853 that contain the Y chromosome as their only human component. Human male and female DNAs were also amplified as Y positive and negative controls.

Products of the correct size (233bp for ml13d10 and 178bp for ml22a3) were obtained for both cDNAs, from male DNA and from the Y containing somatic cell hybrids; there was no amplification from mouse or hamster parental DNA. ml13d10 and ml22a3 appeared to be Y specific as there was no amplification from female genomic DNA (Fig 4.9). These findings are in
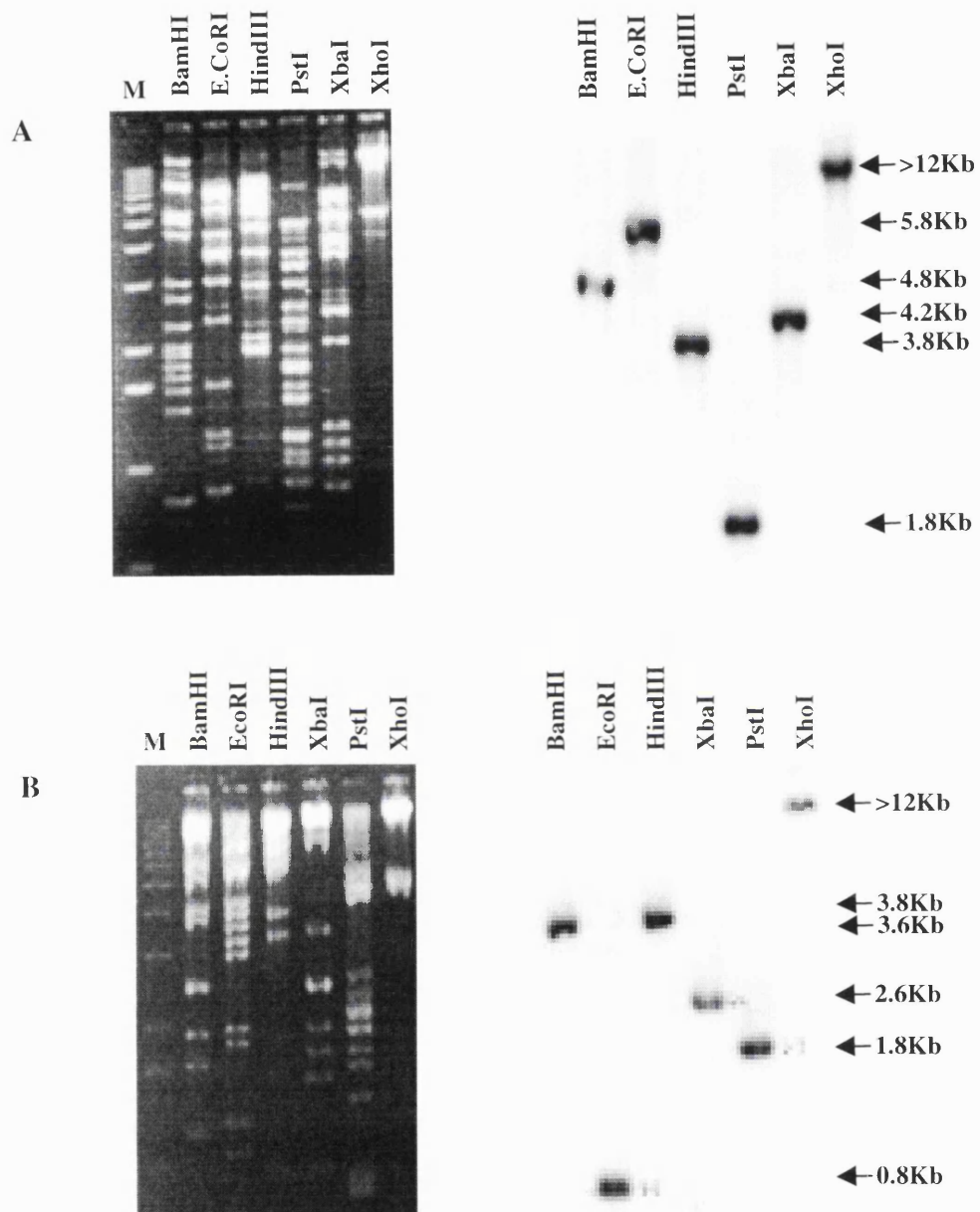
Figure 4.6 Restriction enzyme digest of A.Yco2e9 (ml13d10 specific) and B.Yco2f6 (ml22a3 specific) and Southern blot analysis using the cDNA clones as probes; M=1Kb DNA size marker

A

137230   140982

142

B

140207 **13d10 cDNA** 140603

2975   3374

C

1   549   1676   2393   2995   3417   3752

13d10Hi.F1

13d10Hi.R1

13d10Hi.R2

M13.R

13d10Hi.R3

M13.F

13d10Hi.R4

Figure 4.7 Diagram showing the structural relationship between ml13d10 (B), the AC009491 genomic clone (A) and a subcloned ~3.75Kb fragment fromYco2e9 (C); the sequencing strategy for the 3.75Kb HindIII fragment is shown below as arrows; the primers are listed in Table 2.2; -------- : Sequence not yet obtained;

A

64696   67296

B

64765   m122a3 cDNA   65337

69   644

C

1   1900   2600

M13.R
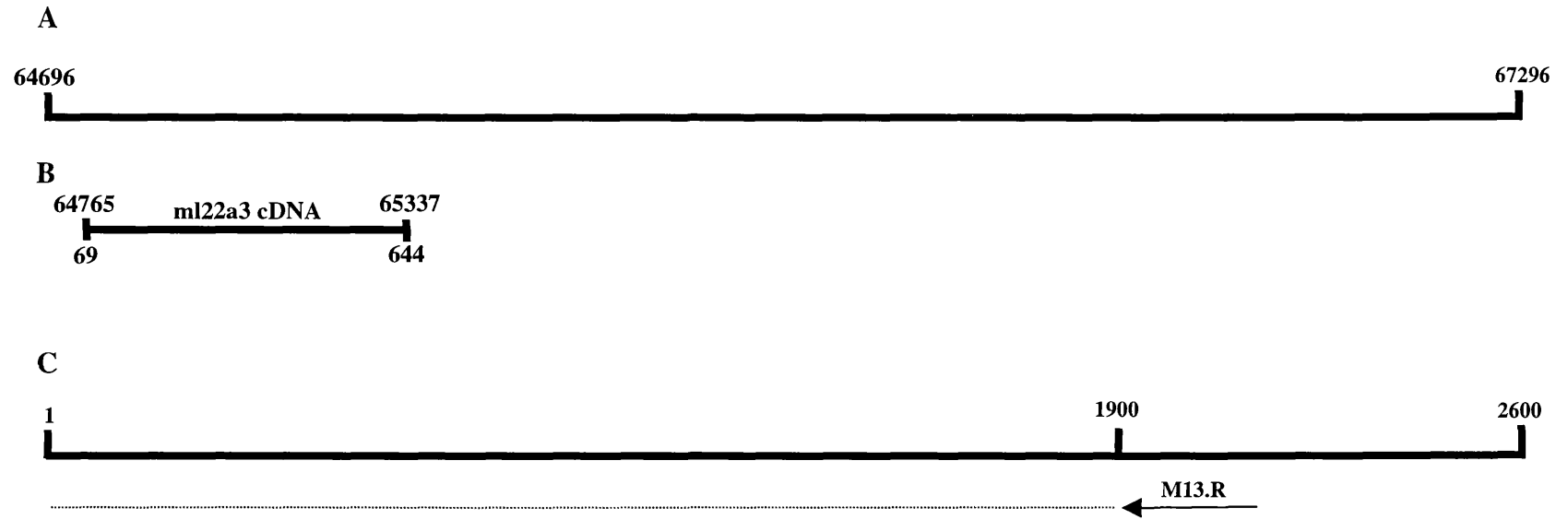
Figure 4.8 Diagram showing the structural relationship between m122a3 (B), the AC023342 genomic clone (A) and a subcloned ~2.6 Kb fragment from Yco2f6 (C); the sequencing strategy for the 2.6Kb XbaI fragment is shown below as arrows; the primers are listed in Table 2.2; -------- : Sequence not yet obtained

A

500bp →
233bp →

M P F IRE3 3E7 Wg3H 853 b1 b2 M

← 500bp

ml13d10

B

500bp →

178bp →

M P F IRE3 3E7 Wg3H 853 b1 b2 M

ml22a3

P	male	IRE3	mouse parent	3E7	Y containing mouse hybrid
F	female	Wg3H	hamster parent	853	Y containing hamster hybrid
B$_{1,2}$	no DNA	M	DNA size marker		

Figure 4.9 Chromosome localisation of cDNA clones A.ml13d10 and B.ml22a3 by gene specific PCR of DNA from two somatic cell hybrids 853 and 3E7, which contain the Y Chr and human male (XY) and female (XX) DNA;

144

line with the suggestion by Lahn and Page (1997) that a gene encoding TTY2 lies in the non-recombining region of the Y chromosome.

Mapping to metaphase chromosomes, by fluorescent in situ hybridisation (FISH) The cosmids Yco2E9 and Yco2F6, specific for ml13d10 and ml22a3 were used as probes for FISH analysis, onto human male metaphase chromosomes (work done in collaboration with Dr. Margaret Fox, MRC Human Biochemical Genetics Unit). Single and double fluorescent labeling was used and 4-6 metaphase spreads were scored all of which showed identical results.

The FISH analysis revealed that Yco2E9 (ml13d10) is located in the distal part of the short arm of the Y chromosome, whereas Yco2F6 (ml22a3) is located on the long arm of the Y chromosome, close to the centromeric region (Fig 4.10A and B). These distinct chromosomal localisations confirm that ml13d10 and ml22a3 cDNAs represent two distinct members of the TTY2 gene family. It was notable that a par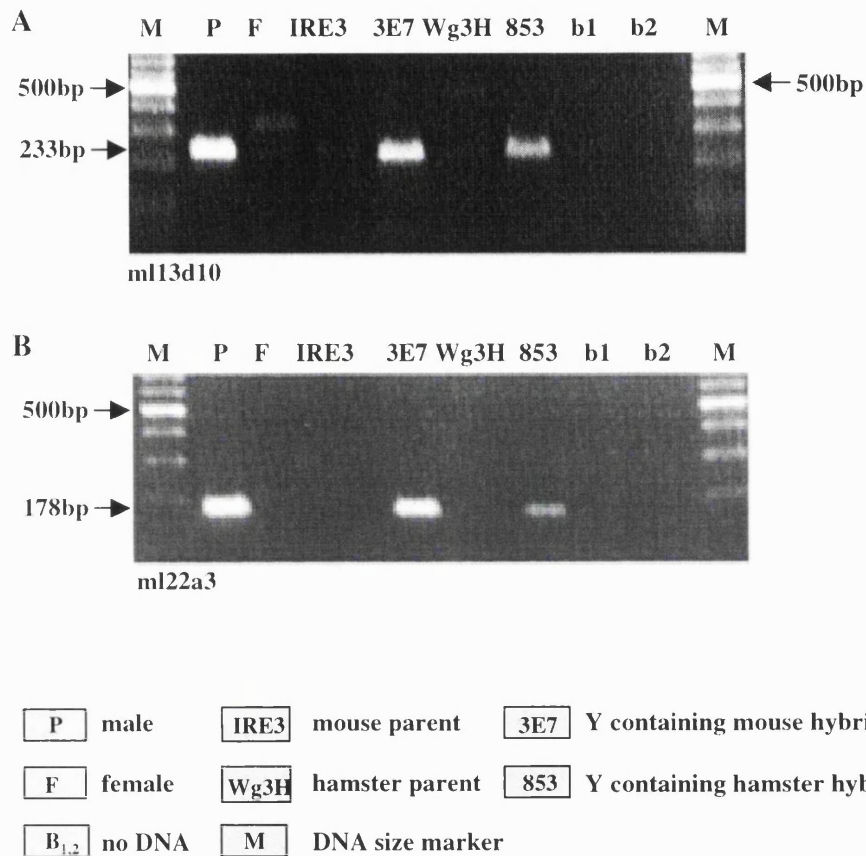ticularly strong FISH signal was seen in both cases, which could indicate that both cosmids contain a Chr Y cluster of related genes or sequences.

Fine mapping by using a Y-specific YAC contig An attempt was made to map ml13d10 and ml22a3 cDNAs more accurately on the Y chromosome, using Y-linked YACs, which formed a tiling path across the entire chromosome (DNA from YACs was kindly donated by Dr. N. Affara

Cambridge University). This tiling path is made up of a total of 97 YACs with insert sizes of around 0.9-1.2Mb, and which have been aligned along the Y chromosome by using 222 STS sites (Jones *et al.*, 1994)

In order to check for the presence of human DNA and the quality/quantity of the YAC DNA, inter *Alu* PCR was performed using a

145

A
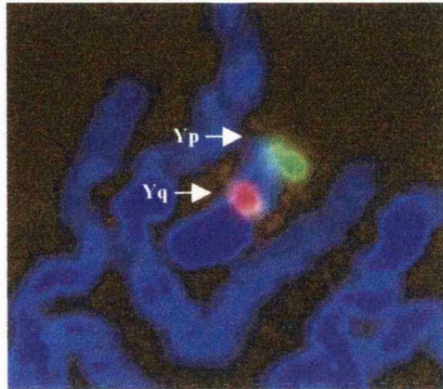
Yp: Yco2E9 (ml13d10)
Yq: Yco2F6 (ml22a3)

B

Yp: cosmid 2e9 (ml13d10)   **Yq:Yco2f6 (ml22a3)**

Figure 4.10 Double FISH analysis using Yco2E9 and Yco2F6 as probes on human metaphase chromosomes with A.double and B.single exposure with the red/green filter turned on/off accordingly. Signal is seen only on the Y chromosome

146

primer specific for *Alu* repeats (*Alu* primer IV) (Young *et al.*, 1990). Inter *Alu* sequences were amplified from all the YACs in the series, except 6H1, 853G9, 827A10 and 218F9, where products were very weak or absent (Fig 4.11). Amplification was then carried out with primers 13d10.2F and 13d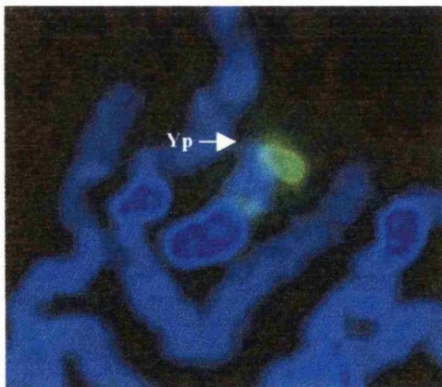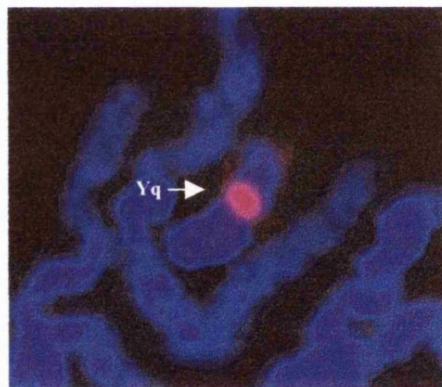10.2R (Table 2.1) specific for ml13d10. When the same conditions as used to amplify human and rodent genomic DNA were applied, resulted in multiple, non-specific products from all the YAC DNAs, except 6H1, 77H2 and 827A10, where no product was seen. Products ranged in size from approximately 220bp to more than 1Kb (Fig 4.12). This suggests that yeast chromosomal DNA includes some sequences, which are prone to be amplified non-specifically by the 13d10.2F and 13d10.2R primers. This lack of specificity was eventually overcome by adjusting empirically the primer/template ratio so that less YAC DNA was added to the PCR mix. Using only ~50-100ng of YAC DNA, a single product of the correct size was amplified from a single YAC, 759G2, using the specific ml13d10 primers (Fig 4.13 A). Similar strategy led to the amplification of a correct size product for ml22a3 from one YAC, 933A6 using 22a3.2F and 22a3.2R primers (Fig 4.13 B). To confirm that the correct sequence had been amplified, the YAC PCR products were gel extracted and sequenced. The sequence of the products, corresponded 100% to that of the two TTY2 like cDNAs (ml13d10 and ml22a3).

YAC 759G2 (ml13d10 positive) has previously been localised to the Yp11.31-11.2 region. This YAC clone covers part of the subinterval 2C, the whole of 3A to 3G and most of interval 4A close to the centromere (as defined by Vergnaud *et al.*, 1986 and Vollrath *et al.*, 1992). However, most of the YAC part that covers intervals 3C to 3G is absent due to internal deletions. Parts of YAC 759G2 overlap with YAC clones 853G9 and 758G1, from which there was no amplification with primers specific for ml13d10. Therefore, it appears that the most likely candidate location for ml13d10 is a
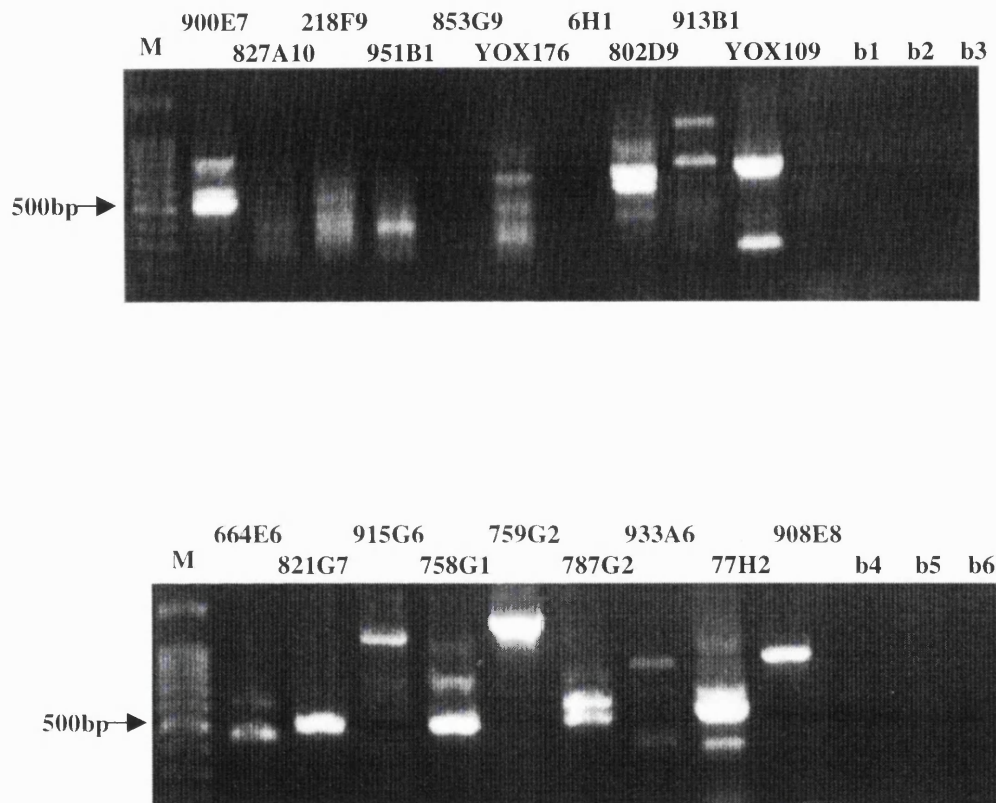
Figure 4.11 PCR amplification of Y chromosome YAC DNA, using the *Alu* IV primer - B27A10, 853G9 and 6H1 failed to amplify; M: DNA size marker; b1-6: PCRs without DNA
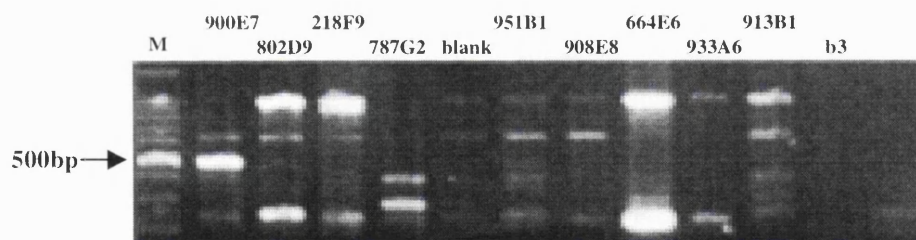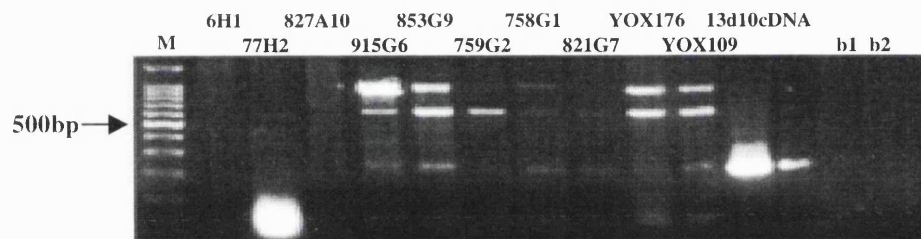
Figure 4.12 Amplification of YAC DNA, with ml13d10 specific primers, using PCR conditions identical to those used for the amplification of human genomic DNA and a high amount of ~200-500ng of YAC DNA. 6H1, 77H2 and 827A10 failed to amplify; M: DNA size marker; b1-3: PCRs without DNA

149
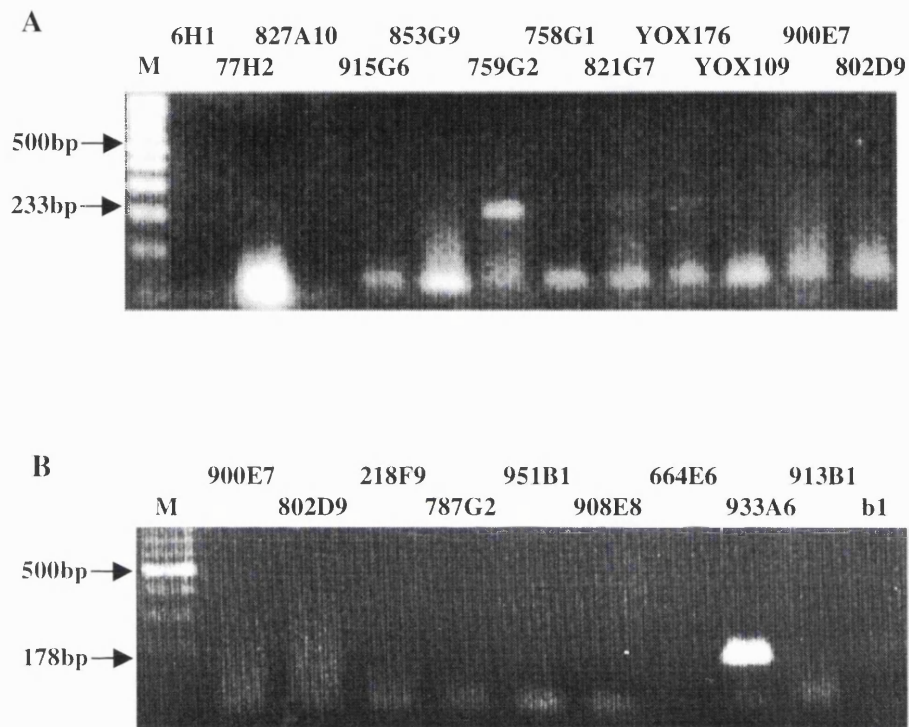
Figure 4.13 Amplification of YAC DNA using less amount in order to increase the specificity (see section 4.3) with A.ml13d10 specific primers and B.ml22a3 specific primers; M: DNA size marker; b1: PCR without DNA

region, which includes part of interval 4A extending between STS markers sY75 and sY62 (Fig 4.14).

YAC 933A6 (ml22a3 positive) has been localised to Yq11.2 and covers the AZFb region and part of the AZFc region, from the 5Q subinterval across the whole of interval 6 (subintervals 6A to 6F). However, PCR amplification showed that ml22a3 cDNA sequence is not present on YACs 664E6 and 913B1, parts of which overlaps with 933A6. According to the diagram in Fig 4.14, this result indicates that ml22a3 clone is located in the proximal 6 interval (including subintervals 6A, 6B, 6C and part of 6D), either within the AZFb region, defined by subintervals 5O to 6B, or the AZFc region defined by subintervals 6C to 6E (Vergnaud *et al.*, 1986; Vollrath *et al.*, 1992).

4.4 RT-PCR expression studies

In order to examine the expression pattern of the cDNA clones ml13d10 and ml22a3, RT-PCR was carried out using RNA prepared from human adult and fetal tissues, including adult testis, prostate and kidney and fetal testis, heart, brain, kidney, intestine, stomach, lung and limbs. The fetal tissues derived from four different individuals, with ages between 11.5 and 16.5 weeks gestation. All tissues were sex determined by DNA PCR amplification of amelogenin (Nakahori *et al.*, 1991; Pertl *et al.*, 1996), using the primers shown in table 2.2. The X and the Y homologues of amelogenin amplify to produce products of different sizes (432bp X; 252bp Y), so that males (XY) show two distinct PCR products and females (XX) a single product (Fig 4.15 A). This experiment allowed the selection and use of male tissues for the preparation of RNAs.

The quality and quantity of the RNA was checked by electrophoresis on an RNA MOPS-formamide gel (Fig 4.15 B). cDNA was prepared using reverse transcriptase and random oligonucleotide primers. PCR amplification
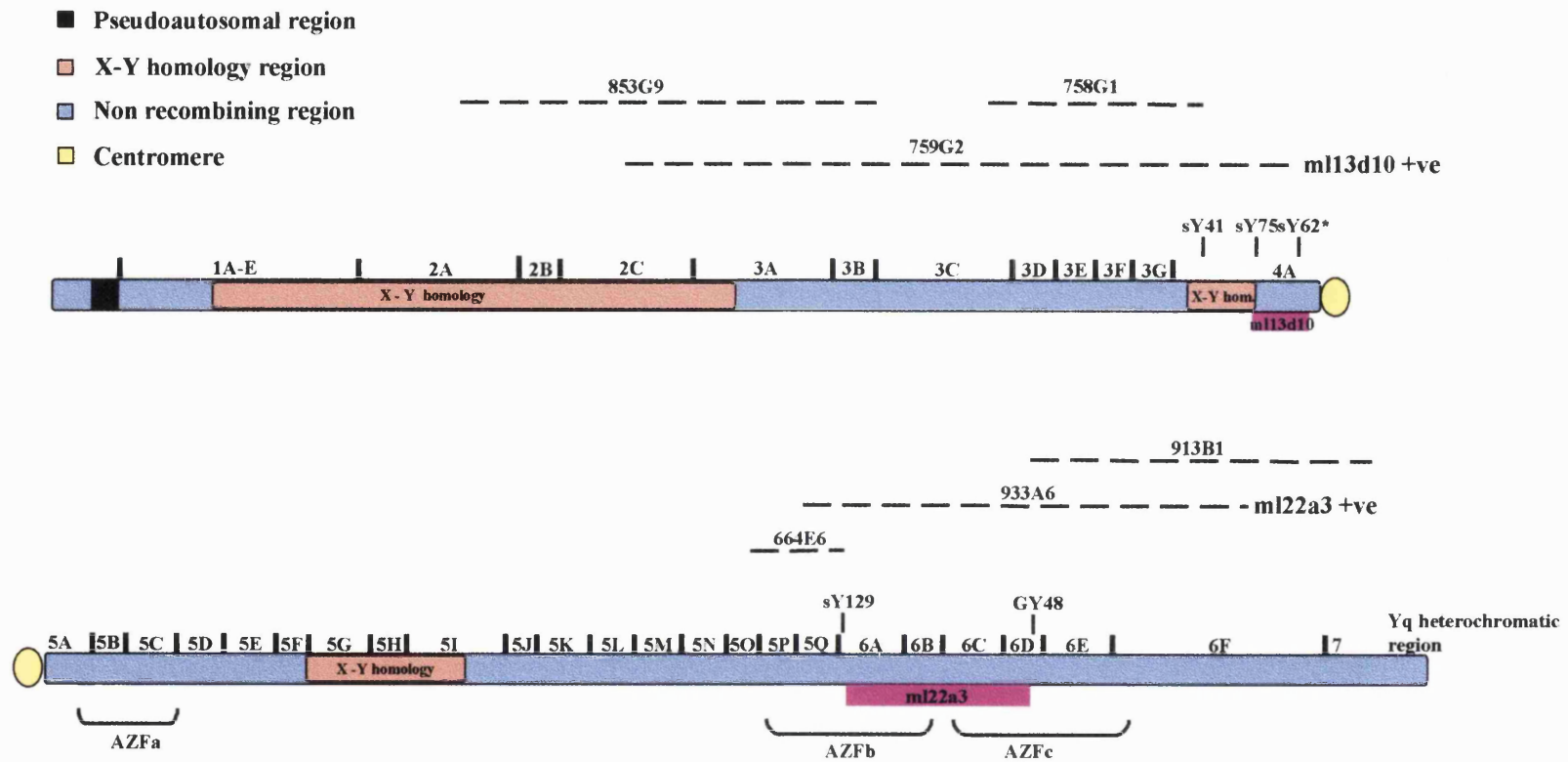
Figure 4.14 Localisation of ml13d10 and ml22a3 along the Y chromosome, deduced by PCR amplification of Y-mapped YAC clones spanning the Y chromosome euchromatic region and part of the Yq heterochromatic region. Dashed lines indicate position of YACs; STSs with an asterisk indicate a multiple copy STS. Adapted from Jones *et al.*, 1994
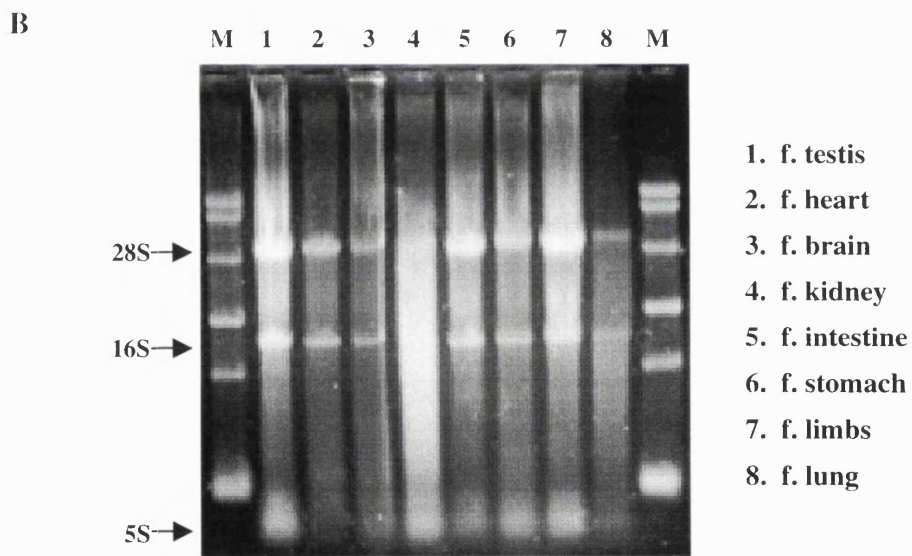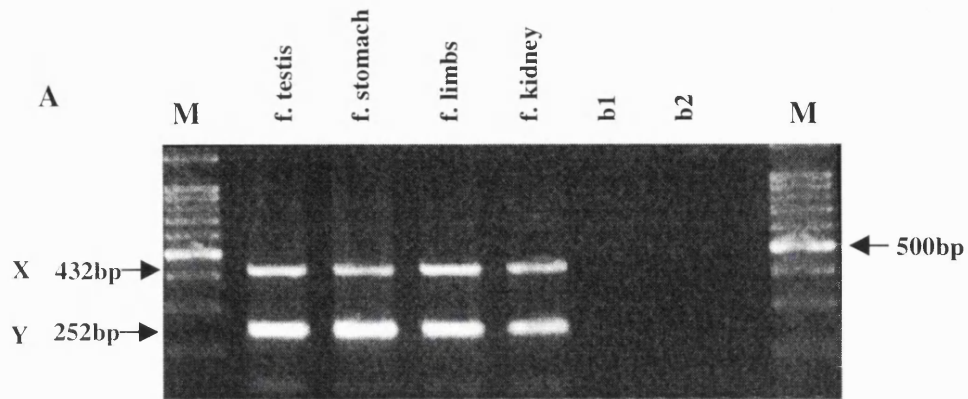
Figure 4.15 A.PCR amplification of genomic DNAs prepared from human fetal tissues, using amelogenin primers; B.RNA isolated from human fetal tissues; M: DNA or RNA size marker; b1-2: PCRs without DNA

using primers for the ubiquitously expressed gene phosphoglucomutase 1 (*PGM1*) (Edwards *et al.*, 1995), gave the expected product of 420bp in all tissues. When cDNA synthesis was carried out in the absence of reverse transcriptase (-) these samples gave no product, confirming that the RNA preparations were not significantly contaminated with DNA (Fig 4.16). Amplification of the cDNA with primers specific for ml13d10, found no expression in fetal kidney and limbs, moderate to weak expression in all other tissues including adult and fetal testis and highest levels of expression in adult kidney, fetal lung and fetal intestine (Fig 4.17 A).

Amplification using ml22a3 specific primers 22a3.2F and 22a3.2R showed a similar distribution, with highest levels in fetal lung and adult kidney and low or absent expression in fetal kidney and limbs; all other tissues showed moderate levels of product (Fig 4.17 B). These results confirmed that both TTY2-like cDNAs are expressed in developing and adult testis. However, it is also clear that testis is not the major site of expression.

In order to confirm that the products detected in non-testis tissue were identical in sequence to that found in testis, PCR products from adult testis, adult kidney and fetal lung for both cDNAs, were gel purified using the QIAGEN gel purification kit and sequenced. In each case the testis sequence was identical to that found in other tissues, confirming that they derived from amplification of the same gene product (Fig 4.18).

4.5 Comparative studies

In order to investigate whether sequences homologous to TTY2-like ml13d10 cDNA occur in other species, "zoo blot" analysis was carried out using genomic DNA samples from human male and female, a number of primates and mouse. ml22a3 was not subject to this kind of investigation, since it was expected that because of the high homology of its sequence with
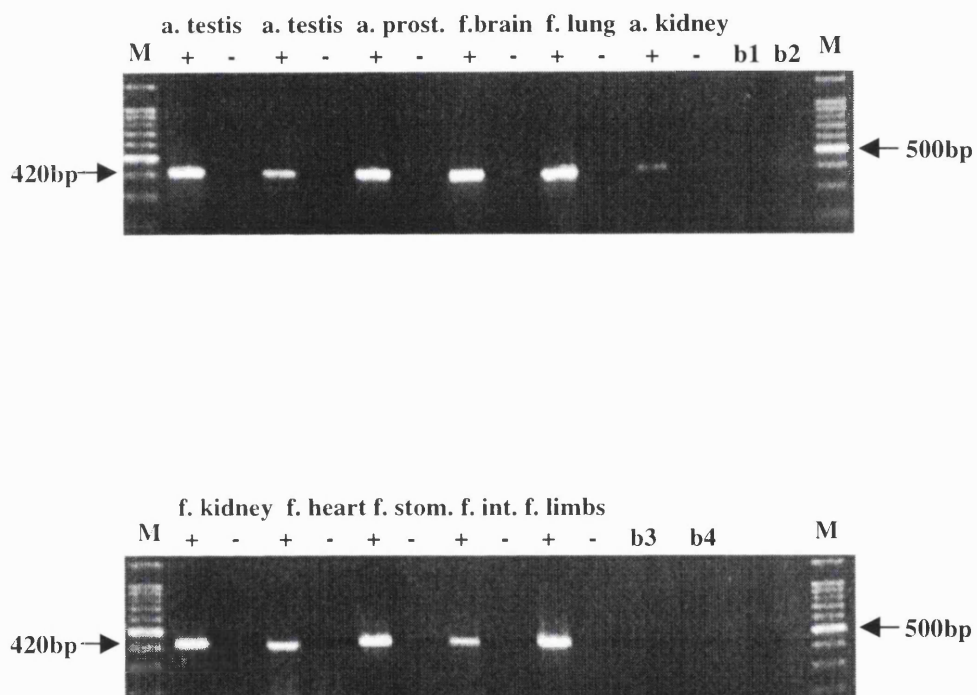
Figure 4.16 PCR amplification of tissue cDNA, using PGM1 primers

+= cDNA synthesis with reverse transcriptase, - = cDNA synthesis

in the absence of reverse transcriptase; M: DNA size marker;

b1-4: PCRs without cDNA

A

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14

500bp→
233bp→

ml13d10

B

M 1 2 3 4 5 6 7 8 9 10 11 12 13 14

500bp→
178bp→

ml22a3

1. adult testis
2. fetal testis
3. adult prostate
4. fetal brain

5. fetal lung
6. adult kidney
7. fetal kidney
8. fetal heart

9. fetal stomach
10. fetal intestine
11. fetal limbs
12-14. no cDNA
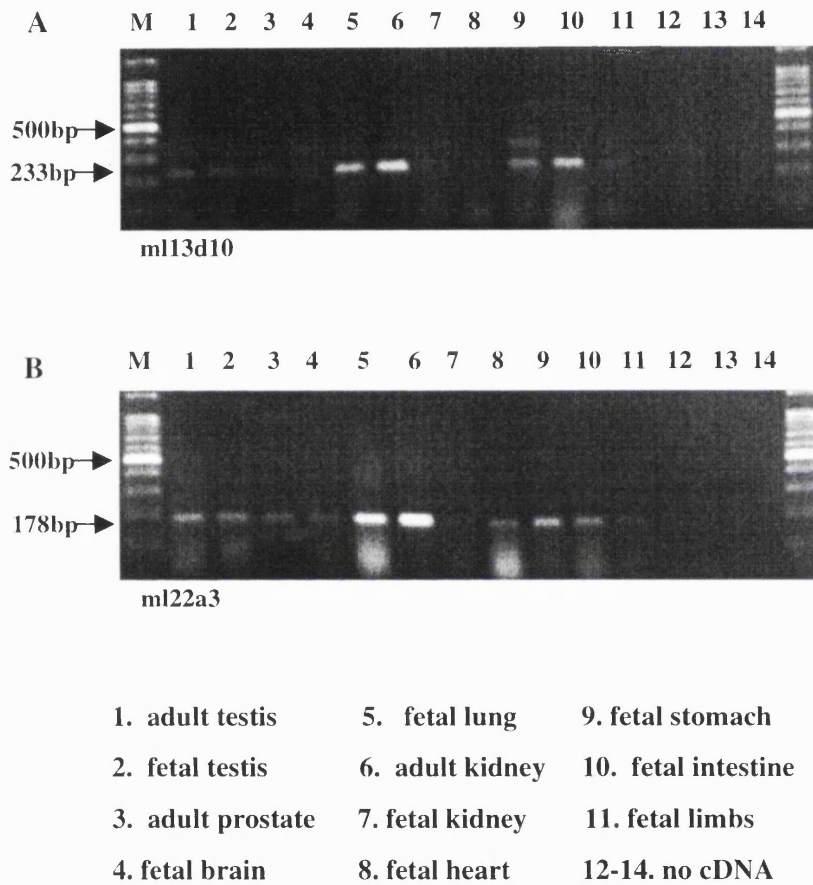
Figure 4.17 PCR amplification of human cDNAs, using primers specific for A.ml13d10 and B.ml22a3; M: DNA size marker
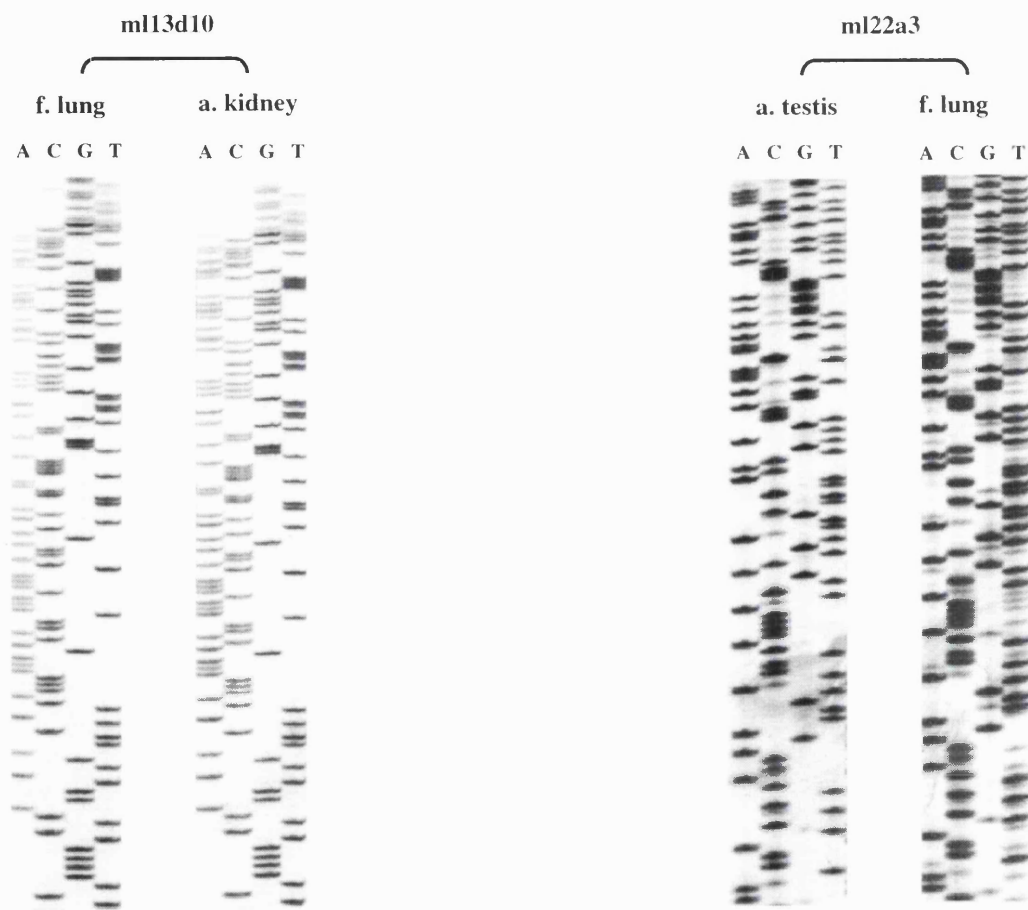
Figure 4.18 Sequence comparison of ml13d10 and ml22a3 RT-PCR products from different tissues

ml13d10, the same banding pattern would be obtained. The DNAs were digested with PstI, electrophoresed and blotted onto nitrocellulose membrane. The blot was hybridised under low stringency (57°C, 0.2xSSC) with ^{32}P labeled PCR products from ml13d10 (primers; 13d10.2F and 13d10.2R).

These analyses showed that ml13d10 hybridised stronger to male human and male chimpanzee, identifying bands between 1 and 2Kb and more weakly in the male gorilla and orangutan (Fig 4.19). Unfortunately, there was probably not sufficient pigmy chimpanzee DNA to see a result so it is difficult to say whether similar size bands are present. At least one band of 1.3Kb (sizes marked in Fig 4.19) seems to be present in humans, both male and female. However, other bands, for example one at 1.8Kb, appear not to be present in the human female DNA, indicating that they may be Y-specific. The multiple bands of different sizes seen in male human, chimpanzee and gorilla indicate that the ml13d10-like sequence either identified more than one exons of the gene, because they contain repetitive elements, or because TTY2 exists in multiple copies. The strong bands seen in male human and chimpanzee are not seen in the mouse, indicating either that the sequence has diverged considerably, or that in the mouse the number of copies of the TTY2 like gene may be very low.

In order to compare the chromosome location of ml13d10 and ml22a3 sequences in other species, FISH analysis onto metaphase spreads from male chimpanzee, pigmy chimpanzee and gorilla was performed, using as probe the cosmids Yco2E9 (ml13d10) and Yco2F6 (ml22a3). This work was again done in collaboration with Dr. Margaret Fox in the MRC Human Biochemical Genetics Unit.

Double fluorescent labelling using as probes both cosmids Yco2E9 red/biotin labeled) and Yco2F6 (green/digoxigenin labeled) was carried out. 8 metaphase spreads were examined for each species. In some cases the
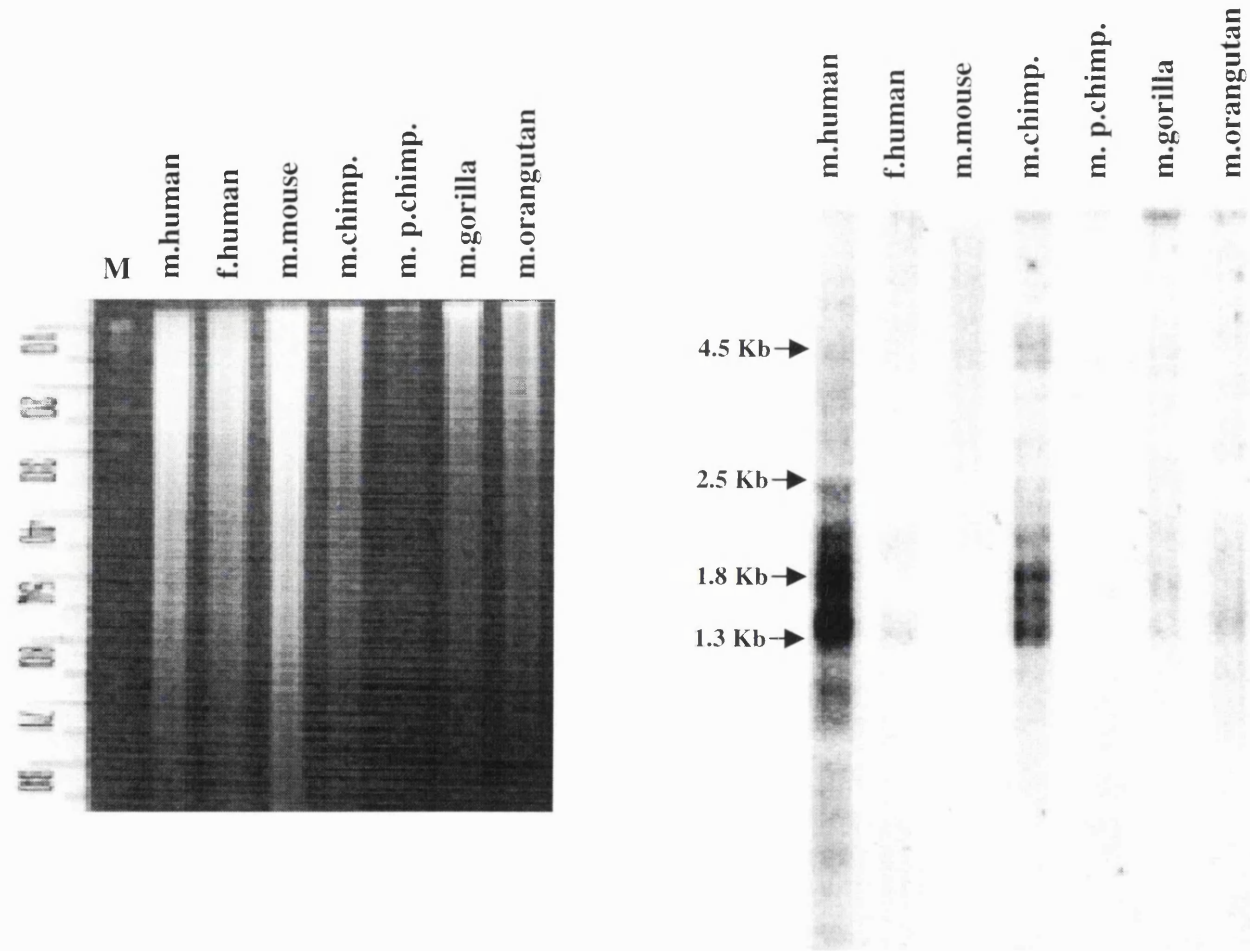
Figure 4.19 Restriction digest of genomic DNAs with PstI enzyme and "zoo blot" hybridisation using ml13d10

sequence as probe;m: male; f: female; M: DNA size marker

background signal was high, but all of the spreads for the pigmy chimpanzee were scorable, whereas only 6 were scorable for chimpanzee and gorilla. These experiments showed that both sequences are present at distinct locations on the short and long arm of Chr Y of chimpanzee, pigmy chimpanzee and gorilla (Fig 4.20).

