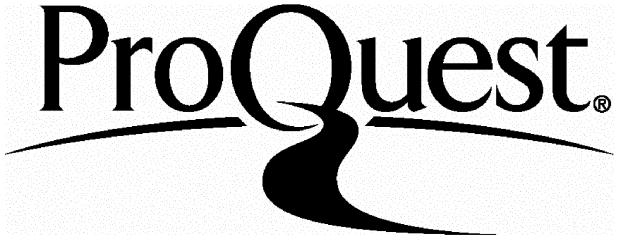
THE IDENTIFICATION AND CHARACTERISATION OF GENES EXPRESSED IN ADULT HUMAN TESTIS

Mohammad Hossein Modarressi

April 2001

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

The Galton Laboratory
Department of Biology
University College London
University of London



ABSTRACT OF THESIS

By the identification and isolation of cDNAs expressed specifically in the human testis I aimed to identify novel genes involved in spermatogenesis and/or spermiogenesis. Using Differential Display Reverse Transcriptase-PCR (DDRT-PCR) on RNA from a range of tissues which were selected to be logically related to testis functions I isolated a number of short cDNA fragments. Tissue-specific expression was confirmed by using RT-PCR with gene-specific primers. Three genes were examined in detail: Entire transcripts were sequenced by primer walking along cDNA templates generated by 5'RACE and the size of full-length cDNAs was confirmed by Northern analysis. The chromosomal locations of the genes were determined by two methods. Firstly by PCR amplification using gene specific primers of a panel of rodent somatic cell hybrids that each contain a single human chromosome Secondly, genomic clones (BAC or a PAC) were hybridised in situ to human metaphase chromosomes (FISH). Using the BAC clones, some genomic gene structure was determined.

In this way I have isolated and characterised three novel genes from testis, *TSGA10*, *GNG2* and *T2G3*. *Testis Specific Gene A10* (*TSGA10*) is not expressed in a variety of tissues functionally related to testis nor is it expressed in either testes of two infertile patients or in foetal testis. The *TSGA10* gene (GenBank Accession no. AF254756) consists of 3038 nucleotides spread over 19 exons and contains an open reading frame of 698 amino acids and one alternative splice site. RT-PCR results using RNA prepared from foetal and tumour tissues showed that transcripts from this gene are also present in squamous cell carcinoma, parotid and thyroid tumours, benign prostate hypertrophy and almost all foetal tissues though with different levels of expression in different tissues. Several ESTs with almost 100% homology to the *TSGA10* transcript exist in the databases but all are probably derived from testis apart from three which were isolated from tumour tissues. This pattern of expression suggests that *TSGA10* has a role in actively dividing cells and possibly in cell division and mitosis. The mouse homologue of this gene was also cloned and characterised.

The gene *GNG2* is a G protein subunit. The gene *T2G3* is another novel gene. Both have been characterised to a lesser extent than *TSGA10*.

ACKNOWLEDGMENTS

I would like to thank my supervisor **Dr. Jonathan Wolfe**, to whom I am most grateful, for his help, support and motivation during the course of this project. I would also like to say a special thank you to my previous supervisor **Dr. Kay Taylor** for reading through my thesis and useful discussions and my second supervisor **Dr Paul Burgoyne** for useful comments. I am extremely privileged to have been under the guidance of above-mentioned people who are patient, enthusiastic, innovative and who encouraged me towards independent thought. I am sure this will be a valuable gift for my scientific career.

I would like to thank **Prof. Sue Povey** for giving me the opportunity to be a member of her group and for kind advice.

In addition, I would like to extend my thanks to everyone in the MRC unit and our lab for their help, encouragement and technical advice, especially **Prof. D. Hopkinson, Dr. Margaret Fox, Wendy Putt, Jessie Cameron, Eleni Makrinou and Sue's group members.**

Last but by no means last thanks to my wife, **Maryam**, for her unconditional and constant support and faith in me.

This thesis is dedicated to my **Mum and Dad**

This work was supported by an Iranian MHME studentship , and also Overseas Research Students (ORS) Award from UK government.

TABLE OF CONTENTS

*Chapter II

Materials and Methods

*Chapter III

*Chapter IV

Cloning, characterisation and mapping of the gene encoding the human G protein gamma 2 subunit

*Chapter V

Identification, characterisation and mapping of a human testis-specific gene, *T2G3*

Figures and tables:

Chapter I

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
c_ot	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_20	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
HCl	hydrochloric acid

HGMP	Human Genome Mapping Project
HnRNP	heterogeneous ribobuncleoproteins
hnRNA	heterogeneous nuclear RNA
ICSI	intracytoplasmic sperm injection
kb	kilobase
kDa	kilodalton
KCl	potassium chloride
l	litre
LH	luteininzing hormone
M	molar
min	minute
ml	millilitre
mm	millimeter
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
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
SCO	sertoli cell only
SDS	sodium dodecyl sulphate
sec	second
SSC	salt sodium citrate
SSH	suppression subtractive hybridisation
SSPE	salt sodium phosphate EDTA
T an	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
µg	microgram
µl	microlitre
µm	micromolar

CHAPTERS

I-V

Chapter I

Introduction

Testis, compared to other organs in the body, is unique in containing many dividing cells in a pathway ranging from undifferentiated germ cells; to well differentiated cells such as sperm. Spermatogenesis in human males is the process by which self renewing stem cells, the spermatogonia, develop into mature spermatozoa, and it occurs within the seminiferous tubules of the testis. Failure of any stage of this process, which involves millions of cell divisions and is highly controlled, will cause the male to become subfertile or infertile and many different environmental and genetic factors have been implicated in these cases.

The aim of this project has been to identify novel genes involved in spermatogenesis by characterising sequences that are expressed either solely or highly in the testis. RNA from a selection of tissues were compared to testis RNA using Differential Display RT-PCR. Tissues were chosen which have similarities to the functions of testis except for spermatogenesis, so DDRT-PCR should be more likely to reveal differences in the testis related to its involvement in spermatogenesis.

Part one of this introduction describes the structure of testis and mechanism of human spermatogenesis. **Part two** discusses different causes of subfertility and infertility in human males. **Part three** shows some of the genes known or thought to be involved in the process of spermatogenesis/spermiogenesis. Advantages and disadvantages of differentially expressed gene discovery techniques in comparison to Differential Display RT-PCR are given in **Part four**.

Part one

1.1 Embryology of testis

1.1.1 Origin of the testes.

Gonads in the early embryo are identical in both sexes and initially appear as swellings in the central portion of the urogenital ridges on either side of the dorsal midline. The cranial part of this region makes the adrenocortical primordial and the caudal part forms the genital ridges, which are identifiable in the fifth week (Carlson, 1994). The genital ridge is morphologically unrecognisable between male and female.

1.1.2 Sex determination and differentiation of the testes. Cells have dynamic properties in gonads development. The first signs of testicular organisation can be recognised at approximately 42 days (after fertilisation in human) when cells in the medial portion of the mesonephros gather into cords of primordial Sertoli cells which come to cover inflowing germ cells to form primordial tubules. These events are almost worked out in terms of the genes involved (Capel, 1995&1996). According to models proposed by several authors, gonadal differentiation is under the influence of the SRY gene (Sinclair *et al.*, 1990). This so called testis determining factor, on chromosome Yp11.3 acts principally as a switch, causing the undifferentiated gonad to become a testis. There is a cascade of genes necessary for guiding the differentiation of the testis, but studies indicate that three of these genes are on stage at the same time, These are *SRY*, *DAX* and *SF-1* (review, Greenfield, 1998). The products of the steroidogenic factor 1(*SF-1*) (Luo at al., 1994; Shen *et al.*, 1994) and Wilms` tumour 1 (*WT1*) genes (Pritchard-Jones *et al.*, 1990), are also essential for mammalian gonadogenesis prior to sexual differentiation. In males, SF-1 has a role during sexual development in regulating expression of Mullerian inhibiting substance (MIS) or anti-Mullerian hormone (AMH) (Cate *et al.*, 1986; Shen *et al.*, 1994; Giuili *et al.*, 1997). WT1 is a zinc-finger transcription factor, which associates and synergizes with SF-1 (Nachtigal *et al.*, 1998). It has been