The low signal seen with ml13d10 on Southern blot raised the question of whether the TTY2-like genes are conserved in the mouse. In order to test this, an attempt was made to amplify ml13d10 and ml22a3 from mouse testis and kidney cDNA using the human specific primers but at an annealing temperature 2-3°C less than that used for human DNA. Very weak or no bands were seen in the correct position on the gel, implying that there is no highly conserved homologues in the mouse (data not shown).

In summary, results from Southern blot suggest that conservation level of ml13d10 and ml22a3 sequence is relatively weak in the mouse. In contrast, FISH analysis showed that both ml13d10 and ml22a3 are Y-linked in primates and no homologues occur on autosomes in the primate genome.

4.6 Proposed gene models for ml13d10 and ml22a3 TTY2-like genes

A more detailed database search was performed with the published TTY2 cDNA sequence and the sequences of the testis selected cDNA clones ml13d10 and ml22a3. This search was done using the GenomeNet BLAST2 server to search Genbank, EST, STS, HTG, dbGSS (Genome Survey Sequences Release) and EMBL databases.

The TTY2 cDNA sequence identified 4 genomic clones with 100% homology (see later, section 4.7, Table 4.7). Contiguous regions of the cDNA matched non-continuous regions of the genomic clones and presumably, represent exons with intervening introns. The entire gene

Chimpanzee **P. chimpanzee** **Gorilla**
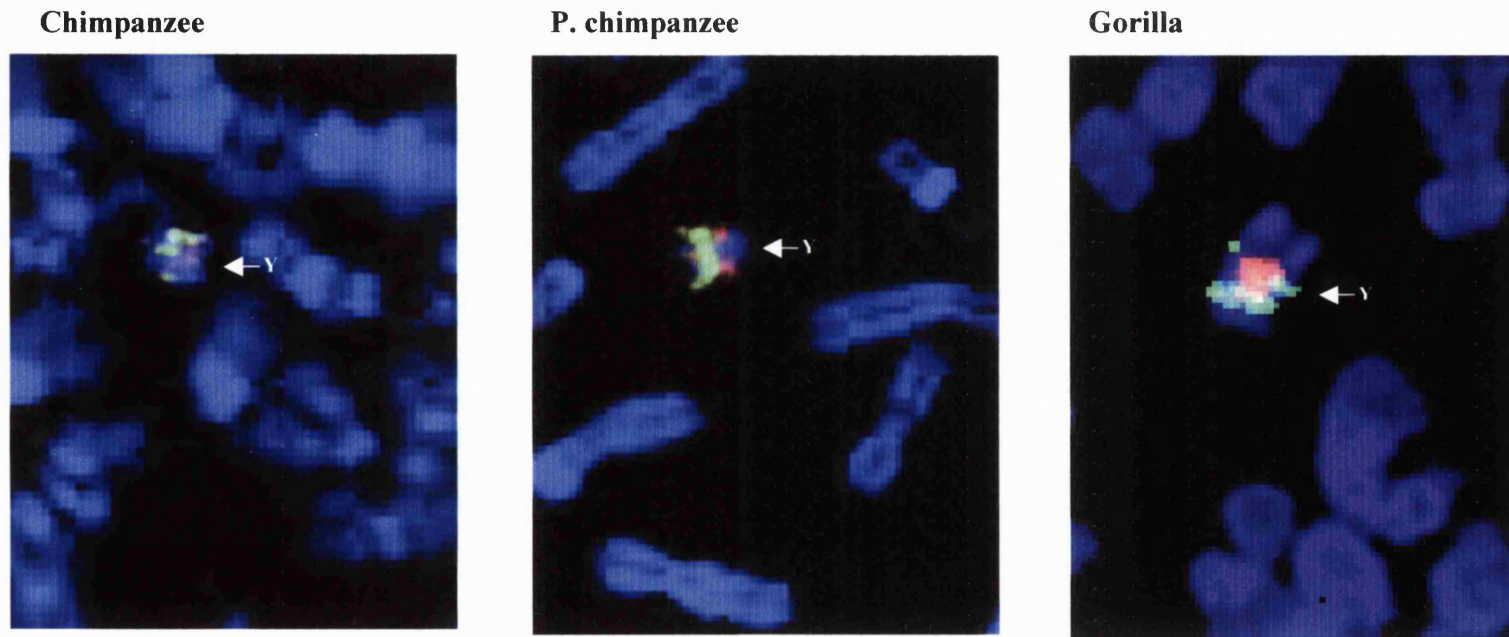
Figure 4.20 Double FISH labelling using Yco2e9 (ml13d10/red) and Yco2f6 (ml22a3/green) as probes on primate metaphase chromosome spreads. Signal is seen only on the Y chromosome

sequence lay within one genomic sequence, AC006335. By locating each portion of the cDNA it could be shown that the TTY2 gene is 17.9Kb and comprises 7 exons and 6 introns, the sizes of which are shown in Table 4.1. However, Lahn and Page claimed to have partial sequence of the TTY2 cDNA (the grounds of this claim are not clear), which suggests that there may be more exons in the gene. Each exon/intron boundary coincides correctly with the appropriate donor/acceptor splice site sequences. In addition, polypyrimidine regions occur in the appropriate position (<20 nucleotides) from each acceptor site.

ml13d10 identified with 100% homology the genomic clone AC009491, which is a working draft sequence of one unordered DNA fragment of 174147 bp released in Sept. 1999. Similarly, ml22a3 identified the genomic clone AC007320, a complete sequence of 191414 bp size that was released in Aug. 1999 and AC023342, a 179841bp sequence which exists at the moment as two unordered fragments with a gap of unknown length at around 10Kb. This clone was released to the database in Feb. 2000.

The information about the exon/intron structure of the TTY2 gene allowed the remaining exons of the ml13d10 gene to be located within the genomic clone AC009491. ml13d10 comprises part of exon 1 and BESTFIT (GCG package) was used to identify cDNA segments of AC009491 that showed the greatest homology to the TTY2 exons. Seven potential exons were identified, which showed between 72-82% homology with the TTY2 exons. The best possible match to each exon/intron boundary was selected. The exon/intron boundaries are shown in Table 4.2. In several cases the exon sizes are not the same, with the ml13d10 cDNA sequence showing nucleotides absent from the TTY2 exons and vice versa. The NIX multiprogram analysis system identified some of the ml13d10 exons (1, 6 and 7), using gene recognition programs (Table 4.3). The predicted ml13d10

Table 4.1 Structural details of the TTY2 gene, derived by alignment of the cDNA with genomic DNA from AC006335 fragment. Each exon/intron junction is illustrated by giving on the left the 3' end of the exon and on the right the 5' sequence of the adjacent exon. The beginning and end of each intron are shown in lower case letters. The sizes of each exon and intron are given in parentheses.

Exon	3' sequence	Intron sequence	5' sequence of next exon
exon 1 (1818 bp)	AAAATGCATTGT	gtgagtattt... (2474 bp) ...ccttttgcag	AGAGGCCCTGCA
exon 2 (87 bp)	CTGGGTTCTCAG	gtaagttcct... (7394 bp) ...ttttttttag	GCTGGCTGACAG
exon 3 (149bp)	AACTCCGTGGAG	gtgagtcagc... (692 bp) ...tgtctttcag	CCAATCCAAGGA
exon 4 (157 bp)	AACAGCAATAAG	gtaataaggt... (2830 bp) ...attggcgtag	TTTCCTTGGTCT
exon 5 (79 bp)	GGATTCTATTGT	gtgagtattt... (120 bp) ...acttctatag	AAAAGCTGCATG
exon 6 (198bp)	AGCAGCAATAAG	gtaagatgcg... (1223 bp) ...ttcctttcag	GCTGGGTCCACA
exon 7 (683 bp)	TGACAAGTCCCT	gtggattgag...	

163

Table 4.2 Structural details of ml13d10 TTY2-like gene derived by aligning ml13d10 and TTY2 cDNAs, with genomic DNA from AC009491 fragment

Each exon/intron junction is illustrated by giving on the left the 3' end of an exon and on the right the 5' sequence of the next exon. The beginning and end of each exon are shown in lower case letters. The sizes of each exon and intron are given in parentheses.

Exon	3' sequence	Intron sequence	5' sequence of next exon
exon 1 (1914 bp)	GGTATCCATTGC	ataagtgttt...(10678 bp)...cctttcccag	AGAAACCCTGTG
exon 2 (86 bp)	CTGGGGTCTCAG	atatgattct...(281 bp)...ctctgcttag	ACAGGCTGACAG
exon 3 (165 bp)	AACACCATGGAG	gtgggttggt...(679 bp)...tgtccttcag	CCAAGCCAGGGA
exon 4 (157 bp)	AGGAGCAATGAG	gacagatagg...(2722 bp)...cgacataggg	TTTCCTGGATCT
exon 5 (79 bp)	GGAATCTATTGT	gtgagtgttt...(1775 bp)...agcacaacaa	AACAAAAAAAC
exon 6 (191 bp)	AGCAGCAATAAG	atcagattgg...(7082 bp)...ttgctttcat	CAGGAATCTACT

164

Table 4.3 List of NIX gene prediction programs (HGMP) with their description, that predicted exons within the region of the AC009491 genomic clone that corresponds to modeled exons of ml13d10 gene.

These programs predicted exons in regions corresponding to ml13d10 gene model exons 1,6 and 7 only.

Gene prediction program	Description	Prediction accuracy	Exons identified on AC009491, that overlap with the ml13d10 gene model exons	Corresponding exons of ml13d10 gene model
Grail exon	Identifies exons in genomic sequences with ORFs, considering the presence of splice junctions, translation starts and non-coding scores of 60bp on either side of a putative exon	>98%	140176-140322 141713-141873	**Exon 1 (140199-142237)**
Genemark	With an algorithm based on Bayes Theorem, it divides the sequence to be analysed into small fragments and predicts protein-coding regions.	~50%	140174-140245 141768-141819 161456-161481 168816-168953	**Exon 1 (140199-142237)** **Exon 6 (161329-161504)** **Exon 7 (168595-169247)**
Fgene	(1) Predicts exons by linear discriminant functions and (2) then provides the optimal exon combination and constructs a complete gene model	(1) 81% (2) 90%	141713-141873 161401-161561	**Exon 1 (140199-142237)** **Exon 6 (161329-161504)**

Gene prediction program	Description	Prediction accuracy	Exons identified on AC009491, that overlap with the ml13d10 gene model exons	Corresponding exons of ml13d10 gene model
Genescan	Uses a probabilistic model that accounts for gene structural and compositional properties and constructs gene models	----	161401-161561 168813-168998	**Exon 6 (161329-161504)** **Exon 7 (168595-169247)**
GeneFinder	It uses statistical criteria (primarily log likelihood ratios) and a dynamic programming algorithm, to identify non-overlapping candidate genes	85%	141713-141787	**Exon 1 (140199-142237)**
HMM Gene	It is based on a hidden Markov model, which is a probabilistic model of the gene structure and it predicts whole genes	----	141713-141787	**Exon 1 (140199-142237)**
Fex	Uses a linear discriminant function to predict internal exons and corresponding discriminant functions for prediction of 5'and 3' exons	70%	141713-141873 161401-161561 168682-168770	**Exon 1 (140199-142237)** **Exon 6 (161329-161504)** **Exon 7 (168595-169247)**

gene appears to span 29Kb, in contrast to TTY2, which spans around 18Kb. This is due to a major difference in the sizes of introns 1, 2 and 6.

A similar approach was used to construct a gene model for ml22a3, which showed homology with exon 1 of the TTY2 gene, within the AC023342 genomic sequence and the exon-intron boundaries are shown in Table 4.4. According to the GT-AG rule, introns should start with a GT and end with an AG, flanking the 5' and 3' ends of exons and this was the case for the TTY2 gene constructed from the published cDNA. However, ml13d10 and ml22a3 introns do not start in all cases with a GT and stop with an AG. Rare introns exist, where the conserved splice donor dinucleotide GT is replaced by AT and the splice acceptor dinucleotide AG is replaced by AC. However, it is quite clear from these alignments that the exon/intron boundaries found in TTY2, are not conserved in the TTY2-like genes ml13d10 and ml22a3. Sequences of the TTY2, ml13d10 and ml22a3 genes are given in appendices I, II and III and a diagram with the exon/intron structure of the TTY2 gene and the proposed gene models for ml13d10 and ml22a3 genes are shown in Fig 4.21.

4.7   Identification of TTY2-like subfamilies

34 Y-linked cosmids were identified as containing TTY2-like sequences (see section 4.2, Fig 4.2). 33 of these (excluding the ml13d10 specific cosmid), were amplified using the primers 13d10.F and 13d10.R, which appear to amplify all TTY2-like sequences (Fig 4.5 A). These primers amplified a PCR product of similar size (~299bp) in all but four of the TTY2-like positive cosmids and 27 of these products were successfully sequenced. Amongst these, there were three pairs of cosmids, where both members of the pair showed identical sequence and one member of each pair was excluded from further analysis.

Table 4.4 Structural details of ml22a3 TTY2-like gene using genomic DNA from clone AC023342 as a template

Each exon/intron junction is illustrated by giving on the left the 3' end of an exon and on the right the 5' sequence of the next exon.

The beginning and end of each exon are shown in lower case letters. The sizes of each exon and intron are given in parentheses.

Exon	3' sequence	Intron sequence	5' sequence of next exon
exon 1 (1951 bp)	GGAATCCATTGT	gtgagtgttt... (3113 bp) ...cctttcctag	AGAGCCCCTGTG
exon 2 (87 bp)	CTGGGGTCTCAG	gtatgattct... (2424 bp) ...ctctgcttag	GTAGGGTGACAG
exon 3 (168 bp)	ATTTCCATGAAA	gtgtgttggt... (5581 bp) ...tttcctgtgg	CCAACCCACAGA
exon 4 (157 bp)	AACAAGTCAGAT	gagtgaagat... (2085 bp) ...cggcctaggg	CTTCCTGGGTCT
exon 5 (79 bp)	GGAATCCAATGC	atgagtgttt... (1859 bp) ...accacaccaa	AAAACAGACATC
exon 6 (191 bp)	GCAGGGAGAAAT	tcagatatgg... (3332 bp) ...tggctttcat	GGGGGATTCACA
exon 7 (677 bp)	TGCCACCTTTGC	ctagtgacaa...	

TTY-2 gene 17.9Kb

ml13d10 gene   26.5Kb

ml22a3 gene 21.7Kb
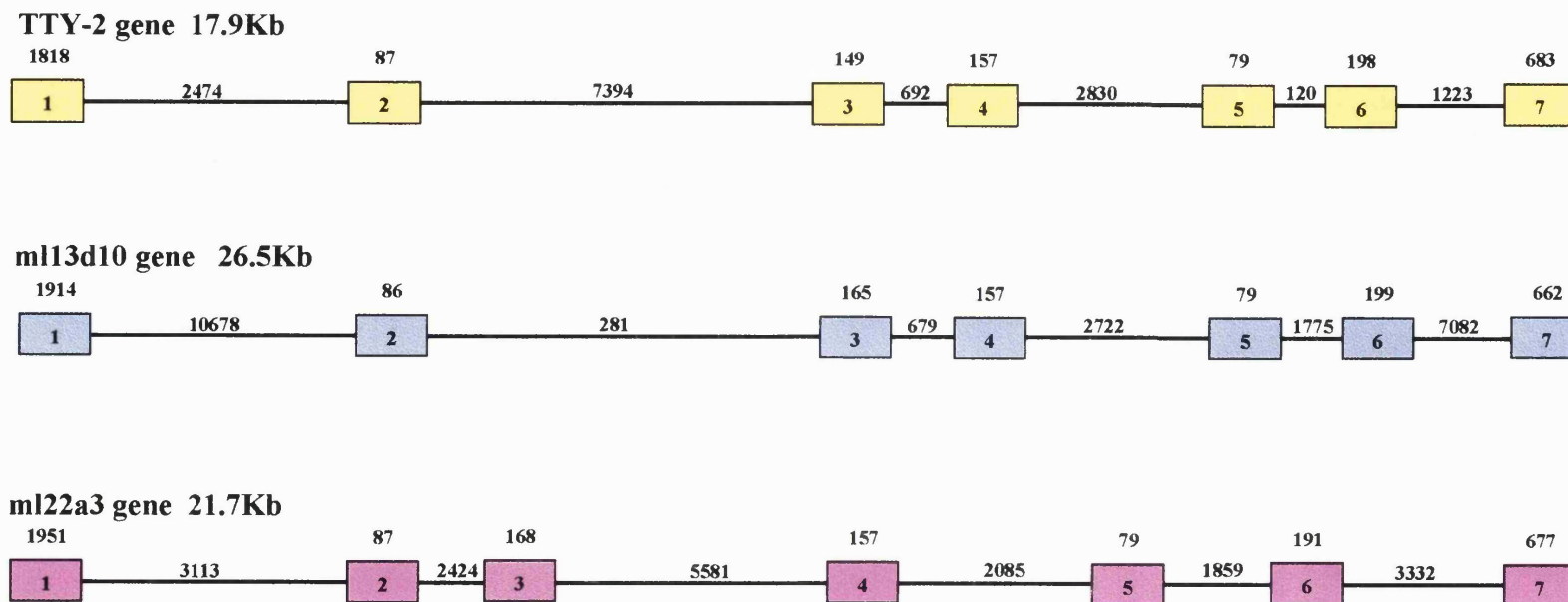
Figure 4.21 Diagram of predicted gene models for the TTY2 gene, ml13d10 gene and ml22a3 gene based on BESTFIT analysis of the genomic clones AC006335, AC009491 and AC023342 respectively.

* Coloured boxes indicate exons, numbers in boxes indicate exon numbers and numbers above exons indicate number of bp

* Solid lines indicates introns and numbers above introns indicate the size in bp

In order to examine their relationship, one to the other, CLUSTALX was used to align and order the 24 sequences and that of ml13d10 and TTY2 (Fig 4.22). This analysis divided the TTY2-like sequences into 14 subfamilies, with each subfamily containing one to five members which share 93-99% identity. Members of different subfamilies have diverged more significantly and share between 55 and 87% homology (Table 4.5). When the cosmid 13d10.F/13d10.R PCR product sequences are aligned with the TTY2 cDNA sequence, they match with similar levels of homology to two different regions of the TTY2, at positions 1bp to 350bp at the 5' end of the cDNA and at 2500bp to 2850bp towards the 3' end (Table 4.6). This finding indicates that the TTY2 cDNA contains an internal repeated motif, an observation that was explored further (see later, section 4.8).

The sequences of all the PCR products (including TTY2, ml13d10 and ml22a3), were checked by translation in all six frames using MAP for the presence of an open reading frame (ORF). Members of six subfamilies (3, 4, 5, 6, 8 and 11) appeared to have an ORF of around 75 amino acids (Fig 4.23). However, only subfamilies 3-2a3 and 6-fd7 appear to have the same ORF, whereas the remaining four, each have different ORFs. For subfamilies 3-2a3 and 6-fd7, the identity of the aa sequences is 58.3% and of the DNA sequences is 72.6%. This indicates that the aa sequence is less conserved than the DNA sequence and probably does not represent true coding region. Thus, amongst these sequences there was no strong evidence that the TTY2 like gene families share a conserved, translated sequence.

A search of the Genbank and EMBL database revealed that each of the TTY2 like sequences, matched with 100% homology to large genomic fragments, mainly BACs (Table 4.7). In some cases, a sequence from a TTY2-like subfamily identified more than one genomic clone, indicating that those particular genomic clones overlap. It was of interest to find that members of different subfamilies sometimes identified the same genomic

Figure 4.22 Alignment of TTY-2 like sequences using CLUSTALX and division into 14

TTY-2 subfamilies. The sequences were derived by PCR amplification of individual cosmids

with primers from a conserved region of the sequence. Boxes separate members of the same

subfamily from other subfamilies. Number on the left denotes each TTY-2 subfamily

```
 1  ed9      ------------------------------TGTTGAGAGGC------CAGTTCTAGGT
    hc8      ----------------------------------AGAGGC------CAGTTCAAGGT
    ff10     ------------------CCCTCTCCTGTACCTGTGAGAGGC------CAGTTCTAGGT
    hb10     ----------------------------------GAGGC------CAGTTCTAGGT
    2e3      ------------------------CTGTACCTGTGAGAGGC------CAGTTCTAGGT
 2  ML22A3   -----------ATGTCTCAACCTCATCTGCCCCTGTAAGAGGC------CAGTTTGAGGT
 3  2a3      ----GCCTGTCCGAGGTATGAAAACACTGTTCCATCTTGGGCT------TGAATTTGT-G
    hb2      ----GCCACTCCGAGGTATGAAAACACTGTTCCATCTTGGGCT------TGAATTTGT-G
 4  eh10     AGAGGCCAGTACCAGGTGTGAGAACACTGCTTCACCTTGGTCT------TGCCATTGT-T
    gg3      AGAGGCCAGTACAAGGTGTGAGAACACTGCTTCACCTTGGTCT------TGCCATTGT-T
    hf5      AGAGGCCAGTACAAGGTGTGAGAACACAGCTCCACCTTGGTCT------TGCCATTGT-T
 5  gb3      AGAGGTCAGTGCAAGATGTGAGAACACAGCTCCATCTTGGACT------TGCCTTTGT-T
 6  fd7      ----------ATGAGGTGGGAGAAGACTGCTCCACCTTGGACT-----TGCCTTAGT-C
 7  2a7      AGAGGCCATTCCGAGATGTGAGAACACTGCTCCACATTAGACT------TGCCTTTGT-T
 8  dd9      AGAGACCAGTCTGAGGTGTGAGACCACTGCTCCACCTTGGGCT------TTCCTTTGT-C
    fa7      AGAGACCAGTCTGAGGTGTGAGACCACTGCTCCACCTTGGGCT------TTCCTCTGT-C
 9  2a6      AGAGACCAGTCTGAGGTGTGAGACAACTGCTCCACCTTGGGCT------TTCCTCGGT-C
10  2d2      AGAGGACAGTCCGAGGTGTGAAAC-ACTGCTCTGCTTTGGACT------TGCCTTTGT-C
    fh9      AGAGGACAGTCCGAGGTGTGAAAC-ACTGCTCTGCTTTGGACT------TGCCTTTGT-C
    2d5      AGAGGACAGTCCGAGGTGTGAAAC-ACTGCTCTGCTTTGGACT------TGCCTTTGT-C
11  2b1      -------AGTACGAGGTGTGAGAACTCTGCTCCTTGTTGGATT------TGCCTTTGTGC
    fd6      -------AGTGCGAGGTGTGAGAACTCTGCTCCTTGTTGGATT------TGCCTTTGTGC
    gg9      -------AGTGCGAGGTGTGAGAACTCTGCTCCTTGTTGGATT------TGCCTTTGTGC
12  TTY-2    AGAGGCCAGTTTGAGGTATGAGAACACTGTTTCAATTTGGACT------TGCCTTTGT-C
13  ml13d10  AGAGGCCAGTTTGAGGTGTGAGAGTACTTCTCCAATTTGGACT------TGCCTTTGT-C
14  ff1      ---AGCCAGTCTGNGTTGTGAAAACACTGCTCCACAATAGAACGTTAAATGTTTGCGTGT
```

```
 1  ed9      GTGGTTCCTGCTTACCCGAGAGAGATCCTGTGAGGTCCAGGATGAAGTGAGGCAC-TGCG
    hc8      GTGGTTCCTGCTTACCCGAGAGAGATCCTGTGAGGTCCAGGATGAAGTGAGGCAC-TGCG
    ff10     GTGGTACCTGCTTACCCGATAGAGATCCTGTGAGGTCCAGGATGAAGTGAGGCAC-TGCG
    hb10     GTGGTTCCTGCTTACCCGAGAGAGATCCTGTGAGGTCCAGGATGAAGTGAGGCAC-TGCG
    2e3      GTGGTTCCTGCTTACCCGAGAGAGATCCTGTGAGGTCCAGGATGAAGTGAGGCAC-TGCG
 2  ML22A3   GTGGTTCCCGCTTACCCGAGAGAGATCCTATCACGCCAAGAATGAAGGGAGGCAC-TGAG
 3  2a3      ATGGTTCCTGTTTCACCCATATAGCCGCTGTTAGGCCTATGATCACATGAGGCAT-TGAA
    hb2      ATGGTTCCTGTTTCACCCATATAGCCGCTGTTAGGCCTATGATCACATGAGGCAT-TGAA
 4  eh10     GTGGTTCGTGCCTGTACCAAAG-GCCACTGCAAAGCCCAGGATGAATGGAGGCATACGAG
    gg3      GTGGTTCGTGCCTGTCCCAAAG-GCCACTGCAAAGCCCAGGATGAATGGAGGCAT-TGAG
    hf5      GTGGTTCGTGCCTGTCCCAAAG-GCCACTGCAAAGCCCAGGATGAATGGAGGCAT-TGAG
 5  gb3      GTGGTTCCTGCCTTCCCCAAAAAGCTCCTGCAAGGCCGAGGATGAAGGGAGGCAG-TGAA
 6  fd7      TTAGTTCCTGCCCTTCCCAAAGAGCCCCTGCAAGGCCCAGGGTGAAGGGAGGCAA-TGAC
 7  2a7      ATGGTTCTTGCCTTTCATAGATAGCCCCTGCAAGGCTCAAAATGAAGGGAGGCAT-TCAG
 8  dd9      GTGGTTCCTGACTTTTCCAGAGGGCCCCTGTGAGGCCCAGGATGAAGGGAGGCCA-TGAT
    fa7      ATGGTTCCTGACTTTTCCAGAGGGCCCCTGTGAGGCCCAGGATGAAGGGAGGCCA-TGAT
 9  2a6      ATGGTTCCTGACTTTTCCAGAGGGCCCCTGTGAGGCCTAGGGTGAAGGGAGGCCA-TGAG
10  2d2      ATGGTTTCTGCCTTTCCCAGAAAGCCCCTGTGAGGCCCAGGGTGAAGAGAGGCAG-TGAG
    fh9      ATGGTTTCTGCCTTTCCCAGAAAGCCCCTGTGAGGCCCACGGTGAAGAGAGGCAG-TGAG
    2d5      ATGGTTTCTGCCTTTCCCAGAAAGCCCCTGTGAGGCCCANGGTGAAGAGAGGCAG-TGAG
11  2b1      ATGGTTTCTGCCTTTTCTAGAGAGCCCCTGTGAGGCCCATGATGAAAACAGGCAG-AGAC
    fd6      ATGGTTTCTGCCTTTTCTAGAGAGCCCCTGCGAGGCCCATGATGAAAACAGGCAG-ACAC
    gg9      ATGGTTTCTGCCTTTTCTAGAGAGCCCCTGTGAGGCCCATGATGAAAACAGGCAG-TGCA
12  TTY-2    TTGGTTCCTGCTTTTCCCAGATGGCACCTACCCAACCCAGGATGAATGAGTGCAG-AGAG
13  ml13d10  ATGTTTCCTGCCTATGTGAGAGAGACCCTGTGAAACCCAGGTAGAAGGGAGGCAG-TAAG
14  ff1      ATGTTGTAGCATTGTGTTGAGGTGTGTTTGCGTGTGTGTATGTGCCTGAGACCAC--CTC
                 *    *            *     *                      *
```

First block

1	ed9	GTTAAGGGCCTGGTCGTCTTTCATTGACACTCATCTCTGGGGTCTCAGGTATGATTCTAT
	hc8	GTTAAGGGCCTGGTCGTCTTTCATTGACACTCATCTCTGGGGTCTCAGGTATGATTCTAT
	ff10	GTTAAGGGCCTGGTCGTCTTTCATTGACACTCATCTATGGGGTCTCAGGTATGATTCTAT
	hb10	GTTAAGGGCCTGGTCGTCTTTCATTGACACTCATCTCTGGGGTCTCAGGTATGATTCTAT
	2e3	GTTAAGGGCCTGGTCGTCTTTCATTGACACTCATCTCTGGGGTCTCAGGTATGATTCTAT
2	ML22A3	GTTAAGGCCCTGGTCATTTTTAATTGACACCCATCTCTGGGGTATCAGGTAGGATTCTAT
3	2a3	GTCAATAGCCTGGCCATCTGTCACTGACACCTGCCTCTCGGGTCTCATATATGATTCTAT
	hb2	GTCAATAGCCTGGCCATCTGTCACTGACACCTGCCTCTCGGGTCTCATATATGATTCTAT
4	eh10	AACAATAGCCTGGCTAGCTTTCAATGACTCCCACCTCTGGGTTTTCAGGTATTATTCTAT
	gg3	AACAATAGCCTGGCTAGCTTTCAATGACTCCCACCTCTGGGTTTTCAGGTATTATTCAAT
	hf5	AACAATAGCCCGGCAAGCTTTCAATGACTCCCACCTCTGGGTTTTCAGGTATTATTCAAT
5	gb3	GTCAAGAGCCTGTCCGTCTTTTATAAACACCTGCCTCTGGGGTTTCAGGTATGATTCTAT
6	fd7	ATCAAAAGCCCANGCATCGTTCACAGACAACAACCTCTGTGGTATCAGGTATGATTCTAT
7	2a7	TTCAAAAGCCCAATCATCTTTTGCTGACATCCACTTCTGGTGTCTCAGGTATGTTTCTGT
8	dd9	GTC--AAGCTCGGACATCTTTTACTGATACCCATGTCTGTGGGGTCACGTGTATTACTTC
	fa7	GTC--AAGCTCGGACATCTTTTACTGATACCCATGTCTGTGGGGTCACGTGTATTACTTC
9	2a6	GTC--AAGCTCGGGCATCTCTCACTGATGCTCACCTCTGGGCTTTCAGGTATAATTCTAT
10	2d2	GTCCAGAGCCCAGCTATCTTTCACTGACACCTAGTTCTGGGGTCTGAGGTGTAATTCAAT
	fh9	GTCCAGAGCCCAGCTATCTTTCACTGACACCTAGTTCTGGGGTCTGAGGTGTAATTCAAT
	2d5	GTCCAAAGCCCAGCTATCTTTCACTGACACCTAGTTCTGGGGTCTGAGGTGTAATTCAAT
11	2b1	ATTAATGGGCCAGCCATCTTT-GCTGACACACGACTCTGGGGTCTGAGGGATTACTTCAC
	fd6	ATTAATGGGCCAGCCATCTTT-GCCGACACACGACTCTGGGGTCTGACGGATTACTTCAC
	gg9	ATTAATGGCCCAGCCATCTTT-GCCGACACATGACTCTGGGGTCCCAGGTATGATTCCAC
12	TTY-2	GTCAAGTGCCAGGCCATCTTTTTGCTGACACCCTTTTCTGGTATTTCAGGTATAAGTCCAT
13	ml13d10	GTTAATGGCCCAGCCATCTTTTTGCTGACCCCCATCTCTGGGGTCTCAGGTATGATTCTAT
14	ff1	TTG--GGTCTCAGATATAATT--CTAAAGACCACT--CAACATCTCACCTGACTACATTC

```
               *            *                    *
```

Second block

1	ed9	CATTCAAAGACCCCTAAACAACCCACCAGACTACATTCCAATCC-CATGG-ACCTGATAA
	hc8	CATTCAAAGACCCCTAAACAACCCACCAGACTACATTCCAATCC-CATGG-ACCTGATAA
	ff10	CATTCAAAGACCCCTAAACAACCCACCAGACTACATTCCAATCC-CATGG-ACCTGATAC
	hb10	CATTCAAAGACCCCTAAACAACCCACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
	2e3	CATTCAAAGACCCCTAAACAACCCACCAGACTACATTCCAATCC-CATGG-ACCTGATAA
2	ML22A3	CACCCAAAGACCCCTAAACAATGCACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
3	2a3	CACCCATGGAACACCACACAATACACCAGACAGCATTCCAATCC-CATGG-ACCTGATA-
	hb2	CACCCATGGAACACCACACAATACACCAGACAGCATTCCAATCC-CATGG-ACCTGATA-
4	eh10	CCATCACAGAACCCTGAAAAACACACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
	gg3	CAATCACAGACCCCTGAAAAACACACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
	hf5	CAATCACACACCCCTGAAAAACACACCAAACTACATTCCAATCC-CATGG-ACCTGATA-
5	gb3	CACCGAAAGAACCCTCAAAAATACACCAGAATACATTCCAATCC-CATGG-ACCTGATAA
6	fd7	CACCCAAGGAACCCTCAAAAACACACCTGACTGCATTCCAATCC-CATGG-ACCTTATA-
7	2a7	CGCCCAAAGAACACTCAACGACACACCAGATTACATTCCAATCC-CATGG-ACCTGATA-
8	dd9	CTNACCCAGAAACC-CAT--
	fa7	CTCACCCAGAAACC-CATCAACACATTACATTCTACTCCC-------------------
9	2a6	CACCCAAAGAACCC-CAACAACACAACAGATTCCATTCCAATCC-CATGG-ACCTGATAA
10	2d2	CACCCAAAGAATCCTCAAAAACACACCAGACTACATTCCAATCC-CATGG-ACCTGATAA
	fh9	CACCCAAAGAATCCTCAAAAACACACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
	2d5	CACCCAAAGAATCCTCAAAAACACACCAGACTACATTCCAATCC-CATGG-ACCTGATA-
11	2b1	CATCCAAACGCTCTTAAACAACTCACCATATGACATTCCAATCC-CATGG-ACCTGATA-
	fd6	CATCCCAACGCTCTTCAACAACTCACCATATGACATTCCAATCC-CATGG-ACCTGATA-
	gg9	TATCCAAACACTTCTCAACAACTCACCAGATGACATTCCAATCC-CATGG-ACCTGATAG
12	TTY-2	CATCCAAAGACTGCTCAACATCTCACCAGAATATATTTCAATCCTCATGGGGCATGATTC
13	ml13d10	CACCCAAAGACACCTAAACAAATCACCAGACTACATTCCAATCCCCATGGGACCTGATTA
14	ff1	CAATCCCATGGACCTGATAAAT--ACCAGACAGCATTCCAATCC-CATGG-ACCTGATA-

172

```
        ed9        ACACACACCANACTACATTCCAATCCCATGGACCTGATA--------------------
        hc8        ----------GAC-ACATTCCAATCCCATGGACCTGATAC-------------------
1       ff10       ----------AAGCACATTCCAATCCCATGGACCTGATAC-------------------
        hb10       ----------AAGCACATTCCAATCCCATGGACCTGATCC-------------------
        2e3        ACA-------------------------------------------------------
2       ML22A3     ----------------------------------------------------------
3       2a3        ----------------------------------------------------------
        hb2        ----------------------------------------------------------
        eh10       ----------------------------------------------------------
4       gg3        ----------------------------------------------------------
        hf5        ----------------------------------------------------------
5       gb3        TA--------------------------------------------------------
6       fd7        ----------------------------------------------------------
7       2a7        ----------------------------------------------------------
        dd9        ----------------------------------------------------------
8       fa7        ----------------------------------------------------------
9       2a6        ----------------------------------------------------------
        2d2        ----------------------------------------------------------
10      fh9        ----------------------------------------------------------
        2d5        ----------------------------------------------------------
        2b1        ----------------------------------------------------------
11      fd6        ----------------------------------------------------------
        gg9        T---------------------------------------------------------
12      TTY-2      TTTCAC----------------------------------------------------
13      ml13d10    TTGCAC----------------------------------------------------
14      ff1        ----------------------------------------------------------
```

Figure 4.23 Members of the TTY-2 subfamilies that appear to have an ORF of around 75 a.a along the sequence that is available. Subfamilies 3 and 6 have the same ORF and red color indicates the identical a.a. ------- next to members of a subfamily indicate that they do not have a significant ORF.

Subfamily 11
```
2b1    YSTHSSQEKNSKGKHMTEAKELSG-RHPGH-HFCASVNIPWGD--KSVCSEPTQPIVEGDLR-EVVWIVNWDWP--GS
fd6    HSTHSSQEKNSKGKHMTEAKELSG-RRPGH-HFCASVNIPWGD--KGVCSEPTQRIVEGDWREEVVWIVNWDWP--GS
gg9    HSTHSSQEKNSKGKHMTEAKELSG-RHPGH-HFCATCNIAWGD--KGVCSEPTGPIIGSDLCKEVVWIVNWDWP--GSL
```

Subfamily 3
```
2a3    GTRPIFVSNWRPSSNTITGTEGMYG-SNPRH-DCSANFDIAQGD--T--VSVQREPRMYSEIVWPVGCLVGSLMG-IGHVQY
hb2    GSRPIFVSNWRPSSNTITGTEGMYG-SNPRH-DCSANFDIAQGD--T--VSVQREPRMYSEIVWPVGCLVGSLMG-IGHVQY
```

Subfamily 6
```
fd7    --HPPLLSSWRPSAKTKTGARGLSG-RCTGP-HLSAIVDFAWAD--N--VSLLRQPILYSEIVWPVRLFVGSQIG-IGHVKY-
```

Subfamily 5
```
gb3  LDTCSTLVSSWRPSAKTTTGAKGLSS-RCPRP-HLSATFDLAQGD—KISVQRQPKLYSEIVWLVRLFVCWFINWVERSRII
```

Subfamily 8
```
dd9  ---SVLRLHSVVAGGQAKRQRPEQSK-WLAG--TLGLIFP-PWS---T--LSPCRKS-IGMDTP-DRTNSGGSVW
fa7  -----------------------------------------------------------------------
```

Subfamily 4
```
eh10   --RPVPGVRTLLHLGLAIVVVRACTKGHCKAQDEWRHTRTIAWLAFNDSHLWVFRYYSIHHRTLKNTPDYIPIPWT-
gg3    -----------------------------------------------------------------------
hf5    -----------------------------------------------------------------------
```

Table 4.5 The TTY2 like subfamilies denoted by number and name of the first listed sequence. The % homology between members of a subfamily and between first members of the nearest subfamilies are shown.

Subfamily	No. of independent clones	Average % homology within subfamily	% homology to nearest subfamily
1-ed9	5	99%	1 and 2 87%
2-ml22a3	1	----	2 and 3 67%
3-2a3	2	99%	3 and 4 72%
4-eh10	3	97%	4 and 5 80%
5-gb3	1	----	5 and 6 81%
6-fd7	1	----	6 and 7 78.5%
7-2a7	1	----	7 and 8 70%
8-dd9	2	99%	8 and 9 84%
9-2a6	1	----	9 and 10 81%
10-2d2	3	99.5%	10 and 11 79%
11-2b1	3	93%	11 and 12 67%
12-TTY-2	1	----	12 and 13 76%
13-ml13d10	1	----	13 and 14 55%
14-ff1	1	----	----

Table 4.6 Alignment of the TTY2-like cosmid sequences with exons 1 and 7 of the TTY-2 cDNA. Subfamily is denoted by the number of the subfamily followed by the name of the first listed sequence.

TTY-2 like subfamily	% homology with TTY-2 cDNA	
	exon 1	**exon 7**
1-ed9	63%	57%
2-ml22a3	70%	70%
3-2a3	62.5%	67%
4-eh10	69%	69%
5-gb3	75%	69%
6-fd7	73%	66%
7-2a7	68.5%	70%
8-dd9	71%	68%
9-2a6	71%	68%
10-2d2	71%	75%
11-2b1	69%	67%
12-TTY-2	100%	100%
13-ml13d10	72%	71%
14-ff1	54%	52%

Table 4.7 Genomic sequences containing TTY2 subfamily cDNA sequences. Subfamily designated by number followed by first listed sequence.

Subfamily no.	genomic clones	size (bp)	region of 100% homology to TTY-2 subfamily sequence	orientation (subfamily/gen. Clone
1-ed9	AC006335	222512	6464-6649	+/+
	AC010154	156021	103282-103464	+/+
	AC017019	162187	71696-71878	+/+
2-ml22a3 gene	AC023342	179841	63549-85243	+/+
	AC007320	191414	185814-...	+/+
	AC008175	205236	...-19816	+/+
	AC007359	100627	...-16104	+/+
3-2a3	------	------	------	------
4-eh10	------	------	------	------
5-gb3	AC010141	120333	45363-45592	+/+

177

subfamily no.	genomic clones	size (bp)	region of 100% homology to TTY-2 subfamily sequence	orientation (subfamily/gen. clone
6-fd7	AC017019	162187	154476-154680	+/+
	AC006335	222512	89254-89458	+/+
	AC010891	176365	53422-53624	+/+
7-2a7	AC007359	100627	35468-35688	+/+
	AC017045	99214	51845-52055	+/+
	AC017019	162187	121834-122054	+/+
	AC010154	156021	153429-153645	+/+
	AC010891	176365	20755-20971	+/+
	AC006335	222512	56613-56833	+/+
	AC008175	205236	39168-39380	+/+
	AC023342	179841	104604-104826	+/+
8-dd9	------	-----	------	------
9-2a6	AC010141	120333	7505-7716	+/+
10-2d2	AC009491	174146	108190-108402	+/+
11-2b1	------	------	------	------

subfamily no.	genomic clones	size (bp)	region of 100% homology to TTY-2 subfamily sequence	orientation (subfamily/gen. clone
12-TTY-2 gene	AC006335	222512	34268-52172	+/+
	AC017019	162187	99498-117392	+/+
	AC010154	156021	54284-72188	+/+
	AC010891	176365	16310-...	+/+
13-ml13d10 gene	AC009491	174146	140201-169229	+/+
14-ff1	------	------	------	------
15-ml18g1	AC017019	162187	95944-96165	+/+
	AC010154	156021	127528-127749	+/+
	AC006335	222512	30713-30934	+/+
16-ml22h10	AC017019	162187	50600-50892	+/-
	AC010154	156021	82186-82478	+/-

179

clone, suggesting that some TTY2-like genes are clustered. By sequence alignment it was possible to place the TTY2-like sequences on individual genomic fragments.

Two clusters were mapped in detail and are shown in Fig 4.24 and 4.25. Figure 4.24 shows the clustering of sequences from 6 TTY2-like sequences including TTY2, two testis selected cDNAs (ml18g1, ml22h10, see section 4.1) and four sequences derived from the TTY2-like containing cosmids, on a contiguous stretch of genomic DNA, 90Kb long. In this diagram, the cDNA clones ml18g1 and ml22h10 are positioned in three overlapping genomic fragments AC006335, AC010891 and AC017019 that make up this contig. The sequence 7-2a7, positioned close to TTY2 also shows 100% homology with four other genomic DNA fragments, which are not shown. These DNA fragments are of interest, because they contain the ml22a3 sequence, but because they are working draft sequences of unordered pieces they could not be positioned accurately within the contig of Fig 4.24. Nevertheless, since TTY cDNA has been mapped within the AZFc region (Lahn and Page., 1997) and ml22a3 sequence has been positioned on Yq11 within either the AZFb or AZFc region (section 4.3), this indicates that the cluster of TTY2-like genes shown in Fig 4.24, which includes the ml22a3 and TTY2 genes, is located on the long arm of the Y chromosome, within AZFc.

Figure 4.25 shows the testis selected cDNA ml13d10 gene and the cosmid-derived sequence 10-2d2, located about 32Kb apart on the genomic DNA AC009491. This indicates that there is a TTY2-like gene cluster at the map position derived for ml13d10, at Yp11.2 (interval 4A).

These observations led to the conclusion that members of the TTY2 gene family have arisen through a series of duplication events that took place during the evolution of the Y chromosome. Transposition of gene copies and further duplication, have resulted in the formation of clusters of TTY2-like
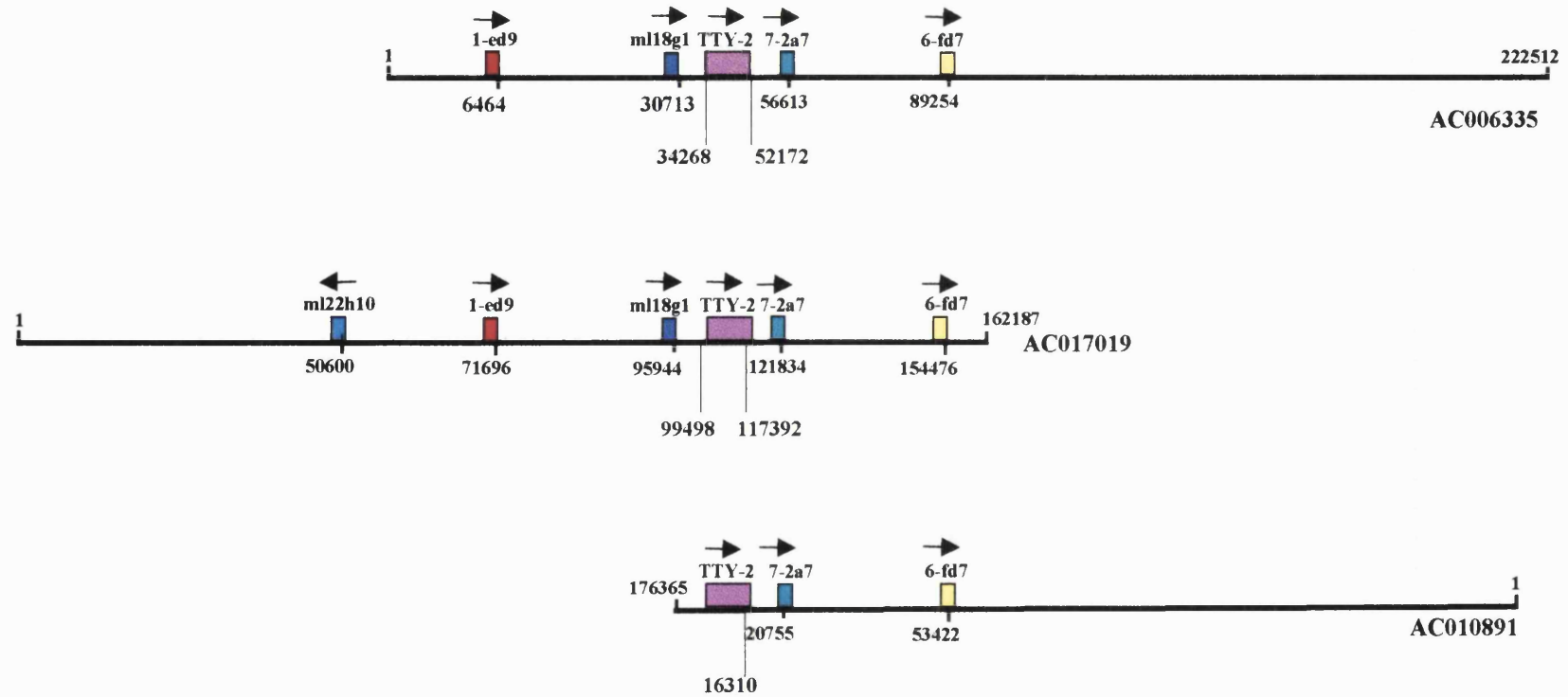
180

Figure 4.24 Schematic representation of three overlapping genomic fragments and localisation of individual members of the TTY2-like subfamilies 1, 6 and 7 and the two TTY2-like cDNA clones ml18g1 and ml22h10; arrows indicate orientation of the genes in relation to the genomic sequence; numbers below the genes indicate the first bp of homology within the genomic fragment
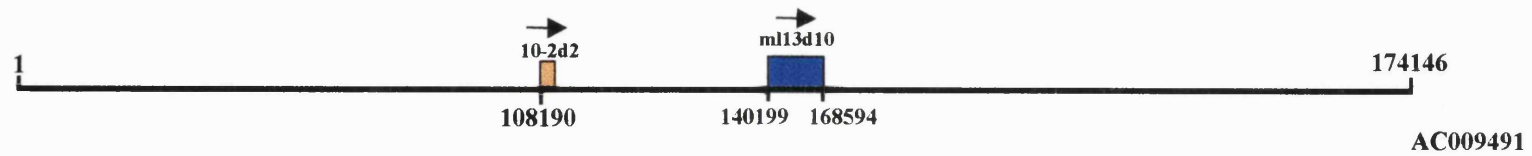
Figure 4.25 Schematic representation of the genomic fragment AC009491 and localisation of the ml13d10 gene and 2d2, a member of the TTY2-like subfamily 10; arrows indicate orientation of the genes in relation to the genomic fragment; numbers below indicate the first bp of homology within the genomic fragment and for ml13d10 gene the 5' and 3' ends of the predicted gene

genes, distributed across the Y chromosome.