shown that DAX-1, which is the product of an X-linked gene and ordinarily acts in the absence of SRY, is important for female sex development (review, Swain and Lovell-Badge, 1997; swain *et al.*, 1998). Most of these genes involved in sex determination are transcription factors and are able to switch on or off other genes. DAX-1 and WT1 oppose each other to modulate SF1 mediated transactivation of male-specific genes. In addition, there are other genes (figure I.1) both X-linked (*SOX3*) and autosomal (*SOX9*) which are involved in sex determination (review, Jimenez and Burgos, 1998). However, the details of communication and interaction between these genes and their products during sex determination still remain unknown.

In the absence of *SRY* gene expression, the gonad eventually differentiates into an ovary. The testis, when *SRY* is expressed, develops more rapidly than the ovary. By the end of the sixth week the human testis shows evidence of differentiation. Primordial germ cells migrate from the posterior wall of the yolk sac into primitive sex cords in the genital ridge. During their migration they proliferate in response to some mitogenic factors such as LIF (leukemia inhibitory factor) and MGF (mastocyte growth factor) (De Felici *et al.*, 1992). The primitive sex cords become enlarged and are better defined, and it is thought that the precursors of the Sertoli cells are derived from them. The outer portion of the testicular sex cords forms the seminiferous tubules and the inner portion is the origin of the efferent ductules. MIS controls the subsequent steps of male sexual differentiation by causing the Mullerian ducts to degenerate and the Wolffian ducts to develop into the seminal vesicles, epididymis and vas deferens.

The pritubular myoid cells and myoepithelial cells surrounding blood vessels are also critical to cord formation (review, Greenfield, 1998).

Leydig cells appear during the eighth week and begin to produce androgenic hormones (testosterone and androstenedione). Differentiation of the male external genitalia at this stage depends on androgenic hormone secreted from the foetal testis (Huhtaniemi 1994). The production of testosterone by the foetal testes is stimulated by chorionic gonadatropin (CGH), a hormone secreted by the placenta. Hormonal activity of the Leydig cells finishes by about 18 weeks. The cells remain

undeveloped and do not reactivated until puberty to stimulate spermatogenesis. During late embryonic and foetal periods and after birth the primordial germ cells in the testis divide by mitosis, very slowly (Carlson, 1994).

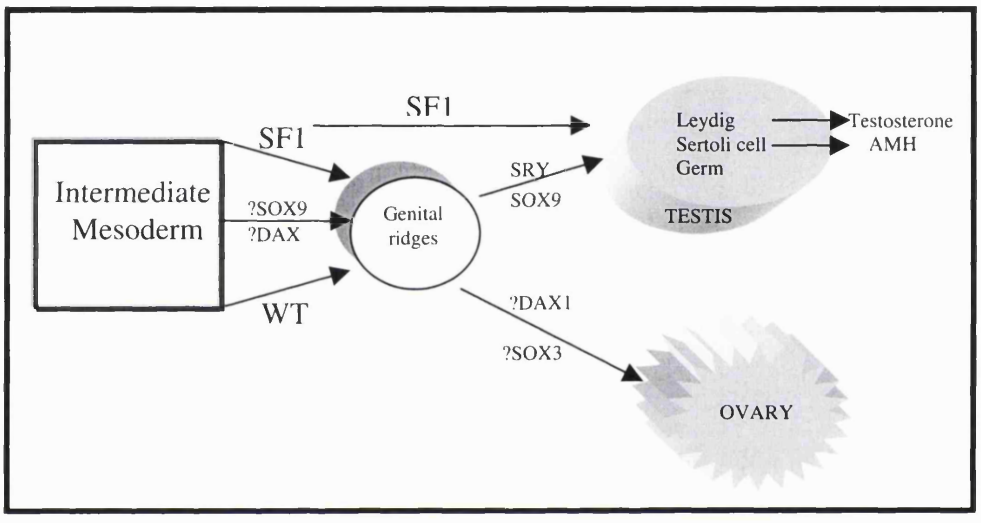


Figure I.1: Effect of genes in the pathway of gonadal development and differentiation. Chromosomal location of candidate genes in the testicular development cascade are:
***SRY* : Yp11.3, *DAX-1* : Xp21-p22, *SF-1* : 9q33, *WT-1* : 11p13, *SOX-9* : 17q24-q25 and *AMH* : 19p13.3**