A better view of the duplication events that took place throughout
evolution of this gene family, is provided by the construction of an unrooted
phylogenetic tree, which shows the branching patterns amongst the TTY2
subfamilies, without specifying the oldest sequence from which the rest were
originated (Fig 4.26). Confidence of the tree was evaluated by 100 bootstrap
replicates of the examined data (Fig 4.27). In order to simplify the
construction of the tree, only sequences that share less than 90% homology
were used. Correspondingly, only the first member from each of the
subfamilies shown in Fig 4.22, were used as representatives and their
sequences were compared and aligned using the PILEUP program. The
pairwise distance matrix was created using DISTANCES and the
phylogenetic tree was produced using the neighbour-joining method with the
program GROWTREE and visualised with the help of PAUPdisplay (GCG
program package at HGMP) and PAUP (Macindosh phylogeny package).

The cladogram displayed in Figs 4.26 and 4.27 cannot accurately
specify which was the ancestral sequence. In addition, bootstrap replication
(Fig 4.27) failed to derive strong confidence values for the tree groups and
therefore collapsed most of the sequences under the same branch.
Nevertheless, the phylogetetic analysis of Fig 4.26 supports the CLUSTALX
subfamily order and shows that subfamilies that have been positioned
adjacent by CLUSTALX, are also members of the same cladistic branch.
One example is that a member of subfamily 6 (fd7), appear to have given
rise to members of subfamilies 4, 5 and 7. In addition, both phylogenetic
trees indicated that ml13d10 and TTY2 and ed9-1 and ml22a3 are relatively
closely related and phylogenetically appear to be closer than with other
TTY2 genes. As yet though, there are no clues about the evolutionary time
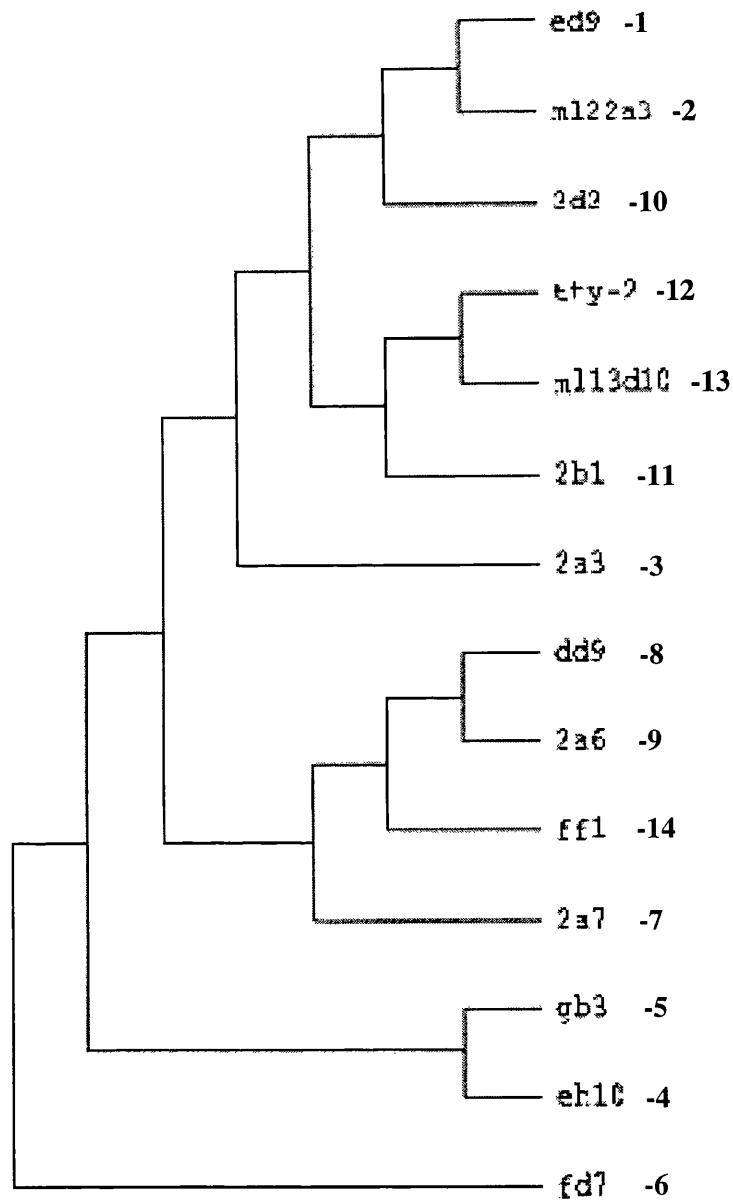that this multiplication and divergence of TTY2-like sequences took place.

Figure 4.26 Phylogenetic tree of TTY-2 and TTY-2-like sequences,

based on the analysis of a single member from each subfamily, created using

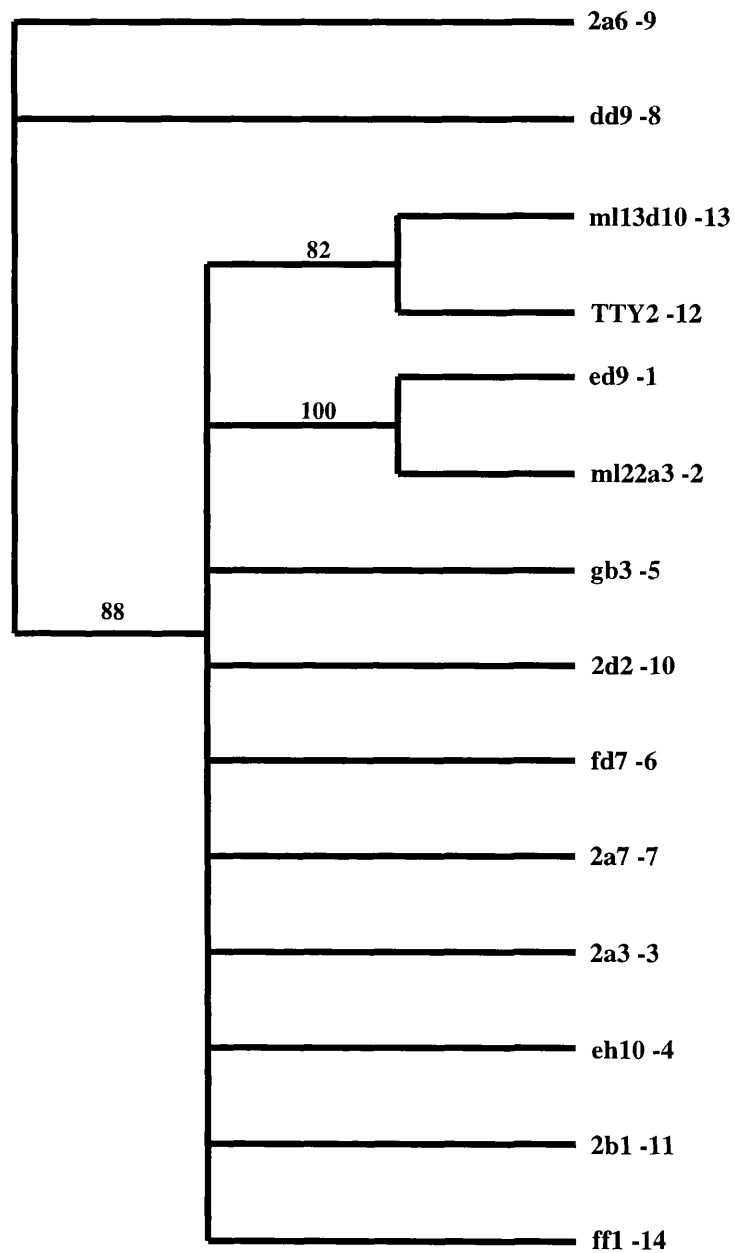GROWTREE and displayed using PAUPDISPLAY programs.

Figure 4.27 Phylogeny of the TTY2 and TTY2-like sequences. This tree was produced using the heuristic method. Percentage support in 100 bootstrap replications is given below the major branches. The tree was created and displayed using PAUP program

4.8 Investigation of an internally repeated structure, characteristic of TTY2-like genes

In order to investigate further the finding that elements of the TTY2 and TTY2-like sequences are repeated within individual genes (see section 4.7), GAP analysis was used to compare the exons of the TTY2 gene, one with another. This showed that exons 2 to 7 overlap with 70-75% homology with sequence from exon 1, which is the largest exon comprising of 1818 bp. In addition, exons 3 and 7 and exons 4 and 6 share 65% and 73% homology respectively.

A DOTPLOT program that graphically aligns sequences was used to align the TTY2 cDNA and the TTY2 predicted gene to themselves. This demonstrated the existence of internal repeated patterns for both the cDNA and the genomic sequence (Figs 4.28 and 4.29). In the cDNA, short fragments of sequence are repeated such that a repeat of exons 2 to 7 lies within exon 1. Exon 6 appears to represent a repeat of exon 4 and exon 7 repeat of exon 2. Analysis of the 17.9Kb TTY2 gene, demonstrated that the gene comprises 6 to 7 tandemly arranged repeats of is around 2.4Kb. Each repetitive unit is made up of 4 smaller repeats, which are arranged at a conserved distances of 300bp, 500bp and 700bp.

The position of the 7 exons in relation to these repeats has been marked in Fig 4.29. The TTY2-like ml13d10 and ml22a3 genes followed a similar pattern of internal repeat units and DOTPLOT analysis for both of the predicted cDNA sequence and predicted gene sequence are illustrated in Figs 4.30, 4.31, 4.32 and 4.33.

The above results indicate the existence of an untranslated multicopy gene family on the Y chromosome, with at least 14 distinct subfamilies and at least 26 individual genes, each member of which is characterised by tandemly arranged repeat units and each gene shows evidence of intragenic duplication.
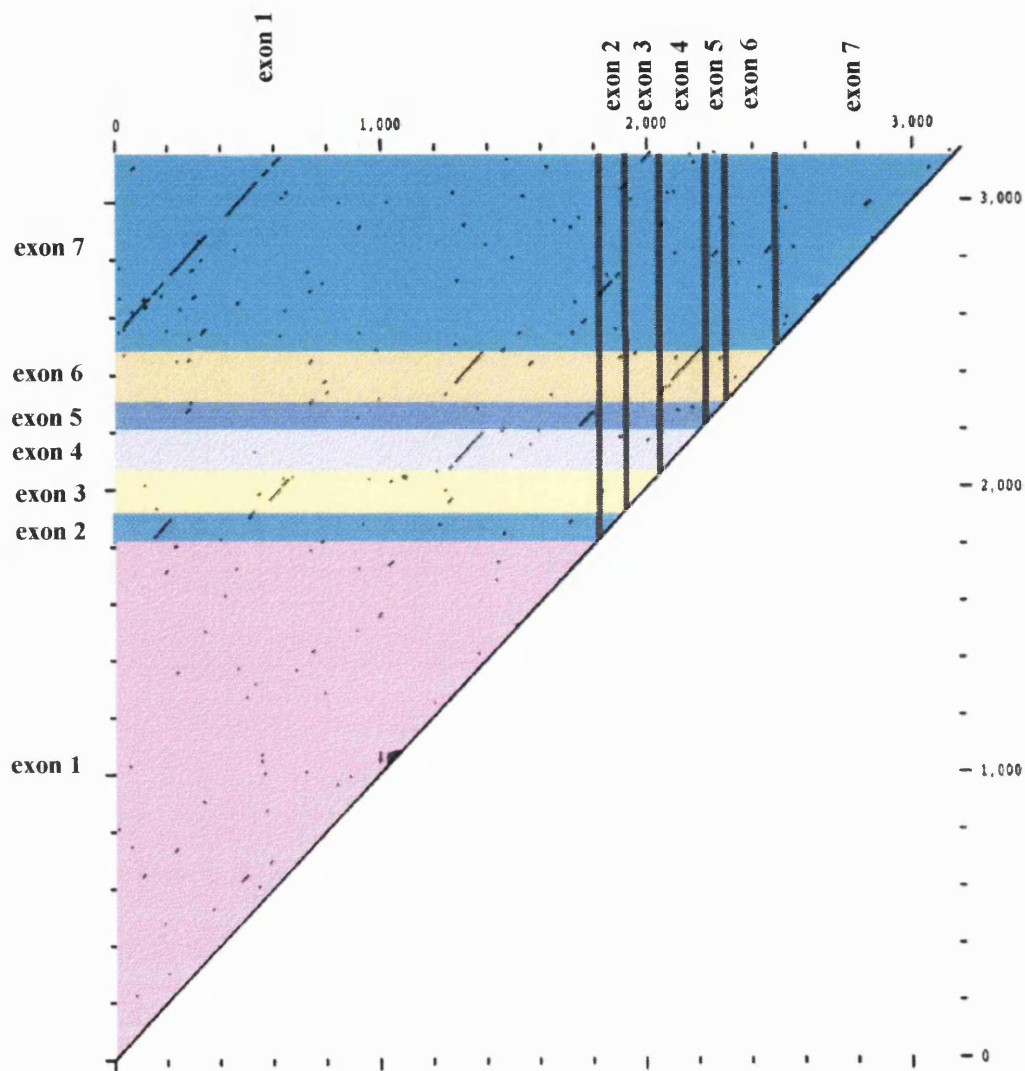
Figure 4.28 DOTPLOT analysis of the TTY2 cDNA (3169bp) to itself;
exons are indicated by coloured blocks; internal repeats appear as
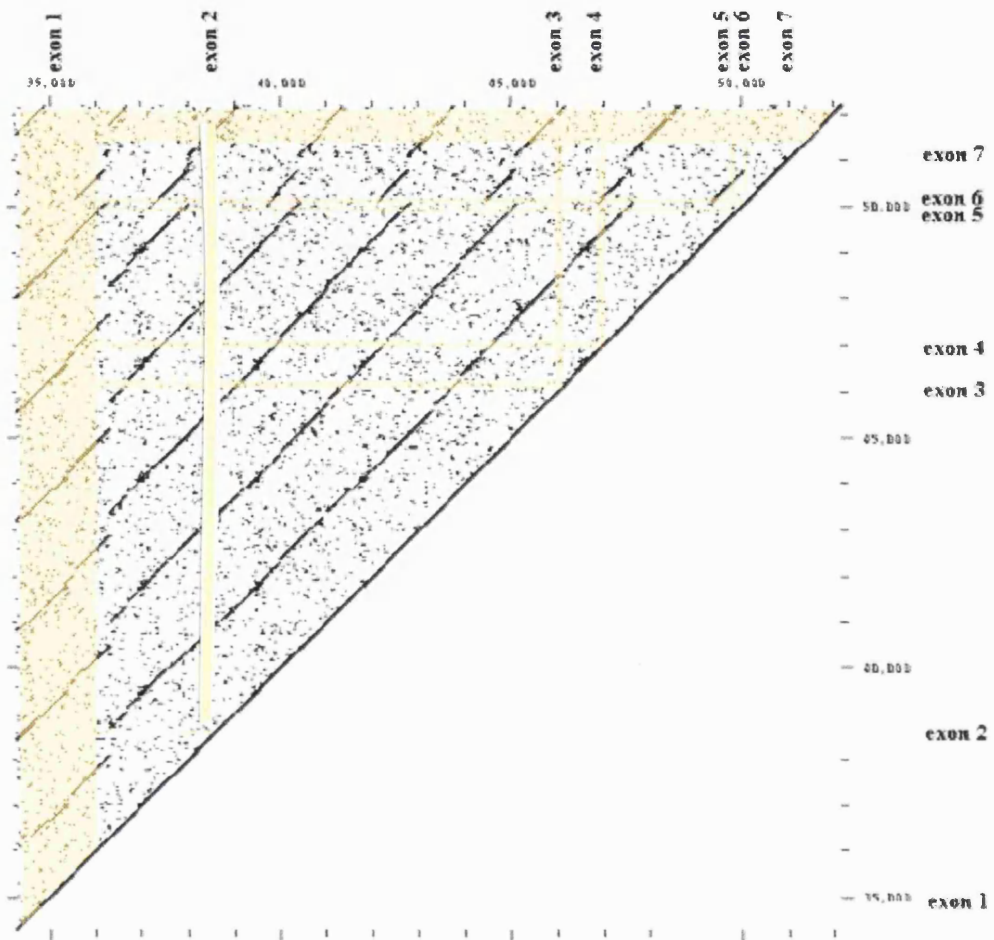diagonal links

Figure 4.29 DOTPLOT analysis of 17.9Kb of sequence from genomic fragment AC006335 that correspond to the TTY2 gene to itself; exons are indicated by coloured blocks; internal repeats appear as diagonal links.
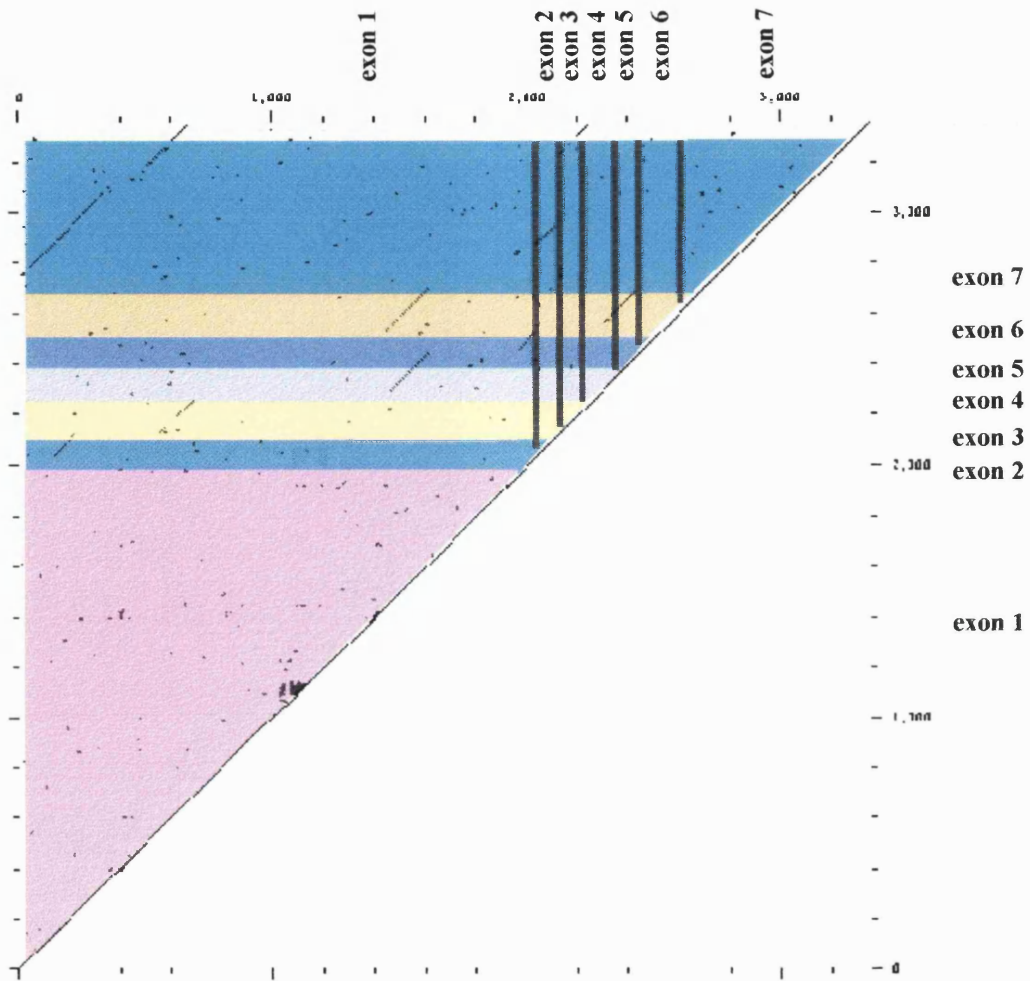
Figure 4.30 DOTPLOT analysis of the ml13d10 cDNA (3262bp) to itself; exons are indicated by coloured blocks; internal repeats appear as diagonal links
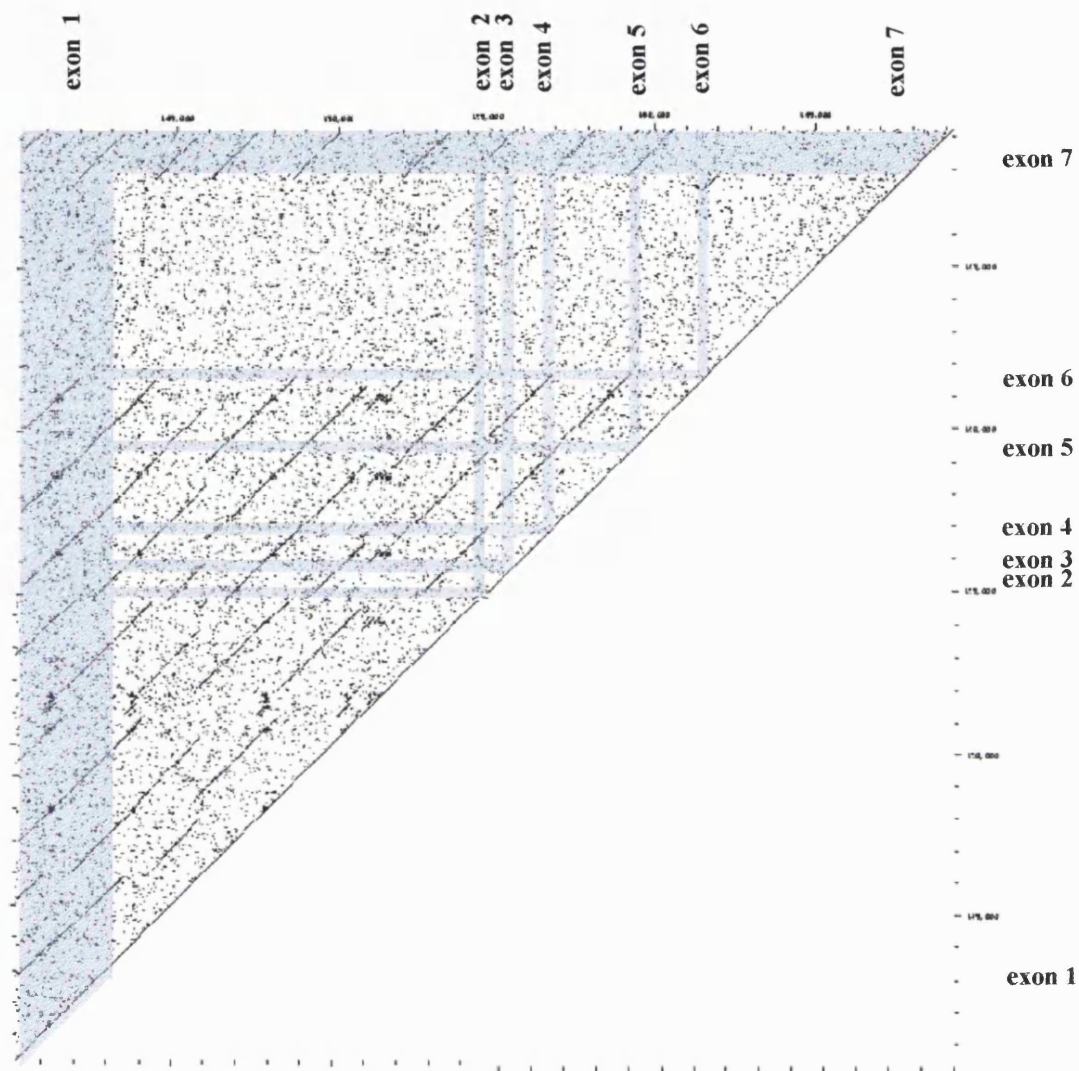
Figure 4.31 DOTPLOT analysis of 26.5Kb of sequence from genomic
fragment AC009491 that correspond to the ml13d10 gene to itself;
exons are indicated by coloured blocks; internal repeats appear as
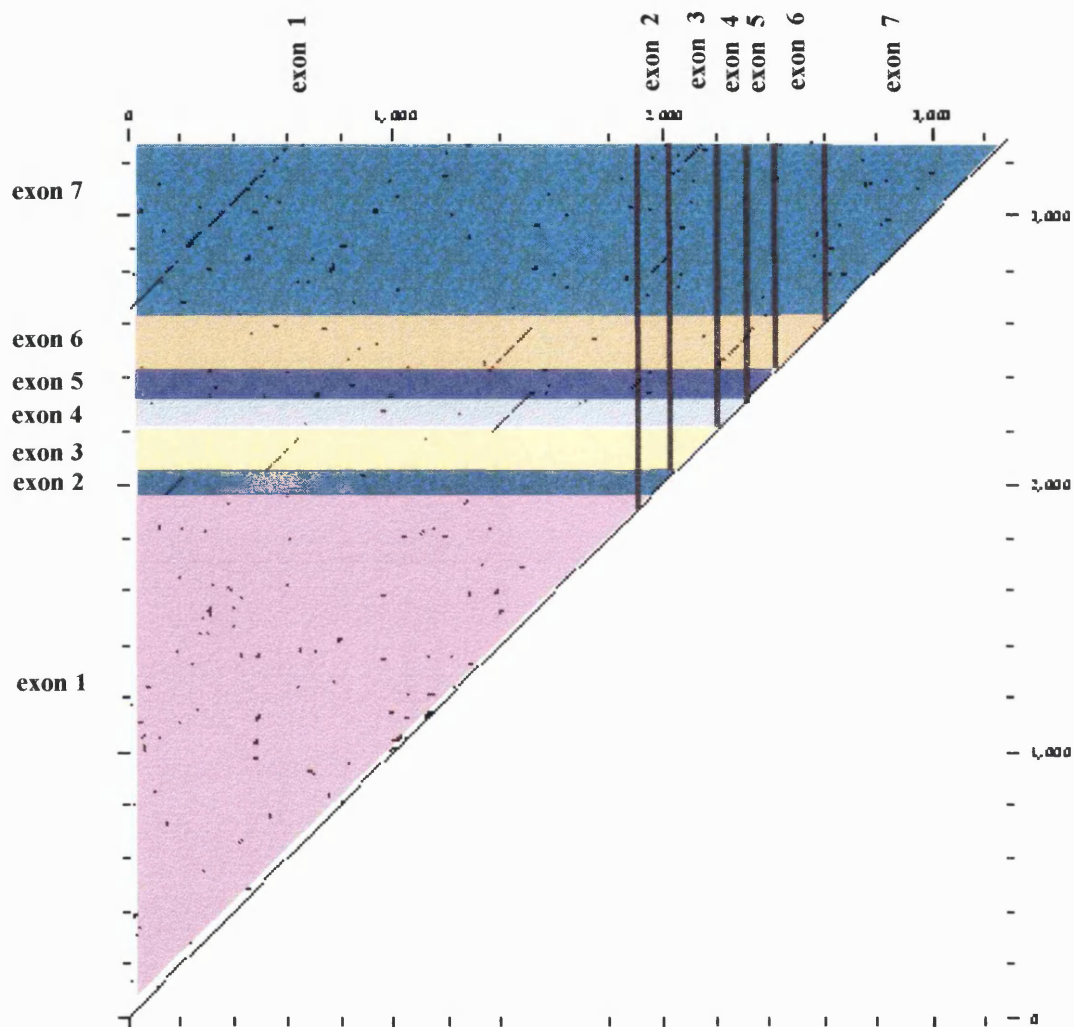diagonal links

Figure 4.32 DOTPLOT analysis of the ml22a3 cDNA (3310bp) to itself; exons are indicated by coloured blocks; internal repeats appear as diagonal links
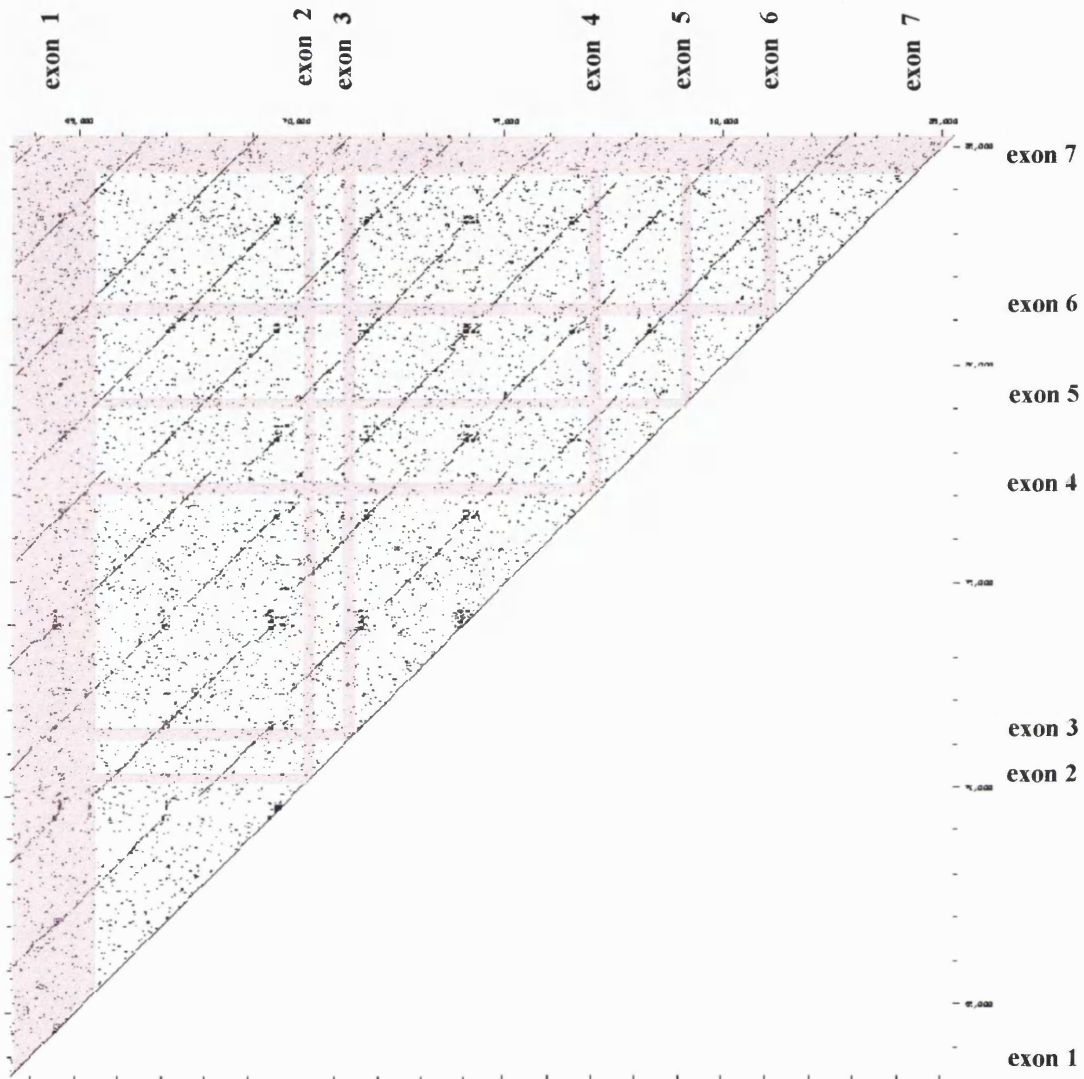
Figure 4.33 DOTPLOT analysis of 21.7Kb of sequence from genomic clone AC023342 that correspond to the ml22a3 gene, to itself; exons are indicated by coloured blocks; internal repeats appear as diagonal links

4.9 Analysis of DNA from azoospermic patients

Several clinical and molecular studies have demonstrated a connection between male infertility and Y chromosome deletions and/or microdeletions. Such studies have tried to define the position and extent of these deletions and the frequency with which they occur.

In order to determine whether deletions of the ml13d10 and ml22a3 TTY2-like sequences occur in infertile patients, DNA from a panel of male individuals (n=52) experiencing fertility problems was analysed by PCR (provided by C. Quilter at the Cytogenetics Unit, UCH hospital). These patients were all selected for idiopathic severe oligozoospermia ($<5 \times 10^6$ sperm/ml of semen) or azoospermia (complete absence of sperm in semen) and showed no evidence of obstructive azoospermia, endocrine deficiencies or any known cytogenetic defects. In addition another set of 30 DNA samples from oligo- and azoospermic patients were provided by Prof. Kleiman at the Maternity Hospital, Tel Aviv. These patients had been previously analysed by Prof. Kleiman for several STS markers on both arms of the Y chromosome (Kleiman *et al.*, 1999).

The quality of the DNA was checked by amplification with ml22d8 primers (22d8.F and 22d8.R), which amplify both Chrs Y and 15 (see section 5.2). One of the DNA samples (no. 48) did not amplify with either ml13d10 or ml22a3 specific primers or ml22d8 primers and was excluded from further analysis. As a positive control and in order to ensure that the sequence of the primer site was not polymorphic in the general population, 40 unrelated CEPH fathers were included in this study. As a negative control for each experiment, PCRs were set up in the absence of DNA.

PCR amplification using 13d10.2F and 13d10.2R primers gave products of the correct size (233bp) for all but 2 DNA samples from UCH no.7 and no.18 (Fig 4.34). PCR products were obtained for these two samples using primers for ml22a3 and the ml22d8 control.

The results for patients no.7 and no.18, could be explained if deletion(s) were present, which eliminate part of the ml13d10 TTY2-like gene. The location of the deletions must include the region corresponding to either/or both ml13d10 specific primers. Until further analysis of the region is carried out, it is difficult to be certain that this interpretation is correct. It might have been possible to design other pairs of primers from the ml13d10 sequence, but it would be necessary to ensure that these primers do not amplify other TTY2-like gene sequences.

Amplification of ml22a3 in the patient panel showed a correct sized product in all the UCH samples. However, amongst the samples from Tel-Aviv, these primers failed to amplify DNA from two patients (no.1 and no.2) (Fig 4.35). In contrast, PCR products of correct size were obtained for both, using primers specific for ml13d10 and for the ml22d8. Kleiman *et al* (1999), have reported that analysis of STS markers and expressed sequences in the DNA of patient no.1 (Kleiman no.95, idiopathic azoospermia), indicates that he lacks the AZFc region only, whereas patient 2 (Kleiman no.102, Sertoli cell-only syndrome) lacks the whole of AZFa-c region. The results presented here, from the PCR amplification of the Tel Aviv patients no.1 and 2 DNA are in agreement with the published data.

In summary, results from screening the panel of 82 oligo- and azoospermic patients for the presence of two of the TTY2-like genes, ml13d10 and ml22a3, suggested that deletions of both genes may occur at low frequency and might contribute to the infertile phenotype. However, the infertile phenotype might be the result of deletions or mutations of other genes, which are present in the region where ml13d10 and ml22a3 have been localised. PCR screening of these patients with either STS primers around the TTY2-like location, or for the presence of mutations in other surrounding genes, would clarify whether TTY2 genes have an impact in male infertility.
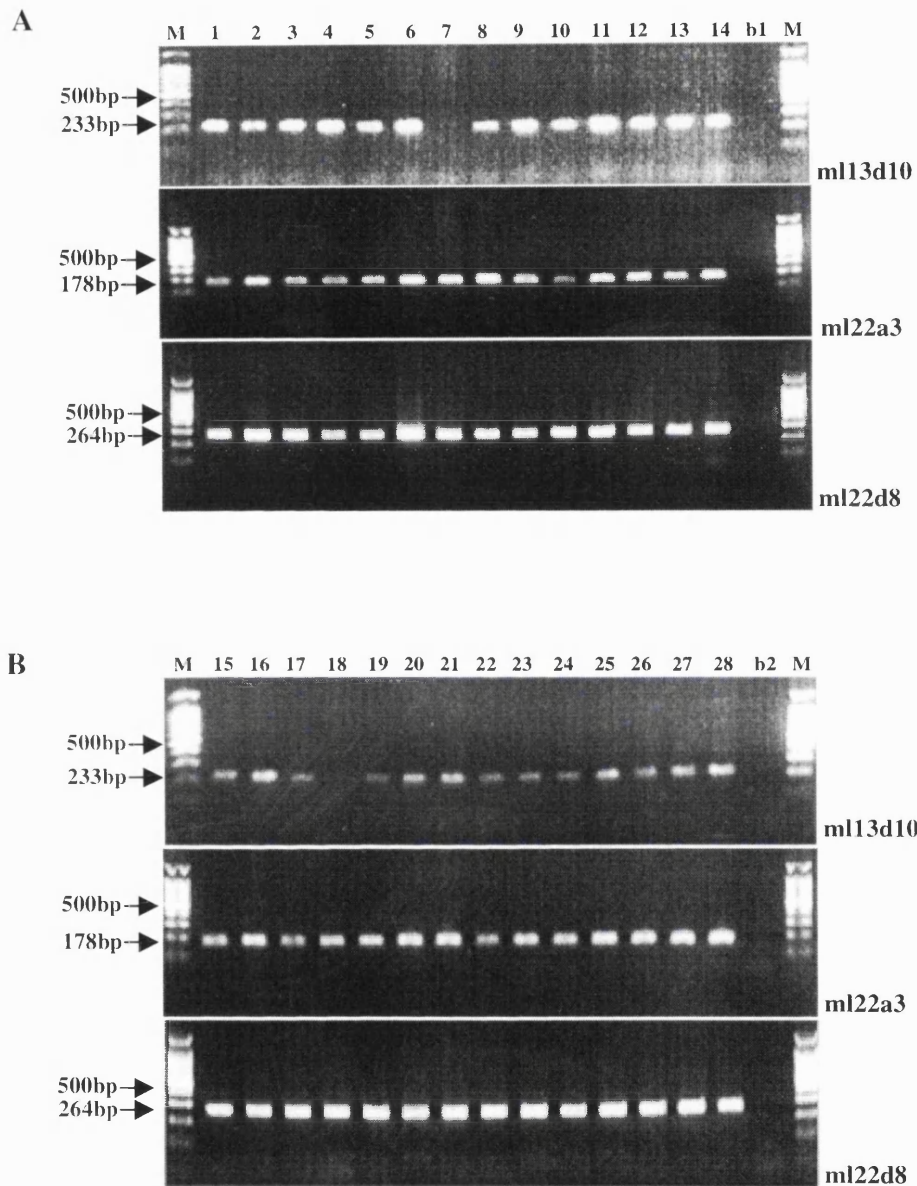
Figure 4.34 Screening of DNA from a panel of oligo- and azoospermic patients with primers specific for ml13d10 and ml22a3. Results for UCH patients 1 to 14 are shown in Fig. A and for UCH patients 15 to 28 in Fig. B. Amplification of the same panel of DNA with primers for ml22d8 (Chrs Y and 15) is also shown as a control; M: DNA size marker; b1-2: no DNA
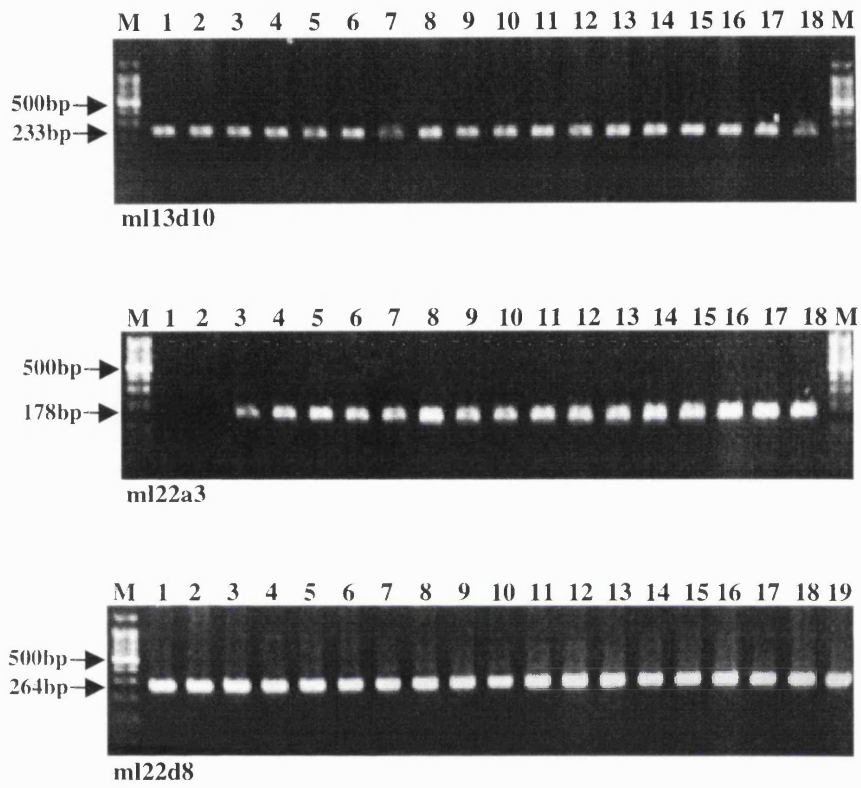
Figure 4.35 Screening of DNA from a panel of oligo and azoospermic patients (Tel Aviv), with primers specific for ml13d10 and ml22a3. Amplification of the same panel of DNA with primers for ml22d8 (Chrs Y and 15) is also shown as a control; M: DNA size marker; b1-2: no DNA

4.10 DISCUSSION

In this section, the TTY2-like genes ml13d10 and ml22a3 and the published TTY2 gene will be discussed further, particularly in the context of how Y chromosome evolutionary process may impact on the characteristics of Y-linked genes like the TTY2 gene family.

4.10.1 TTY2 gene structure

The ancestral TTY2 gene   Mapping studies have shown that members of the TTY2 gene family appear to be Y-specific with no strongly homologous sequences elsewhere in the human and primate genome. However, Southern blot analysis at low stringency (section 4.5) has demonstrated that some bands are common between male and female human DNA, which suggests that there are TTY2-like sequence(s) elsewhere in the genome.

Perhaps this non-Y sequence was the ancestral version, from which via a duplication and transposition event, the first Y-linked TTY2 copy emerged. Failure of FISH analysis to detect an autosomal or X-linked sequence (section 4.5) suggests that the homology between the Y-linked copy and its prototype is low and implies that the transfer of a TTY2 progenitor onto the Y chromosome happened early during mammalian evolution.

The Y-linked RBMY gene seems to have arisen in a similar way. It seems most likely that a single copy of this large genomic family was derived from an autosomal hnRNPG-like ancestor and copied on the X and Y more than 130 million years ago, before the radiation of mammals. This gene accumulated sequence changes during evolution to evolve from an X-Y housekeeping gene, into a male specific gene family that shares very low

197

homology with its X counterpart (Delbridge *et al.*, 1997; 1999, Chai *et al.*, 1998).

Presence of multiple TTY2 copies   The identification of 34 distinct Y-linked TTY2-cosmids and their sequence analysis made it clear that TTY2 sequences are present in many copies on the Y chromosome. A partial sequence analysis revealed that there is a minimum of 26 TTY2-like genes, arranged in 16 subfamilies. Within each subfamily, the level of homology between different copies is high (93-99%), whereas between subfamilies the level of homology drops significantly (55-87%). This indicates that during the evolution of the TTY2 gene family, duplications gave rise to several TTY2 copies that each in turn formed a subfamily. Gene duplications often occur as a result of unequal crossover events and very frequently create a tandemly arranged gene cluster (Smith., 1976).

The gene multiplicity showed by TTY2, is a common characteristic of male specific, Y-linked genes that do not appear to have any close homologues on the X chromosome. Other examples are the RBMY, DAZ and TSPY gene families. This characteristic is associated with the Y chromosome's haploid nature and the evolutionary pathways that led to the formation of the sex chromosomes.

One school of thought is that in the absence of extensive recombination on the Chr Y, gene duplication may serve as a mechanism to produce a reservoir of active copies of the same gene (Nowak *et al.*, 1997, Vogel and Schmidtke., 1998). The resulting gene redundancy would be maintained by selective pressure as such a reserve that would avoid the complete loss of gene activity due to accumulation of mutations. Without selection, it might be assumed that genetic drift would eliminate gene copies by mutations, leading to loss of function.

The loss of recombination between the Y and X chromosomes, is thought to have arisen because early in its evolution, the proto-Y acquired gene(s) that favored male sex determination, whereas absence of this gene(s) led to female differentiation. In stages, alleles of other genes responsible for primary or secondary differentiation of the two sexes evolved and became located near the sex-determining region of the proto-Y (Maxson; 1990). Progressively, a mechanism of selection and protection of these genes resulted in suppression of genetic exchange between the sex chromosomes. This was accompanied by divergence of the proto-X and -Y sequences to include a non-recombining region. Lack of extensive recombination, resulted in the appearance of internal rearrangements on the proto-Y, such as gene duplications, followed by internal structural changes and mutations that were tolerated because of the pressure to maintain sequence homology needed for recombination (Mitchel *et al*; 1998).

The best characterised genes that, like TTY2, are located in the non-recombining portion of the Y chromosome and exist in multiple copies, are DAZ, TSPY and RBMY. Family members include both functional copies (although exact numbers are not clear) and pseudogenes (Saxena *et al*., 1996, Chai *et al*., 1997, Manz *et al*., 1993). Lahn and Page (1997) added a further 7 genes to this list. These are CDY, BPY1, BPY2, XKRY, PRY, TTY1 and the TTY2, all of which are expressed in testis.

Arrangement of TTY2 genes in distinct clusters   In addition to duplications, translocations can occasionally break-up gene clusters and scatter genes, or groups of genes in different places within the genome. This has occurred in the history of the TTY2 genes, which form two clusters, a large group on Yq (AZFc) and an apparently smaller one on Yp. The distance between TTY2-like genes within the cluster (Figs 4.25 and 4.26) varies between 4Kb and 40Kb.

Examples of genes that exhibit a similar organisation pattern include the TSPY and RBM genes, which are tandemly arrayed within clusters in both arms of the Y (Chai *et al.*, 1997, Vogel and Schmidke., 1998, Ratti *et al.*, 2000). Interestingly, the location of the TTY2-like genes coincides with the location of TSPY copies in both arms. It has been suggested that the TSPY/TTY2 containing region on the short arm of the Y chromosome, is particularly prone to chromosomal rearrangements and contains many tandemly repeated sequences (Muller *et al.*, 1989). It seems likely that large genomic segments, which contained both TSPY and TTY2 genes, may have been duplicated and translocated in different regions of the Y.

Divergence of TTY2-like genes   The moderate level of homology between TTY2 subfamilies indicate, either that these genes have diverged to form distinct clusters early during evolution, or that mutations accumulated and fixed in these genes because they are not functional and lack any pressure from natural selection mechanisms. It has been estimated that generally in higher organisms, the rates at which mutations are fixed within the population can range from 10^{-6} to 10^{-4} mutations per gene locus per generation. However, it is difficult to predict an estimate of the time over which the TTY2 family has diversified, since the lack of extensive recombination on the Y chromosome leads to a higher rate of mutation fixation and may rapidly lead to the emergence of non-functional gene copies.

It is expected that many of the TTY2-like genes will be pseudogenes, although, the number of functional copies and of pseudogenes is not known at present. Gene clusters are often characterised by the presence of defective genes, which can contain copies of the exons, introns and promoter sites of the functional genes and represent pseudogenes, whose RNA undergone processing. Such an example includes the α and β-globin gene clusters. They both contain amongst the protein encoding genes, a number of pseudogenes,

such as the *θ1* and *ψα* non-functional genes, which are expressed that is they are detectable as mRNAs, but lack any function and proteins derived from their translation are not assembled into the chain of haemoglobin (Hardison *et al.*, 1986, Clegg., 1987).

Mutations tend to be selected for and fixed in functional genes, at a much lower rate than in pseudogenes, since they may lead to loss of a significant function for the organism. This effect can be seen in the homology between they share functional and non-functional members of the Y-linked RBMY and TSPY gene families. They both have around 40 members and between functional copies the homology is as high as 84-98% for RBMY and 97-99% for TSPY, whereas the level of homology between functional genes and their pseudogene counterparts is lower, 40% for RBMY and 90% for TSPY (Prosser *et al.*, 1996).

Results from the construction of a phylogenetic tree with one member from each TTY2 subfamily as representative (Figs 4.26 and 4.27) resulted in some interestingconclusionsOne of them is that 13-ml13d10 and the published 12-TTY2, mapped to Yp and Yq respectively, appear to be cladistically related, with a bootstrap support of 82%, forming two branches of the same group (Figs 4.26 and 4.27). These two genes share 76% homology. In addition, subfamily 11, represented by 11-2b1 appears to be the closest ancestor of both genes and shares 76% homology with 13-ml13d10 and 67% with 12-TTY2. These data suggest that duplication of a subfamily 11 member, led to the appearance of 13-ml13d10, which in turn duplicated to give rise to 12-TTY2.

In addition, this phylogenetic analysis suggested that a member of the subfamily 10 (2d2), which has been mapped on Yp, is associated with members of subfamilies 1 and 2 (ed9 and ml22a3 respectively), which have been localised on Yq. These two genes are members of the same gene cluster, share 87% homology and the phylogenetic tree has confined them as branches

of the same group, with bootstrap support of 100%. Subfamily 10 (2d2) shares 72.4% and 70% homology with subfamilies 1(ed9) and 2 (ml22a3) respectively. The similar level of homology with both genes suggests that either 10-2d2 is the ancestor of both of them, or that it formed one gene, which in turn duplicated to form the second gene. However, the slightly higher homology levels between these genes suggest that the duplication events may have taken place more recent during the TTY2 evolution. This phylogenetic relation between some of the Yp and Yq mapped TTY2-like, implies that the two TTY2 clusters as a whole, may be the result of a large genomic fragment bearing the whole cluster, or part of it, from one arm of the Y chromosome to the other, rather than translocation and amplification of individual genes. This phylogenetic analysis also suggests that the cluster translocation may have been followed by extra gene duplications within each cluster.

The phylogeny reconstruction of the TTY2 gene family could be a useful tool for providing insights into the evolutionary history of the Y chromosome. However, not all of these genes have been mapped onto Y-linked regions, so there is no strong evidence to support this hypothesis. In addition, since this phylogenetic tree is unrooted, it only represents the branching order, without indicating the root and it is difficult to predict with accuracy the Y-linked origins of each cluster and the order of the following duplication events. Finally, a detailed analysis of this gene family can only take place when extensive sequence for all members of the gene family is available.

An example of a genomic organisation similar to that of the TTY2-like genes, is provided by the vertebrate homeobox (Hox) genes, where four Hox gene clusters have arisen during evolution by means of duplication and translocation of a single ancestral cluster. Within each cluster, the genes can be assigned according to sequence homology, into several different paralogous groups. Phylogenetic analysis suggests that the Hox cluster duplications have

occurred relatively recently in the vertebrate evolution (Prince *et al.*, 1998). Another example is the Olfactory Receptor (OR) genes. These genes belong to the G-protein-coupled receptor superfamily and are encoded by a large multigene family. OR genes appear to be organised in clusters that arose during evolution through multiple duplications of large genomic fragments and are present on almost every chromosome, (Rouquier., 1998). As for TTY2, there are more than one clusters on the same chromosome. On chromosome 3 there are at least three clusters, two on 3p and one on 3q. These are believed to have arisen by cluster duplication and translocation within the same chromosome (Trask *et al.*, 1998, Brand-Arpon *et al.*, 1999).

At present, it is not known whether there are more than two TTY2 gene clusters. Southern blot analysis carried out by Lahn and Page (1997), indicated that copies of the TTY2 gene exist in three chromosomal positions, Yp, at interval 4A and Yq, within intervals 6A-6D as found in this present study and at Yp, interval 3C. Unfortunately, YAC 759G2 (ml13d10 specific), which lay in the region of this third cluster and was examined in the course of this study, is deleted for the 3C (Jones *et al.*, 1994).

TTY2 genes and internal repeats   Study of the TTY2 gene structure revealed the presence of an internal repetitive structure (section 4.8), which is similar for all three genes (TTY2, ml13d10 and ml22a3). It comprises several distinct but overlapping repeat units with sizes ranging between 0.3-2.4Kb that are tandemly arranged across the gene. This structure suggests that the present day, TTY2 genes have emerged from a series of intragenic duplications.

As is the case for TTY2, the Y-linked DAZ and RBMY genes contain repetitive motifs within their sequence, although these are generally small. RBMY contains a tandemly repeated motif of 37bp called SRGY, which may be involved in protein-protein interactions. DAZ also contains around seven tandemly arranged "DAZ specific"repeat units of 72bp that differ from each

other by a few nucleotides and are polymorphic within the population. It has been suggested that the variety in the number of DAZ repeats, may be associated with different sperm counts amongst individuals (Reijo et al., 1995, Yen et al., 1997).

It has recently been proposed (Fatyol et al., 2000) that TTY2 is a repeat element with retroposon-like features, sharing homology with a novel centromeric satellite family described in primates. This theory, is supported by the presence of 0.7Kb direct repeats at both ends of the published TTY2 gene sequence, which are thought to have evolved from a retrotransposon, since they appear to be functionally equivalent to their long terminal repeats (LTRs). Fatyol et al (2000), speculated that the testis-specific expression of TTY2 is analogous to the hypomethylation-dependent transcription of several retroelements, which are expressed in gametes and early embryonic cells (review, Yoder et al., 1997). There are examples of retroelements with elevated, or specific testis expression and these include an X-linked glycerol kinase (GK) gene in humans, the intracisternal A-particle (IAP) sequences in mice and the copia retrotransposon element in Drosophila melanogaster (Sargent et al., 1994, Dupressoir et al., 1995, Dupressoir and Heidmann., 1996, Pasyukova et al., 1997). However, it has not been proved that TTY2 is a retrotransposon. It is true that around 625bp at either end of the TTY2-like genes ml13d10 and ml22a3 show 70% homology. However, the DOTPLOT analysis shown in this present study, demonstrated that similar levels of homology are observed between regions of exon 1 and the rest of the TTY2 exons, as well as between exons 4 and 6 and between exon 2 and 7. This structure is more commonly regarded as the result of intragenic duplication and is not a characteristic feature of a retrotransposon, which is present in the genome as a cDNA copy.

In addition, it is clear that the TTY2 gene itself has an exon/intron structure and is processed and this is not a feature of retroposon-like sequences.

TTY2 exon/intron splicing junctions   Sequence analysis has revealed that TTY2-like sequences are genes with both exons and introns and their size varies from at least 18 to 26.5Kb. The published TTY2 gene structure conforms to the AG-GT rule of exon/intron splicing junctions, but this rule does not hold at all boundaries in the ml13d10 TTY2-like and ml22a3 TTY2-like genes. Could this finding indicate that these two RNAs are not spliced? Unprocessed mRNAs are spliced by the spliceosome, a large complex of small nuclear ribonucleoproteins (snRNPs) and other factors that recognise and bind to the AG-GT splice sites. However, there are several examples of genes including the nucleolar protein P120 and the cartilage matrix protein (CMP) that are excepted from this general rule. These genes appear either to have different splicing junctions, like AT-AC, or other novel and rare splice site motifs, such as GC instead of GT (Wu *et al.*, 1999), or CAC instead of AG (Jackson., 1991).

Interestingly, around 0.1% of all mRNAs have replaced the AG-GT splicing junctions by AT-AC (Mount., 1996). These exon/intron boundaries are recognised and spliced by a spliceosome, in which the U11 and U12 snRNPs are replaced by U1 and U2 snRNPs. There are also cases where both AG-GT and AC-CT splicing junctions co-exist, like in the mouse *pale ear* (*ep*) gene (Feng *et al.*, 1997). Comparative studies have shown that AT-AC introns are conserved in distinct vertebrate species and within gene families, which suggests that they may play a significant role in tissue specific regulation or developmental gene expression (Wu and Krainer., 1996, Feng *et al.*, 1997).

Exon/intron splicing junctions of both ml13d10 and ml22a3 indicated the presence of AG-GT splice sites for intron 3 of ml13d10 and for introns 1 and 2 of ml22a3. The remainders of the introns, in both cases, appear to have either an AT or an AC splice site, together with various other forms, like three Gs or three Ts that have not been described in the literature. These unusual splice junctions, may be explained by sequencing mistakes within the genomic fragments used for the construction of the two gene models, (sequence retrieved from database), but it seems unlikely that those errors could account for all changes.

Alternatively, it should not be excluded that there may be some differences in the exon structures of different members and that the true exon/intron splicing junctions have not been correctly identified. While the splice junctions chosen are those that align with those in the TTY2 gene, there are other potential AG-GT sites, some with features closely related to exon/intron splicing junctions, like a stretch of pyrimidines close to the splice acceptor site, or other particular nucleotides adjacent to the donor/acceptor splicing sites, have been highlighted and are shown in Appendix xx.

However, a third possibility is that these TTY2-like members are pseudogenes. The transition between a functioning gene into a pseudogene is usually gradual and therefore, at early stages, the gene can be no longer functional, but continues to be expressed at the RNA level, such as the θ gene of the α-globin gene family (Clegg., 1987).

Comparative studies-TTY2 genes   FISH analysis of chimpanzee, pigmy chimpanzee and gorilla metaphase spreads, indicates that both ml13d10 and ml22a3 sequences have been Y-linked and Y-specific, since humans diverged from gorillas, 7 million years ago (for review of primate evolution see Goodman *et al.*, 1998, Goodman., 1999).

In addition, the relatively high intensity and widespread distribution of the fluorescent signal, especially for the ml22a3 gene, suggests that primate homologues of human TTY2 genes are present in multiple copies and in several, widely spread positions. The fluorescent signal observed on pigmy Y chromosome was especially intense and allowed the localisation of both genes on the long arm of the Y chromosome, with ml13d10 closer to the centromere. However, the fluorescent signal for the chimpanzee and gorilla was rather diffuse and was not possible to accurately localise the primate homologues.

Archidiacono et al., (1998), examined and compared the structure and evolution of several primate Y chromosomes by FISH analysis, using as probes 17 human Y-specific YACs. These YACs, belonged to the same panel of YACs, members of which were used in this present study for mapping of the ml13d10 and ml22a3 TTY2-like genes. YAC 821G7 includes the ml13d10 sequence (Yp11.2, intervals 3G-4A). Reviewing Archidiacono et als data, (1998) shows that in chimpanzees a ml13d10 homologous region lies at Yq12.1-12.2 and in gorillas at Yp11.2-11.1. Thus, ml13d10 is located in both humans and gorillas on the short arm of the Y, whereas in chimpanzees and pigmy-chimpanzees it is present on the long arm. These data suggest that after the divergence of the human and chimpanzee lineages, around 3-4 Mya, an internal rearrangement, such as a translocation and reinsertion, may have occurred in the chimpanzee/pigmy-chimpanzee ancestor, which transferred a genomic fragment from Yp to Yq.

These data imply that the gorilla Y chromosome structure is closer to humans than chimpanzees and in the context, it is of interest to note that the Y linked banding patterns of human and gorilla show more resemblance than those of chimpanzees (Pearson et al., 1971). In addition, some genes are present on the Y chromosome of orangutan, gorilla and human, but not chimpanzee, for example the GMGXY12 locus (Lambson et al., 1992). These

findings suggest that the Y chromosome of chimpanzee has undergone a greater degree of rearrangement, since its divergence from the human lineage.

Despite the high degree of genetic similarity (~99%) between humans and hominoid apes, their Y chromosomes show significant variability, with different and complex fluorescent banding patterns (Miller., 1977), but similar gene content (Schemmp *et al.*, 1995). These differences account for most of the different size of primate Y chromosomes (Yunis and Prakash., 1982, Ried *et al.*, 1993). More specifically, the Y chromosome of gorilla is longer and more metacentric than that of humans (Egozcue *et al.*, 1973), whereas those of chimpanzee and pigmy-chimpanzee are much smaller than the human (Darlington and Haque., 1955), with a very small short arm. Examples of Y chromosome rearrangements during evolution have been demonstrated by FISH comparative studies for other Y-linked genes like the RBM and TSPY (Schempp *et al.*, 1995), which show that although these genes remain Y-specific, the number of copies and their location(s) may vary amongst the different primate species.

4.10.2 TTY2 genes function

ml22a3 expression levels   Random sequencing of clones from the cDNA selection library identified 19 TTY2-like cDNA sequences, of which the majority, (16) were ml22a3 (section 4.1). This suggests that ml22a3 mRNA is far more abundant in the testis mRNA pool than mRNAs from other members of the TTY2 family. While not carried out in a quantitative fashion, it was also observed that the RT-PCR products from both foetal and adult tissue cDNAs (section 4.4), were more intense for ml22a3 than for ml13d10 (Fig 4.17) and also suggested elevated expression levels for ml22a3 in a range of cell types. This might suggest that ml22a3 encodes a functional product,

whereas the remainder of the TTY2-like cDNAs may be pseudogenes, at an intermediate stage of degradation.

There are several cases of mRNAs transcribed from non-functional pseudogenes being both transcribed and processed. Typical examples on the Y chromosome include the X-Y homologous genes ARSE/D, STS and XGPY (Yen *et al.*, 1988, Ellis *et al.*, 1994, Meroni *et al.*, 1996), which appear to be expressed, although their genomic structure favors their classification as pseudogenes. It has been proposed that such genes have two functional X-linked homologues, both of which escape X inactivation in order to compensate for loss of the functional Y-linked copy (section 2.4, Table 2.2). Some members of the DAZ, TSPY and RBMY gene families, are pseudogenes at various stages of deterioration that appear nevertheless to be transcribed (Glaser *et al.*, 1998, Prosser *et al.*, 1996, Vogt *et al.*, 1997). This is not a feature solely of the Y chromosome; there are several examples of expressed pseudogenes with autosomal locations and some of these include the human fertilin-α gene, two testis specific genes, *TPTE* and PGK and members of the olfactory receptor gene family (McCarrey and Thomas., 1987, Crowe *et al.*, 1996, Jury *et al.*, 1997, Chen *et al.*, 1999). In general, most of these copies have been characterised as pseudogenes, either because they are intronless, because they lack an ATG start codon, or because of small insertions, deletions and termination codons that destroy the expected reading frame. Thus, it is not clear whether the pseudogenes with introns have exon/intron splicing junctions that obey the AG-GT rule.

Interestingly, ml22a3 is located within one of the functionally active AZF regions (AZFc), where the functional members of the DAZ, RBMY and TSPY gene families are located (review, Cooke and Elliot., 1997, Chai *et al.*, 1997, Ratti *et al.*, 2000). There may be a selective pressure to retain functional copies of genes in this region. Two other TTY2-like genes, ml18g1 and ml22h10, occur within the same AZFc cluster as ml22a3 (section 4.7, Fig

4.25) but at present, there are no data regarding their cell/tissue mRNA expression or structure.

Analysis of RT-PCR expression studies   Lahn and Page (1997) characterised TTY2 expression as testis specific, on the basis of a Northern blot analysis of RNA from adult spleen, thymus, prostate, testis, ovary, intestine, colon and leukocyte. In contrast, the RT-PCR expression studies of ml13d10 and ml22a3 TTY2-like genes suggest that these genes are more widely expressed in both adult and foetal tissues.

The RT-PCR studies used mainly RNA from foetal tissues, with the exception of testis, prostate and kidney, while Northern blot analysis was of only adult tissues (Lahn and Page., 1997) and did not include kidney, which was found to express high levels of both ml22a3 and ml13d10. As a result, a direct comparison of the two methods cannot be made and the results may be explained if we consider different levels of expression in different developmental stages.

It should also be noted that RT-PCR is a very sensitive technique, known to detect even low levels of mRNA and even extremely low levels ascribed to "illegitimate transcription" by Gala *et al* (1998). It is not clear that TTY2 would be entirely confined to the testis. There is no information about expression in development or in a broad range of different adult tissues. It is also possible that the promoters of the three TTY2-like genes (ml13d10, ml22a3 and TTY2) have diversified and now transcribe each TTY2-like gene in a different pattern of expression.

There are several examples in the literature, where illegitimate transcription has been described. For example the proacrosin and protamine 2 genes are specifically transcribed in spermatozoa during spermatogenesis, but transcripts are also detected at low levels in blood lymphocytes by RT-PCR (Slomski *et al.*, 1991). Similarly, cystic fibrosis conductance regulator gene (CFTR), a

gene expressed abundantly in the lung and intestine has low levels of mRNA in lymphoid cells that are detectable by RT-PCR (Fonknechten et al., 1992).

However, the high expression levels of ml13d10 and ml22a3 in lung and kidney cannot be explained as a result of ectopic transcription and strongly suggest that not all members of the TTY2 gene family are testis specific. It is interesting to speculate that the adult kidneys high levels of expression may be connected with testis expression, since both the reproductive and urinary systems in males arise from the nephros and are developmentally and phylogenetically connected (Carlson., 1994). The testis together with the epididymis, vas deferens and seminal vesicles, arise from the primitive gonad on the medial surface of the embryonic mesonephros, which becomes the primitive kidney. Migration of cells from the mesonephros brings at least three somatic cell types into the male gonad; peritubular myoid cells, endothelial cell and cells associated with the endothelium (Martineau et al., 1997). An explanation of this possibility could be made in the future using in situ mRNA hybridisation.

There are examples of Y specific genes that demonstrate ubiquitous expression and SRY is one of these. This gene, despite being essential to testis determination, appears to be present in several non-testis, embryonic and adult tissues (Clepet et al., 1993).

The expression sites of ml13d10 and ml22a3 have not yet been identified. However, both are expressed in adult and foetal testis with similar intensity. In future work, RNA in situ hybridisation into sections of testis and other tissues at different developmental can be used to localise the sites of TTY2-like gene expression.

There are several genes that are thought to play a significant role in sex determination and which are expressed in both testis and kidney (section 1.3, Table 1.4). One of them is WT-1, a zinc finger transcription factor, which is associated with genitourinary malformations and with an important role in

211

genitourinary development. WT- is expressed in the embryonal mesonephric kidney and in the metanephric permanent kidneys. In kidney, WT-1 is expressed in the glomerular podocytes, metanephric blastema and glomeruloid structures and adjacent cells.

In testis, WT1 is expressed during early development and after birth. Sites of expression include Sertoli cells and their precursors, the tunica albuginea and seminiferous tubules united that join the epididymis (Pritsard-Jones., 1990, Mundlos *et al.*, 1993, review, Little *et al.*, 1999).

The SRY gene is expressed in many tissues including adult and foetal testis and kidney (Harry *et al.*, 1995). Experiments with XXY transgenic mice that lacked the *Sry* gene and XX mice that had the *Sry* gene, showed that the Sry protein is critical for initiating mesonephric cell migration, since mesonephros derived cells are detectable in the gonads (Capel *et al.*, 1999).

TTY2 genes as non-coding RNAs   What can be said about the functional role of the TTY2 gene products? One possibility, which should be considered, is that these genes might exist as functional, non-coding RNAs (Table 4.8).

Functional non-coding RNAs either lack any defined ORFs, or have very small ORFs that are not conserved between species. They produce transcripts that lack any protein coding capacity and as a result, appear to be immune to frameshift or non-sense mutations (review, Eddy 1999). Their sizes can range from 0.9Kb to 40Kb, some present as multiple redundant copies and most lack introns and are located in the nucleus as non-processed RNAs (Table 4.8). However, a small proportion of these genes produce large, spliced

Table 4.8 Examples of non-coding functional RNAs

Symbol	Name	Structure	expression	copies	Internal repeats	Refs.
XIST	X inactivation-specific transcript	Intronless	During X inactivation in all cells; spermatogenesis	Single	Tandem repeats of 43 and 300bp	Brockdorff *et al.* 1992, Brown *et al.* 1992, Salido *et al.*, 1992
Tsix	X inactivation-specific transcript-antisense	------	Before X inactivation	Single	None	Lee *et al.* (1999) Heard *et al.*, 1999
X1sirt / Xenopus		------	oocytes	Tandem multiple copies	3-13 interspersed repeat units, each ~80bp long	Kloc *et al.*, 1993 Klock and Etkin., 1994
H19		5 exons, separated by 4 unusually small introns	during embryonic development and adult skeletal and cardiac muscle	Single	None	Leibovitch *et al.*, 1991 Hao *et al.*, 1993 Pfeifer *et al.*, 1996
IPW	imprinted gene in the Prader-Willi syndrome region	intronless	ubiquitous	Single	Multiple copies of a 147bp repeat	Wevrick *et al.*, 1994
HIS-1		Spliced and polyadenylated	myeloid leukemia cells; low levels in epithelial cells, stomach, prostate, seminal vesicle	Single	None	Askew *et al.* (1991)

213

Symbol	Name	Structure	expression	copies	repeats	Refs.
NTT	noncoding transcript in t cells	intronless	Activated T cells	Single	Contains Alu and MER repeats; also a 74bp repeated fragment	Liu *et al.*, 1997
G90		intronless	Intestine, testis and kidney	single	------	Krause *et al.*, 1999
DGCR5	Digeorge syndrome critical region gene	spliced with introns	Ubiquitous	Series of alternatively spliced RNAs from single gene	------	Sutherland *et al.*, 1996
CMPD	Campomelic Dysplasia-associated RNA	------	Adult testis	single	------	Ninomiya *et al.*, 1996
Bsr Rat		intronless	CNS at late stages of differentiation	Multiple repetitive units with 87-100% homology that spann 150Kb	Tandem repeats of 100/1500 copies; cover over 50Kb	Komine *et al.*, 1999)

and polyadenylated RNAs, with a cytoplasmic location, for example the H19 gene.

In some cases, functional roles have been proposed (reviews, Erdmann *et al.*, 1999, Erdmann *et al.*, 2000) (Table 4.8). One of these is that some RNAs function as regulators of other genes, perhaps by either blocking, or help initiating their translation. For example, T*six* mRNA is thought to regulate the expression of X*ist* and events at the onset of X-inactivation, such as X chromosome counting and/or choise of the active X chromosome (Lee *et al.*, 1999).

In addition, a number of other functions have been proposed for several other RNA transcripts. The product of the H19 gene is thought to act as tumor suppressor, since tumor cell lines transfected with H19 protein, showed growth retardation and morphological changes (Hao *et al.*, 1993). The imprinted gene in Prader-Willi syndrome (IPW), has been isolated using a direct selection approach, it is expressed in several adult and foetal tissues from the paternal allele and is thought to play a role in the imprinting process (Wevrick *et al.*, 1994). Finally, the X1sirts, a family of repeat transcripts in the Xenopus genome that are homologous to Xist gene are thought to be structural components of the Xenopus vegetal cortex and help to localise the expression of some other mRNAs in this region (Kloc *et al.*, 1993, Klock and Etkin., 1994).

These studies showed that the published TTY2 gene has exon/intron splicing junctions that obey the GT-AG rule and the isolation of the TTY2 cDNA (Lahn and Page., 1997) shows that at least the product of this particular TTY2 gene, if it is polyadenylated, may be located in the cytoplasm of cells. The lack of any large, distinct ORF indicates that this gene may function at the RNA level and does not encode a protein product. ml13d10 and ml22a3 TTY2-like genes also lack any distinctive ORFs. However, their exon/intron structure was obtained after their genomic sequence was compared with the

TTY2 identified exons (section 4.6) and exon/intron boundaries may not be correct (Discussion, exon/intron splicing junctions). As a result, it cannot be certified at present that these two TTY2-like genes exist as processed mRNAs, or simply as pseudogenes at some stage of degradation that happen to be expressed. An experiment that would produce evidence of splicing of ml13d10 and ml22a3, would be RT-PCR amplification across their cDNA, from one exon to the other. In addition, evidence for polyadenylation could be produced by RT-PCR amplification using an unchored poly(A) primer and a primer specific for the ml13d10 and ml22a3 3' UTR region. A similar procedure is used in the differential display technique (DDRT) and the specific TTY2 primer would increase the stringency of the reaction.

Some non-coding RNAs, contain repeated elements, in their structure. Examples are the X inactivation-specific transcript XIST (Salido *et al.*, 1992), the paternally imprinted gene in Prader-Willi syndrome IPW (Wevrick *et al.*, 1994), a transcript present in T cells called NTT (non-coding transcript in T cells) (Liu *et al.*, 1997) and a gene expressed in the brain of rats, called Bsr (Komine *et al.*, 1999) (Table 4.8).

Around 50% of the transcripts described in Table 4.8, appear to have a repetitive pattern within their structure, as the TTY2 genes have (section 4.8). Their repetitive structure may be related with their function. It is believed that some non-coding RNAs regulate the expression of other genes and such repetitive units may help the non-coding RNA to adopt a conformation that facilitates its interaction with the gene's promoter site. However, the repetitive unit of the TTY2 genes corresponds to a genomic structure that includes both exons and introns (Figs 4.29, 4.31 and 4.33), whereas RNAs with internal repeats are either intronless, or their structure is not yet known. Because of the different structure of the repeat units present at the TTY2 genes and at the non-coding RNAs a direct comparison cannot be made.

4.12 TTY2 deletions in azoospermic patients

Results from screening a panel of 82 azoo- and oligozoospermic patients with specific primers designed within exon 1 of ml13d10 and ml22a3, showed that the TTY2-like corresponding region was absent in some patients. The absence of ml22a3 from patients no. 1 and no.2 (Tel Aviv) was expected, since both of them lacked the Yq region (AZFc), where ml22a3 has been mapped. It has been suggested that deletions and microdeletions in the AZF regions at Yq11, appear to occur frequently as de novo mutation events in men with idiopathic azoo- or oligozoospermia. Various labs have estimated that 3-18% of infertile patients have deletions in the AZF regions (a, b and c) and the genes located within (Kobayashi *et al.*, 1994, Rejio *et al.*, 1995, Vogt *et al.*, 1996).

As a result it is believed that genes located within these three regions play an important role in the normal process of spermatogenesis and are designated as "AZF candidate genes" (Vogt., 1998). The best investigated genes are the RBMY (AZFb), TSPY and DAZ (AZFc) gene families. All these three genes are expressed in testis and their function is related to the germ cell proliferation (TSPY, Schnieders *et al.*, 1996), nuclear metabolism of testis mRNA (RBMY, Elliot *et al.*, 1997) and normal function of spermatids and mature sperm (DAZ, Haberman *et al.*, 1998). ml22a3 is also located within an AZFc cluster and although the screening studies were inconclusive, may well represent another gene that could be associated with male infertility.

ml13d10 appeared to be absent from the DNA of two azoospermic patients (UCH, Cytogenetics unit) (2.4%). ml13d10 has been mapped in Yp, within interval 4A. To date, there is no report that associates deletions within this region with infertility, although genes associated with spermatogenesis, like TSPY members, have been mapped within interval 4A. Both patients have primary idiopathic azoospermia. However, further investigation of the

region where ml13d10 is located, would point out whether the infertile phenotype is related to the ml13d10 TTY2-like gene, or is symptomatic and appears due to deletions of other genes located within this region.

CHAPTER 5

ml22d8 like sequences

5.1 Description of ml22d8, ml1g1 and ml22c6 cDNA clones

In this chapter, three further cDNAs, ml22d8, ml1g1 and ml22c6, isolated from the cDNA selection library will be described. Sequence analysis showed that these clones shared no homology. However, database BLAST search showed that all three, showed homology with closely spaced regions of the same genomic clones and furthermore two of them, ml22d8 and ml1g1 matched with 97-99% homology ESTs derived from germ cells and germ cell tumor cDNA libraries.

Further evidence that these three cDNAs were close together in the genome came when they were individually used to screen the panel of 500 Y-cosmids. The same two cosmids Yco2C4 and YcoDG11 hybridised to all three cDNAs. ml22d8 additionally, identified a third cosmid, YcoGE10. The presence of ml22d8 sequence in these cosmids was confirmed by PCR amplification using primers 22d8.F and 22d8.R, specific for ml22d8 (listed in Table 2.1). In all three cases a single PCR product of correct size (264bp) was amplified (Fig 5.2A). Digestion of maxi-prep cosmid DNA using 6 different restriction enzymes showed that Yco2C4 and YcoDG11 overlapped considerably in sequence, whereas the restriction pattern of YcoE10 was significantly different. These various data raised the question; are ml22d8, ml1g1 and ml22c6 different exons of the same gene?

A BLAST search of the cDNA sequences using the NCBI dbEST database, showed that ml22c6 sequence failed to identify any EST matches. In contrast, ml22d8 and ml1g1 matched with 97-100% homology several ESTs of unknown chromosomal origin. These ESTs are partial sequences of cDNA clones from several types of libraries, including a pooled germ cell tumor library and a compound library containing a mixture of 21 normalised

libraries, but reselected for germ cell expressed sequences. Perhaps significantly, none of the ESTs were the same for each cDNA. Besides these EST matches, ml1g1 identified two ESTs expressed in adult brain but with lesser homology (90%). In the case of ml22d8, various matching ESTs overlapped, but the boundary extended only by 30bp at the 5' end of the known sequence, whereas the boundary of the ml1g1 EST contig, extended by 191bp the 5' end and by 352bp at the 3' end, giving a total of 862bp (Fig 5.1).

If ml22d8 and ml1g1 were adjacent exons, then cDNA sequence from both should identify the same contig of overlapping ESTs. However, this was not the case and furthermore, comparison of the ml22d8 sequence with the extended ml1g1 sequence, failed to identify any homology. These observations indicate that these cDNAs may be short distinct expressed sequences.

ml22d8 cDNA was one of the earliest cDNAs to be sequenced from the cDNA selection library and was subject to further analysis.

5.2 ml22d8 chromosomal location

Localisation to the Y chromosome using cell hybrids   In order to define the chromosomal localisation of ml22d8, PCR amplification using 22d8.F and 22d8.R primers, was carried out on DNA from human male and female DNA and two somatic cell hybrids, 3E7 and 853 that contain the Y chromosome as their only human component.

Products of the correct size (264bp) were obtained from male DNA and from the Y containing somatic cell hybrids, indicating that the ml22d8 sequence is Y-linked. In addition, primers amplified a product from female DNA, suggesting the presence of a ml22d8 homologue either on the X chromosome or an autosome (Fig 5.2B).
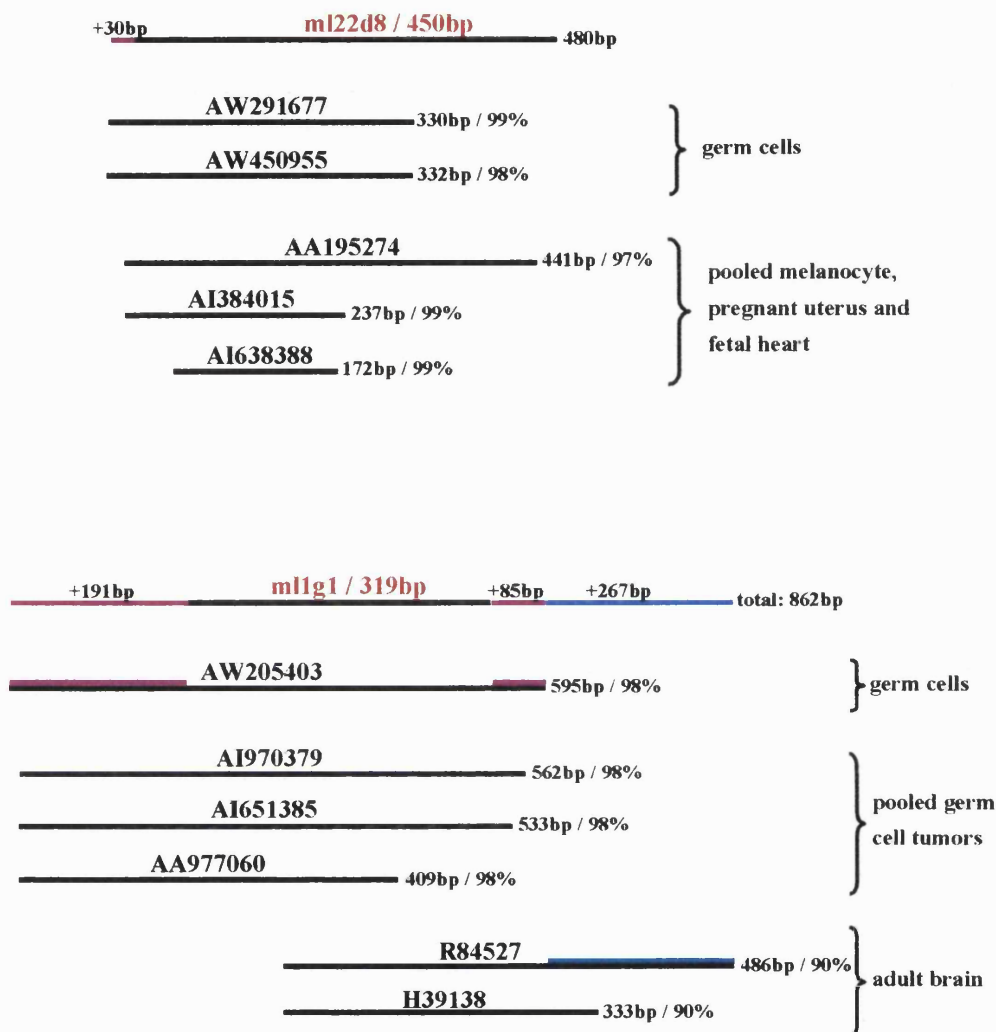
Figure 5.1 Contigs of ESTs around the ml22d8 and ml1g1 regions. Pink and blue lines indicate how the EST contig extended the cDNA sequence. The EST accession number, sizes in bp and % homology with either ml22d8 or ml1g1 cDNAs are indicated. Bracketed ESTs have originated from the same cDNA library and the tissue of origin is shown.
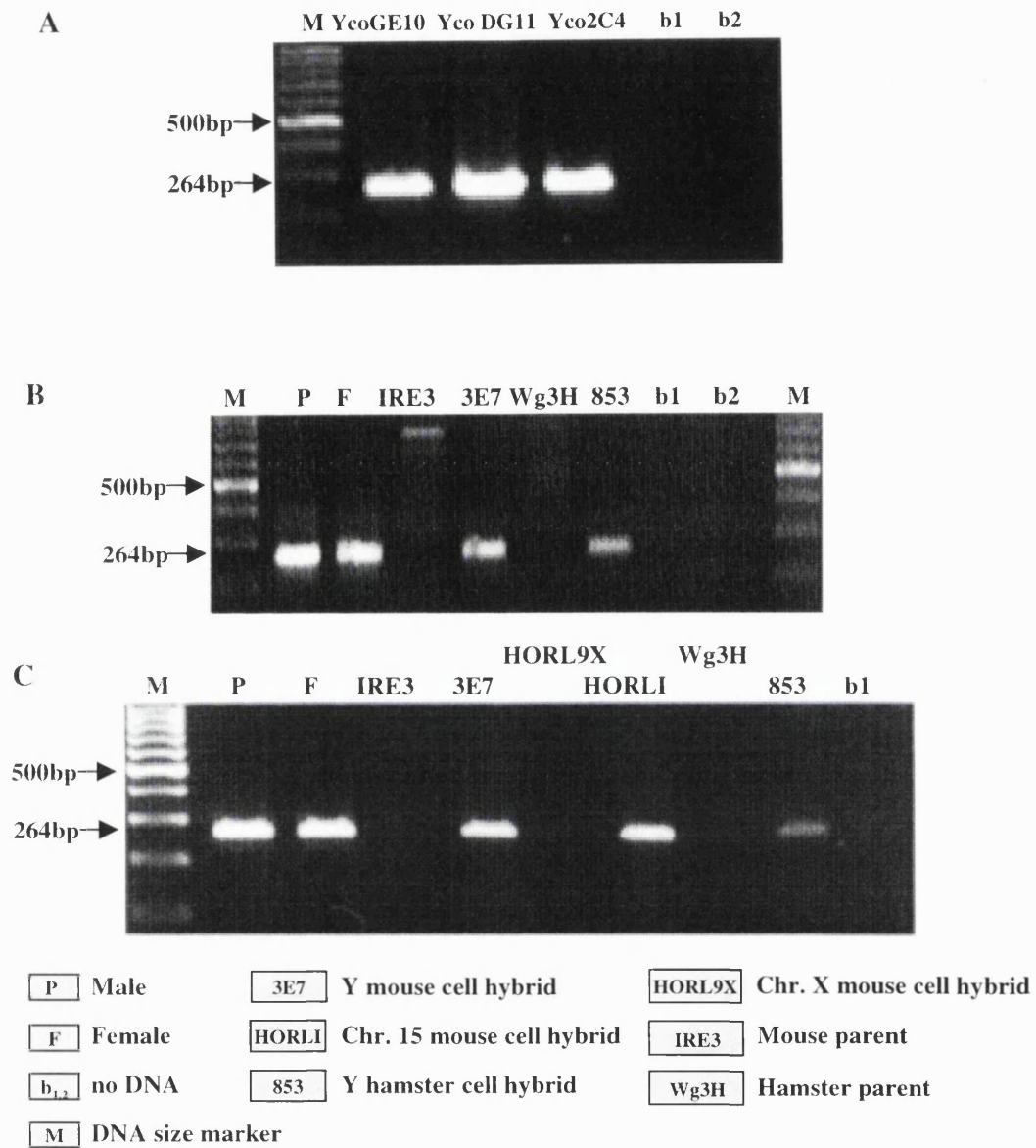
Figure 5.2 A. PCR amplification of ml22d8 sequence from YcoGE10, YcoDG11 and Yco2C4; B. PCR amplification of ml22d8 from human male and female DNA, human Y-containing somatic cell hybrids and rodent parents; C. PCR amplification of ml22d8 from somatic cell hybrids containing human Chrs Y and 15

**Mapping by fluorescent in situ hybridisation (FISH)**   FISH analyses

of human metaphase chromosomes, using the cosmids YcoDG11, Yco2C4

and YcoGE10 as probes, showed that all three cosmids hybridised to a

location on the long arm of the Chr Y and in addition each hybridised to two

locations on Chr 15, giving a strong signal at 15q23-25 and a weaker one at

15q11-q13 (Figs 5.3 and 5.4 A, B and D). 4 metaphase spreads were scored

using Yco2C4 and YcoDG11 as probes and 7 with YcoGE10. YcoGE10

showed a general high level of background and also appeared to give a weak

signal at the tip of the short arm of the X chromosome (Fig 5.4 C).

To explore these observations further, DNA from somatic cell hybrids

HORLI, HORL9X, 3E7 and 853 containing human chromosomes 15, X and Y

as their sole human chromosomal components, were PCR amplified using the

ml22d8 specific primers. The correct sized product was amplified from both

Chrs 15 and Y containing hybrids (HORLI, 3E7 and 853 respectively),

indicating that the ml22d8-like sequences represent a genome duplication

event. No product from the X chromosome hybrid (HORL9X) was detected,

confirming that these sequences do not occur on the X Chr (Fig 5.2 B).

Since the ml22d8 primers amplify sequence present in the YcoGE10 (Fig

5.2 A), but not amplify from the X chromosome, it can be assumed either that

the X-linked sequences detected with YcoGE10 are only weakly homologous

to ml22d8, or that YcoGE10 contains additional sequence, that is not ml22d8

related.

The ml22d8-like sequences on Chrs 15 and Y, were compared by

sequence analysis of the 264bp PCR products from the Chr Y and 15-

containing hybrids. The sequences obtained are compared in Fig 5.5 and show

approximately 96% homology across a region of approximately 230bp. It

seems likely that the Chr 15 sequence corresponds to the strong signals at

A.
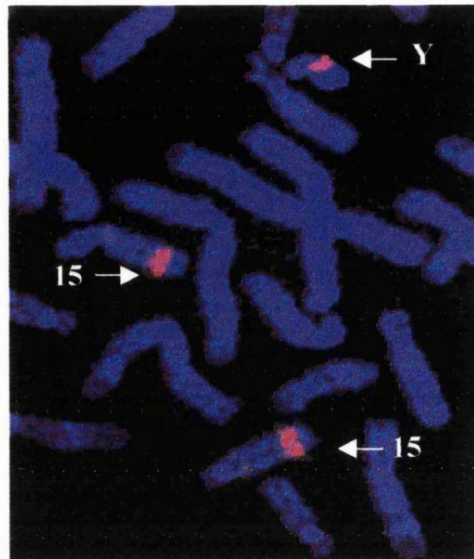
Yq, 15q: Yco2C4

B.
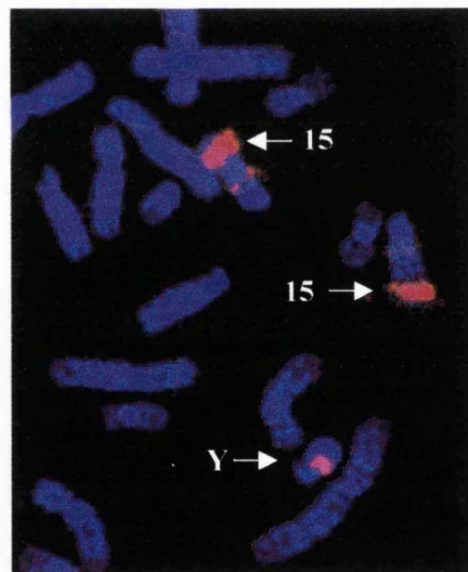
Yq, 15q: YcoDG11

Figure 5.3 FISH analysis using Yco2C4 (A) and YcoDG11 (B) DNA as probe on human metaphase chromosomes. Signal is seen in both Chrs Y and 15

224

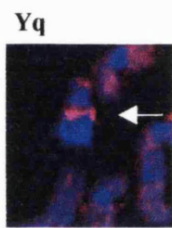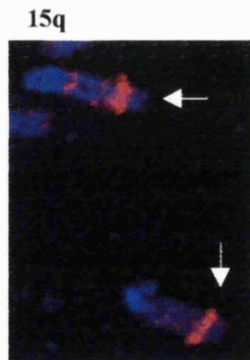**Figure 5.4** FISH analysis using A. YcoGE10 as probe. B, C and D are the same metaphase spread, seen at higher magnification
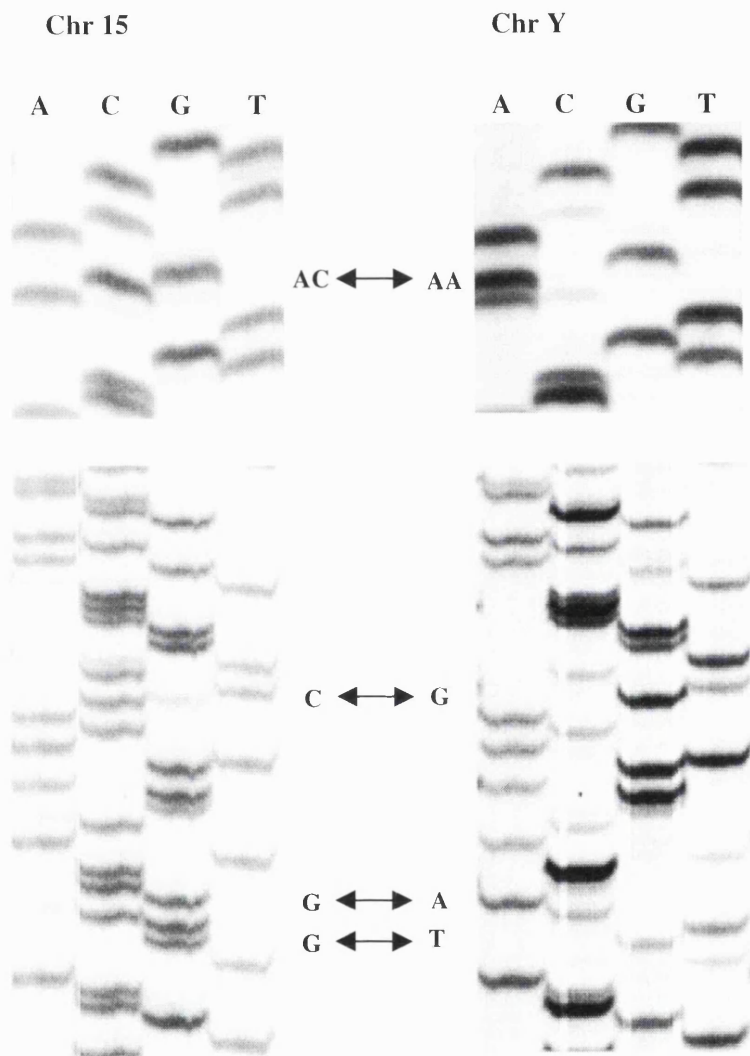
Figure 5.5 Sequence analysis of ml22d8 sequences amplified from
Chrs 15 and Y. Examples of differences are indicated by arrows.

15q23-q25 but it is not possible to be entirely certain it does not represent the weaker 15q11-q13 signal.