1.1.3 Descent of the testis. The testes are retroperitoneal structures. They migrate behind the peritoneal epithelium from an intra-abdominal location into the scrotum. This process starts during the seventh month and may not be completed until birth but after the completed descent the left testis is usually located about 1cm lower than the right one in the scrotum. The action of testosterone and the guidance

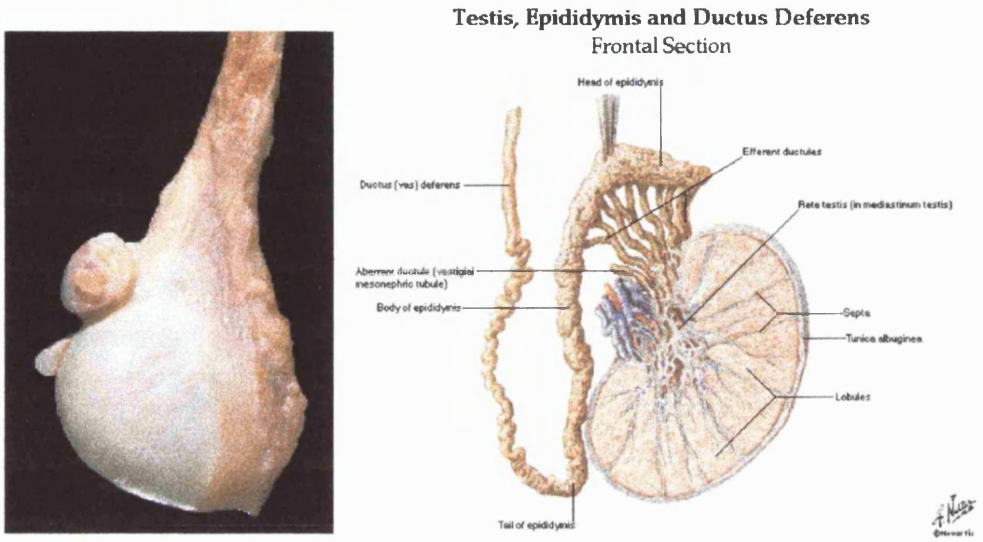
of the inguinal ligament mesonephros are very important in controlling the process of testis decent (Carlson, 1994).

1.2 Anatomy and histology of testis

1.2.1 Macro-anatomy. The testis is the paired male gonad located outside the abdominal cavity, peritoan, in the scrotum, a cutaneous fibromuscular sac which is temperature sensitive. Each testis is roughly egg-shaped, and in the adult, measures approximately 4 cm between the rounded superior and inferior poles, 3 cm wide and 2.5 cm deep with average weight of 10.5-14 g (figure I.2-A)(Gray's anatomy; 1995).

1.2.2 Micro-anatomy. The testis lies with in a serous sac, the *tunica vaginalis* and is covered by a smooth white capsule, the *tunica albuginea*, from which septa extend into the organ to divide it into 200-300 pyramidal lobules (in human). Each contains one to three tightly packed and highly convoluted *seminiferous tubules*, which contain gametes in various stages of development and supportive cells, Sertoli cells. The tubules in each lobule join to form a short, straight tube called the *straight rectus tubule*. Towards the posterior portion of the testis, *straight rectus tubules* from all lobules form a collecting network of interconnecting tubes called the *rete testis*, which empties into highly convoluted *efferent ductules* (figure I.2-B). Attached to the posterior aspect of the testis and running longitudinally along its postero-lateral aspect are the conical shaped lobules of the *epididymis*, which is made up of a tightly coiled tube (figure I.2). The *efferent ductules* drain into the duct of the *epididymis* in the upper portion of the epididymis (called the head or *caput*) which descends in the body (*corpus*) of the *epididymis* to the tail (*cauda*) at the inferior pole where it becomes the *ductus* or *vas deferens*. The *vas* ascends behind the testis into the spermatic cord, which also contains the artery, vein, nerves and lymphatics and all are surrounded by muscle (figure I.2-B).