Fine mapping of Y-linked ml22d8, using a Y-specific YAC contig   In order to map the Y-linked ml22d8 more accurately, the YACs tiling pathway across the Y chromosome were investigated (see chapter 4, section 4.3.3) by PCR amplification using the ml22d8 specific primers. A strong PCR product was amplified from YAC 913B1 and in addition, a faint product was obtained from YAC 951B1(Fig 5.6 A). To confirm that the correct sequence had been amplified, the PCR products from the YAC PCRs were gel extracted and sequenced. The sequence of the PCR product from YAC 913B1 corresponded 100% to that of ml22d8 cDNA. However, PCR product from YAC 951B1 was much more faint and the DNA extracted was not sufficient to obtain good quality sequencing.

YAC 913B1, like YAC 933A6, is located within the Yq11 region. However, the ml22d8 sequence is not present on the overlapping YAC 933A6, indicating that ml22d8 is localised in the most distal part of Yq11.23, within part of subinterval 6F and the adjacent section of interval 7, which corresponds to the quinacrine-bright heterochromatic distal part of Yq (Vergnaud *et al*; 1986) and lies within region enclosed by the STS markers GY28 and sY159. YAC clone 951B1, which gave a weak ml22d8 signal, covers part of subinterval 5I directly adjacent to an X-Y homologous region and extends up to 5K, containing the STS markers OX3 and GY17 (Fig 5.6 B). However, a PCR product was not amplified from YAC 802D9, which overlaps with part of the clone 951B1, which suggests that the second ml22d8 homologue is located within the genomic region that covers part of 5J, and 5K.
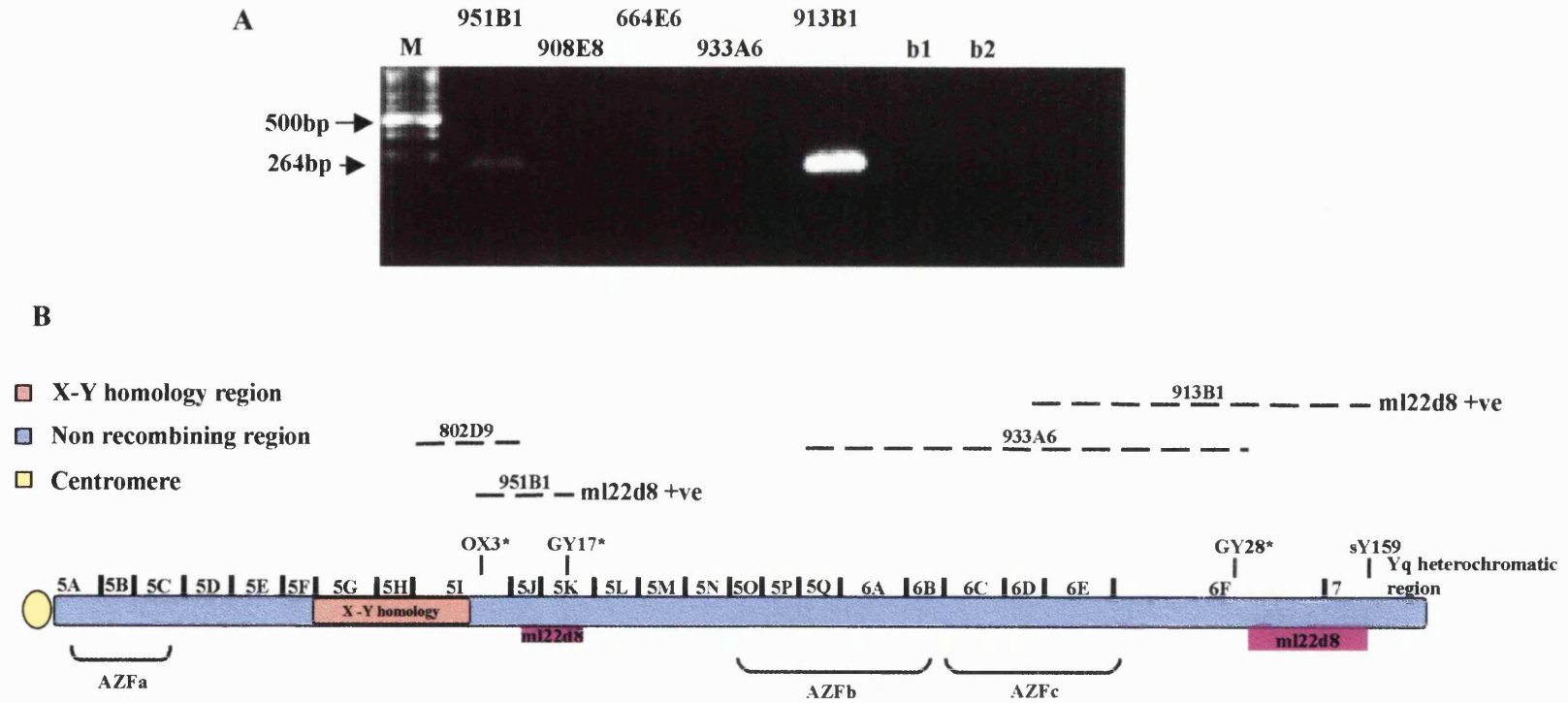
Figure 5.6 A. PCR amplification of Y-mapped YAC clones spanning the Y chromosome euchromatic region and part of the Yq heterochromatic region with 22d8.F and 22d8.R primers. B. Schematic representation of two possible localisations for ml22d8 on the Y chromosome. Dashed lines indicate position of YACs; STSs with an asterisk indicate a multiple copy STS. Adapted from Jones et al; 1994

5.3   ml22d8 expression studies

The expression pattern of ml22d8 was examined by RT-PCR, using the same set of male sex determined adult and foetal tissues, used for the expression studies of the TTY2-like genes ml13d10 and ml22a3 (see chapter 4, section 4.4). RT-PCR amplification using 22d8.F and 22d8.R primers, showed that ml22d8 is ubiquitously expressed at moderate to high levels in all tissues. Adult kidney and foetal lung showed the highest levels of expression, while foetal kidney, heart and intestine showed the lowest (Fig 5.7). These results confirmed the expression of ml22d8 in testis, both in the adult and during development as well as several other tissues, demonstrating that testis is not the major site of expression. Sequencing of RT-PCR products from adult testis, adult kidney and foetal lung showed that the sequence obtained was identical for all three tissues, confirming that they have derived from amplification of the same gene product.

5.4   ml22d8 comparative studies

The presence of sequences homologous to ml22d8 in other species, was investigated using a "zoo blot" containing DNA from human male and female, a number of primates and mouse. The DNAs were digested with PstI, electrophoresed and blotted onto nitrocellulose membrane. The blot was hybridised under low stringency conditions (57° C, 0.2xSSC) with ^{32}P labelled PCR products from ml22d8 (264bp).

This analysis showed similar banding patterns in human male and female DNA and gorilla with moderately strong bands of 0.9Kb, 3.2Kb, 3.5Kb, 5.5Kb and 6.5Kb shared in common (Fig 5.8). The chimpanzee also shared the 0.9Kb band, but other bands of 3.8Kb, 6.8Kb and 20Kb had no counterpart in human DNA; the 6.8Kb and 20Kb band were however also present in the chimpanzee. A band of 6Kb seen in man might represent a Y
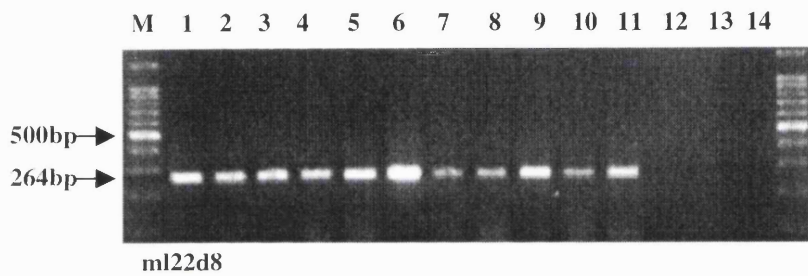
Figure 5.7 PCR amplification of human cDNAs, using primers specific for ml22d8.
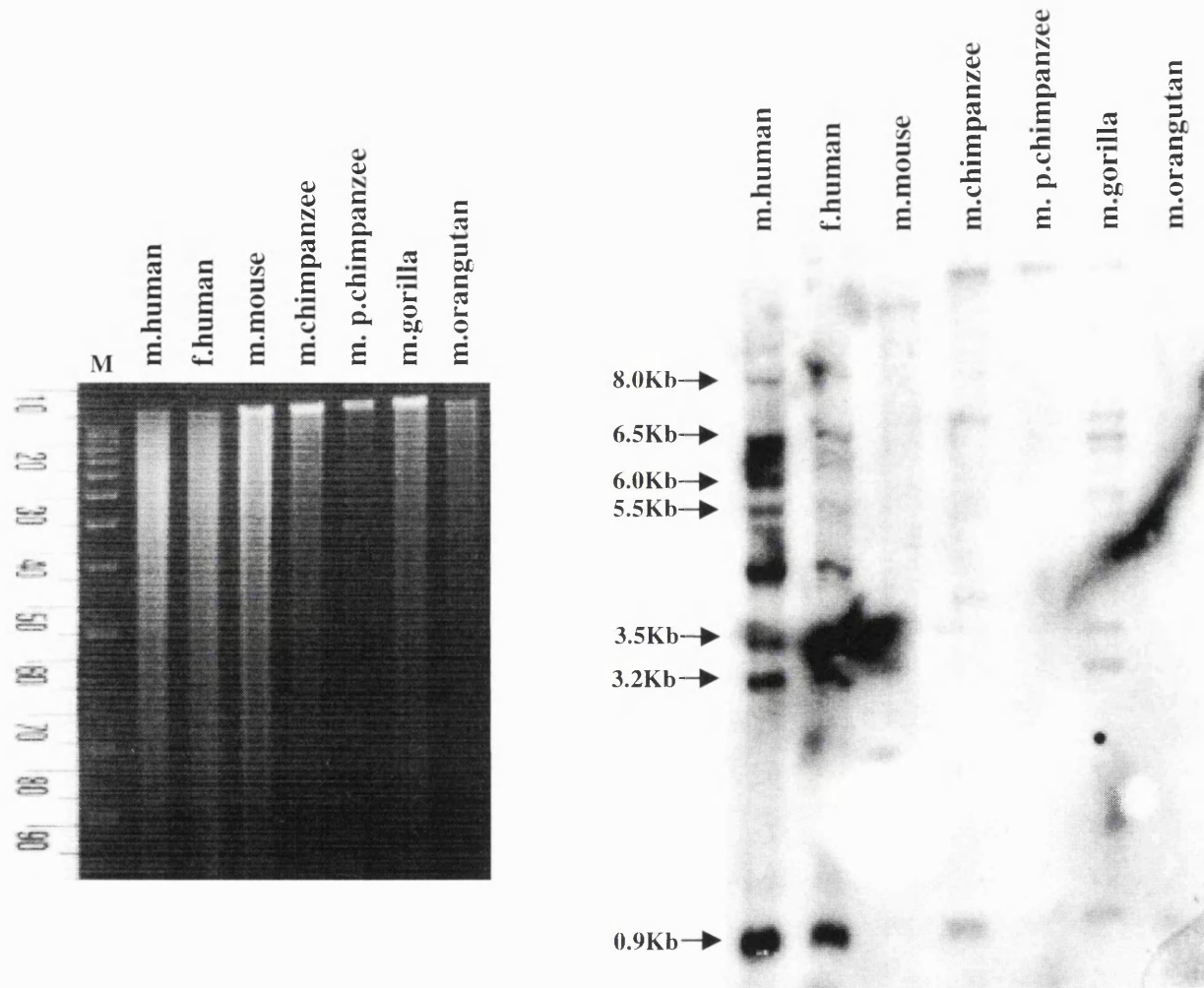
Figure 5.8 Restriction digest of genomic DNAs with PstI enzyme and "zoo blot" hybridisation using ml22d8 sequence as probe;m: male; f: female; M: DNA size marker

specific band. A large number of bands ranging in size from 0.9Kb to 8Kb are detected by the ml22d8 cDNA. It seems unlikely that these represent a single gene and suggest that the ml22d8 sequence occurs in multiple copies.

In order to define the chromosome location of ml22d8 -like sequences in other species and to compare them with the human chromosomal locations, FISH analysis (in collaboration with Dr. Margaret Fox) onto metaphase spreads from male chimpanzee, pigmy chimpanzee and gorilla was performed, using as probe YcoDG11.

Fluorescent labelling identified an homologous sequence on the long arm of an autosome for all three primates. On the gorilla autosome, like in humans, YcoDG11 gives two signals, a faint one closer to the centromere and a stronger signal which is located more distantly from the centromere (Fig 5.9). It was difficult to be certain if the signal appears twice on the chimpanzee's autosome. A very faint signal was detectable on the Y chromosome of both chimpanzee and pigmy chimpanzee. However, in gorilla, unlike humans, there is no evidence of a signal on the Y chromosome.

Since the "zoo" blot analysis gave very little evidence for the presence of ml22d8-like sequence(s) in mouse, further investigation included an attempt to amplify ml22d8 from mouse testis and kidney cDNA using the human 22d8.F and 22d8.R primers but at an annealing temperature 2-3°C less than that used for human DNA. However, as described for the TTY2 –like genes, very weak or no bands were seen in the correct position on the gel.

Chimpanzee; Y, autosome
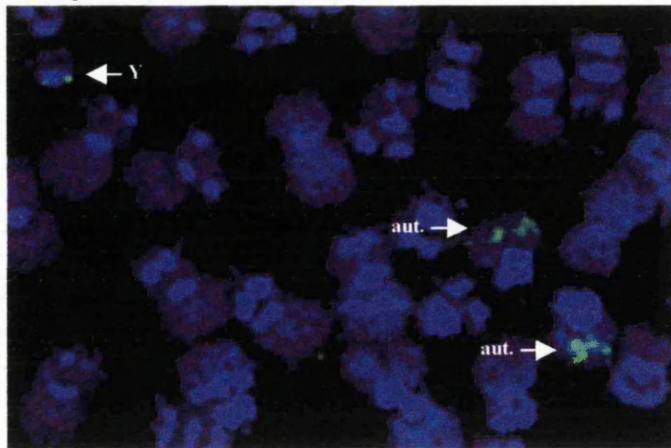
Pigmy chimpanzee; Y, autosome
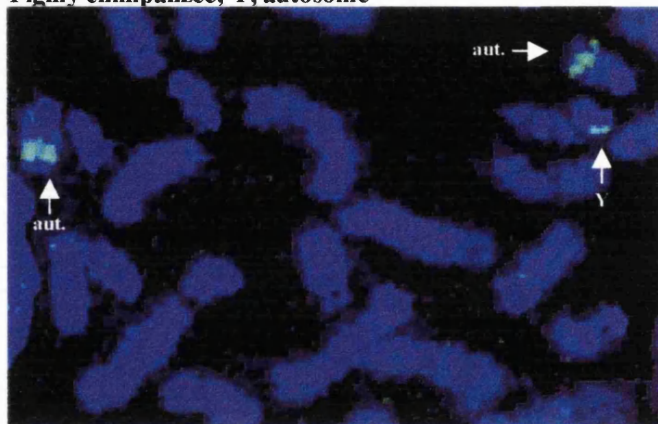
Gorilla; autosome
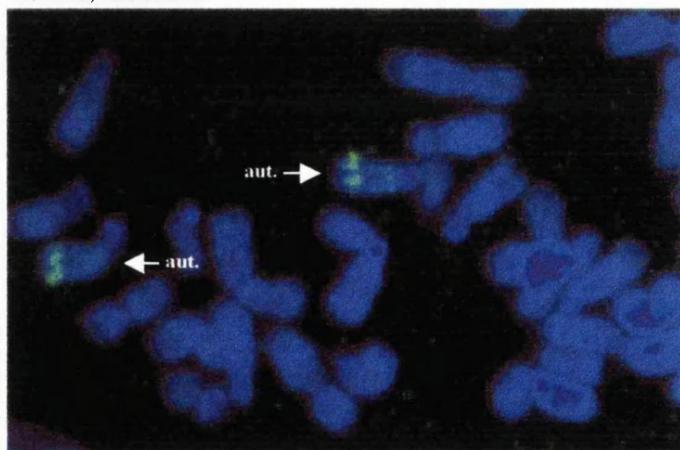
Figure 5.9 FISH of primate metaphase chromosome spreads using

YcoDG11 (ml22d8) as probe

In summary, comparative studies indicated that sequences homologous to ml22d8 are not present in the mouse. This implies that ml22d8 cDNA sequences are unique to primates, where both autosomal and Y-linked copies occur. It seems likely that the Y chromosome copy is the result of a recent duplication and translocation event that took place within the last 3-4 million years, since this copy occurs on humans and chimpanzee, but not on the gorilla Y.

5.5 Analysis of ml22d8 genomic sequences derived from the databases

The localisation of ml22d8 on Chrs 15 and Y was also verified by finding a 92-99% match of ml22d8 and of ml1g1 and ml22c6, with several large genomic clones one of which, AC005630 is located on Chr 15 and the other, AC006328 is located on Chr Y (for the sake of clarity these clones will be referred in the text as AC005630-15 and AC006328-Y). All three cDNAs align with separate, but relatively close regions of these two genomic clones. Along the ~137.7Kb region of AC006328-Y, the ml22d8 and ml1g1 like sequences occur twice and ml22c6 once. Similarly, across the ~177.8Kb of AC005630-15 ml22d8 occurs twice, ml1g1 three times and ml22c6 once. Moreover, they appear in some instances to be in the opposite orientation to one another; this also suggests that these three cDNAs may not be exons of the same gene but a cluster of distinct short expressed sequences. The location and orientation of ml22d8, ml1g1 and ml22c6 cDNA sequences in AC005630-15 and AC006328-Y are shown in Table 5.1 and in Fig 5.10.

The presence of a similar configuration of expressed sequences on two different chromosomes suggests that a partial chromosomal duplication occurred in the evolutionary past, which was accompanied by translocation from one chromosome to another. The homologies that the three cDNAs share with the Y-genomic fragment is not significantly different (92-98%)

Table 5.1 Localisation and % homology of ml22d8, ml1g1 and ml22c6 to the genomic clones AC005630-15 and AC006328-Y.

	ml22d8	ml1g1	ml22c6
AC005630, 137699 bp Chr 15	1) 42079-42530, 98% 2) 120841-121292, 98%	1) 45320-45005, 98% 2) 76601-76916, 99% 3) 124102-123787, 99%	1) 63146-62808, 98%
AC006328, 177769 bp Chr Y	1) 89128-89579, 92% 2) 158752-158301, 95%	1) 128917-128602, 98% 2) 155686-156001, 98%	1) 146661-146334, 96%
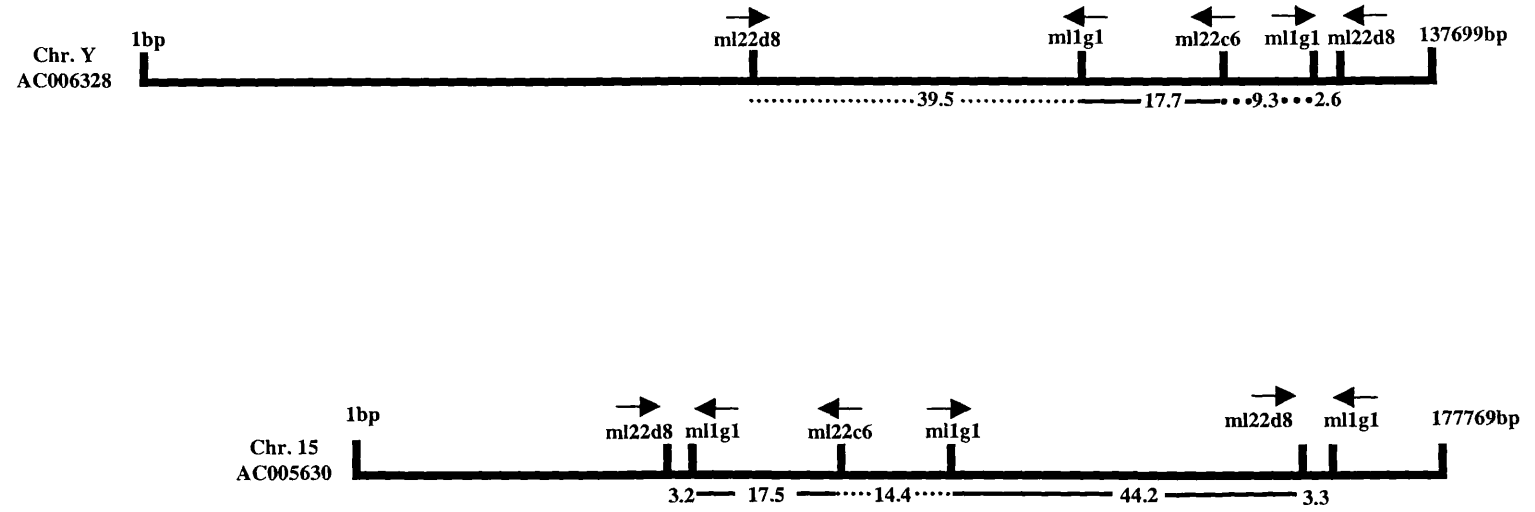
Figure 5.10 Localisation of the ml22d8, ml1g1 and ml22c6 sequences in the genomic clones AC005630-15 and AC006328-Y. Figures underneath represent distance in Kb between the positions of the cDNA sequences.

from that shared with the Chr 15-genomic fragment (98-99%), which suggests that this gene duplication took place fairly recently in human evolution. However, the order that these events took place on Chr Y and 15 is not clear at present.

NIX analysis using genomic clones AC005630-15 and AC006328-Y as a template suggested the presence of potential exon regions corresponding to ml22d8 and ml1g1 sequences in both sense and antisense DNA strands. These programs failed to identify any exons close to the ml22c6 region.

5.6   Identification of the ml22d8 / 3.4Kb testis cDNA

During the course of this work, database BLAST searches were conducted regularly and recently, ml22d8 identified with 98% homology a ~3.4Kb testis cDNA released to the database in Feb. 2000 (accession number AL137524). This sequence was obtained from an adult testis cDNA library and sequenced as part of the german human genome project and is freely available from the German Cancer Research Center. The complete, polyadenylated cDNA sequence is present in the database. ml1g1 and ml22c6 sequences were not included within the 3.4Kb testis cDNA. There has been no further characterisation of this cDNA by the German team who recovered it.

A GAP comparison between the whole of 3.4Kb testis cDNA (will be designated as ml22d8/3.4 TT) and the genomic fragments AC005630-15 and AC006328-Y, showed that the cDNA exists as a contiguous sequence with no introns. On Chr Y there is one complete copy of the ml22d8/3.4 TT and 3 fragments in various orientations; on Chr 15 there are two complete copies in the same orientation and two fragments in the opposite orientations. A further BLAST search with the entire ml22d8/3.4 TT found homologous sequences in three genomic clones that had been assigned to Chrs 9, 12 and 22. Here the homologues were less well conserved (84.5-87.7%) and only one copy was present on each chromosome (Table 5.2). In some cases only fragments of the

237

Table 5.2 Localisation and % homology of the 3.4Kb testis cDNA to sequences on various genomic clones that have been mapped on Chrs 15, Y, 9, 12 and 22. + or - indicates the orientation of the 3.4Kb testis cDNA-like sequences; F: incomplete fragment of the 3.4Kb testis cDNA.

	clone location (bp)	part of 3.4Kb testis cDNA present (bp)	orientation	% homology
AC005630 / chr. 15	41235-44579	1-3379	-	98.4%
	119988-123363	1-3379	-	99.5%
	57723-58309	F 1570-2131	+	87%
	77336-78518	F 1-1202	+	93.6%
AC006328 / chr. Y	156238-159599	1-3379	+	95%
	88266-91026	F 625-3379	-	95.1%
	141320-141866	F 1557-2094	+	87%
	127338-128365	F 1-1043	-	93.3%
AC006312 / chr 9	188331-191516	F 1-3197	-	87.7%
AC010203 / chr 12	62452-65225	F 421-3197	-	84.5%
HSDJ47A17 / chr 22	88975-90735	F 421-2274	-	84.9%

238

ml22d8/3.4 TT were present. These findings suggest that during evolution, several duplication and translocation events have taken place, the most recent being that which gave rise to the copies on Chrs 15 and Y.

The presence of multiple copies of ml22d8 in the genome was confirmed by the identification of a further 25 cDNA clones with high homology to ml22d8 in the cDNA selection library. The segment ml22d8 contains two Cac8I restriction sites, which could be used to categorise these multiple cDNAs. Restriction digest of 20 clones with Cac8I, showed that they could be classified into at least three distinct groups (Fig 5.11). The three groups seem likely to represent at least three closely related genes. 16 of the ml22d8 cDNA clones hybridised to ^{32}P labelled RNA/cDNA from mouse testis RNA confirming that ml22d8 is expressed in testis.

A dbEST database BLAST search with the ml22d8/3.4 TT identified 33 ESTs with high homology, but none of these extended the cDNAs towards the 5' end and it can be assumed that the 3.4Kb sequence is complete. In addition, it was found that the ml22d8/3.4Kb testis cDNA shares 91% homology with the dynamin 1 cDNA (DNM1), across 590bp. DNM1 cDNA is 3187bp long and has been mapped to 9q34 (Van Der Bliek *et al.*, 1993, Newman-Smith *et al.*, 1997). It is a member of a subfamily of GTP-binding proteins and is involved in the production of microtubule bundles and vesicular trafficking processes, particularly endocytosis. DNM1 transcripts are subject to alternative splicing and are ubiquitously expressed with highest expression in the brain (review; Van Der Bliek., 1999).

By using as template genomic DNA from the Chr 9 clone (AC006312), most of the DNM1 exons could be identified as non-continuous regions of the genomic DNA. It was then possible to localise the ml22d8/3.4 TT in relation to the DNM1 gene. From the diagram shown in Fig 5.12 it is clear that the entire ml22d8/3.4 TT comprises sequence that corresponds with 91% homology to part of the last DNM1 intron, 3' UTR and part of the 3' flanking
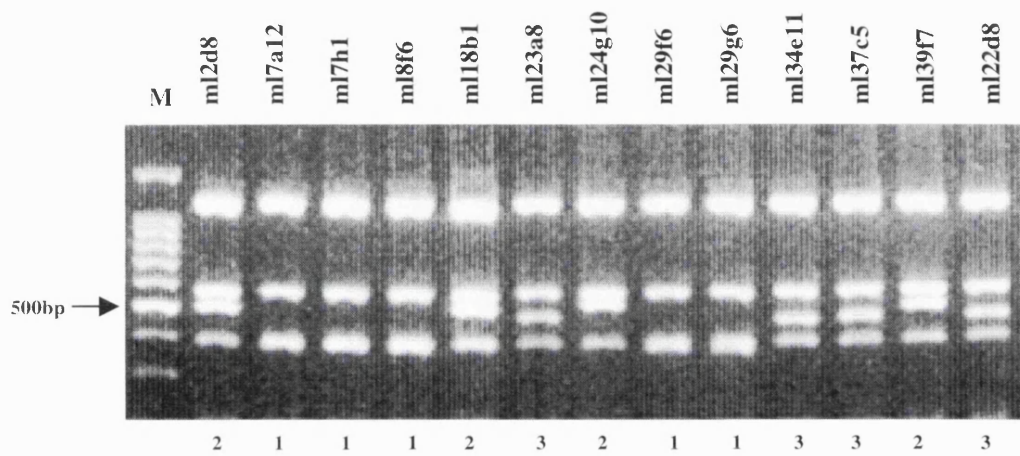
Figure 5.11 Cac8I restriction digest of cDNA clones identified in the cDNA selection library by screening with the ml22d8 sequence. Three distinct restriction patterns are seen (as indicated); M: DNA size marker

8 Kb fragment of AC006312 / chr 9 genomic DNA

Figure 5.12 Localisation of the ml22d8 and ml1g1 cDNA sequences and the 3.4Kb testis cDNA, in relation to the dynamin 1 (DNM1) gene. Numbers indicate the sizes in bp. The last 194bp of the 3.4Kb testis cDNA (total=3391bp) showed no homology with the AC006312 / chr 9 clone and are shown as a broken line. The ORF of 207 aa is also indicated

region. The ml22d8 sequence corresponds to a portion of the 3' flanking sequence. Interestingly, ml1g1 is also part of the DNM1 gene, corresponding to the last intron of DNM1. In contrast, ml22c6 sequence did not correspond to any part of DNM1, or adjacent Chr 9 sequence, suggesting that ml22c6 sequences are present only on Chrs 15 and Y.

Apparently, a fragment of the DNM1 gene was duplicated and translocated to another chromosome. During this process the DNA fragment acquired sequences which allowed parts of it to be expressed. However, a search for promoter elements and transcription factor binding sites (TESS, NNPP / Eukaryotic) associated with the ml22d8/3.4 TT sequences Chrs Y and 15, failed to find a well defined promoter region. Several potential transcription factor binding sites, including the C/EBP and Sp1 were present, but there were no cis-acting promoter elements like TATA box or CAAT box, within 450bp of 5' flanking sequence.

The ml22d8/3.4Kb testis cDNA, does not contain a substantial ORF frame, although starting at 1710bp is a small ORF of 207 aa, which is present in part of the ml22d8 cDNA sequence but shows no significant homologies to sequences present in databases (Fig 5.12).

The conclusion that can be drawn is that a fragment of the genome that contains sequence homologous to ml22d8, ml1g1 and ml22c6 was duplicated at some point during evolution and inserted itself into other chromosomal regions. This duplicated region contained, from earlier duplication/insertion event, a fragment of a dynamin-like gene corresponding to its 3' end. Some/all of these dynamin-like fragments appear to have acquired sequences, which allow its expression, probably binding sites for transcription factors. Within this expressed, intronless sequence lies the ml22d8 sequence. mlig1 and ml22c6 represent distinct expression units. All cDNAs are as yet of unknown function.

5.7 DISCUSSION

5.7.1 22d8/3.4 TT structure

Investigation of the cDNA ml22d8, showed that it corresponds to a 3.4Kb testis transcribed mRNA. The full sequence of this cDNA was deposited in the EMBL database in Feb. 2000, by members of the German human genome mapping project. Curiously, this 3.4Kb testis RNA shows high homology (91%) to the non-coding 3' end of the DNM1 gene, which is located on Chr 9, at 9q34. In addition, mapping studies and sequence analysis revealed that this transcript, designated ml22d8/3.4Kb TT, has related sequences on several chromosomes, including 9, 12, 15 and Y (85-99% homology). This multiplicity is attributed to fragment duplication and translocation.

Since ml22d8/3.4Kb TT corresponds to a fragment of the DNM1 gene, it seems likely that ml22d8/3.4Kb TT arose as a truncated gene copy by duplication. However, such sequences are often found within clustered gene families as the result of unequal crossover or unequal sister chromatid exchange. Examples are found in the class I HLA gene family, which is located in 6p21.3 and contains at least 17 members, many of which are either non-processed pseudogenes or truncated gene copies (Geraghty et al., 1992).

Another possibility is that ml22d8/3.4Kb TT could be part of a "duplicon", or otherwise a duplicated fragment of low-copy repetitive sequence (reviewed by Eichler, 1998, in response to the National Institutes of Health Meeting, "Genomic alterations in Genetic disease"). Duplicons are paralogous, highly homologous (95-99%), genomic segments of 50-200Kb that have >2-10 copies in the genome and are usually present in subtelomeric and pericentromeric chromosomal locations. Such regions are considered to be unstable and prone to chromosomal rearrangements and include chromosomes 2p11, 10p11, 15q11, 16p11 and 22q11.

Interestingly, ml22d8/3.4Kb TT has copies that share high homology (85-99%) in several chromosomal locations, including chromosomes 15q11 and 22. However, the duplicons are considered to be large genomic segments and sequence comparison showed that in the case of ml22d8/3.4 TT sequence only a fragment of around 4Kb has been duplicated and translocated, whereas there is no evidence that sequence outside this 4Kb fragment is located elsewhere in the genome. In addition, sequence analysis and mapping studies showed that its location at least on Chrs 9, 15 and Y is neither subtelomeric, nor pericentromeric. Consequently, although ml22d8/3.4Kb TT bears some of the "duplicon" characteristics, like a small number of copies in the genome that share high homology and do not contain any repeats, it appears that it cannot be classified as one.

Because the ml22d8/3.4Kb TT has no introns, it would seem likely that it represents an expressed, processed gene or pseudogene that arose through retroposition. This class of elements, exist in the mammalian genome in a rather low proportion (5%) and appear to lack the specific structural and coding properties of transposable elements, like long terminal repeats (LTRs) and a reverse transcriptase coding region. They usually contain a poly-(A) tail, which ml22d8/3.4 TT does not appear to have in the genomic sequence and are thought to arise by integration into chromosomes of an mRNA sequence, generated by reverse transcriptase.

An example of an intronless processed pseudogene, is the testis-specific pyruvate dehydrogenase (PDHA2) isoform, which is located in 4q22-q23 and was generated by reverse transcription of the PDHA1 gene, a member of the pyruvate dehydrogenase (PDH) complex, located in Xp22.2 (Dahl et al., 1990).

An example of a gene family that has many functional members without introns in their structure, is the G-protein coupled receptor family (GPCRs). These are a large and diverse family of receptors, scattered in

244

several chromosomal locations that mediate signalling between extra- and intracellular environment. It is of interest that most members of this large gene family lack any introns in their structure and are though to have arisen in the genome by retroposition (Brosius., 1999b; Gentles and Karlin., 1999).

The ml22d8/3.4Kb TT genomic sequence, has no poly-(A) tail, although a poly-(A) tail is added during the course of mRNA processing. Furthermore, although the sequence has some homology to the DNM1 gene, the similarity is to the 3' end of it, as it is present in genomic DNA, rather than RNA. Thus, it seems unlikely that ml22d8/3.4Kb TT arose by reinsertion of DNM1 RNA into the genome. However, its presence in several chromosomal locations indicates that it is a mobile element that possibly arose by reinsertion of part of the DNM1 DNA in a region(s), where promoter elements, which have either evolved by chance, or originally regulated another gene, allow its processing and expression.

Lahn and Page (1997; 1999) have described two intronless Y-linked genes that have arisen by retroposition, CDY1 and CDY2. Both genes are intronless and are thought to have derived during primate evolution by retroposition of their autosomal homologue CDYL (CDY-like) mRNA on chromosome 13, followed by duplication of the Y-linked retroposed gene. It is of interest that Lahn and Page have mapped the CDY genes in Yq, in two distinct locations (intervals 5L and 6F), which appear to be very similar to the two locations of ml22d8/3.4 TT, within intervals 5J-5K and 6F-7. Sequence analysis has shown no homology between CDY1/CDY2 genes and the ml22d8/3.4 TT. At present, there is no evidence indicating whether or not, these sequences arose by independent, transposition-mediated events.

Interestingly, it is believed that transposition is a mechanism that for many species has contributed to the gene content of the Y chromosome and during evolution led to its decay, since insertion and accumulation of mobile

transposable elements, causes disruption of functional genes and generates spontaneous mutations (Steinemann and Steinemann., 1992a).

Investigation of *Drosophila miranda* neo-Y chromosome, revealed the presence of a dense cluster of 4 copies of the larval cuticle protein genes (*Lcp*), which have been incorporated into the genome in the form of retroposons. From those, only *Lcp3* is moderately expressed, whereas the rest of the copies have been silenced, possibly by neighbouring, unrelated transposable elements (Steinemann and Steinemann, 1992b; Steinemann et al., 1993).

Similar to the *Lcp1-4* retrotransposons in *Drosophila miranda*, the neighboring location of CDY1/2 and ml22d8/3.4 TT, may suggest that there are Y-linked regions that favor the insertion of transposable elements, which have the ability to regulate the expression of one another.

Further sequence analysis, would provide data regarding the presence of any DNA conformations in the region that can may relate and/or interfere with the expression, or silencing of these two nearby genes. Interestingly, expression studies of ml22d8 have shown that its is ubiquitously expressed and the elevated expression levels suggest that more than one copy is functional. ml22d8/3.4 TT appears to retain a housekeeping function that allows it to be expressed in all cell types. Prediction of its potential protein structure (ORF of 207 aa), using variety of protein analysis programs, might propose a functional role, but this has not yet been carried out.

5.7.2 Chromosome Y and 15 ml22d8/3.4 TT homologues

FISH analysis using YcoDG11 as a probe (section 5.2), indicated that sequence corresponding to ml22d8 lies on 15q23-25, and another possibly at 15q11-q13, where the signal was weak. In addition, FISH analysis and mapping studies identified a homologous sequence present on Yq11.23 between the AZFc and the Y-heterochromatic region.

Several investigations have shown that chromosomes Y and 15 are prone to exchange of genetic material and there are several reports of prenatal diagnoses where such rearrangements have been demonstrated by FISH analysis (Schempp et al., 1985; Alitalo et al.; 1988; Wilkinson et al., 1993; Verlinsky et al., 1998; Reddy., 1998). Some of these translocations lead to an abnormal pathological phenotype, including azoospermia, whereas others are not connected with any phenotypic abnormalities and the outcome of pregnancy is healthy male babies. The apparent high levels of homology between 15 and Y-linked ml22d8/3.4Kb TT copies, as judged by the level of FISH hybridisation signal and sequence comparisons, and the lower levels of homology between these and homologues on other autosomes, suggest that the duplication and translocation involving Chrs 15 and Y are the most recent in the mammalian lineage.

It is of interest that 15q11-q13, the region showing a weak YcoDG11 (ml22d8) signal, is considered as a region rich in *Alu* elements and other repetitive sequences that is prone to genomic rearrangements, such as deletions, duplications and translocations (Donlon et al., 1986; Robinson et al., 1991; Huq et al., 1997; Christian et al., 1999). It is also considered as the critical region for two imprinting disorders, Prader-Willi syndrome (PWS) and Angelman syndrome (AS) for which ~70% of cases are associated with interstitial deletions, due to the presence of several recombination breakpoints (Christian et al., 1995).

5.7.3 Comparative studies

Comparative FISH studies showed that the Y-linked copies of ml22d8, are the result of a duplication and translocation event that took place 3-4 Mya, after the divergence of gorillas from chimpanzees and humans (section 5.4) (Goodman et al., 1998, review, Goodman, 1999).

YAC 913B1, which carries the ml22d8/3.4 TT sequence, was used for comparative FISH analysis by Archidiacono et al (1998). His data allow an estimate of the position of ml22d8/3.4 TT homologues on chimpanzee and gorilla Chr Y. In chimpanzees, YAC 913B1 maps to both arms, at Yp11.2 and Yq11.1, whereas in gorillas, there is only one chromosomal location, at Yq12.2. These data show that humans and gorillas ml22d8/3.4 TT, have a similar location, whereas in chimpanzees, it appears that further duplication has taken place to give signal in two distinct locations. These results are rather similar to those obtained for the two TTY2-like genes ml13d10 and ml22a3 (section 4.10) and suggest that structural rearrangements have happened in the chimpanzee Y chromosome after its divergence from the gorilla lineage.

However, some caution is required in interpretation, since the ml22d8/3.4Kb testis transcript showed homology with only part of YAC 913B1. YAC 913B1 contains around 1.6Mb of genomic material and FISH analysis cannot indicate whether the whole or part of it, correspond to the identified regions (Archidiacono et al., 1998). Consequently, although interpretation of the data does not exclude the possible presence of ml22d8/3.4Kb testis transcript within the identified Y-linked locations, it does not provide direct evidence of the estimated location.

CHAPTER SIX

Final conclusion - Remarks

Computational analysis of the genome has progressed significantly over the last two decades and a large number of genomic sequence data have accumulated (Claverie, 1997). During this period, the genome sequence of two species, *C.elegans* (1998) and *D.melanogaster* (2000) was completed and very recently (June 2000) the first working draft of the human genome was published. In addition to the general sequencing progress, analysis of ESTs estimates the presence of approximately 120,000 genes in the human genome (Liang *et al.*, 2000).

The human Y chromosome, with a size of 59Mb and an estimated content of 100 genes, is amongst the smallest human chromosomes, alongside Chrs 21 (50Mb) and 22 (56 Mb), where at least 190 and 304 genes and pseudogenes (GDB database) have respectively been identified. The whole of chromosome 22 and most of Chr 21 sequence are now available (Lapenta *et al.*, 1998; Hildmann *et al.*, 1999; Dunham *et al.*, 1999), whereas sequencing and transcriptional mapping of Chr Y has proven very slow. According to NCBI database, at present there are 44 contigs on the Y chromosome totalling for ~13.3 Mb of genomic sequence. Whitehead institute (MIT) has sequenced to date 1.65 Mb, which corresponds to 5.5% of the estimated complete coverage of the Y chromosome euchromatic region. EMBL is also participating in the sequencing of the human Y chromosome and at present it has finished around 1/3 of the sequence as a working draft with 12.6% as complete sequence. Several independent research groups have also obtained valuable information regarding the Y chromosome's physical mapping (Wolfe *et al.*, 1984; Affara *et al.*, 1986, Buckle *et al.*, 1987; Vollrath *et al.*, 1992; Foote *et al.*, 1992; Taylor *et al.*, 1996).

Genes on the Y chromosome, although small in number, are of particular interest, since they can provide us with knowledge upon its origin and its function in sex determination and differentiation. The Y chromosome, together with the X, are thought to have originally derived from a homologous pair of autosomes in a vertebrate ancestor. Y-linked genes appear to fall into three distinct categories (Burgoyne, 1998). The first group includes genes that were present in the original prototypic sex chromosomes, examples of which are the Y-linked SRY and its X-linked counterpart SOX3; evolutionary, these are among the most ancient known X-Y gene pairs in humans (Collignon *et al.*, 1996; Lahn and Page, 1999). The second category includes highly homologous genes that were more recently generated, by successive additions made via the two PARs and the X-Y homologous regions (see sections 1.2.3 and 1.2.4). Finally the third category includes genes that correspond to recent additions of autosomal material, directly onto the non-recombining region of the Y chromosome that have acquired a male specific function. An example of such gene is DAZ that was added to the Y during the evolution of primates, with an autosomal progenitor that has a testis specific expression (Saxena *et al.*, 1996; Yen *et al.*, 1996).

This thesis involved the identification and further characterisation of Y-linked genes, using a cDNA selection approach. Amongst several potential novel Y-linked cDNAs further studies involved expressed sequences that fall into two groups; cDNAs that showed homology (~75%) to the published TTY2 cDNA (Lahn and Page, 1997) and the ml22d8/3.4 TT low copy genes.

Investigation of the TTY2 gene family showed that it falls into the third category of Y-linked genes and exhibits features similar to those of other known and more well characterised gene families mapped in the non-recombining region, like RBMY, DAZ and TSPY (Lahn and Page, 1997; Burgoyne, 1998). As for these latter families, TTY2 exist as a multicopy gene family with members expressed in testis, although not exclusively (section

4.3). Although they have no close relatives on the X chromosome, they possibly have originated from an autosomal or X-linked ancestor (section 4.10). However, either because a TTY2 copy was translocated very early during the evolution of the Y chromosome, or because there is no selective pressure due to lack of recombination in this region, the Y-linked TTY2 genes have diverged significantly from their autosomal or X-linked prototype.

ml22d8/3.4 TT, the second subject of investigation in this thesis, although it exists in the non-recombining region of the Y chromosome, it does not appear to fit into any of the common Y-linked gene groups. It is present as a low copy intronless, possibly transposable element of unknown function, located on several chromosomes, including the Y. These data propose that some genes on the Y chromosome apparently have arisen due to retroposition, such as the CDY1 and CDY2 transcripts (Lahn and Page, 1999). Such genes, although they exhibit some features characteristic of genes in the third category (exist as more than one copies, have an autosomal ancestor) they may be considered as members of a fourth group of Y-linked genes.

Lahn and Page (1997), following a similar cDNA selection approach identified 12 Y-linked genes and at the time almost doubled the known Y chromosome gene content. Evaluation of the cDNA selection library used in the project described in this thesis, found that ~61% of the sequenced cDNAs were Y-linked. Although a portion of these clones is expected to be redundant, the remaining Y-linked cDNAs, offer raw material for the identification of potentially novel genes. Thus, although almost half of the gene content of the Y chromosome has been defined (47 genes and pseudogenes, GDB database), it is clear that more Y-linked genes will eventually be identified.

The Y chromosome, due to lack of extensive recombination, has been the subject of major structural rearrangements during evolution, a price paid for the acquisition of the sex determining factor(s). However, it is of interest that although it has accumulated a large number of repeats and mutations in its

genes, it has been selected and retained in the genome of most species, because it has acquired an important function(s) in male sex determination and differentiation. Nevertheless, the presence of the Y chromosome in the genome of several species, is a relatively recent event in terms of evolution and perhaps its progress should be considered as ongoing. There are species that lack sex determining genes and have developed other ways of sex determination, like reptiles, in which the sex-determining mechanism depends upon the egg incubation temperature (Lance, 1997, Western *et al.*, 1999). It is not certain that the Y chromosome is now so invaluable that it will remain an essential feature of the animal genome. Instead, it may be the case that eventually, its genomic isolation will lead to rejection and replacement by other ways of sex determination.

In mammalian systems the Y chromosome is known to bear the male sex determining factor and therefore although is not considered necessary for the viability of the individual (e.g. XO, Turner syndrome), it is considered important for the survival of the species. Although a great amount of work regarding its sequencing and gene identification needs yet to be done, at present, with the help of more advanced technology and availability of computational analysis, we are on the verge of isolating more Y-linked genes and understanding the function of at least some of them.

Appendix I

Predicted gene structure of TTY2, based on BLAST analysis and BESTFIT
comparison of the published TTY2 cDNA, with sequence from the genomic clone
AC006335. Blue colour indicates the exons; black colour indicates the introns; red
colour shows the exon/intron boundaries; pink colour shows nucleotide sequences
that usually are associated with exon/intron splicing junctions; numbers on the left
side of the sequence indicate nucleotide number on the genomic clone AC006335.
Dotted line in between intron sequence corresponds to intronic sequence that was
not included.

```
34101   GTTGTTTGCA TTTTGAAAGA GGCTTTTTAT TTCTGTGACA GGAATATTGG

34151   TACATTGCTT GGACTCTAGC ACAAGTCAGC TCGTTCTTTC AGATGATGCT

34201   TGATGTTTCT TGGCTTTCAT GGGGGATTCA CATTGCCACT CAACCACACT

34251   ACTGGATACA CTTTTCAGGC TTGCCATCAC CACAGATGGC CTCTGAGACA

34301   CTGTTTGAAC CACATCTGCA CCTGTGAGAG GCCAGTTTGA GGTATGAGAA

34351   CACTGTTTCA ATTTGGACTT GCCTTTGTCT TGGTTCCTGC TTTTCCCAGA

34401   TAGCACCTAC CCAACCCAGG ATGAATGAGT GCAGAGAGGT CAAGTGCCAG

34451   GCCATCTTTT GCTGACACCC TTTTCTGGTA TTTCAGGTAT AAGTCCATCA

34501   TCCAAAGACT GCTCAACATC TCACCAGAAT ATATTTCAAT CCTCATGGGG

34551   CATGATTCTT TCACAAAACC CCTTTCAGGA ATGGAGTCAG AAGAGTAGTT

34601   TCCAGAGACA ACCTCACAGT CTTGAAACGG CTCTGCCTCC CATGTGATCT

34651   GACCATGGAG ATGGCATATA AGGGCCCTAA GTTTGAGACT TTTAGGGTAC

34701   TGCAATGCGT TATCACAGGC AGCCTTTATC CTGATACCAA GCCAGCTCTG

34751   CCTGTACCAT TTTCCTCTGC TTAGGCAGGC TGACAGCCCT GACACCCTGG

34801   TGCTCCAGTT TGAGTCACTA TATGTGGATG TGCTAGTCTT AGGGCAATGG

34851   ACCTGAGCTG TGAGCTGTAG CTAGTGTCAC AATGAATGCC AGCTTTGCTA

34901   GTAACAATTC CCTTTGGCTT GGTAGAGAAG GAGACCTCTG TGGAGGTACA

34951   ATGGTGGTGC ACTGTCACCT GTCTTCTCTG TGGGATCCAT GGGACAGTTC
```

253

```
35001   CATGATCCTA GGAGAGGGTA GATGTGAGCC AGCCTGAAGA AATGTCAAGC

35051   AGAGCCCCAG GAATGAAGCA CAAAATCACT ACAGATCCAA AAGGATCTGC

35101   AGAATTTGTC AGGCCTGCCT AGACATTGTA GGGGTTAGTC TTATTGAAAT

35151   GTGTCCCACT GTAATTTCCA ACTTCAGCCT CCTGTGTTC CCAGCAGTTT

35201   CTCTCTCCCA GGTGGGGCTT TCTGCAGAAT GACACAGCCT CAGAAGCTAC

35251   TGGGCTGTGT GTTACTGTGG GAGTGTTGCG AGTGTTGGAT GTCAGCATGT

35301   GTGTGTGGCT TTGTGTGTTT GTGTATGTGT CTGTGTGTGT GTATGTAAAT

35351   GAATTCTGTG GATCAGGAAT CAGCAATGAC TAGTTAAGCT GTCTGTGACC

35401   AGCCGGGTTC CCCATCGTCT GCCCCTGCCA AAAAAACAGG TACTCTTCTA

35451   CAAAGAAGAG GAGAGCACCA CACCCAAGAA CAGACATCTC CCAGTGTTGC

35501   ATTATAAAGC AGCCAACCCA CAGACACTAG CACTCTGGTC TGCATAGCCC

35551   CTTTAATTTA CCTAGAATTC AGTTCCCAGC CAAGTAGGTG CTTCATGTCC

35601   TGAGGGTGCA ATCCTCCATC ATCTTGAGAT TTCATGCTGG TACAGAGAGT

35651   GTGACAGCAA TAAGGTCAGA TAGGGGTGAG TATACAACCT GGTGAAGGGT

35701   GGATGGGGTC CCGTACCTTC ACCAGCAAAA AGGGTGAAAA TAGATGACAC

35751   AGAATGTGCT TCCAACTCCA TCCCCACATT CCCATAATTG CAAAATCAGT

35801   CAACAACATG GCCTGGTGTT TAGGTGGGAG TACTCCAACC TGCAGGAAAA

35851   ATTTGGAGTG CAAATTGTGG CCAATCTGGA AAACTCCTGG TTTGAGGGTT

35901   TTAATACCTG TAGTCAAATG GAAGTGGAAT AGATTGATGC TGGGTGGGTT

35951   GTGGCCTCCA CATTTGTGTC CTCTTTTACT GACTTCCATT GTCCTCATTG

36001   GTGTAGGGCT TCCTGGATCT GGCTCAACAT CTTCCACACT AAACTCTTCC

36051   CTGTTCACAG AAGACCATCA TAAAAATGCA TTGTGTGAGT ATTTACTTGT

36101   AAACACTGTC ACGATTTAAT GACTGAGTTT CTGTGATACT TTTAAAATCA

36151   TAAACTACTA TTACAACCAG CAACAGGGAA ACTTTTGTTC TCCCACCGTT

36201   ATCAGAGGGC TGCAGGGTTC GTGAAGGAGG AGAACCAGGC AGCCATGTCT

36251   GGCTTTTGCC TTGTAATCTA GGCTCTGTTT CCCTTCATCT GCAGACTATC

. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

38301   ATGAGTCATT GAACATTGCT TGGACTCCAT CACAAGGTAG CTCATTCTTT
```

254

38351   CAGGTGAGCA TTGATTTTTC TTTGCTTTCA TGGGGAATCC ACATTGCCCC

38401   TCAACAGTAG TACTGGACAC CATTTTCAGG CTTACCATCA CCACAAACAG

38451   CCTCTTAGAC ACTGTCTCAA CCTCATCTGC ACCTGTGAGT TGTGAGCCCC

38501   AGGTGTCAGC ACACTGCTTT ACTGTGGACT ATCCTTTGTC GTGGTTCCTG

38551   CCTTTTGCAG AGAGGCCCTG CAAGGACCAG GATGAAGGGA GACAGTGAGG

38601   TCAAGAGCCC AGCCATCTTT CACTGACACC CACTTCTGGG TTCTCAGGTA

38651   AGTTCCTATC ACCAAAATAA CCCTCAACAA CATACCAGAC TATATTCCAA

38701   TCCCCAAGCG ACCTGATTCT TACACAGCCT CTTTTCAGAA TGGAGTCAGA

38751   AGGGCAGTTT CCAGAGACTC ACTCACAGTC ATGAACACC TCCTTCTCCA

· ·· · ·

45751   ATGATTCCAC CACCCAAAGA GCCCTCAACA ACTCACCAGA ATATATTCTA

45801   ACCTCCATCT GTCCAGACTC TTGCACACAG CCTTTCAGGA ATGGAATCAG

45851   AGTAACACTT TCCAAAGACC ACCTCGTAGT CTCGAATCAC CTCCTTCTCC

45901   AGTGGGACCT GGCCACAGGA ATGACTTGTA GAGGCACTAC AGTTGAGACA

45951   TTTATGGTCC CTCATTGAGT TATTGTAGGC AGCATTTTTT TCAATACCGT

46001   GCCGGCTCTT CCTGTATCTT TTTTTTTCTT TTTTTTTTAG GCTGGCTGAC

46051   AGGTCTGACA GCCCATCACG AAAGCCTGCA TACTCTTAGA CACAAGGACT

46101   GAGCTATGGG CTCCAGCTAG CATCACAATG AAGGCCACCA TTGCCTAGGG

46151   ATAAGTCCCT GTGACTTTGT GGATAAGAAC TCCGTGGAGG TGAGTCAGCG

46201   GTGGACTCTC GCCTATCTTC CCTGTGGGAT TCCCACGATA GTCCCATGGG

46251   CCTAAGGAGA GGGAGCAGGT GAGTCAGCCT GAAGAAAAGT CAAGCACAGC

46301   CCCAGGAATA AGCCACAAAA TCCGTACAGA TCCAAAGAA TCTGCAGGAT

46351   GCCTCAAGCC TCCCAGACTT TGTCGGGATG AGTCTTTTTA AAACTTACCA

46401   TACTGTAATT TCTAGGCACA GGCTGCCTGT GTTCCCTGGG GTTGCTCTTT

46451   CAAAGGCAGG GCTTCCTGCA GAACCAAGAA GCCTAAGAAG CTGCTGGGCT

46501   GTGTGTTTTT GTGAGAGTGT TGCAAATGTT GGAAATCTGC ATGTATGTGG

46551   TAGTGTGTAT GTGTGATTGA GTGTGTGTGT GCCTGTTAAG TGGAGTCTGC

46601   TTAAAGGAAT GTGGCTAATG CACTGCAGTG ATTCTTTTTT CTTTTTTTTT

255

```
46651   TTTTGAGTCT CCAAACATTG TAGTGGCCTG TATGTGTGGG TCTACTTGGG

46701   CTGCTGGGAT CCATATTCTT TATTTTTCTG AGGATCATGA ATCTGCAGTG

46751   AATTGGCAAG CTGGCTGAGA AACACCACTG TCCAAATCAC CTTCCTTTGC

46801   CAAAAAGCC ACTCTTCTAG AAAGAACAGA AGAACACCAC AACCAAGAAC

46851   AGAAATATAC CAGTGTTTAA TTGTCTTTCA GCCAATCCAA GGAGAGACAC

46901   TATCATTCAT CTCTACCGGG CCTATTAAAT TTACCTCGAA TTTGATTCCC

46951   AGAGGAGTTG GTGCTTCACA TCATCAGGGG GAACTTCTCC ATTGTCTTGG

47001   GATTTCAGTC TGGGATAGAG ACTTTGAACA GCAATAAGGT AATAAGGTCA

47051   GATAGGGGTG GGGATACCCC TCTGGTGAGG GGTGGATGCC ATGCTGTACC

47101   TTCACCTGCA AAAAAGAAA AAAAAAAAG AAGACAGATA ACACAGAAGT

        •••••••••••••••••••••••••••••••••••••••••••••••••••••••• ••

49651   ATCACAAGCT GTCCACACTA TGGCCCCATT TTCATGTGGG AGTACTCTAA

49701   CACACAGGGA ACATTTGGAG TGCAAACTGG GGCCATCCTG GCAAACTCTC

49751   AATTTGAGGG CTTTCAAACG CAGAGCCAAA TGTGAGTGGG ATGGATTGAT

49801   ACTGGGTGGG ATGCAGCTTC CACAATTGCC TCATCTTTTC CTGACTTCTT

49851   TGTTCGTCAT TGGCGTAGTT TCCTTGGTCT GGCTCAACGT CTTCTAAACT

49901   CAACATTCCC CAGTTCATGG AAAATGATCC TCATGGGATT CTATTGTGTG

49951   AGTATTTCCC CCTAAACACT GCCATGTTTT AATGACTGGG CAGCTGTGAT

50001   ACTTTTAAAA CTGTAAATTC ATGCAACAGC CACAAACAAG GAAACTCTTG

50051   TTCTCTCACT TCTATAGAAA AGCTGCATGA TTCCTGGAGA ATGTTTTATT

50101   GTCCTGAAGT CAACCCAAAA ATCAACACTA CCAGTCATGT TGCAGAGCTC

50151   CTTGAATTAA CCTTGAATTC AGTTTCCAGC TGAGCAGCTG CTTCACGTTG

50201   TGAGGGGGCA ATCCTCCATC ATCCTGGGAT TTCATTCTGG GACACAGAGT

50251   TTGAGCAGCA ATAAGGTAAG ATGCGGGTGA GGATGCAATC TGGTGAAGGG

50301   TGGATGGGGT CCCACACCTT CACCTGCAGA AATGGTAAAA ACAGATGACA

50351   CAGAAGGTGC TTCCAACTGC ATTTAAGCAT TCCCTTAATT GCAAAAGGAT

50401   TCCACACCAT GGCCCTGTGT TCTGGTGTGA CTACTCCAAG TGGCAGGAAA

50451   CAATGTTAGT GCAAACTTTG GCAATGCTGG CACATTCCCT ATTGGAGGGC
```

256

```
51201   ATCTGCAAGT CTTGTCTCAT TGTGGAGGGC ATATAGCATT GAATGTTGCT

51251   AGATGGGAAT GTCTCTAAAT ACAGATTTTG CTGCCACTGA TTTCAGAAAG

51301   CAAAAATGAC TTTGCATAGG CTGGCTGCCT TCCAGCTTGT GGTTCGTGGT

51351   CTCATTGTGA GAGCTGAGGT TGTTTGCAGT TCTCAGGAGG CTTTTGGTTC

51401   CTCTGACAGG AACCTTTGAA AGTTTCTTAT ACTCCAGCTC AAGCCAGCTC

51451   CTTCTCTCAA GCAAGCCTTG ATTTTTCTTT CCTTTCAGGC TGGGTCCACA

51501   TTGCCCCTCA ACAGCATTAG TGGACATGAT TGTCAGACTT GCAATTTCCG

51551   CAGACACCTT CTGTGAACAT TTTTCAACAT CATCTACATG AGTGAGAGAC

51601   CCGTTCGACA TGTAAGAATA CTGCTTGACT TTGGACCTGC CTTTGTCGTG

51651   GTTCCTGCCT TTCTCATAGA TCCCCTGCCA GGCCCAGGAT GATAGGAGGC

51701   AATGAAGTCA AGGGCCGAGC CCCATTCATT GAAAGCTGAC TCTGGGGTCT

51751   CGGGTATAAT TCCATCACAT AAAATCCCCT CAACAACTCA CCAGACTATA

51801   TTCCAATCTC CATGGAACCT GATTCTTGCA CACAGCCTCT TTCAGGAATG

51851   GAGTCAGAAG AGCAGTCTTC AGAGACCACC TCAGTTTGGA AAAGCCTCCT

51901   CCTTCAGTGG TTCCCAGCCA TGGAGTCATC GTGAAGGGGC TCCATGGTCA

51951   ATAATTTTAC GGTACTGCAC TTGGTTATCA CAGACAGACT TTTTCATGAT

52001   AGCATGCCAT CTCTGTCTAT ATCATTTTCC TCTGCTTAGG CAGGCTGACA

52051   ACTCTGACAG CCAGGGGCCC GAATCTACCT GGCAAATGTG CATGCTCTAC

52101   TCTCAGTGCA AAAGGCCTGT TTGGGAGTTC TGACTAGTGT CACAATAAAT

52151   GCCGCCATTG CCTAGTGACA AGTCCCTGTG GATTGAGGGA GGAAACCTCC

52201   ATGTAGGGGC ATTGGTAGTG GACTCTCACC TGTCTTCTAA GTAAAATGCA
```

APPENDIX II

Predicted gene structure of ml13d10, based on BESTFIT comparison of the TTY2 identified exons, with sequence from the genomic clone AC009491. Blue colour indicates the exons; black colour indicates the introns; red colour shows the exon/intron boundaries; pink colour shows nucleotide sequences that usually are associated with exon/intron splicing junctions; green colour indicates potential exon/intron splicing junctions; numbers on the left side of the sequence indicate nucleotide number on the genomic clone AC009491. Dotted line in between intron sequence corresponds to intronic sequence that was not included.

```
142601    GAGATTGTTT GCACATTGTA GGAGGCTTTT AGGTCCTCTG ATGGGAATCA

142651    TTGAACATTG CTTTGACTCC AGCAGAAGGC AGCTCATTCT TTCAGGCGAG

142701    TCTTGATTAT TCTTTGCTTT TGTGGGTAAT TCACAGTGAA TCTTGACAGC

142751    ACTAATATTC AGCATTTTCA TCCTTGCCAT CACCACAGAT GGTCTCGGAT

142801    ACACTGTCTG AACCTCATCT GCACCCTTGA GAGGACAGTC CAAGGTGTGA

142851    GAACATTACT CCACCTTGGA CTTGCCTTTG TCATGTTCCT GCCTTTCCCA

142901    GAGAGCCAAT GGGAGGCCTG AGATGAAGAC AGTCAGTGAT GTCAAGAGCC

142951    CAGCCATCTT TCACTGTCAC ATGCCTCTGG GGTCTCAGGT AAGATTCTAT

143001    CCCTCAAAGA ACACTCAACT ACACACCAAA CTATATTCCA ATCCGCATGT

143051    GACCCAATTC TTGCACACAC ACATTCTCTT TTGGGAATAT AGTGAGAAGA

143101    TCAGTTTCTA GCAACCACCT CACAGTCTTG AAATGCCTCC TCCTCCAGTG

143151    AAAACTGACC ATGGAGATGG CCCGAATTGG TCCTGATGTC TAGACTTTTA

143201    GGGGCCTGCA GTGGGTTTTT GCAGTCAGAA TTTTTCTGAT ACCAGGCTGG

143251    CTCTGCCTGC ACTATTTTCC TCTGCTTAGA CAGGATGACA GCTCTGACAA

143301    ATGGGCACCA GAGCCTGCCT CACGAATGTG CATGTTCTAG TCTCAGGGCA

143351    CCAGGACTGA TTGTGAGCCC TTGCTAGCAT CACAGTGATT GTGACTGGTG

143401    CATAGTGACA AGTGCCTGAA GCTTGGCAGA GGAGACTTTT GTGGAGGTGC

143451    ATCGGTGGTG GACTCTTGCC TATCTTCTCT GTAGATCCAT GGGATAGTCC
```

```
143501   CACGAATCTA GGAGGGAGCA GACACGAGCC AGCTTGAATA AACCTCAAGC

143551   CCAGCCCCCG GAATAAACCG CGGAATCCCT AAGGATCTAA AACGTTCTTC

143601   AGGATTCCTC AGACCTACCT AAATGTTGTA GGAGTGAGGT TTCTTGAAAC

143651   TTGTCCCACT GTGATTTCTA GGTACATCCT GACTGTGTTA GCTGGGGTTG

143701   CTCTCCCCGA CGTGGGGCTA CCTGCAGAAC CACACAGCTT CAGGAGCTGC

143751   TTGGCTATGT GTTTCTGTGA GAGTGTTTTC TGCATATGTG TTTAGCATTT

143801   TGTGTGTGTT TGTGTGTGCC TTTTAGCGGA GTCTTCTTAA AGGAACATGG

143851   TTAAAGCACT TCAGTGCTTC TTTTTTTTGA GACTCCTAAC CTTTTGGTGG

143901   CTGCCCTGTG TGGCTTTTCT TGGGCTCTGG GGTCTTGTTT TCTTCATTTA

143951   TCTGTGCATT ATAAATCCAC TGTGAATTGG GAGGCATGCC AAGACCCATC

144001   AGCATCCAAA TAACCTCCCT CTGGAAAAAA AAAAAGCCAC TCTTTTAGAA

144051   AAAAGAGGAA GACACCACAC CAAAAAACAG ACATTTTTCT GTATTTTATT

144101   GTCCTGCAGC CAACCCAGAC AGACAGACTA GGCATCCTAT CTGCTGGGCC

144151   CCTGTAATTT ACCTTGAATT AGGTTCCCAT CTGATCAGGT ACTTCACTTT

144201   GTGAAGAGGC ACTCCTCCAT CATCTTAGGA TTGCATCCTG GGACATAGAG

144251   TGTTAGAAGT AATATGGTGA AATAGGTGTG AGGATACAAT CTATTGAGGG

144301   GTGGATTGGG TCCTGCAACT TCACCTGCAA TGCAAAGGAA GACAGATGAC

144351   AAAGAAGGTG CTTTCAACAC CATTCCCACA TTCCTTCAAT TGCACAAGCA

144401   GTCCACACCA TGGTCTGATG TTCAGGTGCA TGAACCTGCA AGAAACTTTT

144451   GGTGTGTAAA TTGGGGCCAT CCTAGCTAAC TCCTGATTTG ATAACTTTCA

144501   TACCCAGAGT CAAATGGGAG TGGAATGGAT TGATGCTGTG TGGAATTTGG

144551   CCTCCAGACT TTCCTCTTCT TTTTCTGACT ACCATTTTTC TCATTGACCT

144601   AGGGTTTCCT GAGTCTAGCT CAATTAATTC TACACTAAAC ATTTCCCAGT

144651   TCATGAAAAA CGACCCTCAG GGGTATCCAT TGCATAAGTG TTTTCTTCTT

144701   AACACTGGCT CTCTCACTTC CATTGGAGGG CTGCAAGATT CCTGTAGGAT

144751   AAGAAGCAGG AAGCAATGCC TGGCTTCTGT CTGTTAATCT GGCCTCTTAT

144801   TCTTTTCATC TTCATGGCCT TCTCATTGTG GAGGGGCTCT TTCATTGGTC

         •.........................................................................................•..•

155101   CAGTAGTCTT TGAATATTAC TTGCACTCCA GCACAATTTA CCTTCTTCTT
```

259

155151   TCAGGTGAGC CTTGATTTTT CTTTACTTTT ATTGGGGGTC AAAATTGCCC

155201   TTCAACAGCA CTACTGGACA TCCCTTTCAG GCTTGCAATG GCCACAGATG

155251   GCCACTGAGG TTCTGCCTCA ACTCCATATG CATCCTTGAG AGGCCAGTTC

155301   AAGGTGTGAG AACACTGCTC CACCCAGGAT GTGCCTTTGT CGTGGTTCCT

155351   GCCTTTCCCA GAGAAACCCT GTGAGTCTCA GCATGAAGGG AGGCAGTGAG

155401   GTCAAGAGTC CGACCATCTT TCTCTGACAC CCACCTCTGG GGTCTCAGAT

155451   ATGATTCTAT CACCCAGAGA ATTCTAAACA ACACTCCAGC CTTAATCCAA

155501   TTTCCATGGG AAATTCTTTC ATACACCCTC TTTCAGGAAT GGAGTCAGAA

155551   AAGCAGTTTT CAGTGACTTT GTCAGTCTTG AAATGCCTCC TCCTCCAGTA

155601   AGACCTCACC ACTGAGAAGG CACAAAGGCG CCCTGAGGTA AAAAATTTTA

155651   GGATCCCGGG ATGGGTTTTG ACCGACAGCT TTTTTCTGAT AACAGGCCAG

155701   CTCTGCCTGT ACCATTTTCC TCTGCTTAGA CAGGCTGACA GCTCTGATAG

155751   TCAGGTACCC TGACTTCTTC ATGAATTGGC ATGCATATTC TCAGGGCCCC

155801   AGGCCTGATT CTGAGCTGTG TCTAGCATCA CAATGAATGT CATCGTTGCC

155851   TAGAGACAAG TCCGTTGGTC TTGGTGAAGA AGAACACCAT GGAGGTGGGT

155901   TGGTGGTGGG CTTTTGTCCA TCTTCTTTAT GAGATCCATA AGATAGTTCC

155951   ATGATTCTAG GGGAGGGCAG ACATGAGACA ACCTGAAGAA ACATCAAGCA

    •...• •

156401   GGGGCACATT CTTATTTTAC CATGAATCAT GAATCTGAAG TGAAAGGAGG

156451   GGTGTGCCAA GACCCGCCGA CGTCCTAGTC ACCTCTCCCT GCAAAAGAAG

156501   CCACTCTCCT AGGAAGAAAA GGAGCACAAC ACACAAAATA ACAAACATAT

156551   CCCAGTGTTT CATTGTCCTT CAGCCAAGCC AGGGAGAGAC AGTAGCAGTC

156601   CTCTCCACAG GGGCTCTTGA ATTTAACTCG AATTCTATCC CCAGCAGAGC

156651   AGGTGCTTCA CATCATGAGG TGGTATTCCT CCATCGCCTT GGGATTCCAT

156701   CCTGGGACAT AGAATGTGAG GAGCAATGAG GACAGATAGG GGTGAGGATA

156751   CAATATCCTG AGTGGTGGAT GAGGTCTTGC AGCTTCACTT GCAAAAAAAG

156801   AAATGAGGAC AGATGACAGA GAAGACACTT CCAACCCCAT TCCCACATTC

    •...• •

```
159251   TTGGCCCTGT TCTCACATGG GAGTACTCCA ATGTGCCAGG AATATTTGGA

159301   GGGCAAACTG AGGCTATCCT GGCAAAATCC CAAATTGAGG GCTTTCATAC

159351   TCGGAGCCAA AAATGAGTGG AAAAGACTGA TGCTAGGTTG GGTGTGGCCT

159401   CTACACTTGC CTGTTCTTTT CCTGACTTCC ATGTTCCTCA TCGACATAGG

159451   GTTTCCTGGA TCTGGCTCAG CAACTTCCAC AATAAACATT TCCCAGTTCA

159501   TGGAGAACGA CCCTCATGGG AATCTATTGT GTGAGTGTTT TTTTGTAAAC

159551   ACTATCACGT TTTAATGACT GGGAAGCTTT GATATATTTA CAACCATAAC

159601   TTTCCATTAG AGCCATCAAC AAATAAACTT GTTCTTCCAC TTCTATTGGA

........................................................................

161101   CAAGCGATTC TTTTATATTA AGTCTCAAAA CCTTTTGTTG TCCTGTCTGT

161151   GTGGCTCTCC TTGGGATGCA GGGCTCCATG TTCTTTATTT TTCTGTTGAT

161201   GATGAATCTG CGGTGATTTG GGAGGCATGC CATTACCCAC AAGCATCCAA

161251   GTCACTTTTA CCTTCAAAAA AAGCAACTTT TCTAGAGAAA AGAGGAGCAC

161301   AACAAAACAA AAAAAACAGA CATCTCCCAG TGTTTCTTTT TCCTGCACCA

161351   AAACATGGAG AGATACTAGC AGTCCTAAAT GCATGCCCCC TTGTATTTAC

161401   CTTGAATTCA CTTCCCAGCT GAGCAGGTGC TTCATATCAT GAGGGGGCAA

161451   TCCTTCATCA TCTTGGGATT TCATCTTGAT ACATAGAGTT TGAGCAGCAA

161501   TAAGATCAGA TTGGGTGAGA ATACATTCTG ATGAGGGGTG GATGGGGTCC

161551   CGCAACTTCA GCTGCAAAAA CAAAATAAAT GAAGACAGAT GACACAGAAG

161601   CTTCCAGCAA ATCTGACCTC ACATTCTCTT GTAGCAGGA TGAGCTGCAG

161651   AGAAAATCCG TCAGACACCA GATTAAAGAA GTAAATGGCT TTATTCAGCT

........................................................................

168351   GTGGGGCTGC CTGTTACCAC AGATTATTTA GCTGGAAGGG ATTTCAGACA

168401   GCAAAACGGA CTTCAGGTAG GCTGGCTGTG CTCCAGCATG CGGGTGATTG

168451   TCTCATTTCC AGGGCTGAGG CTGTTTGCAC TTTGCAGTGG ACTTTTGGGT

168501   CCTTTAAGAG GCATCATTGA ACATTGCTTG GACCCAGCAC CAGTTAGCTC

168551   ATTCTCTCAT GTGAGCCTTG GCCTTTCTTT GCTTTCATCA GGAATCTACT

168601   GTGCCCCTCA ACTGCACTAC TGGAAACATT TTCAGCCTGG CCGTCAACAC
```

261

```
168651  AGACAATCTC TGAGACAGGT CTCAACCTCA CCTGCATGCC TAAGAGGCCA

168701  GTCCAAGTTG TGAGAATACC GCTTCACCTT GGACTTGCCA TTGTCCTGGT

168751  TCCTGCCTTT CTTAGAAAGC CTCTGTGAGG CCCAGATGAA GGGGAGCAGT

168801  GAGGTCAACA ACCTGGCCAT CTTTCACTGA CAATAGCTTC TGAGGTCTGA

168851  GGTATGATTC TATCACCCAA AGAACATGCA ACAGCACACC AGACAATATT

168901  CCAATTCCCA TTGGACCTGA TTCTTGCACA CAGCGTTTTT CAGGAGTGGA

168951  GTCAGAAGAG CAGTTTCCAG AGACAACCTC ATAGTCTTCA AATGCCTCCT

169001  GACACAGCAG GATTCAACCA TGGGGTCCAC CCAAAAAGGC CCTGAGCAGT

169051  GGATTTTCAC AGGCAGCTTT TTTCCTGATA CCAGGATGGG TTTGGCTGTA

169101  CCACTTTCCT CTGCTTAGGC AGGCTGACAG CTCTCAAAGC CTGACTCCTG

169151  AGTCTGCCTC AGAAATGGAC ATGTGCTAGT TTGAGGGGAC TAGGCCTGAT

169201  TGTGAGCACT GGTTAGCATC ACAATGAATG TCACCGTTGC CAAGAGACAA

169251  GTCCCTGTGG CTTGGCAGAG AAGATGTCTG TGGAGGTGCA TTGGTGTTGG

169301  ACTCACCTGT CTTCTCTGTG GGATCCATGG GTTTGATCCA TGATCCTAGG

169351  AGAGGGCAGA CATGAGCCAG CCTCAAGAAA TGTCAAACAG AGCCTGAGGA

169401  ATAAACCATG AAATCCCTAA GAATCCAAGA AGATCTGCAG GATTCCTCAG

169451  GCCTGCCTAG ATGTTGTAAA GGTGAGTCTT TTTGAAACTT GTGCTACTGT
```

262

APPENDIX III

Predicted gene structure of ml22a3, based on BESTFIT comparison of the TTY2
identified exons, with sequence from the genomic clone AC023342. Blue colour
indicates the exons; black colour indicates the introns; red colour shows the
exon/intron boundaries; pink colour shows nucleotide sequences that usually are
associated with exon/intron splicing junctions; green colour indicates potential
exon/intron splicing junctions; numbers on the left side of the sequence indicate
nucleotide number on the genomic clone AC023342. Dotted line in between intron
sequence corresponds to intronic sequence that was not included.

```
63501   CAGTGAATCT CAACGGCACT AATATTCACC TTTTTTAGGT TTGCCATCAC
63551   CACAGATGGC CTCAGAGACA CTGTCTCAAC CTCGTCTGTG CCCGTTGGAG
63601   GCCAGTCTGA GGTGTGAGAA CACTGCTTCA CCTTGTAGTT GCCTTTGTCA
63651   TTGTAACTGC CTTTCCCAGA GAGCCCCTGT TAGGCCCAAG ATGAAGGGAG
63701   GCAGTGAGAT CAAGATCTCA GCCATCTTTT GCTGACACAA GCCTCTGGGG
63751   TTGTGTCAGC AAAATAATCC TTGACAACAC ACCAGACTAT ACTGCAATCC
63801   CCATGTGACC TGATTCTTGC ACACACACAT TCTTATTTGG GAATGCAGTC
63851   AGAGAAGCAG TTTGCAGCGA CTATCTCCCA GTCTCAAAAT GCCTCCTCCT
63901   CCGATGAAAA ACTGACCACA GAGACAGCCT GAAGGGGCTC TGATGTAGAG
63951   ACTTTTAGGG TCCCGCAGTG GGTTTTTGCA GTCATATTTT TTCCGATACC
64001   AGGCTGGCTC TGCCTGTATT ATTTTTCTCT CCTTAAGCAG ACTGACTGCT
64051   GTGACAAATA GGCACCTGAG CCTGCTTCAC GAATGCACAT ACACTAGTCT
64101   CAGGGCACCA GGACTGATTG TGAGACCTTG CTAGCATCAC AATGACTGTG
64151   GCTGCTGCAT AGTGAGAAGT CCCTGTGGCT TGGCAGAGAA AGAGACCTTT
64201   GTGGAGGTGC ATCAGTGTTG AACTCTCACC TCTCTTCTCT GTGAGATACA
64251   TGGCATAGTC CCATGATCCT AGGAGAGGGC AGACTTCAGC CAGCCTGAAG
64301   AAATGTCAAG CTCAGCCCCA GGAATAAACC ACAAAATCCC TAAGATCCAA
64351   AAGGATCTGC AGAATTCCTC AGGCCTGCTA GACATTGTAG GGGTGAATCT
```

```
64401   TCTTGAAACT TGCCCCACTA TGATATCTAG GTACATCCCG CCTGTGTTCA

64451   CTGGGGTTGC TCTCTCCCAG GTGGGACATC CTGCAGAATC ACACAGCTTC

64501   AGGTGCTGCC AGGATATTTC TTTCTGTGGG ATTGTTGTGA CTGTTGGATG

64551   TCTGTGTTCG TGTGTGGCAC TGTGTGTGTT TTGTATTTGT GTGTGCCCGT

64601   TAGTGGAGTC TGCTTAAATG CTTAAAAGAA TGTGGCTAAC ACACGTCAGT

64651   GCTTCTTTTT TTTTTTTTTT TTTTTTTTTT TTAGACTGTA ACCCTTTTGT

64701   GGGCACCCTA TATGGCTTTG CTTGGGCTGC GGGACCTCAT GTTCTTGATT

64751   TTTCTGTGGA TCATGAATCC ACTGTGAATT GGGAGGCATG CCAAGACCTG

64801   CCGGCATCCA AATCAACTCC CCCTGCAAAA GAAAAGCCAC TCTTCTAGAA

64851   AGAAGAGGAA CACACCCCAC CAAAAAACAG ACATCTTCCT GTATTTTATT

64901   GTCCTGCAGC CATCTCAGAG AGAGACACTA GCAATCCTGT CCACTGGACC

64951   CCTGGAATTT ACCTCAAATT AGATTCCTAT CTGAGCAGGT ACTTTACATC

65001   ATGAGGGGAC CCTCCTCCAT CATCTTAGGA TTTCATCCTG GGACATACAG

65051   TGTAAGAAGC AATACAGTCA GATGGGGTGA GGATACAATC TGGTGAGAGA

65101   TGGATTGGGT CCTACAACTT CACCTGCAAC ACTGAGAAAG ACAGATGACA

65151   CAGAAGGTGC TTCCAACACC ATTCGCACAT TCCTTTAATT GCACAAGCAG

65201   TCCATACAAT GGCCCGATGT TGAGGTGGGA GTACTCCAAC ACGCCAAAAA

65251   CATTTGGTGT GCAAATTGGG TCCATCCTGG CTAACTCCTG ATTTGATGGC

65301   TTTCATACCC AGAGTCAAAC GGGAGTGCAA CGGAATGACG CTGTGTGGGA

65351   TGTGGCCTCC ACACTTTCCT CTTTGTTTCT GACTTCCATT TTTCTCATCG

65401   GTCTAGGGTT TCCTGAGTTT GGCTCAATGA ATTCCACACT AAACATTTCC

65451   CAGTTCATGA AAATCTGCCC TCATGGGAAT CCATTGTGTG AGTGTTTCCT

65501   TCTAAACTCT GTCACATTTT AATGAATGGA TCTCATTGAT ACTGTTAAAA

65551   GCATAAATTC CTGTTACAGC CTCCAAAAAG AAATTTTAGT TCTCTCACTT

..................................................................

68351   GAACATTGCA TAGACTCCAG CACAAGGTAG CTCATTCTCT CAGGCAAGCC

68401   TTGGGTTTTC TTCGTTTTCA TGGGGAATCC ACATTGCCCC TTAACAGCAC

68451   TACTGGACAC CATTTTCAGG CTGGCCATCC CAACAAACGG CCTCTGAGAC
```

264

```
68501   ACTGTTTCAA CCTCATCTGC ACCCAAGAGA GGGCAGTACA AGGTGGGAGA

68551   ACAGTGATCC AACTCTGACT TGCCTTTGTT GTAGTTTCTG CCTTTCCTAG

68601   AGAGCCCCTG TGAGGTCAAA GTTGAAGGAA GGCAGTGAGG TCAAGAGTCT

68651   GGCCATCTTT TCATGATATC CACTTCTGGG GTCTCAGGTA TGATTCTATC

68701   ACACAAAGAA CCCTCAACAA CACACCAGTC TATATTTTAA TCCCCATGGT

68751   ACTCAATTCT TGCACACAGC CTTTTTCCAG AATGAAATCT GAAGAGCAGT

68801   TTCTAGAAAC TATCTCACAG TCACAAAATG CCTCCTCCTC CAGTGGAAGT

•··········································································································································..·

70801   ACACCCACCT CTGGGGTCTC AGGTATTATT CTATCACTGA AAGAGCTGTC

70851   AACAACACAG CAGACTATAT TCCAATCCCT GCTGGAACGG ATTCTTGCAT

70901   ATAGATCTTT CTTGAATGGA GTCAGAAGAG CAGTTTCCTG TGACCACCTC

70951   ACAGTAACAA AATGCTTCTG CCTCCAGCAG GACCCGACCA CCGAGACAGC

71001   CTGGAGTGTT CACCAGTTTG ACAGTTTTAG GGTCCTGAAG TGAGTGTTTG

71051   CAGGCAGCGT TTTTTTTTTT ACACCAGCAC GGCTCTGCCT TTACCATTTT

71101   CCTCTGCTTA GGTAGGGTGA CAGCTCTGAC AGCACTGCAC CCGAGCCTCC

71151   CTCACGAATG TGCATGTGCT AGTCTTAGGG CACCAGGCCT GATTTTTGAG

71201   TTCTGGCTAG CATCACAGTG CATGCCACCG TTGCCTAGCA ACAAGTCCCT

71251   GCTACTTGGT GGAGAAGATT TCCATGAAAG TGTGTTGGTT TTGATCTCTC

71301   ACCTGTCTTC TCTGTGGAAT CCATGACATA ACCCACATTC CTAGGAGGGG

71351   GCAGACGTGA GCCAGCTTGA AGAGACATCA AGCACAGCCC CAGGAATAAA

71401   CCATAATATC CCTAAGGATG CAAAAGATCT TCAGGATTCC TCAGGACAGC

•··········································································································································..·

76551   CACGCGCGTT TTTATGTGTG TGCCTGTAAG TGGAATCTCC TTAAAGGAAT

76601   GTGGCTAACG AATTTCAGCA CTTTTTTTTT TTTGAGTCTC AGAAACTTTT

76651   GGTGGCCTCT CTGTGTGGGT CTGCTTGGGC TCCAGAGGTC CATGTTCTTT

76701   ATATTTCTGT GGACTATTAA TCCGCAGTGA ATTTGAAGGT GAGCTGATAC

76751   ATGTGGGCAT CCAAGTCACC TCACCCTGCA AGAAAATTGA CTCCTTTAGA

76801   AAGAAAAGGG GCACACCACA CCAAAAAAGA GGCATCTCTC AGTGTTTTAT

76851   TTTCCTGTGG CCAACCCACA GAGAGACAAT AGCAGTCCTG TCTGCAGGTC
```

265

```
76901   CCCTTGAATG TACCTTGAAT TCGGTTCCCA GCCCAGCAGG TGCTTCACGT

76951   CATGAGGAAA CACTCCTCCA TCATTTTGGG ATTTCCTTCT GTGAAATAGA

77001   TTGTGAACAA GTCAGATAGG AGTGAAGATA CAATTTGGTG AAGAGTAAAT

77051   GGTGTCCTGC AACTTCACCT GCCAAAAAAA AAAAAAAAAA AATGAAGGCA

77101   GATGACACAG AAGTGGTTTC CCACTCCATC CCCACATTCT TTTAATTGCA

77151   CAAGCAGTCC ACGCAATGGC CTGGTGTTCA AGTGGAGTCC TGCAACGTGC

77201   CAGGAAAAGT TGGAGTGCAA ATTGGGGCCA TTCTGACAAA CTCCAGATTT

77251   GGGGTTTTTC ATACCTGGAG CCAAATGGGA CTGGAAAGGA CTGATGCTGA

78851   TACTTCCAAC TGCATGCCCA CATTCCCTTA ATTACACAGG CAGTTCATAT

78901   TATGTCCTGG TGTTGAGGTG GGAGTAATAC AATGTTGAGG GAACATTTGG

78951   AGTGCAAGCT GGGGCGGTTC TAACAAGCTC CCGATTTGAG AGCTCTCATA

79001   CCCTGAGCCA AATGAGAGTG GGATGTGTTG ATGCTGGGTG GGATGTGCCC

79051   TGCACACTTG TCTCTTCTTT TCCTGACTTC CATGTTCCTC ATCGGCCTAG

79101   GGCTTCCTGG GTCTGGCTCC ACATCTTCCA CACTAAGCAA TTCACAGTTC

79151   ATGGGGGATG ACCTTCATGG GAATCCAATG CATGAGTGTT TTCTTCTAAA

79201   CACTGTCACA TTTTAAAGAC TGGAAAGATG TGATAATTTT AACACTGTAA

79251   GTTTTCGTTA CAGCCACCAA TAAAGAAACT CTTGTTCTCC CACTTTTATC

79301   AGAGGGCTGC ATGATTCCTC TAGCATGAAA AGCAGGCAGC TGTGTTGGGC

79351   TTTTGCCTGG TAATCTAGCC CCTGTTTCAT TGCATCTGCC AGTCCTTTCT

79401   CAATGCGAAT GAGGTCTTTC ATTGGGCTGT TGCTGGATGG GGTTGCCTCT

80851   AGTCTCCCAA TCATTTGTTG GCCTGTCTGT GTGGCTCTGC TTCAGTTGTG

80901   AGGCTTTGTG TTATTTATTT TTGTGTGGAT CATGAGTCTG CAGCAAATTG

80951   AGAGGTGGGC AGAGACCTGC CAATGTCCAA ATCACCTGCC CCTGCAAAGA

81001   AAACCACACT TCTGTAAAGA AGAGAAGCAC ACCACACCAA AAAACAGACA

81051   TCTCCCAGTG TTTCATTGTC CTGTGTCCCA CCCAGAAAGA AACACTAGCA

81101   GTCCTCTGCA GAAACCCCTG AATTTACATC GAATTTGGCT CCCAGCTGAG
```

266

81151   CGGTTCTTCA TGTCATGATG GGGCACTCCT CCTTTGTCTT GGGATATCAT

81201   CCTGGGACAT AGAATATGAG CAGGGAGAAA TTCAGATATG GGTGAGGATA

81251   CAATCTGGTG AGGAGTGGAT GGGGCTCTGC AAATTCATCT GCAAAAAAAA

81301   AATAAATACA AATGACAAAG ATCCTTCTTT AAACTCCATC CCTGCATTTC

81351   CTTAATTACA CAAGCTGTCC ATACCTTGGC CCAGTGTTCA GGTGGGAGTA

81401   CTCTAATGTG CAAAAAACAT TTGGACTGCA AATTGAAGCC ATCTTGGTAA

81451   ACCCTGGATT TGAGGCCTTT TACACCCAGA GGCAAATGGG AGTGGAATGG

81501   GTTGATGCTG TGTGGGATGT GGCCTCCACA CTGGCCTCTT CTTTTCTGAC

•..•.•

84201   GAATGATTCC TGTATGATGA GAAGCAGGCA GCCGTGTCTA GATTTAGCCT

84251   GGTAATATAG CCTCTGCTTC ATTTCATCTG CTCATCCTTC TCTTTGTGGA

84301   TGGGCTCTTT CATTGGCCTG TTGCTGGATG GGACTGCCTC TTGCCACAGA

84351   TATTTTCGTT GCCAGGAATT TCAGGAAGAA AAATCGTTGT AGAGTAGACT

84401   GGCTGCACGT TGGGTTTTCA ATCATTATCT TGTGGGAGCT GAGGTTGTTT

84451   GCATTTTGAA AGAGGCTTCT TCTTTCTGTG ACAGGAATAT GGAAATGTTG

84501   CTTGGACCCC AGCACAAGTC AGCTCTTTCT TTCAGGTGAT GCTTGATGTT

84551   TCTTGGCTTT CATGGGGGAT TCACATTGCC CCTCAACCAC ACTACTGGAT

84601   ACACTTTTCA GGATTGCAAT CTCCACAGGT GGCCTCTGAG ACACTGTCTG

84651   AACCTCATCT GCACCTGTGA GAGACCAGTT CGAGGTCTGA GAACATTGCT

84701   TCAATTTGGA CGTGCCTTTG TCATGGTTCC TGCTTTTCCC AGATAGCACC

84751   TGAGCAACCC AGAATGAAGG AATGCAGTGA GGTCAAGGCC AGGCCATCTT

84801   TCACTGACAA CTTTTTTCTG GTATTTCAAG TAAGAGCCCA TCACCCAAAG

84851   AACCCTCAAC ATCTCACCGT AATATATTCC AATCCCCATG GGACAGGATT

84901   CTTGCACAAA GCCTCTTTCA GGAATGGAGT CAGAAGAGTA GTTTCCAGTG

84951   ACAACCTCAC AGTCTTGAAA TGGCTCTGCC TCCAGTGTGA CCTGACCATG

85001   GAGACAGCGT ATAAGGGCCC TAAGGTTGAG ACTTTTAGGA TCCTGCAATG

85051   GGTTATCACA GGCAGCCTTT TTCCTGATAC CAGGCCAACT CTGTGTGTAC

85101   CATTTTCCTC TGCTTAGGCA GGCTGACAGC CCTGACACCC TGGTGCTCTA

85151   ATTTGAGTCA CTAATGTGGA TGTGCTAGTC TTAGGGCAAT GGGCCTGAGC

```
85201   TGCGAGCTCT GGCTAGCATT ACAATGAATG CCACCTTTGC CTAGTGACAA

85251   TTCTCTTTGG CTTGATGGAG AAGGAGATCT CTGTGGAGGT GCATCGGCAG

85301   TAGACTCTCA CCTGTCTTCT CCGTGGGATC CATGGGCTAG TCCCATGATC

85351   CTAGGAGAGG GCAGATGTGA GCCTGCCTGA AGAAATGTCA AGCAGAGCCC

85401   CAGGAATAAA GCACAAAATC CGTACAGATC CAAAAGGATC TGCAGAATTC

85451   TTCAGGCCTG CCTAGACATT GTAGCAGTTT GTCTTATTGA AATGTGTCCC

85501   ACTGTAATTT CTAAGTTTAG CCTTCCTTTG TTTTCCAAGC TTTCTCTCTC

85551   CCAGGAGGGG CTTCCTGCAG AATGAAGCAG CCTCAGAAGC TACTGGGCTG

85601   TGTGTTACTG CGGGAGTATT GCAAGTGTTG GATGTCTGCA TGTGTGTGTG

85651   GCTTTATGTG TTTATGTGTG TGTGTCTGTG TGTGTGCATG TAAGTGAATT

85701   CGGCTTAAAG GAATGTAACT AACACACTGC AGTGCTTTTT TATTTTTTAT
```

References

Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaioulo, A., Nigro, V., Simeone, A. & Boncinelli, E. (1989). The human HOX gene. **Nuc. Acids Res. 25**, 10385-10402.

Achermann, J. C.; Ito, M.; Ito, M.; Hindmarsh, P. C. & Jameson, J. L. (1999). A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. **Nature Genet. 22**: 125-126.

Affara, N.A., Florentin, L., Morrison, N., Kwok, K., Mitchell, M., Cook, A., Jamieson, D.,Glasgow, L., Meredith, L., Boyd, E., et al. (1986). Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. **Nucleic Acids Res. 14**, 5353-73.

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 and females with XY karyotypes.
Nucleic Acids Res. 15, 7325-42.

Affara, N.A., Chambers, D., O'Brien, .J., Habeebu, S.S., Kalaitsidaki, M., Bishop, C.E,. & Ferguson-Smith, M.A. (1989). Evidence for distinguishable transcripts of the putative testis determining gene (ZFY) and mapping of homologous cDNA sequences to chromosomes X,Y and 9. **Nucleic Acids Res. 17**, 2987-99.

Affara, N., Bishop, C., Brown, W., Cooke, H., Davey, P., Ellis, N., Graves, J.M., Jones, M., Mitchell, M., Rappold, G., Tyler-Smith, C., Yen, P. & Lau, Y-F.C. (1996). Report of the second international workshop on Y chromosome mapping 1995. **Cytogenet. Cell Genet**. 73, 33-76.

Agulnik, A.I., Mitchell, M.J., Mattel, M-G., Borsani, G., Avner, P.A., Lerner, J.L. & Bishop, C.E. (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. **Hum. Mol. Genet. 3**, 879-884.

Agulnik, A.I., Zharkikh, A., Boettger-Tong, H., Bourgeron, T., McElreavey, K. & Bishop, C.E. (1998). Evolution of the DAZ gene family suggests that Y-linked DAZ plays little, or a limited, role in spermatogenesis but underlines a recent African origin for human populations. **Hum Mol Genet. 7**, 1371-7.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J.D. (1989). **Molecular biology of the cell**. (Garland Publishing, Inc., New York & London).

Alitalo, T., Tiihonen, J., Hakola, P. & de la Chapelle, A. (1988). Molecular characterisation of a Y;15 translocation segregating in a family. **Hum. Genet. 79**, 29-35.