The *Tonica vasculosa* contains a plexus of blood vessels and delicate loose connective tissue, extending over the internal aspect of the tunica albuginea and covering the septa and all the testicular lobules (Gray's anatomy; 1995).

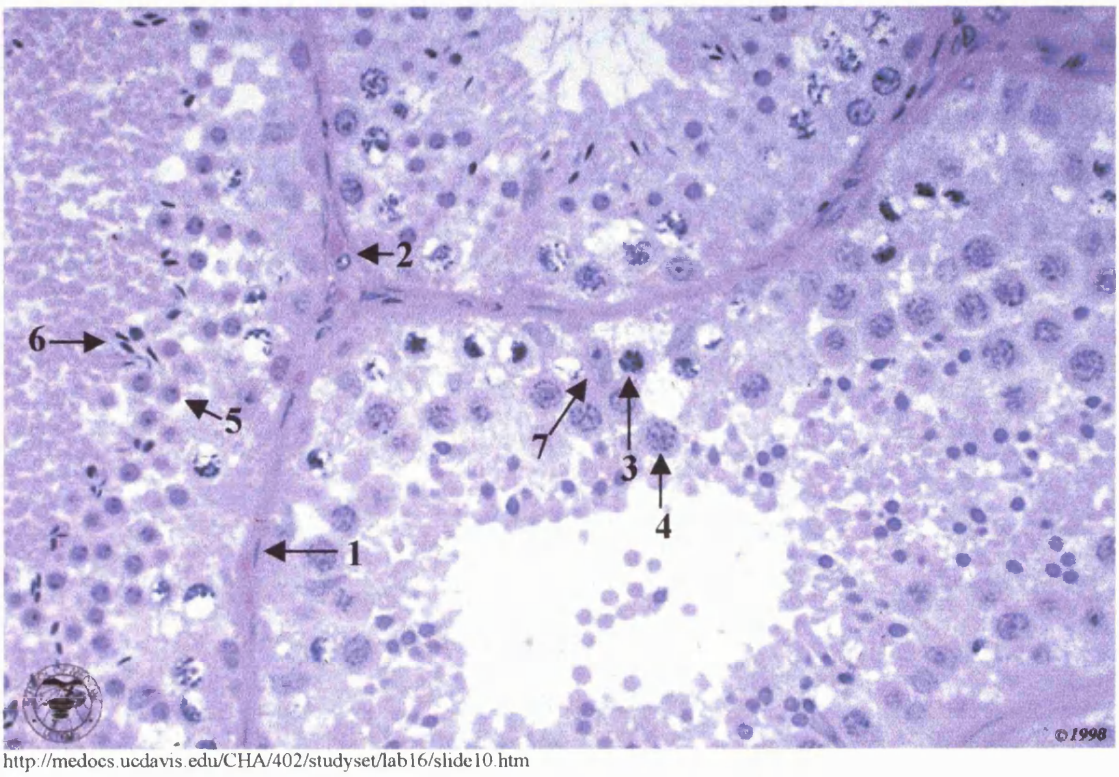


A B

Figure I.2: A) Normal testis, epididymis and spermatic cord. B) Frontal section through the testis and epididymis, showing the arrangement of the ducts of the testis. (Figure A obtained from site http://www-medlib.med.utah.edu/WebPath/MALEHTML/MALE130.html.)

1.2.3 Histology. 85-90% of the testis is composed of semniferous tubules (figure I.3). Other compartments between tubules are vessels including veins, arteries and lymphatics; nerves and containing the Leydig cells. Each seminiferous tubule is surrounded by a well-defined cellular basement membrane. This contains myoid cells and has a loose coat of interstitial fibrocytes separating the interstitial and basal components. No blood and lymphatic vessels or nerves penetrate into the seminiferous tubules (figure I.3).