Andersson, S. & Moghrabi, N. (1997). Physiology and molecular genetics of 17 beta-hydroxysteroid dehydrogenases. **Steroids. 62**, 143-7.

Andolz, P., Bielsa, M.A. & Vila, J. (1999). Evolution of semen quality in North-eastern Spain: a study in 222 759 infertile men over a 36 year period. **Hum. Reprod. 14**, 731-735.

Archidiacono, N., Storlazzi, C.T., Spalluto, C., Ricco, A.S., Marzella, R. & Rocchi, M. (1998). Evolution of chromosome Y in primates. **Chromosoma 107**, 241-246.

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. **Nucleic Acids Res. 15**, 8713-8724.

Arnemann, J., Jakubiczka, S., Thüring, S. & Schmidtke, J. (1991). Cloning and sequence analysis of a human Y-chromosome derived, testicular cDNA, TSPY. **Genomics 11**, 108-114.

Askew, D.S., Bartholomew, C., Buchbert, A.M., Valentine, M.B., Jenkins, N.A., Copeland, N.G. & Ihle, J.N. (1991). His-1 and His-2: identification and chromosomal mapping of two commonly rearranged sites of viral integration in a myeloid leukemia. **Oncogene 6**, 2041-2047.

Baccetti, B. Burrini, A.G. Collodel, G., Piomoni, P. & Renieri, T. (1991). A "miniacrosome" sperm defect causing infertility in two brothers. **J. Androl. 12**, 104-111.

Behlke, M.A., Bogan, J.S., Beer-Romero, P. & Page, D.C. (1993). Evidence that the SRY protein is encoded by a isngle exon on the human Y chromosome. **Genomics 17**, 736-739.

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

Bishop, C., Weissenbach, J., Casanova, M., Bernheim, A., & Fellous, M. (1985) DNA sequences and analysis of the human Y chromosome. **Liss A.R, INC., New York**. 141-176.

Blendy, J.A., Kaestner, K.H., Weinbauer, G.F., Nieschlag, E. & Schutz, G. (1996). Severe impairment of spermatogenesis in mice lacking CREM gene. **Nature 380**, 162-165.

Bostock, C.J., Gosden, J.R., & Mitchell, A.R. (1978). Localisation of a male-specific DNA fragment to a sub-region of the human Y chromosome. **Nature 272**, 324-8.

Brand-Arpon, V., Rouquier, S., Massa, H., de Jong, P.J., Ferraz, C., Ioannou, P.A., Demaille, J.G., Trask, B.J. & Giorgi, D. (1999). A genomic region ecompassing a cluster of olfactory receptor genes and a myosin light chain kinase (MYLK) gene is duplicated on human chromosome regions 3q13-q21 and 3p13. **Genomics 56**, 98-110.

Brandell, R.A., Mielnik, A., Liotta, D., Ye, Z., Veeck, L.L., Palermo, G.D. & Schlegel, P.N. (1998). AZFb deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a prognostic genetic test. **Hum. Reprod. 13**, 2812-2815.

Braun, R.E. (1998). Every sperm is sacred - or is it? **Nat. Genet. 18**, 202-204.

Bremner, W.J. (1989). Inhibin: from hypothesis to clinical application. **New Eng. J. Med. 321**, 826-827.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S. & Rastan, S. (1992). The product of the mouse Xist gene is a 15kb inactive x-specific transcript containing no conserved ORF and located in the nucleus. **Cell 71**, 515-526.

Brody, L.C., Abel, K.J., Castilla, L.H., Couch, F.J., McKinley, D.R., Yin, G., Ho, P.P., Merajver, S., Chandrasekharappa, S.C., Xu, J., et al. (1995). Construction of a transcription map surrounding the BRCA1 locus of human chromosome 17. **Genomics. 25**, 238-47.

Brosius, J. (1999). RNAs from all categories generate retrosequences that may be exapted as a novel genes or regulatory elements. **Gene 238**, 115-134.

Brosius, J. (1999b). Many G-protein-coupled receptors are encoded by retrogenes. **TIG 15**, 304-305.

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

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J. & Willard, H.F. (1992). The human XIST gene: analysis of a 17kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. **Cell 71**, 527-542.

Boguski, M.S., Lowe, T.M.J. & Tolstoshev, C.M. (1993). dbEST - database for expressed sequence tags. **Nat. Genet. 4**, 332-333.

Bowles, J., Cooper, L., Berkman, J. & Koopman, P. (1999). Sry requires a CAG repeat domain for male sex determination in Mus musculus. **Nat. Genet. 22**, 405-408.

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.

Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. & Housman, D.E. (1991). Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. **Proc. Natl. Acad. Sci. USA 88**, 4005-4009.

Bujan, L., Mansat, A., Pontonnier, F. & Mieusset, R. (1996). Time series analysis of sperm concentration in fertile men in Toulouse, France between 1977 and 1992. **BMJ 312**, 471-472.

Bull, J.J. (1983). Evolution of sex chromosome differences: models and theory. In **Evolution of sex determining mechanisms**. Cummings, B., Park, M. ed). California.

Burgoyne, P.S. (1982). Genetic homology and crossing over in the X and Y chromosomes of Mammals. **Hum Genet. 61**, 85-90.

Burgoyne, P.S. (1998). The mammalian Y chromosome: a new perspective. **Bioessays. 20**, 363-6.

Burk, R.D., Szabo, P., O'Brien, S., Nash, W.G., Yu, L. & Smith, K.D. (1985). Organization and chromosomal specificity of autosomal homologs of human Y chromosome repeated DNA. **Chromosoma. 92**, 225-233.

Cameron, F. J. & Sinclair, A. H. (1997). Mutations in SRY and SOX9: testis-determining genes. **Hum. Mutat. 9**, 388-395.

Capel, B. (1995). New bedfellows in the mammalian sex determination affair. **TIG 11**, 161-163.

Capel, B., Albrecht, K.H., Washburn, L.L. & Eicher, E.M. (1999). Migration of mesonephric cells into the mammalian gonad depends on Sry. **Mechanisms of Development 84**, 127-131.

Caporale, L.H. (2000). Mutation is modulated: implications for evolution. **BioEssays 22**, 388-395.

Carani, C., Gromoll, J., Brinkworth, M.H., Simoni, M., Weinbauer, G.F. & Nieschlag, E. (1997). CynDAZLA: a cynomolgus monkey homologue of the human autosomal DAZ gene. **Mol. Hum. Reprod. 3**, 479-483.

Carlsen, E., Giwercman, A., Keiding, N. & Skakkeback, N. (1992). Evidence for decreasing quality of semen during past 50 years. **BMJ 305**, 609-612.

Carlson, B.M. (1994). The Urogential system. In: **Human Embryology and Developmental Biology**. 340-371.

Casau, A.E., Vaughan, J.E., Lozano, G. & Levine, A.J. (1999). Germ cell expression of an isolated human endogenous retroviral long terminal repeat of the HERV-K/HTDV family in transgenic mice. **J. Virol. 73**, 9976-9983.

Chai, N.N., Salido, E.C. & Yen, P.H. (1997). Multiple functional copies of the RBM gene family a spermatogenesis candidate on the human Y chromosome. **Genomics 45**, 355-361.

Chai, N.N., Zhou, H., Hernandez, J., Najmabadi, H., Bhasin, S. & Yen, P.H. (1998). Structure and organisation of the RBMY genes on the human Y chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. **Genomics 49**, 283-289.

Chandley, A.C. (1979). The chromosomal basis of human infertility. **Br. Med. Bull. 35**, 181-186.

Chandley, A.C. (1998). Genetic contribution to male infertility. **Hum Reprod. 3**, 76-83.

Chang, P.L., Sauer, M.V. & Brown, S. (1999). Y Chromosome microdeletion in a father and his four infertile sons. **Hum. Reprod. 14**, 2689-2694.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. **Proc. Natl. Acad. Sci. USA 75**, 5618-5622.

Charlesworth, B. (1991). The evolution of sex chromosomes. **Science 251**, 1030-1033.

Chelly, J., Concodet, J-P., Kaplan, J-C. & Khan, A. (1989). Illegitimate transcription: transcription of any gene in any cell type. **Proc. Natl. Acad. Sci. USA 86**, 2617-2621.

Chemes, H.E., Puigdomenech, E.T., Carizza, C., Brugo Olmedo, S., Zanchetti, F. & Hermes, R. (1999). Acephalic spermatozoa and abnormal development of the head-neck attachment: a human syndrome of genetic origin. **Hum. Reprod. 14**, 1811-1818.

Chen H., Rossier, C., Morris, M.A., Scott, H.S., Gos, A., Bairoch, A. & Antonarakis, S.E. (1999). A testis-specific gene, TPTE, encodes a putative transmembrane tyrosine

phosphatase and maps to the pericentromeric region of human chromosomes 21 and 13 and to chromosomes 15, 22 and Y. **Hum. Genet. 105**, 399-409.

Cheng, J-F., Boyartchuk, V. & Zhu, Y. (1994). Isolation and mapping of human chromosome 21 cDNA: progress in constructing a chromosome 21 expression map. **Genomics 23**, 75-84.

Chiannilkulchai, N., Pasturaud, P., Richard, I., Auffray, C. & Beckmann, J.S. (1995). A primary expression map of the chromosome 15q15 region containing the recessive form of limb-girdle muscular dystrophy (LGMD2A) gene. **Hum. Mol Genet. 4**, 717-725.

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. **Anal Biochem. 162**, 156-9.

Christian, S.L., Robinson, w.p., Huang, B., Mutiranngura, A., Line, M.R., Nakao, M., Surti, U., Chakravarti, A. & Ledbetter, D.H. (1995). Molecular characterisation of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. **Am. J. Hum. Genet. 57**, 40-48.

Christian, S.L., Fantes, J.A., Mewborn, S.K., Huang, B. & Ledbetter, D.H. (1999). Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome regions (15q11-q13). **Hum. Mol. Genet., 8**, 1025-1037.

Clark, A.G. (1994). Invasion and maintenance of a gene duplication. **Proc. Natl. Acad. Sci. USA 91**, 2950-2954.

Claverie, J.M.& Makalowski, W. (1994). Alu alert. **Nature. 371**, 752.

Claverie, J.M. (1997). Computational methods for the identification of genes in vertebrate genomic sequences. **Hum Mol Genet. 6**, 1735-44.

Clegg, J.B. (1987). Can the product of the the o gene be a real globin? **Nature 329**, 465-466.

Clepet, C., Schafer, A.J., Sinclair, A.H., Palmer, M.S., Lovell-Badge, R. & Goodfellow, P.N. (1993). The human SRY transcript. **Hum Mol Genet. 2**, 2007-12.

Collingnon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N. & Lovell-Badge, R. (1996). A comparison of the properties of *Sox*-3 with *Sry* and two related genes, *Sox*-1 and *Sox*-2. **Development 122**, 509-520.

Cooke, H,. (1976) Repeated sequence specific to human males. **Nature. 262**, 182-6.

Cooke, H.J. & Smith, B.A. (1986). Variability at the telomeres of the human X/Y pseudoautosomal region. **Cold Spring Harb Symp Quant Biol. 51**, :213-9.

Cooke, H.J., Brown, W.R.A. & Rappold, G.A. (1985). Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. **Nature 317**, 687-691.

Cooke, H.J., Brown, W.A.R. & Rappold, G.A. (1985). Closely related sequences on human X and Y chromosomes outside the pairing region. **Nature 311**, 259-261.

Cooke, H.J., Lee, M., Kerr, S. & Ruggiu, M. (1996). A murine homologue of the human DAZ gene is autosomal and expressed only in male and female gonads. **Hum. Mol. Genet. 5**, 513-516.

Cooke, H.J & Elliott, D.J. (1997). RNA-binding proteins and human male infertility. **Trends Genet. 13**, 87-9.

Cooke, H.J., Hargreave, T. & Elliott, D.J. (1998). Understanding the genes involved in spermatogenesis: a progress report. **Fertil Steril. 69**, 989-95.

Cooper, T.G. (1990). In defense of a function for the human epididymis. **Fertil. Steril. 54**, 965-975.

Corbo, L., Maley, J.A., Nelson, D.O. & Caskey, C.T. (1990). Direct cloning of human transcripts with HnRNA from hybrid cell lines. **Science 249**, 652-655.

Cross, S.H., Charlton, J.A., Nan, X. & Bird, A.P. (1994). Purification of CpG islands using a methylated DNA biding column. **Nat. Genet. 6**, 236-244.

Crowe, M.L., Perry, B.N. & Connerton, I.F. (1996). Olfactory receptor-encoding genes and pseudogenes are expressed in humans. **Gene 169**, 247-249.

Dahl, H.-H. M.; Brown, R. M.; Hutchison, W. M.; Maragos, C.; Brown, G. K. (1990). A testis-specific form of the human pyruvate dehydrogenase E1-alpha subunit is coded for by an intronless gene on chromosome 4. **Genomics 8**, 225-232.

Darlington, C.D. & Haque, A. (1955). Chromosomes of monkeys and men. **Nature, 175**, 32.

Dawson, C. & Whitfield, H. (1996). ABC of Urology: Subfertility and male sexual dysfunction. **B.M.J. 312**, 902-905.

Dechend, F., Schubert, S., Nanda, I., Vogel, T., Schmid, M. & Schmidtke, J. (1998). Organization and expression of rat Tspy. **Cytogenet Cell Genet. 83**, 270-4.

D'Esposito, M., Ciccodicola, A., Gianfrancesco, F., Eposito, T., Falgiello, L, Mazzarella, R., Schlessinger, D. & D'Urso, M. (1996). A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. **Nature Genetics 13**, 227-229.

Deininger, P.L., Jolly, D.J., Rubin, C.M., Friedmann, T. & Schmid, C.W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones. **J. Mol. Biol. 151**, 17-33.

De Jonge, C.J. & Barratt, C.L.R. (1999). WHO manual Who should care? **Hum. Reprod. 14**, 2431-2433.

De Kretser, D.M (1993). Molecular biology of the male reproductive system. In: **testis molecular aspects**. (Academic Press, San Diego, London).

De Kretser, D.M. (1997). Male fertility. **Lancet 340**, 787-790.

De la Chapelle, A., Tippett, P.A., Wetterstrand, G. and Page, D.C. (1984). Genetic evidence of X-Y interchange in a human XX male. **Nature. 307**, 170-1

Delbridge, M.L., Harry, J.L., Toder, R., Waugh O'Neill, R.J., Ma, K., Chandley, A.C. & Marshall Graves, J.A. (1997). A human candidate spermatogenesis genes, RBM1 is conserved and amplified on the marsupial Y chromosome. **Nat. Genet. 15**, 131-136.

Delbridge, M.L., Lingenfelter, P.A., Disteche, C.M. & Marshall Graves, J.A. (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. **Nat. Genet. 22**, 223-224.

Del Castillo, I., Cohen-Salmon, M., Blanchard, S., Lutfalla, G. & Petit, C. (1992). Structure of the X-linked Kallmann syndrome gene and its homologous pseudogene on the Y chromosome. **Nature Genet. 3**, 305-310.

Delobel, B., Djelati, R., Gabriel-Robez, O., Croquette, M-F., Rousseaux-Prevost, R., Rousseaux, J., Rigor, J-M. & Rumpler, Y. (1998). Y-autosome translocation and infertility: usefulness of molecular, cytogenetic and meiotic studies. **Hum. Genet. 102**, 98-102.

Del Mastro, R.G., Wang, L., Simmons, A.D., Gallardo, T.D., Clines, G.A., Ashley, J.A., Hilliard, C.J., Wasmuth, J.J., McPherson, J.D. & Lovett, M. (1995) Human chromosome specific cDNA libraries: new tools for gene identification and genome annotation. **Genome Res. 5**, 185-194.

Del Mastro, R.G & Lovett, M. (1997). Isolation of coding sequences from Genomic regions using direct selection. Chapt. 13. In: **Methods in Molecular Biology, Vol 68**: Gene isolation & Mapping Protocols. Eds J Boultwood, Human Press inc. NJ. USA.

Denny, P., Swift, S., Brand, N., Dabhade, N., Barton, P. & Ashworth, A. (1992). A conserved family of genes related to the testis determining gene, SRY. (1992). **N.A.R. 20**, 2887.

Diatchenko, L, Lau, Y-F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lokyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. & Siebert, P.D. (1996). Suppression subtractive hybridisation: a method for generating differentially regulted or tissue-specific cDNA probes and libraries. **Proc. Natl. Acad. Sci. USA 93**, 6025-6030.

Dix, D.J., Allen, J.W., Collins, B.W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E.H. & Eddy, E.M. (1996). Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis and male infertility. **Proc. Natl. Acad. Sci. USA, 93**, 3264-3268.

Donlon, T.A., Lalande, M., Wyman, A., Bruns, G. & Latt, S.A. (1986). Isolation of molecular probes associated with the chromosome 15 instability in the Prader-Willi syndrome. **Proc Natl Acad Sci U S A. 83**, 4408-12.

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K.N., Beasley, O., Bird, C.P., Blakely, S., Bridgeman, A.M., Buck, D., Burgess, J., Burrill, W.D., O'Brien, K.P. et al. (1999). The DNA sequence of human chromosome 22. **Nature 402**, 489-495.

Del Castillo, I.; Cohen-Salmon, M.; Blanchard, S.; Lutfalla, G.; Petit, C. (1992). Structure of the X-linked Kallmann syndrome gene and its homologous pseudogene on the Y chromosome. **Nature Genet. 2**, 305-310.

Dupressoir, A., Puech, A. & Heidmann, T. (1995). IAP retrotransposons in the mouse liver as reporters of ageing. **Biochim Biophys Acta. 1264**, 397-402.

Dupressoir, A. & Heidmann, T. (1996). Germ line specific expression of intracisternal A-particle retrotransposons in transgenic mice. **Mol. & Cell Biol. 16**, 4495-4503.

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. USA 87**, 8995-8999.

Eberhart, C.G., Maines, J.Z. & Wasserman, S.A. (1996). Meiotic cell cycle requirement for a fly homologue of human deleted in azzospermia. **Nature 381**, 783-785.

Eddy, S.R. (1999). Noncoding RNA genes. **Curr. Opin. Genet. & Dev. 9**, 695-699.

Edwards, Y.H., Putt, W., Fox, M. & Ives, J.H. (1995). A novel human phosphoglucomutase (PGM5) maps to the centromeric region of chromosome 9. **Genomics 30**, 350-3.

Edwards, R.G. & Beard, H.K. (1999). Hypothesis: sex determination and germline formation are committed at the pronucleate stage in mammalian embryos. **Mol. Hum. Reprod. 5**, 595-606.

Egozcue, J., Aragones, J., Caballin, MR. & Goday, C. (1973). Banding patterns of the chromosomes of man and gorilla. **Ann Genet. 16**, 207-10.

Eichler, E.E. (1998). Masquerading repeats: paralogous pitfalls of the human genome. **Genome Res. 8**, 758-762.

Elliot, D.J., Ma, K., Keer, S.M., Thakrar, R., Speed, R., Chandley, A.C. & Cooke, H. (1996). An RBM homologue maps to the mouse Y chromosome and is expressed in germ cells. **Hum. Mol Genet. 5**, 869-974.

Elliot, D.J. & Cooke, H.J. (1997). The molecular genetics of male infertility. **BioEssays 19**, 801-809.

Elliot, D.J., Millar, M.R., Oghene, K., Ross, A., Kiesewetter, F., Pryor, J., McIntyre, M., Hargreave, T.B., Saunders, P.T.K., Vogt, P.H., Chandley, A.C. & Cooke, H. (1997). Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. **Proc. Natl. Acad. Sci. USA 94**, 3848-3853.

Ellis, N. & Goodfellow, P.N. (1989). The mammalian pseudoautosomal region. **TIG 5**, 406-410.

Ellis, N. A.; Tippett, P.; Petty, A.; Reid, M.; Weller, P. A.; Ye, T. Z.; German, J.; Goodfellow, P. N.; Thomas, S.; Banting, G. (1994). PBDX is the XG blood group gene. **Nature Genet. 8**, 285-290.

Erdmann, V.A., Szymanski, M., Hochberg, A., de Groot, N. & Barciszewski, J. (1999). Collection of mRNA-like non-coding RNAs. **Nucleic Acids Res. 27**, 192-195.

Erdmann, V.A., Szymanski, M., Hochberg, A., Groot, N. & Barciszewski, J. (2000). Non-coding, mRNA-like RNAs database Y2K. **Nucleic Acids Res. 28**, 197-200.

Esposito, T., Gianfrancesco, F., Ciccodicola, A., Montanini, L., Mumm, S., D'Urso, M. & Forabosco, A. (1999). A novel pseudoautosomal human gene encodes a putative prtoein similar to Ac-like transposases. **Hum. Mol. Genet. 8**, 61-67.

Fatyol, K., Illes, K., Daimond, D.C., Janish, C. & Szalay, A.A. (2000). Mer22-related sequence elements form pericentric repetitive DNA families in primates. **Mol. Gen. Genet. 262**, 931-939.

Feig, L.A., Bellve, A.R., Horbach Erickson, N. & Klagsbrun, M. (1980). Sertoli cells contain a mitogenic polypeptide. **Proc. Natl. Acad. Sci. USA 77**, 4774-4778.

Feinberg, A.P. & Vogelstein, B..(1984). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. **Anal Biochem. 137**, 266-7.

Feng, G.H., Bailin, T., Oh, J. & Spritz, R.A. (1997). Mouse pale ear (ep) is homologous to human Hermansky-Pudlak syndrome and contains a rare 'AT-AC' intron. **Hum. Mol. Genet. 6**, 793-798.

Ferlin, A., Moro, E., Onisto, M., Toscano, E., Bettella, A. & Foresta, C. (1999). Absence of testicular DAZ gene expression in idiopathic severe testiculopathies. **Hum. Reprod. 14**, 2286-2292.

Ferrari, S., Finelli, P., Rocchi, M. & Bianchi, M.E. (1996). The active gene that encodes human high mobility group 1 protein (HMG1) contains introns and maps to chromosome 13. **Genomics 35**, 367-371.

Fincham, A.G., Bessem, C.C., Lau, E.C., Pavlova, X. Shuler, C., Slavkin, H.C. & Snead, M.L. (1991). Human developing enamel proteins exhibit a sex-linked diamorphism. **Calcif Tissue Int. 48**, 288-290.

Fisch, H. & Goluboff, E.T. (1996). Geographic variations in sperm counts: a potential cause of bias in studies of semen quality. **Fertility & Sterility 65**, 1044-1046.

Fisch, H.. Goluboff, E.T., Olson, J.H., Feldshuh, J., Broder, S.J. & Barad, D.H. (1996). Semen analyses in 1,283 men from the United States over a 25-year period: no decline in quality. **Fertility & Sterility 65**, 1009-1014.

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.

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.

Fonknechten, N., Chelly, J., Lepercq, J., Kahn, A., Kaplan, J-C., Kitzis, A. & Chomel, J-C. (1992). CFTR illegitimate transcription in lymphoid cells: quantification and applications to the investigation of pathological transcripts. **Hum. Genet. 88**, 508-512.

Foote, S., Vollrath, D., Hilton, A. & Page, D.C. (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region. **Science 258**, 60-6.

Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y-L. & Postelthwait, J. (1999). Preservation of duplicate genes by complementary , degenerative mutatons. **Genetics 151**, 1531-1545.

Foster, J.W., Brennan, F.E., Hampikian, G.K., Goodfellow, P.N., Sinclair, A.H., Lovell-Badge, Selwood, T., Renfree, M.B., Cooper, R.W. & Marshall Graves, J.A. (1992). Evolution of sex determination and the Y chromosome: SRY-related sequences in marsupials. **Nature 359**, 531-533.

Foster, J.W. & Marshall Graves, J.A. (1994). An SRY related sequence on the marsupial X chromosome: implications for the evolution of the mammalian testis-determining gene. **Proc. Natl. Acad. Sci. USA 91**, 1927-1931.

Foster, J.W., Dominguez-Steglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanovic, M., Weissenbach, J., Mansour, S., Young, I.D., Goodfellow, P.N., Brook, J.D. & Schafer, A.J. (1994). Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. **Nature 372**, 525-530.

Foulkes, N.S., Mellstrom, B., Benusiglio, E. & Sassone-Corsi, P. (1992). Developmental switch of CREM function during spermatogenesis: from antagonist to activator. **Nature 355**, 80-84.

Foulkes, N.S., Schlotter, F., Pevet, P. & Sassone-Corsi, P. (1993). Pituitary hormone FSH directs the CREM functional switch during spermatogenesis. **Nature 362**, 264-267.

Franco, B., Meroni, G., Parenti, G., Levilliers, J., Bernard, L., Gebbia, M., Cox, L., Maroteaux, P., Sheffield, L., Rappold, G.A., Andria, G., Petit, C. & Ballabio, A. (1995). A cluster of sulfatase genes on Xp22.3: mutations in chondrodysplasia punctata (CDPX) and implications for warfarin embryopathy. **Cell 81**, 15-25.

Frayne, J., Dimsey, E.A.C., Jury, J.A. & Hall, L. (1999). Transcripts encoding the sperm surface protein tMDC II are non-functional in the human. **Biochem. J. 341**, 771-775.

Freije, D., Helms, C., Watson, M.S. & Donis-Keller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. **Science 258**, 1784-1787.

Frommer, M., Prosser, J. & Vincent, P.C. (1984). Human satellite I sequences include a male specific 2.47 kb tandemly repeated unit containing one Alu family member per repeat. **Nucleic Acids Res. 12**, 2887-900.

Fuster, C.D., Farrell, D., Stern, F.A. & Hecht, N.B. (1977). RNA polymerase reactivity in bovine spermatozoa. **J. Cell Biol. 74**, 698-706.

Gala, J-L., Heusterpreute, M., Loric, S., Hanon, F., Tombal, B., Van Cangh, P., De Nayer, P. & Philippe, M. (1998). Expression of prostate specific antigen and prostate-specific membrane antigen transcripts in blood cells: implications for the detection of hematogenous prostate cells and standardization. **Clin. Chem. 44**, 472-481.

Gantla, G., Bakker, C.T.M., Deocharan, B., Thummala, N.R., Zweiner, J., Sinaasappel, M., Chowdhury, J.R., Bosma, P.J. & Chowdhury, N.R. (1998). Splice-site mutations: a novel genetic mechanism of Crigler-Najjar Syndrome type 1. **Am. J. Hum. Genet. 62**, 585-592.

Gentles, A.J. & Karlin, S. (1999). Why are human G-protein-coupled receptors predominantly intronless? **TIG 15**, 47-49.

Geraghty, D.E., Koller, B.H., Hansen, J.A. & Orr, H.T. (1992). The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. **J. Immunol. 15**, 1934-1946.

Gianfrancesco, F., Esposito, T., Montanini, L., Ciccodicola, A., Mumm, S., Mazzarella, R., Rao, E., Giglio, S., Rappold, G. & Forabosco, A. (1998). A novel pseudoautosomal gene encoding a putative GTP-binding protein resides in the vicinity of the Xp/Yp telomere. **Hum. Mol. Genet. 7**, 407-414.

Glaser, B. Yen, P.H. & Schempp, W. (1998). Fibre-fluorescence in situ hybridization unravels apparently seven DAZ genes of pseudogenes clustered within a Y-chromosome region frequently deleted in azoospermic males. **Chromosome Res. 6**, 481-486.

Goodfellow, P., Darling , S. & Wolfe, J. (1985) The human Y chromosome. **J. Med. Genet. 5**, 329-44.

Goodfellow, P. N. & Lovell-Badge, R. (1993). SRY and sex determination in mammals. **Annu. Rev. Genet. 27**, 71-92.

Goodman, M., Porter, C.A., Czelusniak, J., Page, S.L., Schneider, H., Shoshani, J. Gunnell, G. & Groves, C.P. (1998). Towards a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. **Mol. Phylogenet. & Evol. 9**, 585-598.

Goodman, M. (1999). The genomic record of humankind's evolutionary roots. **Am. J. Hum. Genet. 64**, 31-39.

Gough, N.M., Gearing, D.P., Nicola, N.A., Baker, E., Pritchard, M., Callen, D.F. & Sutherland, G.R. (1990). Localization of the human GM-CSF receptor gene to the X-Y pseudoautosomal region. **Nature 345**, 734-736.

Graves, J.A.M. (1998). Interactions between SRY and SOX genes in mammalian sex determination. **Bioessays 20**, 264-269.

Graves, J.A.M., Disteche, C.M. & Toder, R. (1998). Gene dosage in the evolution and function of mammalian sex chromosomes. **Cytogenet. Cell Genet. 80**, 94-103.

Graves, J.A.M., Wakefield, M.J. & Toder, R. (1998). The origin and evolution of the pseudoautosomal regions of human sex chromosomes. **Hum. Mol. Genet. 76**, 1991-1996.

Grays Anatomy. 38th Edition. Churchill Livingstone.

Griswold, M.D. (1995). Interactions between germ cells and Sertoli cells in the testis. **Biol Reprod. 52**, 211-6.

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.

Guimera, J., Pucharcos, C., Domenech A., Casas, C., Solans, A., Gallaardo, T., Ashley, J., Lovett, TM., Estivill, X. & Pritchard, M. (1997). Cosmid contig and transcriptional map of three regions of human chromosome 21q22: identification of 37 novel transcripts by direct selection. **Genomics 45**, 59-67.

Guttenbach, M., Muller, U. & Schmid, M. (1992). A human moderately repeated Y-specific DNA sequence is evolutionarily conserved in the Y chromosome of the great apes. **Genomics 13**, 363-7.

Haber, D. A.; Sohn, R. L.; Buckler, A. J.; Pelletier, J.; Call, K. M.; Housman, D. E. (1991). Alternative splicing and genomic structure of the Wilms tumor gene WT1. **Proc. Nat. Acad. Sci. 88**, 9618-9622:

Habermann, B., Mi, H.F., Edelmann, A., Bohring, C., Backert, I.T., Kiesewetter, F., Aumuller, G., Vogt, P.H. (1998). DAZ (Deleted in AZoospermia) genes encode proteins located in human late spermatids and in sperm tails. **Hum Reprod. 13**, 363-9.

Hacker, A., Cape,l B., Goodfellow ,P. & Lovell-Badge, R. (1995). Expression of Sry, the mouse sex determining gene. **Development. 121**, 1603-14.

Haqq, C.M., King, C-Y., Donahoe, P.K. & Weiss, M.A. (1993). SRY recognizes conserved DNA sites in sex-specific promoters. **Proc. Natl. Acad. Sci. USA 90**, 1097-1101.

Haqq, C.M., King, C-Y., Ukiyama, E., Falsafi, S., Haqq, T.N., Donahoe, P.K. & Weiss, M.A. (1994). Molecular basis of mamalian sexual determination: activation of Mullerian inhibiting substance gene expression by SRY. **Science 266**, 1494-1500.

Hao, Y., Crenshaw, T. Moulton, T., Newcomb, E. & Tycko, B. (1993). Tumour-suppressor activity of H19 RNA. **Nature 365**, 764-767.

Hara, E., Kato, T., Nakada, S., Sekiya, S. & Oda, K. (1991). Subtractive cDNA cloning using oligo(dT)30-latex and PCR: isolation of cDNA clones specific to undifferentiated human embryonal carcinoma cells. **Nuc. Acids Res. 19**, 7097-7104.

Hardison, R.C., Sawada, I., Cheng, J.F., Shen, C.K. & Schmid, C.W. (1986). A previously undetected pseudogene in the human alpha globin gene cluster. **Nuc Acid Res. 25**, 1903-1911.

Harry, J.L., Koopman, P., Brennan, F.E., Marshall Graves, J.A. & Renfree, M.B. (1995). Widespread expression of the testis-determining gene SRY in a marsupial. **Nat. Genet. 11**, 347-349.

Hassold, T.J., Sherman, S.L., Pettay, D., Page, D.C. & Jacobs, P.A. (1991). XY chromosome nondisjunction in man is associated with diminished recombination in the pseudoautosomal region. **Am J Hum Genet.;49**, 253-60.

Heard, E., Lovell-Badge, R. & Avner, P. (1999). Anti-Xistentialism. **Nat. Genet. 21**, 343-344.

Hecht, N.B. (1998). Molecular mechanisms of male germ cell differentiation. **BioEssays 20**, 555-561.

Hendriksen, P.J., Hoogerbnrugge, J.W., Baarends, W.M., de Boer, P., Vreeburg, J.T., Vos, E.A., van der Lende, T. & Grootegoed, J.A. (1997). Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. **Genomics 41**, 350-359.

Henke, A., Wapenaaar, M., van Ommen, G-J., Maraschio, P., Camerino, G. & Rappold, G. (1991). Deletions within the pseudoautosomal region help map three new markers and indicate a possible role of this region in linear growth. **Am. J. Hum. Genet. 49**, 811-819.

Hildmann, T., Kong, X., O'Brien, J., Riesselman, L., Christensen, H.M., Dagand, E., Lehrach, H. & Yaspo, M.L. (1999). A contiguous 3-Mb sequence-ready map in the S3-MX region on 21q22.2 based on high- throughput nonisotopic library screenings. **Genome Res. 9**, 360-372.

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

Hull, MG.R., Gazener, C.M.A., Kelly, N.J., Conway, D.I., Foster, P.A., Hinton, R.A.,Coulson, C., Lambert, P.A., Watt, E.M. & Desai, K.M. (1985). Population study of causes, treatment and outcome of fertility. **BMJ 291**, 1693-1697.

Huq, A.H.M.M., Sutcliffe, J.S., Nakao, M., Shen, Y., Gibbs, R.A. & Beaudet, A.L. (1997). Sequencing and functional analysis of the SNRPN promoter: in vitro methylation abolishes promoter activity. **Genome Res. 7**, 642-648.

Hurst, L.D. & Smith, N.G.C. (1999). Molecular evolutionary evidence that H19 mRNA is functional. **Trends in Genetics, 15**, 134-135.

Irvine, D.S. (1997). Declining sperm quality: a review of facts and hypotheses. **Ballieres Clin. Obs & Gynae. 11**, 655-671.

Irvine, S., Cawood, E., Richardson, D., MacDonald, E. & Aitken, J. (1996). Evidence of deteriorating semen quality in the United Kingdom: birth cohort study in 577 men ain Scotland over 11 years. **BMJ 312**, 467-471.

Jackson, I.J. (1991). A reappraisal of non-consensus mRNA splice sites. **Nucleic Acids Res. 19**, 3795-3798.

Jakubiczka, S. Schnieders, F. & Schmidtke, J. (1993). A bovine homologue of the human TSPY gene. **Genomics 17**, 732-735.

Jegalian, K. & Page, D.C. (1998). A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. **Nature 394**, 776-780.

Jequier, A.M. (1993). Male infertility. **Brit. J. Obs. & Gynae. 100**, 612-614.

Jequier, A.M. (1997). Clinical assessment of male infertility in the era of intracytoplasmic sperm injection. **Ballieres Clinical Obs. & Gynae. 11**, 617-627.

Jimenez, R., Sanchez, A., Burgos, M. & de la Guardia, R.D. (1996). Puzzling out the genetics of mammalian sex determination. **TIG 12**, 164-166.

Jimenez, R. & Burgos, M. (1998). Mammalian sex determination: joining pieces of the genetic puzzle. **BioEssays 20**, 696-699.

Johnson, A.D. & Gomes, W.R. (1977). The Testis. In: **Advances in physiology, biochemistry and function**. Eds Johnson, A.D., Gomes, W.R. & Vandemark, N.L. Academic Press. New York.

Johnson, M.D. (1998). Genetic risks of intracytoplasmic sperm injection in the treatment of male fertility: recommendations for genetic counselling and screening. **Fertility & Sterility 70**, 397-411.

Jones, M.H., Khwaja, O.S.A., Briggs, H., Lambson, B., Davey, P.M., Chalmers, J., Zhou, C-Y., Walker, E.M., Zhang, Y., Todd, C, Ferguson-Smith, M.A. & Affara, N.A. (1994). A set of ninety-seve overlapping yeast artificial chromosome clones spanning the human Y chromosome euchromatin. **Genomics 24**, 266-275.

Jones, M.H., Davey, P.M., Aplin, H. & Affara, N.A. (1995). Expression analysis genomic structure and mapping to 7q31 of the human sperm adhesion molecule gene SPAM1. **Genomics 29**, 796-800.

Jones, M.H., Furlong, R.A., Burkin, H., Chalmers, I.J., Brown, G.M., Khwaja, O. & Affara, N.A. (1996). The drosophila developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. **Hum. Mol. Genet. 5**, 1695-1701.

Jones, M.H., Zhang, Y., Tirosvoutis, K.N., Davey, P.M., Webster, A.R., Walsh, D., Spurr, N.K. & Affara, N.A. (1997). Chromosomal assignment of 311 sequences transcribed in human adult testis. **Genomics 40**, 155-167.

Juan, V., Crain, C. & Wilson, C. (2000). Evidence for evolutionarily conserved secondary structure in the H19 tumor suppressor RNA. **N.A.R. 28**, 1221-1227.

Jury, J.A., Frayne, J. & Hall, L. (1997). The human fertilin α gene is non functional: implications for its proposed role in fertilisation. **Biochem. J. 321**, 577-581.

291

Kamischke, A., Gromoll, J., Simoni, M., Behre, H.M. & Nieschlag, E. (1999). Transmission of a Y chromosomal deletion involving the deleted in azoospermia (DAZ) and chromodomain (CYD1) genes from father to son through intracytoplasmic sperm injection: case report. **Hum. Reprod. 14**, 2320-2322.

Kent-First, M., Juallem, A., Shultz, J., Pryor, J., Roberts, K., Nolten, W., Meisner, L., Chandley, A., Gouchy, G., Jorgensen, L., Havighurst, T. & Grosch, J. (1999). Defining regions of the Y chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. **Mol. Reprod. Dev. 53**, 27-41.

Kierszenbaum, K.L. & Tres, L.L. (1975). Structural and transcriptional features of the mouse spermatid genome. **J. Cell Biol. 65**, 258-270.

King, C.Y. & Weiss, M.A. (1993). The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: a topological mechanism of sequence specificity. **Proc Natl Acad Sci U S A 90**, 11990-4.

Kipling, D., Wilson, H.E., Thomson, Lee, M., Perry, J., Palmer, S., Ashworth, A. & Cooke, H.J. (1996). Structural variation of the pseudoautosomal region between and within inbred mouse strains. **Proc. Natl. Acad. Sci. USA 93**, 171-175.

Kleiman, S.E., Yogev, L., Gamzu, R., Hauser, R., Botchan, A., Lessing, J.B., Paz, G. & Yavetz, H. (1999). Genetic evaluation of infertile men. **Hum Reprod. 14**, 33-8.

Klink, A., Schiebel, K., Winkelmann, M., Rao, E., Horsthemke, B., Ludecke, H-J., Claussen, U., Scherer, G. & Rappold, G. (1995). The human protein kinase gene PKX1 on Xp22.3 displays Xp/Yp homology and is a site of chromosomal instability. **Hum. Mol. Genet. 4**, 869-878.

Kloc, M., Spohr, G. & Etkin, L.D. (1993). Translocation of repetitive RNA sequences with the germ plasm in Xenopus oocytes. **Science 262**, 1712-4.

Kloc, M. & Etkin, L.D. (1994). Delocalization of Vg1 mRNA from the vegetal cortex in Xenopus oocytes after destruction of Xlsirt RNA. **Science. 265**, 1101-3.

Kobayashi, K., Mizuno, K., Hida, A., Komaki, R., Tomita, K., Matsushita, I, Namiki, M., Iwamoto, T.,Tamura, S., Minowada, S., et al. (1994). PCR analysis of the Y chromosome long arm in azoospermic patients: evidence for a second locus required for spermatogenesis. **Hum Mol Genet. 3**, 1965-7.

Kold Jensen, T., Henrik, N., Hollund, I., Brink Henriksen, T., Scheike, T., Kolstad, H., Giwercman, A., Ernst, E., Bonde, J.P., Skakkebaek, N.E. & Olsen, J. (1998). Does moderate alcohol consumption affect fertility? Follow up study among couples planning first pregnancy. **BMJ 317**, 505-510.

Koller, P.C. & Darlington, C.D. (1934). The genetical and mechanical properties of the sex chromosomes. **J. Genet. 29**, 159-173.

Komine, Y., Tanaka, N.K., Yano, R., Takai, S., Yuasa, S., Shiroishi, T., Tsuchiya, K. & Yamamori, T. (1999). A novel type of non-coding RNA expressed in the rat brain. **Mol. Brain Res. 66**, 1-13.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P.N. & Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for SRY. **Nature 351**, 117-121.

Kramer, J.A. & Krawetz, S.A. (1997). RNA in spermatozoa: implications for the alternative haploid genome. **Mol. Hum. Reprod. 3**, 473-478.

Krause, R., Hemberger, M., Himmelbauer, H., Karscheuer, V. & Fundele, R.H. (1999). Identification and characterization of G90, a novel mouse RNA that lacks an extensive open reading frame. **Gene 232**, 35-42.

Krausz, C. & McElreavey, K. (1999). Y chromosome and male infertility. **Front Biosci. 15**; E1-8.

Kreidberg, J.A., Sariola, H., Loring, J.M., Maeda, M., Pelletier, J., Housman, D. & Jaenisch, R. (1993). WT-1 is required for early kidney development. **Cell 74**, 679-691.

Kremer, J.A.M., Tuerlings, J.H.A.M., Meuleman, E.J.H., Schoute, F., Mariman, E., Smeets, D.F.C.M., Hoefsloot, L.H., Braat, D.D.M. & Merkus, H.M.W.M. (1997). Microdeletions of the Y chromosomes and intracytoplasmic sperm injection: from gene to clinic. **Hum. Reprod. 12**, 687-691.

Kumar, G., Patel, D. & Naz, R.K. (1993). c-MYC mRNA is present in human sperm cells. **Cell Mol. Biol. Res. 39**, 111-117.

Kurinczuk, J.J. & Bower, C. (1997). Birth defects in infants conceived by intracytoplasmic sperm injection: an alternative interpretation. **BMJ 315**, 1260-1265.

Kvaloy, K., Galvagni, F. & Brown, W.R.A. (1994). The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. **Hum. Mol. Genet. 3**, 771-778.

Kwok, C., Weller, P.A., Guioli, S., Foster, J.W., Mansour, S., Zuffardi, O., Punnett, H.H., Dominguez-Steglich, M.A., Brook, J.D., Young, I.D., Goodfellow, P.N. & Schafer, A.J. (1995). Mutations in SOX9, the gene responsible for campomelic dysplasia and autosomal sex reversal. **Am. J. Hum. Genet. 57**, 1028-1036.

Lagerstrom, M., Dahl, N., Nakahori, Y., Nakagome, Y., Backman, B., Landegren, U. & Pettersson, U. (1991). A deletion in the amelogenin gene (AMG) causes X-linked amelogenesis imperfecta (AIH1). **Genomics 10**, 971-975.

Lahn, B.T. & Page, D.C. (1997). Functional coherence of the human Y chromosome. **Science 278**, 675-680.

Lahn, B.T. & Page, D.C. (1999). Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. **Nat. Genet. 21**, 429-433.

Lahn, B.T. & Page, D.C. (1999). Four evolutionary strata on the human X chromosome. **Science 286, 964-967**.

Lala, D. S.; Rice, D. A. & Parker, K. L. (1992). Steroidogenic factor 1, a key regulator of steroidogenic enzyme expression, is the mouse homolog of fushi tarazu-factor 1. **Molec. Endocr. 6**, 1249-1258.

Lalonde, L., Langlais, J., Antaki, P., Chapdelaine, A., Roberts, K.D. & Bleau, G. (1988). Male infertility associated with round-headed acrosomeless spermatozoa. **Fertil. Steril. 49**, 316-321.

Lambson, B., Affara, N.A., Mitchell, M. & Ferguson-Smith, M.A. (1992). Evolution of DNA sequence homologies between the sex chromosomes in primate species. **Genomics 14**, 1032-1040.

Lance, V.A. (1997). Sex determination in reptiles: an update. **Am. Zool. 37**, 504-513.

Lankenau, S., Corces, V.G. & Lankenau, D.H. (1994). The Drosophila micropia retrotransposon encodes a testis specific antisense RNA complementary to reverse transcriptase. **Mol. Cell Biol. 14**, 1764-1775.

Lapenta, V., Sossi, V., Gosset, P., Vayssettes, C., Vitali, T., Rabatel, N., Tassone, F., Blouin, J.L., Scott, H.S., Antonarakis, S.E., Creau, N. & Brahe, C. (1998). Construction of a 2.5-Mb integrated physical and gene map of distal 21q22.3. **Genomics 49**, 1-13.

Lahrini, D.K. & Nurnberger, J.I. (1991). A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. **Nucleic Acids Res. 19**, 5444.

Lau, E.C., Mohandas, T.K., Shapiro, L.J., Slavkin, H.C. & Snead, M.L. (1989). Human and mouse amelogenin gene loci are on the sex chromosomes. **Genomics 4**, 162-168.

Lau, Y-F.C. (1999). Sex chromosomes genetics '99: Gonadoblastoma, testicular and prostate cancers, and the TSPY gene. **Am. J. Hum. Genet. 64**, 921-927.

Laval, S.H., Glenister, P.H., Rasberry, C., Thornton, C.E., Mahadevaiah, S.K., Cooke, H.J., Burgoyne, P.S. & Cattanach, B.M. (1995). Y chromosome short arm-Sxr recomination in XSxr/Y males causes deletion of Rbm and XY female sex reversal. **Proc. Natl. Acad. Sci. USA 92**, 20403-20407.

Lee, M., Haugen, H.S., Clegg, C.H. & Braun, R.E. (1995). Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. **Proc. Natl. Acad. Sci. USA. 92**, 12451-12455.

Lee, J.D., Kamiguchi, Y. & Yanagimachi, R. (1996). Analysis of chromosome constitution of human spermatozoa with normal and aberrant head morphologies after injection into mouse oocytes. **Hum. Reprod. 11**, 1942-1946.

Lee, J.T., Davidow, L.S. & Warshawsky, D. (1999). Tsix, a gene antisense to Xist at the X-inactivation centre. **Nat. Genet. 21**, 400-404.

Leibovitch, M.P., Nguyen, V.C., Gross, M.S. Sholhonne, B., Leibovitch, S.A. & Bernheim, A. (1991). The human ASM (adult skeletal muscle) gene: expression and chromosomal assignment to 11p15. **Biochem. Biophys. Res. Comm. 180**, 1241-1250.

Leroy, P., Seboun, E., Mattei, M.G., Fellous, M. & Bishop, C.E. (1987). Testis-specific transcripts detected by a human Y-DNA derived probe. **Development 101**, 177-183.

Liang, P. & Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. **Science 257**, 967-971.

Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L. & Quackenbush, J. (2000). Gene index analysis of the human genome estimates approximately 120,000 genes. **Nature Genetics 25**, 239-240.

Liao, D. (1999). Molecular evolution '99. Concerted evolution: Molecular mechanism and biological implication. **Am. J. Hum. Genet. 64**, 24-30.

Liao, D., Pavelitz, T., Kidd, J.R., Kidd, K.K. & Weiner, A.M. (1997). Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare intechromosomal gene conversion. **The Embo J**. 16, 588-598.

Lifschytz, E. & Lindsley, D.L. (1972). The role of X-chromosome inactivation during spermatogenesis. **Proc. Natl. Acad. Sci. USA, 69**, 182-186.

Lin, T., Calkins, J.K., Morris, P.L., Vale, W. & Bardin, C.W. (1989). Regulation of Leydig cell function in primary culture by inhibin and activin. **Endocrinology 125**, 2134-2140.

Lindsay, S., & Bird, A.P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. **Nature 327**, 336-8.

Little, M., Holmes, G. & Walsh, P. (1999). WT1: what has the last decade told us? **BioEssays 21**, 191-202.

Liu, P., Legerski, R. & Siciliano, M.J. (1989). Isolation of human transcribed sequences form human-rodent somatic cell hybrids. **Science 246**, 813-815.

Liu, A.Y., Torchia, B.S., Migeon, B.R. & Siliciano, R.F. (1997). The human NTT gene: identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. **Genomics 39**, 171-184.

Lloyd, F.H., Powell, P. & Murdoch, A.P. (1996). Lesson of the Week: Anabolic steroid abuse by body builders and male subfertility. **BMJ 313**, 100-101.

Lovell-Badge, R. & Robertson, E.. (1990). XY female mice resulting from a heritable mutation in the primary testis-determining gene, Tdy. **Development 109**, 635-646.

Lovell-Badge, R. (1992). Testis determination: soft talk and kinky sex. **Current Opin. Genet. & Dev. 2**, 596-601.

Lovett, M., Kere, J. & Hinton, L.M. (1991). Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. **Proc. Natl. Acad. Sci. USA 58**, 9628-9632.

Lovett, M. (1994). Fishing for complements: finding genes by direct selection. **TIG 10**, 353-357.

Lucchesi, J.C. (1978). Gene dosage compensation and the evolution of sex chromosomes in diploid organisms compensation evolved gradually and led to heteromorphic sex chromosomes. **Science 202**, 711-716.

Lundberg, Y., Ritzen, M., Harlin, J. & Wedell, A. (1996). Novel missense mutation (P131R) in the HMG box of SRY in XY sex reversal. **Mutation Notes** S328 - S329.

Luo, X., Ikeda, Y. & Parker, K.L. (1994). A cell specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. **Cell 77**, 481-490.

Ma, K., Sharkey, A., Kirsch, S., Vogt, P., Keil, R., Hargreave, T.B,, McBeath, S. & Chandley, A.C. (1992). Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. **Hum Mol Genet. 1**, 29-33.

Ma, K., Inglis, J.D., Sharkey, A., Bickmore, W.A., Hill, R.E., Prosser, E.J., Speed, R.M., Thomson, E.J., Jobling, M., Taylor, K., Wolfe, J., Cooke, H.J., Hargreave, T.B. & Chandley, A.C. (1993). A Y Chromosome gene family with RNA-binding protein homology: candidates for the Azoospermia factor AZF controlling human spermatogenesis. **Cell 75**, 1287-1295.

MacLean, H., Warne, G.L. & Zajact, J.D. (1997). Intersex disorders: shedding light on male sexual differentiation beyond SRY. **Clinical Endocrinology 46**, 101-108.

Maeda, T., Ohno, M., Ishibashi, A., Samejima, M. & Sasaki, K. (1976). Ring Y chromosome: 45,X/46,Xr(Y) chromosome mosaicism in a phenotypically normal male with azoospermia. **Hum. Genet. 34**, 99-102.

Manz, E., Schnieders, F., Muller Brechlin, A. & Schmidtke, J. (1993). TSPY related sequences represent a microheterogeneous gene family organized as constitutive elements in DYZ5 tandem repeat units on the human Y chromosome. **Genomics 17**, 726-731.

Marshall Graves, J.A. (1995). The origin and function of the mammalian Y chromosome and Y-borne genes - an evolving understanding. **BioEssays 17**, 311-320.

Marshall Graves, J.A. (1995). The evolution of mammalian sex chromosomes and the origin of sex determining genes. **Phil. Trans. R. Soc. Lond. 350**, 305-312.

Marshall Graves, J.A. (1998). Interactions between SRY and SOX genes in mammalian sex determination. **BioEssays 20**, 264-269.

Marshall Graves, J.A. & Watson, J.M. (1991). Mammalian sex chromosomes: evolution of organization and function. **Chromosoma 101**, 63-68.

Martin, S.J., Rassoulzadegan, M. & Cuzin F. (1998). Germ cell specific enhancer activity of a repeated element in a variable region of the mouse genome. **Gene 221**, 85-92.

Martineau, J., Nordqvist, K., Tilmann, C., Lovell-Badge, R. & Capel, B. (1997). Male-specific cell migration into the developing gonad. **Current Biology 7**, 958-968.

Massagué, J. (1998). TGF-ß signal transduction. **Annu. Rev. Biochem. 67**, 753-791.

Matsubara, K. & Okubo, K. (1993). cDNA analyses in the human genome project. **Gene 135**, 265-274.

Maxson, S.C. (1990). The evolution of the mammalian Y chromosome. **Behaviour Genetics 20**, 109-123.

Mazeyrat, S. & Mitchell, M.J. (1998). Rodent Y chromosome TSPY gene is functional in rat and non-functional in mouse. **Hum. Mol. Genet. 7**, 557-562.

McCarrey, J.R. & Thomas, K. (1987). Human testis specific PKG gene lacks introns and possesses characteristics of a processed gene. **Nature 326**, 501-505.

McCarrey, J.R., Berg, W.M., Paragioudakis, S.J., Zhang, P.L., Dilworth, D.D., Arnold, B.L., Rossi, J.J. (1992). Differential transcription of Pgk genes during spermatogenesis in the mouse. **Dev Biol. 154**, 160-8.

McLaren, A. (1988). Sex determination in mammals. **Trends Genet. 4**, 153-7.

Menke, D.B., Muti, G.L. & Page, D.C. (1997). Expression of DAZ, an Azoospermia factor candidate, in human spermatogonia. **Am. J. Hum. Genet. 60**, 237-241.

Meroni, G., Franco, B., Archidiacono, N., Messali, S., Adolfi, G., Rocchi, M. & Ballabio, A. (1996). Characterisation of a cluster of sulfatase genes on Xp22.3 suggests gene duplications in an ancestral pseudoautosomal region. **Hum. Mol. Genet. 5**, 423-431.

Meschede, D., Froster, U.G., Bergmann, M. & Nieschlag, E. (1994). Familial pericentric inversion of chromosome 1 (p34q23) and male infertility with stage specific spermatogenic arrest. **J. Med. Genet. 31**, 573-575.

Meschede, D., Nieschlag, E. & Horst, J. (1995). The importance of clinical documentation in genetic studies of male infertility. **Hum. Genet. 96**, 500-501.

Meschede, D. & Horst, J. (1997). The molecular genetics of male infertility. **Mol. Hum. Genet. 3**, 419-430.

Miller, D. (1997). RNA I the ejaculate spermatoozoon: a window into molecular events in spermatogenesis and a record of the unusual requirements of haploid gene expression and post-meiotic equilibrium. **Mol Hum. Reprod. 3**, 669-676.

Miller, D., Briggs, D., Snowden, H., Hamlington, J., Rollinson, S., Lilford, R. & Krawetz, S.A. (1999). A complex population of RNAs exist in human ejaculate spermatozoa: implications for understanding molecular aspects of spermiogenesis. **Gene 237**, 385-392.

Miller, D.A. (1977). Evolution of primate chromosomes. **Science 198**, 1116-24.

Mitchell, M.J., Wilcox, S.A., Watson, J.M., Lerner, J.L., Woods, D.R., Scheffler, J., Hearn, J.P., Bishop, C.E. & Marshall Graves, J.A. (1998). The origin and loss of the ubiquitin activating enzyme gene on the mammalian Y chromosome. **Hum. Mol. Genet. 7**, 429-434.

Mitchell, M.J., Woods, D.R., Wilcox, S.A., Graves, J.A. & Bishop, C.E. (1992). Marsupial Y chromosome encodes a homologue of the mouse Y-linked candidate spermatogenesis gene Ubely. **Nature 359**, 528-531.

Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. (1987). Male driven molecular evolution: a model and nucleotide sequence analysis. **Cold Spring Harb. Symp.Quant Biol. 52**, 863-867.

Mohandas, T.K., Speed, R.M., Passage, M.B., Yen, P.H., Chandley, A.C., & Shapiro, L.J. (1992). Role of the Pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion in distal Xp. **Am. J. Hum. Genet. 51**, 526-533.

Moosani, N., Pattinson, H.A., Carter, M.D., Cox, D.M., Rademaker, A.W. & Martin, R.H, (1995). Chromosomal analysis of sperm from men with idiopathic infertility using sperm karyotyping and fluorescence in situ hybridization. **Fertil. Steril.. 64**, 811-7.

Morais de Silva, S. Hacker, A., Harley, V., Goodfellow, P., Swain, A. & Lovell-Badge, R. (1996). Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds. **Nat. Genet. 14**, 62-67.

Morgan, J.G., Dolganov, G.M., Robbins, S.E., Hinton, L.M. & Lovett, M. (1992). The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. **Nuc. Acids Res. 20**, 5173-5179.

Mount, S.M. (1996). AT-AC introns: An Attack on Dogma. **Science 271**, 1690-1692.

Mulhall, J.P., Reijo, R., Alagappan, R., Brown, L., Page, D., Carson, R. & Oates, R.D. (1997). Azoospermic men with deletion of the DAZ gene cluster are capable of completing spermatogenesis: fertilization normal embryonic development and pregnancy occur when retrieved testicular spermatozoa are used for intracytoplasmic sperm injection. **Hum. Reprod. 12**, 503-508.

Müller, U., Lalande, M., Donlon, T.A. & Heartlein, M.W. (1989). Breakage of the human Y-chromosome short arm between two blocks of tandemly repeated DNA sequences. **Genomics 5**, 153-156.

Mundlos, S., Pelletier, J., Darveau, A., Bachmann, M. Winterpacht, A. & Zabel, B. (1993). Nuclear localization of the protein encoded by the Wilms' tumor gene WT1 in embryonic and adult tissues. **Development 119**, 1329-1341.

Nadeau, J.H. & Sankoff, D. (1997). Comparable rates of gene loss and functional divergence after genome duplication early in vertebrate evolution. **Genetics 147**, 1259-1266.

Najmabadi, H., Chai, N., Kapali, A., Subbarao, M.N., Bhasin, D., Woodhouse, E., Yen, P. & Bhasin, S. (1996). Genomic structure of a Y-specific ribonucleic acid binding motif-containing gene: a putative candidate for a subset of male infertility. **J. Clin. Endocrin. & Metab. 81**, 2159-2163.

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. **Nucleic Acids Res. 14**, 7569-80.

Nakahori, Y., Hamano, K., Iwaya, M. & Nakagome, Y. (1991). Sex identification by polymerase chain reaction using X-Y homologous primer. **Am J Med Genet. 39**, 472-3.

Nakahori, Y., Kuroki, Y., Komaki, R., Kondoh, N., Namiki, M., Iwamoto, T., Toda, T. & Kobayashi, K. (1996). The Y chromosome region essential for spermatogenesis. **Horm Res. 46 Suppl 1**, 20-3.

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**, 74-7.

Nantel, F., Monaco, L., Foulkes, N.S., Masquiller, D., LeMeur, M., Henriksen, K., Dierich, A., Parvinen, M. & Sassone-Corci, P. (1996). Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. **Nature 380**, 159-161.

Nasrin, N., Buggs, C., Kong, X.F., Carnazza, J., Goebl, M. & Alexander-Bridges, M. (1991). DNA binding properties of the product of the testis-determining gene and a related protein. **Nature 354**, 317-320.

Newman-Smith, E.D., Shurland, D.L. & van der Bliek, A.M. (1997). Assignment of the dynamin-1 gene (DNM1) to human chromosome 9q34 by fluorescence in situhybridization and somatic cell hybrid analysis. **Genomics 41**, 286-9.

Nicholls, R.D., Saitoh, S. & Horsthemke, B. (1998). Imprinting in Prader-Willi and Angelman syndromes. **TIG 14**, 194-200.

Ninomiya, S., Isomura, M., Narahara, K., Seino, Y. & Nakamura, Y. (1996). Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. **Hum. Mol. Genet. 5**, 69-72.

Nishi, S., Hoshi, N., Kasahara, M., Ishibashi, T. & Fujimoto, S. (1999). Existence of human DAZLA protein in the cytoplasm of human oocytes. **Mol. Hum. Reprod. 5**, 495-497.

Nowak, M.A., Boerlijst, M.C., Cooke, J. & Maynard Smith, J. (1997). Evolution of genetic redundancy. **Nature 388**, 167-170.

Ogura, A., Matsuda, J. & Yanagimachi, R. (1994). Birth of normal young after electrofusion mouse oocytes with round spermatids. **Proc. Natl. Acad. Sci. USA 91**, 7460-7462.

Okubo, K., Hori, N., Matoba, R., Niyama, T. & Matsubara, K. (1991). A novel system for large-scale sequencing of cDNA by PCR amplification. **DNA Sequence-DNA Sequencing & Mapping 2**, 137-144.

Osbourne-Lawrence, S., Welcsh, P.L., Spillman, M., Chandrasekharappa, S.C., Gallardo, T.D., Lovett, M. & Bowcock, A.M. (1995). Direct selection of expressed sequences within a 1 Mb region flanking BRCA1 on human chromosome 17q21. **Genomics 25**, 248-255.

Osman, R.A., Andria, M.L., Jones, A.D. & Meizel, S. (1989). Steroid induced exocytosis: the human sperm acrosome reaction. **Biochem Biophys Res Commun. 160**, 828-33.

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. (1987). Linkage, physical mapping, and DNA sequence analysis of pseudoautosomal loci on the human X and Y chromosomes. **Genomics 1**, 243-56.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. & Brown, L.G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. **Cell. 51**, 1091-104.

Page, D.C., Silber, S. & Brown, L.G. (1999). Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection, but are likely to transmit the deletion and infertility. **Hum. Reprod. 14,** 1722-1726.

Palermo, G.D., Schlegel, P.N., Sills, E.S. Veeck, L.L., Zaninovic, N., Menendez, S. & Rosenwaks, Z. (1998). Brief Report: Births after intracytoplasmic injection of sperm obtained by testicular extraction from men with nonmosaic Klinefelter's syndrome. **New Eng. J. Med. 338,** 588-590.

Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D. & Weissman, S.M. (1991). CDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. **Proc. Natl. Acad. Sci. USA 88,** 9623-9627.

Parimoo, S., Patanjali, S.R., Kolluri, R., Hu, X., Wei, H. & Weissman, S.M. (1995). CDNA selection an other approaches in positional cloning. **Analytical Biochem. 228,** 1-17.

Parker, K.L., Schedl, A. & Schimmer, B.P. (1999). Gene interactions in gonadal development. **Ann. Rev. Physiol. 61,** 417-433.

Pasyukova, E., Nuzhdin, S., Li, W. & Flavell A.J. (1997). Germ line transposition of the copia retrotransposon in Drosophila melanogaster is restricted to males by tissue specific control of copia RNA levels. **Mol. Gene. Genet. 255,** 115-124.

Paulsen, C.A., Berman, N.G. & Wang, C. (1996). Data from men in greater Seattle area reveals no downward trend in semen quality: further evidence that deterioration of semen quality is not geographically uniform. **Fertility & Sterility 65,** 1015-1020.

Pearson, P.L., Ellis, J.D. & Evans, H.J. (1970). A gross reduction in chiasma formation during meiotic prophase and a defective DNA repair mechanism associated with a case of human male infertility. **Cytogenetics, 9,** 460-467.

Pearson, P.L., Bobrow, M., Vosa, C.G. & Barlow, P.W. (1971). Quinacrine fluorescence in mammalian chromosomes. **Nature. 231**, 326-9.

Pertl, B., Weitgasser, U., Kopp, S., Kroisel, P.M, Sherlock, J. & Adinolfi, M. (1996). Rapid detection of trisomies 21 and 18 and sexing by quantitative fluorescent multiplex PCR. **Hum. Gen. 98**, 55-59.

Pfeifer, K., Leighton, P.A. & Tilghman, S.M. (1996). The structural H19 gene is required for transgene imprinting. **Proc. Natl. Acad. Sci. USA 93**, 13876-13883.

Polymeropoulos, M.H., Xiao, H., Sikela, J.M., Adams, M., Venter, J.C. & Merril, C.R. (1993). Chromosomal distribution of 320 genes of a brain cDNA library. **Nat. Genet. 4**, 381-386.

Poulat, F., Desclozeaux, M., Tuffery, S., Jay. P., Boizet, B. & Berta, P. (1998). Mutation in the 5' noncoding region of the SRY gene in an XY reversed patient. **Hum. Mut. Suppl I**, S192-S194.

Prince, V.E., Joly, L., Ekker, M. & Ho, R.K. (1998). Zebrafish Hox genes: genomic organization and modified colinear expression patterns in the trunk. **Development 125**, 407-420.

Pritchard-Jones, K.; Fleming, S.; Davidson, D.; Bickmore, W.; Porteous, D.; Gosden, C.; Bard, J.; Buckler, A.; Pelletier, J.; Housman, D.; van Heyningen, V.; Hastie, N. (1990). The candidate Wilms' tumour gene is involved in genitourinary development. **Nature 346**, 194-197.

Propst, F., Rosenberg, M.P. & Vande Woude, G.F. (1988). Proto-oncogene expression in germ cell development. **TIG 4**, 183-187.

Prosser, J., Inglis, J.D., Condie, A., Ma, K., Kerr, S., Thakrar, R., Taylor, K., Cameron, J.M. & Cooke, H.J. (1996). Degeneracy in human multicopy RBM (YRRM), a candidate spermatogenesis gene. **Mamm. Genome 7**, 835-842.

Rao, E., Weiss, B. Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., Nordsiek, G., Heinrich, U., Bueuning, M.H., Ranke, M.B., Rosenthal, A., Ogata, T. & Rappold, G.A. (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner Syndrome. **Nature Genetics 16**, 54-63.

Rappold, G.A. (1993). The pseudoautosomal regions of the human sex chromosomes. **Hum. Genet. 92**, 315-324.

Ratti, A., Stuppia, L., Gatta, V., Fogh, I., Calabrese, G., Pizzuti, A. & Palka, G. (2000). Characterisation of a new TSPY gene family member in Yq (TSPYq1). **Cytogenet. Cell Genet. 88**, 159-162.

Reddy, K.S. (1998). Clinical management of a rare de novo translocation 46,X,t(Y;15) (p11.2~11.3;q11.2).ish t(Y;15)(DYZ3+,AMELY+, SNRPN+;D15Z+) found prenatally. **Prenatal Diagn. 18**, 294-297.

Reijo, R., Seligman, J., Dinulos, M.B., Jaffe, T., Brown, L.G., Disteche, C.M. & Page, D.C. (1996). Mouse autosomal homolog of DAZ a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. **Genomics 35**, 346-352.

Reijo, R., Lee, T-Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., de la Chapelle, A., Silber, S. & Page, D.C. (1995). Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. **Nat. Genet. 10**, 383-393.

Renfree, M.B., Harry, J.L. & Shaw, G. (1995). The marsupial male: a role model for sexual development. **Philos Trans R Soc Lond B Biol Sci. 350**, 243-51.

Reubinoff, B.E., Abeliovich, D., Werner, M. Schenker, J.G., Safran, A. & Lewin, A. (1998). A birth in non-mosaic Klinefelter's syndrome after testicular fine needle aspiration, intracytoplasmic sperm injection and preimplantation genetic diagnosis. **Hum. Reprod. 13**, 1887-1892.

Rice, W.R. (1994). Degeneration of a nonrecombining chromosome. Science 263, 230-232.

Richter, W., Dettmer, D & Glander, H-J. (1999). Detection of mRNA transcripts of cyclic nucleotide phosphodiesterase subtypes in ejaculated human spermatozoa. **Mol. Hum. Reprod. 5**, 732-736.

Ried, T., Arnold, N., Ward, D.C. & Wienberg, J. (1993). Comparative high resolution mapping of human and primate chromosomes by fluorescence in situ hybridisation. **Genomics 18**, 381-386.

Ried, K., Rao, E., Schiebel, K. & Rappold, G.A. (1998). Gene duplication as a recurrent theme in the evolution of the human pseudoatuosomal region 1: isolation of the gene ASMTL. **Hum. Mol. Genet. 7**, 1771-1778.

Robertson, K.M., O'Donnell, L., Jones, M.E.E., Meachem, S.J., Boon, W.C., Fisher, C.R., Graves, K.H., McLachlan, R.I. & Simpson, E.R. (1999). Impairment of spermatogenesis in mice lacking a functional aromatase (cyp 19) gene. **Dev. Biol. 96**, 7986-7991.

Robinson, W.P., Bottani, A., Xie, Y.G., Balakrishman, J., Binkert, F., Machler, M., Prader, A. & Schinzel, A. (1991). Molecular, cytogenetic, and clinical investigtions of Prader-Willi syndrome patients. **Am. J. Hum. Genet. 49**, 1219-1234.

Ron-El, R., Friedler, S., Strassburger, D., Komarovsky, D., Schachter, M. & Raziel, A. (1999). Birth of a healthy neonate following the intracytoplasmic injection of testicular spermatozoa from a patient with Klinefelter's syndrome. **Hum. Reprod. 14**, 368-370.

Rouquier, S., Trask, B.J., Taviaux, S., van den Engh, G., Diriong, S., Lennon, G.G. & Giorgi, D. (1995), Direct selection of cDNAs using whole chromosomes. **Nuc. Acids Res. 23**, 4415-4420.

Rouquier, S., Taviaux, S., Trask, B.J., Brand-Arpon, V., van den Engh, G., Demaille, J. & Giorgi, D. (1998). Distribution of olfactory receptor genes in the human genome. **Nat Genet. 18**, 243-50.

Rouyer, F., de la Chapelle, A., Andersson, M. & Weissenbach, J. (1990). An interspersed repeated sequence specific for human subtelomeric regions. **EMBO J. 9**, 505-514.

Salas-Cortez, L., Jaubert, F., Barbaux, S., Nessmann, C., Bono, M.R., Fellous, M., McElreavey, K. & Rosemblatt, M. (1999). The human SRY protein is present in fetal and adult Sertoli cells and germ cells. **Int. J. Dev. Biol. 43**, 135-140.

Salido, E.C., Yen, P.H., Koprivnikar, K., Yu, L-C. & Shapiro, L.J. (1992). The human enamel protein gene ameliogenin is expressed from both the X and the Y chromosomes. **Am. J. Hum. Genet. 50**, 303-316.

Salido, E.C., Yen, P.H., Mohandas, T.K. & Shapiro, L.J. (1992). Expression of the X-inactivation-associated gene XIST during spermatogenesis. **Nat. Genet. 2**, 196-199.

Sanger, F., Nicklen, S. & Coulson, A.R. (1997). DNA sequencing with chain-terminating inhibitors. **Proc. Natl. Acad. Sci. USA 74**, 5463-5467.

Sargent, C.A., Young, C., Marsh, S., Ferguson-Smith, M.A. & Affara, N.A. (1994). The glycerol kinase gene family: structure of the Xp gene, and related intronless retroposons. **Hum. Mol. Genet. 3**, 1317-1324.

Sargent, C.A., Briggs, H., Chalmers, I.J., Lambson, B., Walker, E. & Affara, N.A. (1996). The sequence organization of Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. **Genomics 32**, 200-209.

Sassone-Corsi, P. (1997). Transcriptional checkpoints determining the fate of male germ cells. **Cell 88**, 163-166.

Saunders, P.T.K., Fisher, J.S., Sharpe, R.M. & Millar, M.R. (1998). Expression of oestrogen, receptor beta (ERB) occurs in multiple cell types, including some germ cells, in the rat testis. **J. Endocrin. 156**, R13-R17.

Saxena, R., Brown, L.G., Hawkins, T., Algappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M. & Page, D.C. (1996). The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. **Nat. Genet. 14**, 292-299.

Seuanez, H.N., Carothers, A.D., Martin, D.E., Short, R.V. (1977). Morphological abnormalities in spermatozoa of man and great apes. **Nature. 270**, 345-7.

Seyanez, H.N., Carothers, A.D., Martin, D.E. & Short, R.V. (1977). Morphological abnormalities in spermatozoa of man and great apes. **Nature 270**, 345-347.

Schafer, A.J. & Goodfellow, P.N. (1995). Sex determination in humans. **BioEssays 18**, 955-963.

Schafer, M., Nayernia, K., Engel, W. & Schafer, U. (1995). Translational control in spermatogenesis. **Dev. Biol. 172**, 344-352.

Schempp, W., Weber, B., Serra, A., Neri, G., Gal, A. & Wolf, U. (1985). A 45,X male with evidence of a translocation of Y euchromatin onto chromosome 15. **Hum. Genet. 71**, 150-154.

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

Schempp, W., Binkele, A., Arnemann, J., Gläser, B., Ma, K., Taylor, K., Toder, R., Wolfe, J., Zeitler, S. & Chandley, A.C. (1995). Comparative mapping of YRRM and STPY sequences in man and hominoid ape. **Chr. Res. 3**, 227-234.

Schiebel, K., Winkelmann, M., Mertz, A., Xu, X., Page, D.C., Weil, D., Petit, C. & Rappold, G.A. (1997). Abnormal XY interchange between a novel isolated protein kinase gene, PRKY and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. **Hum. Mol. Genet. 6**, 1985-1989.

Schmid, M., Guttenbach, M., Nanda, I., Studer, R. & Epplen JT. (1990). Organization of DYZ2 repetitive DNA on the human Y chromosome. **Genomics. 6**, 212-8.

Schnieders, F., Dörk, T. Arnemann, J., Vogel, T., Werner, M. & Schmidtke, J. (1996). Testis-specific protein, Y encoded (TSPY) expression in testicular tissues. **Hum. Mol. Genet. 5**, 1801-1807.

Schwartz, A., Chan, D.C., Brown, L.G., Alagappan, R., Pettay, D., Disteche, C., McGillivray, B., de la Chapelle, A. & Page, D.C. (1998). Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. **Hum. Mol. Genet. 7**, 1-11.

Seboun, E., Barbaux, S., Bourgeron, T., Nishi, S., Algonik, A., Egashira, M., Nikkawa, N., Bishop, C., Fellous, M., McElreavey, K. & Kashara, M. (1997). Gene sequence localization and evolutionary conservation of DAZLA, a candidate male sterility gene. **Genomics 41**, 227-235.

Sedlacek, Z., Korn, B., Konecki, D.S., Siebenhaar, R., Coy, J.F., Kioschis, P., & Poustka, A. (1993). Construction of a transcription map of a 800 kb region around the human G6PD locus by direct cDNA selection. **Hum. Mol. Genet. 2**, 1865-1869.

Sedlacek, Z., Konecki, D.S., Siebenhaar, R., Kioschis, P. & Poustka, A. (1993). Direct selection of DNA sequences conserved between species. **Nuc. Acids Res. 21**, 3419-3425.

Setchell, B.P. (1978). The mammalian testis. In: Reproductive biology handbooks.

Shan, Z., Hirschmann, P., Seebacher, T., Edelmann, A., Jauch, A., Morell, J., Urbitsch, P. & Vogt, P.H. (1996). A SPGY copy homologous to the mouse gene Dazla and the Drosophila gene boule is autosomal and expressed only in the human male gonad. **Hum. Mol. Genet. 5**, 2005-2011.

Sharpe, R.M. & Skakkebaek, N.E. (1993). Are oestrogens involved in falling sperm counts and disorders of the male reproductive tract? **Lancet 341**, 1392-1395.

Sharpe, R.M., Atanassova, N., McKinnell, C., Parte, P., Turner, K.J., Fisher, J.S., Kerr, J.B., Groome, N.P., Macpherson, S., Millar, M.R. & Saunders, P.T. (1998). Abnormalities in functional development of the Sertoli cells in rats treated neonatally with diethylstilbestrol: a possible role for estrogens in Sertoli cell development. **Biol Reprod. 59**, 1084-94.

Shimmin, L.C., Chang, B.H-J. & Li, W-H. (1993). Male driven evolution of DNA sequences. **Nature 362**, 745-747.

Simmons, A.D., Goodart, S.A., Gallardo, T.D., Overhauser, J. & Lovett, M. (1995). Five novel genes form the cri-du-chat critical region isolated by direct selection. **Hum. Mol. Genet. 4**, 295-302.

Simmons, A.D., Overhauser, J. & Lovett, M. (1997). Isolation of cDNAs from the Cri-du-chat critical region by direct screening of a chromosome 5 specific cDNA library. **Genome Res. 7**, 118-127.

Simmler, M-C., Rouyer, F., Vergnaud, G., Nystom-Lahti, M., Ngo, K.Y., de la Chappelle, A. & Weissenbach, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. **Nature 317**, 692-697.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A-M., Lovell-Badge, R. & Goodfellow, P.N. (1990). A gene

from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. **Nature 346**, 240-244.

Skakkebaek, N.E., Giwercman, A. & de Kretser, D. (1994). Pathogenesis and management of male infertility. **The Lancet 343**, 1473-1479.

Slee, R., Grimes, B., Speed, R.M., Taggart, M., Maguire, S.M., Ross, A., McGill, N.I., Saunders, P.T.K. & Cooke, H.J. (1999). A human DAZ transgene confers partial rescue of the mouse Dazl null phenotype. **Proc. Natl. Acad. Sci. USA 96**, 8040-8045.

Slomski, R., Schloesser, M., Chlebowska, H., Reiss, J. & Engel, W. (1991). Detection of human spermatid specific transcripts in peripheral blood lymphocytes of males and females. **Hum. Genet. 87**, 307-310.

Smith, G.P. (1976). Evolution of repeated DNA sequences by unequal crossover. **Science. 191**, 528-35.

Smith-Ferguson, M.A., Affara, N.A. & Magenis, R.E. (1987). Ordering of Y-specific sequences by deletion mapping and analysis of X-Y interchange males and females. **Development 101**, Suppl:41-50.

Solari, A.J. (1980). Synaptosomal complexes and associated structures in microspread human spermatocytes. **Chromosoma. 81**, 315-37.

Soulard, M., Valle, V.D., Siomi, M.C., Pinol-Roma, S., Codogno, P. Bauvy, C., Bellini, M., Lacroix, J-C., Monod, G., Dryfuss, G. & Larsen, C-J. (1993). hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. **N.A.R. 21**, 4210-4217.

Speed, R.M. & Chandley, A.C. (1990). Prophase of meiosis in human spermatocytes analysed by EM microspreading in infertile men and their controls and comparisons with human oocytes. **Hum Genet. 84**, 547-54.

Speed, R.M., Faed, M.J.W., Batstone, P.J., Baxby, K. & Barnetson, W. (1991). Persistence of two Y chromosomes through meitotic prophase and metaphase I in an XYY man. **Hum. Genet. 87**, 416-420.

Stehle, J.H., Foulkes, N.S., Molina, C.A., Simmonneaux, V., Pevet, P. Sassone-Corsi, P. (1993). Adrenergic signals direct rhythmic expression of transcriptional repressor CREM in the pineal gland. **Nature 365**, 314-320.

Steinemann, M. & Steinemann, S. (1992a). A duplication including the Y allele of *Lcp2* and the TRIM retrotransposon at the *Lep* Locus on the degenerating *neo*-Y chromosome of *Drosophila miranda*: molecular structure and mechanisms by which it may have arisen. **Genetics 134**, 497-505.

Steinemann, M. & Steinemann, S. (1992b). Degenerating Y chromosome of Drosophila miranda: a trap for retrotransposons. **Proc. Natl. Acad. Sci. USA 89**, 7591-7595.

Steinemann, M., Steinemann, S. & Lottspeich, F. (1993). How Y chromosomes become genetically inert. **Proc. Natl. Acad. Sci. USA 90**, 5737-5741.

Stevanovic, M.; Lovell-Badge, R.; Collignon, J. & Goodfellow, P. N. (1993). SOX3 is an X-linked gene related to SRY. **Hum. Molec. Genet. 2**, 2013-2018.

Strachan, T. & Read, A.P. (1999). Human Molecular Genetics 2. Chapt. 10, **Physical & Transcript mapping**. P260-266.

Su, H. & Lau, Y-F.C. (1993). Identification of the transcriptional unit, structural organization, and promoter sequence of the human sex-determining region Y (SRY) gene, using a reverse genetic approach. **Am. J. Hum. Genet. 52**, 24-38.

Sutherland, H.F., Wadey, R., McKie, J.M., Taylor, C., Atif, U., Johnstone, K.A., Halford, S., Kim, U-J., Goodship, J., Baldini, A. & Scambler, P.J. (1996). Identification of a novel transcript disrupted by a balanced translocation associate with DiGeorge Syndrome. **Am. J. Hum. Genet. 59**, 23-31.

Swain, A.; Narvaez, V.; Burgoyne, P.; Camerino, G. & Lovell-Badge, R. (1998). Dax1 antagonizes Sry action in mammalian sex determination. **Nature 391**, 781-786,

Swain, A. & Lovell-Badge, R. (1999). Mammalian sex determination: a molecular drama. **Genes & Dev. 13**, 755-767.

Taylor, K., Hornigold, N., Conway, D., Williams, D., Ulinowski, Z., Agochiya, M., Fattorini, P., de Jong, P., Little, P.F. & Wolfe, J. (1996). Mapping the human Y chromosome by fingerprinting cosmid clones. **Genome Res. 6**, 235-48.

Tagle, D.A., Swaroop, M., Lovett, M. & Collins, F.S. (1993). Magnetic bead capture of expressed sequences encoded large genomic segments. **Nature 361**, 751-756.

Tajima, T., Nakae, J., Shinohara, N. & Fujieda, K. (1994). A novel mutation localized in the 3' non-HMG box region of the SRY gene in 46,XY gonadal dysgenesis. **Hum. Mol. Genet.** 1187-1189.

Tani, M., Shindo-Okada, N., Hashimoto, Y., Shiroishi, Takenoshita, S., Nagamachi, Y. & Yokota, J. (1997). Isolation of a novel Sry-related gene that is expressed in high metastatic K-1735 murine melanoma cells. **Genomics 39**, 30-27.

Tassone, F., Xu, H., Burkin, H., Weissman,, S. & Gardiner, K. (1995). CDNA selection from 10 mb of chromosome 21 DNA: efficiency in transcriptional mapping and reflections of genome organization. **Hum. Mol Genet. 4**, 1509-1518.

Tiemessen, C.H., Evers, J.L. & Bots, R.S. (1996). Tight-fitting underwear and sperm quality. **Lancet 347**, 1844-5.

Tiepolo, L. & Zuffardi, O. (1976). Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. **Hum. Genet. 34**, 119-124.

Tilmann, C. & Capel, B. (1999). Mesonephric cell migration induces testis cord formation and Sertoli cell differentiation in the mammalian gonad. **Development 126**, 2883-2890.

Touchman, J.W., Bouffard, G.G., Weintraub, L.A., Idol, J.R., Wang, L., Robbins, C.M., Nussbaum, J.C., Lovett, M. & Green, E.D. (1997). 2006 expressed-sequence tags derived from human chromosome 7-enriched cDNA libraries. **Genome Res. 7**, 281-92.

Trask, B.J., Massa, H., Brand-Arpon, V., Chan K., Friedman, C., Nguyen, O.T., Eichler, E., van den Engh, G., Rouquier, S., Shizuya, H. & Giorgi, D. (1998). Large multi-chromosomal duplications encompass many m,embers of the ofactory receptor gene family in the human genome. **Hum. Mol. Genet. 7**, 2007-2020.

Tsuchiya, K., Reijo, R., Page, D.C. & Disteche, C.M. (1995). Gonadoblastoma: Molecular definition of the susceptibility region on the Y chromosome. **Am. J. Hum. Genet. 57**, 1400-1407.

Tsunekawa, N., Matsumoto, M., Tone, S., Mishida, T. & Fujimoto, H. (1999). The Hsp70 homolog gene, Hsc70t, is expressed under translational control during mouse spermiogenesis. **Mol. Reprod. Dev. 52**, 383-391.

Tucker, P.K. & Lundrigan, B.L. (1993). Rapid evolution of the sex determining locus in Old World mice and rats. **Nature 364**, 715-717.

Tyler-Smith, C., Taylor, L. & Muller, U. (1988). Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. **J Mol Biol. 203**, 837-48.

Van Assche, E., Bonduelle, M., Tournaye, H., Joris, H., Verheyen, G., Devroey, P., Van Steirteghem, A. & Liebaers, I. (1996). Cytogenetics of infertile men. **Hum. Reprod. 11**, 1-24.

Van der Bliek, A.M., Redelmeier, T.E., Damke, H., Tisdale, E.J., Meyerowitz, E.M. & Schmid, S.L. (1993). Mutations in human dynamin block an intermediate stage in coated vesicle formation. **J Cell Biol. 122**, 553-63.

Van der Bliek, A.M. (1999). Functional diversity in the dynamin family. **Trends Cell Biol. 9**, 96-102.

Van Steirteghem, A.C., Nagy, Z., Joris, H., Liu, J., Staessen, C., Smitz, J., Wisanto, A. & Devroey, P. (1993). High fertilization and implantation rates after intracytoplasmic sperm injection. **Hum. Reprod. 8**, 1061-1066.

Vendrell, J.M., Garcia, E., Veiga, A., Calderon, G., Egozcue, S. Egozcue, J. & Barri, P.N. (1999). Meiotic abnormalities and spermatogenic parameters in severe oligoasthenozoospermia. **Hum. Reprod. 14**, 375-378.

Vereb, M., Agulnik, A.I., Houston, J.T., Lipschultz, L.I., Lamb, D.J. & Bishop, C.E. (1997). Absence of DAZ gene mutations in cases of non-obstructed azoospermia. **Mol. Hum. Reprod. 3**, 55-59.

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 hybridization. **Am J Hum Genet. 38**, 109-24.

Verlinsky, Y., Ginsberg, N., Chmura, M., White, M., Strom, C & Kuliev, A. (1998). Detection of translocations involving the Y-chromosome in prospective prenatal screening of common chromosomal aneuploidies by fish. **Prenatal Diagn. 18**, 390-392.

Vermeesch, J.R., Petit, P., Kermouni, A., Renauld, J.C., Van Den Berghe, H. & Marynen, P. (1997). The IL-9 receptor gene, located in the Xq/Yq pseudoautosomal region, has an autosomal origin, escapes X inactivation and is expressed from the Y. **Hum Mol Genet. 6**, 1-8.

Viguie, F., Domenjoud, L., Rousseau-Merck, M-F., Dadoune, J-P. & Chevaillier, P. (1990). Chromosomal localisation of the human protaminie genes, PRM1 and PRM2, to 16p13.3 by in situ hybridization. **Hum. Genet. 85**, 171-174.

Vine, M.F., Margolin, B.H., Morrison, H.I. & Hulka, B.S. (1994). Cigarette smoking and sperm density: a meta analysis. **Fertility & Sterility 61**, 35-43.

Vogel, T., Borgmann ,S., Dechend, F., Hecht, W. & Schmidtke, J. (1997). Conserved Y-chromosomal location of TSPY in Bovidae. **Chromosome Res. 5**, 182-5.

Vogel, T., Dittrich, O., Mehraein, Y., Dechend, F., Schnieders, F. & Schmidtke, J. (1998). Murine and human TSPYL genes: novel members of the TSPT-SET-NAP1L1 family. **Cytogenet. Cell Genet. 81**, 265-270.

Vogel, T. & Schmidtke, J. (1998). Structure and function of TSPY, the Y chromosome gene coding for the "testis-specific protein". **Cytogenet. Cell Genet. 80**, 209-213.

Vogt, P., Chandley, A.C., Hargreave, T.B., Keil, R., Ma, K., & Sharkey, A. (1992). Microdeletions in interval 6 of the Y chromosome of males with idiopathic sterility point to disruption of AZF, a human spermatogenesis gene. **Hum Genet. 89**, 491-6.

Vogt, P. H.; Edelmann, A.; Kirsch, S.; Henegariu, O.; Hirschmann, P.; Kiesewetter, F.; Kohn, F. M.; Schill, W. B.;Farah, S.; Ramos, C.; Hartmann, M.; Hartschuh, W.; Meschede, D.; Behre, H. M.; Castel, A.; Nieschlag, E.; Weidner, W.;Grone, H.-J.; Jung, A.; Engel, W. & Haidl, G. (1996): Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. **Hum. Molec.Genet. 5**, 933-943.

Vogt, P.H., Rappold, G.A. & Lau, C.Y.F. (1997). Third international workshop on human Y chromosome mapping. **Cytogenet. Cell Genet. 79**, 1-20.

Vogt, P. (1998). Human chromosome deletions in Yq11, AZF candidate genes and male infertilty: history and update. **Mol. Hum. Reprod. 4**, 739-744.

Vollrath, D., Foote, S., Hilton, A., Brown, L.G., Beer-Romero, P., Bogan, J.S. & Page, D.C. (1992). The human Y chromosome: a 43-interval map based on naturally occurring deletions. **Science 258**, 52-9.

Wagner, A. (1998). The fate of duplicated genes: loss or new function? **BioEssays 20**, 785-788.

Wagner, T., Wirth, J., Meyer, J., Zabel, B., Heid, M., Zimmer, J., Pasantes, J., Bricarelli, F.D., Keutel, J., Hustert, E., Wolf, Ul, Tommerup, N., Schempp, W. & Scherer, G. (1994). Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. **Cell 79**, 1111-1120.

Walker, W.H., Sanborn, B.M. & Habener, J.F. (1994). An isoform of transcripton factor CREM expressed during spermatogenesis lacks the phosphorylation domain and represses cAMP-induced transcription. **Proc. Natl. Acad. Sci. USA 91**, 12423-12427.

Weber, B., Weissenbach, J. & Schempp, W. (1987). Conservation of human-derived pseudoautosomal sequences on the sex chromosomes of the great apes. **Cytogenet. Cell Genet. 45**, 26-29.

Weller, P.A., Critcher, R., Goodfellow, P.N., German, J. & Ellis, N.A. (1995). The human Y chromosomes homologue of XG: transcription of a naturally truncated gene. **Hum. Mol. Genet. 4**, 859-868.

Wevrick, R., Kerns, J.A. & Francke, U. (1994). Identification of a novel paternally expressed gene in the Prader-Willi syndrome region. **Hum. Mol. Genet. 3**, 1877-1882.

Western, P.S, Harry, J.L, Graves, J.A, Sinclair, A.H. (1999). Temperature-dependent sex determination in the american alligator: AMH precedes SOX9 expression. **Dev Dyn. 216**, 411-419.

Whitfield, L.S., Lovell-Badge, R. & Goodfellow, P.N. (1993). Rapid sequence evolution of the mammalian sex-determining gene SRY. **Nature 364**, 713-714.

Wieacker, P. & Jakubiczka, S. (1997). Genetic causes of male infertility. **Andrologia 29**, 63-69.

Willison, K. & Ashworth, A. (1987). Mammalian spermatogenic gene expression. **TIG 3**, 351-355.

Wilkinson, T.A. & Crolla, J.A. (1993). Molecular cytogenetic characterization of three familial cases of satellited Y chromosomes. **Hum. Genet. 91**, 389-391.

Witkin, S.S., Korngold, G.C. & Bendrich, A. (1975). Ribonuclease-sensitive DNA-synthesising complex in human sperm heads and seminal fluid. **Proc. Natl. Acad. Sci. USA 72**, 3295-3299.

Wolfe, J., Erickson, R.P., Rigby, P.W.J. & Goodfellow, P.N, (1984). Cosmid clones derived from both euchromatic and heterochromatic regions of the Y chromosome. **The EMBO Journal. 3**, 1997-2003.

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.

Wright, E., Hargrave, M.R., Christiansen, J., Cooper, L., Kun, J., Evans, T., Gangadharan, U., Greenfield, A. & Koopman, P. (1995). The Sry related gene Sox9 is expressed during chondrogenesis in mouse embryos. **Nat. Genet. 9**, 15-20.

Wu, Q. & Krainer, A.R. (1996). U1-mediated exon definition interactions between AT-AC and GT-AG introns. **Science 274**, 1005-1008.

Wu, Q. & Krainer, A.R. (1999). AT-AC Pre-mRNA splicing mechanisms and conservation of minor introns in voltage gated ion channel genes. **Mol. & Cell. Biol. 19**, 3225-3236.

Wykes, S.M., Nelson, J.E., Visscher, D.W., Djakiew, D. & Krawetz, S.A. (1995). Co-ordinate expression of the PRM1, PRM2 and TNP2 multigene locus in human testis. **DNA & Cell Biol. 14**, 155-161.

Xu, H., Wei, H., Kolluri, R. & Weissman, S.M. (1995). A simplified approach for isolating oligo(dT) primed cDNA clones with probes generated by cDNA selection. **N.A.R. 23**, 4528-4529.

Yen, P.H., Marsh, B., Allen, E., Tsai, S.P., Ellison, J., Connolly, L., Neiswanger, K. & Shapiro, L.J. (1988). The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. **Cell 55**, 1123-1135.

Yen, P.H., Chai, N.N. & Salido, E.C. (1996). The human autosomal gene DAZLA: testis specificity and a candidate for male infertility. **Hum. Mol. Genet. 5**, 2013-2017.

Yen, P.H., Chai, N.N. & Salido, E.C. (1997). The human DAZ genes, a putative male infertility factor on the Y chromosome, are highly polymorphic in the DAZ repeat regions. **Mamm. Genome 8**, 756-759.

Yoder, J.A., Walsh, C.P. & Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. **Trends Genet 13**, 335-340.

Yovich, J.L. & Matson, L. (1995). Male subfertility: concepts in 1995. **Hum. Reprod. 10**, 3-9.

Yu, Y.T. & Steitz, J.A. (1997). Site-specific crosslinking of mammalian U11 and U6atac tot he 5' splice site of an AT-AC intron. **Proc. Natl. Acad. Sci. USA. 94**, 6030-6035.

Yunis, J.J. & Prakash, O. (1982). The origin of man: a chromosomal pictorial legacy. **Science 215**, 1525-1530.

Zhang, J. S.; Yang-Feng, T. L.; Muller, U.; Mohandas, T. K.; de Jong, P. J.; Lau, Y.-F. C. (1992). Molecular isolation and characterization of an expressed gene from the human Y-chromosome. **Hum. Molec. Genet. 1**, 717-726.

Zhao, G.Q., Deng, K., Labosky, P.A., Liaw, L. & Hogan, B.L. (1996). The gene encoding bone morphogenetic protein 8B is required for the initiation and maintenance of spermatogenesis in the mouse. **Genes Dev. 10**, 1657-1669.

Zhu, W. & Liang, P. (1997). Detection and isolation of differentially expressed genes by differential display. Chapt. 15 p211-220. In: Methods in Molecular Biology, Vol 68: **Gene isolation & Mapping Protocols**. Eds J Boultwood, Human Press inc. NJ. USA.

Zinn, A.R., Tonk, V.S., Chen, Z., Flejter, W.L., Gardner, H.A., Guerra, R., Kushner, H., Schwartz, S., Sybert, V.P., Van Dyke, D.L. & Ross, J.L. (1998). Evidence for a Turner Syndrome locus or loci at Xp11.2-p22.1. **Am. J. Hum. Genet. 63**, 1757-1766.