The epithelium of seminiferous tubules is composed of two basic cell types: the supporting cells (sustentocytes) called the Sertoli cells, and spermatogenic or germ cells. The Sertoli cells are non-proliferating elongated cells with cytoplasm extending from the basement membrane to the lumen of the tubule. They mechanically support and protect the spermatogonia and have biochemical

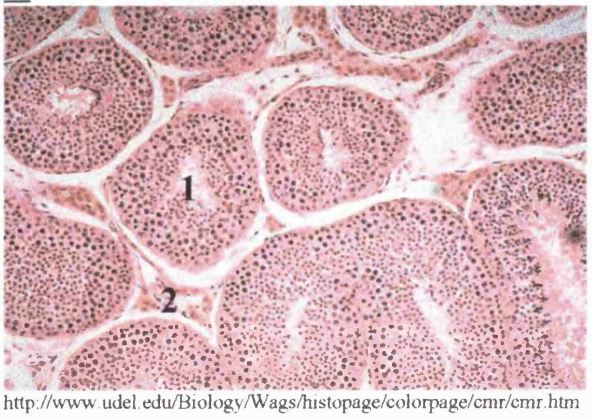


http://medocs.ucdavis.edu/CHA/402/studyset/lab16/slide10.htm

A

Figure I.3: A) Section through the wall of a seminiferous tubule in the testis showing the various cell types that make up the epithelium. Interstitial compartments are interstitial fibrocytes and myoid cells (1), Leydig cells (2), large, polygonal cells with round nuclei and abundant eosinophilic cytoplasm, which may contain rectangular crystalloids (Reinke crystals) and an arteriol. Arrows (3) and (4) point to a spermatogonium and a spermatocyte respectively, which are located close to the basement membrane. Spermatocytes lie lower in the seminiferous epithelium than the round spermatids (5). Elongating spermatids (6) at the apex are about to be released into the lumen. Sertoli cells (7) (elongated cells with oval nuclei and prominent nucleoli) are attached to the basement membrane and stretch into the lumen of the tubule.

B



http://www.udel.edu/Biology/Wags/histopage/colorpage/cmr/cmr.htm

B) As the germ cells proliferate and undergo maturation, they move toward the lumen of the seminiferous tubules (1) such that more differentiated forms are nearer the lumen. Interestitial tissues shown by (2).

interaction and a metabolic influence in relation to germinal cells (review, Griswold, 1998). These cells are also phagocytic (Carr *et al.*, 1968). They are responsible for the blood-testis barrier by virtue of the tight junctions between them. This barrier preserves a constant favourable environment for the process of spermatogenesis and prevents germ cells from immunologically provoking a response. This isolation is important because spermatozoa are produced during puberty, long after the period of self-recognition by the immune system. The Sertoli cells also have responsibility for nourishing spermatocytes (Johnson and Gomes 1977, review, Kerr 1991).

Before onset of sexual maturity at puberty, the tubules contain only small numbers of the most immature germ cell, the spermatogonia.

The spermatogenic cells are actively replicating cells at various stages of a complex differentiating process called spermatogenesis (figure I.4). After puberty, two basic groups of spermatogonia, type A (dark and pale) and type B have been described (Dym, 1994). Type A (dark) divides to maintain a constant source of spermatogonia, producing some type A (pale) cells which divide and differentiate into type B, to become the spermatogonia which undergo mitosis to produce primary spermatocytes. It has been shown *in vivo* that at least two thirds of the type A cells do not give rise to other type A cells or to the differentiated intermediate, type B spermatogonia, but instead, die by a process known as programmed cell death or apoptosis (figure I.5)(Allan *et al.*, 1992; Dym, 1994).

The primary spermatocytes with a diploid chromosome content undergo the first meiotic division to produce transient secondary spermatocytes, which in turn undergo the second meiotic division to form haploid spermatids. The spermatids then mature into spermatozoa, during the spermiogenesis process which takes place in the epididymis tubules (figure I.3 and I.4). The most dramatic morphological changes take place during this stage. This transformation includes nuclear condensation, acrosome formation, loss of most of the cytoplasm, elongation, development of a tail and arrangement of the mitochondria into the upper part of the tail to power it. The sperm head contains several novel structures, for example the acrosome and calyx and contains specialised enzymes necessary for sperm-egg

interaction and fertilisation. In the mature spermatozoon most of the cytoplasm has been lost. It consists largely of a nucleus, mitochondria and a flagellum.

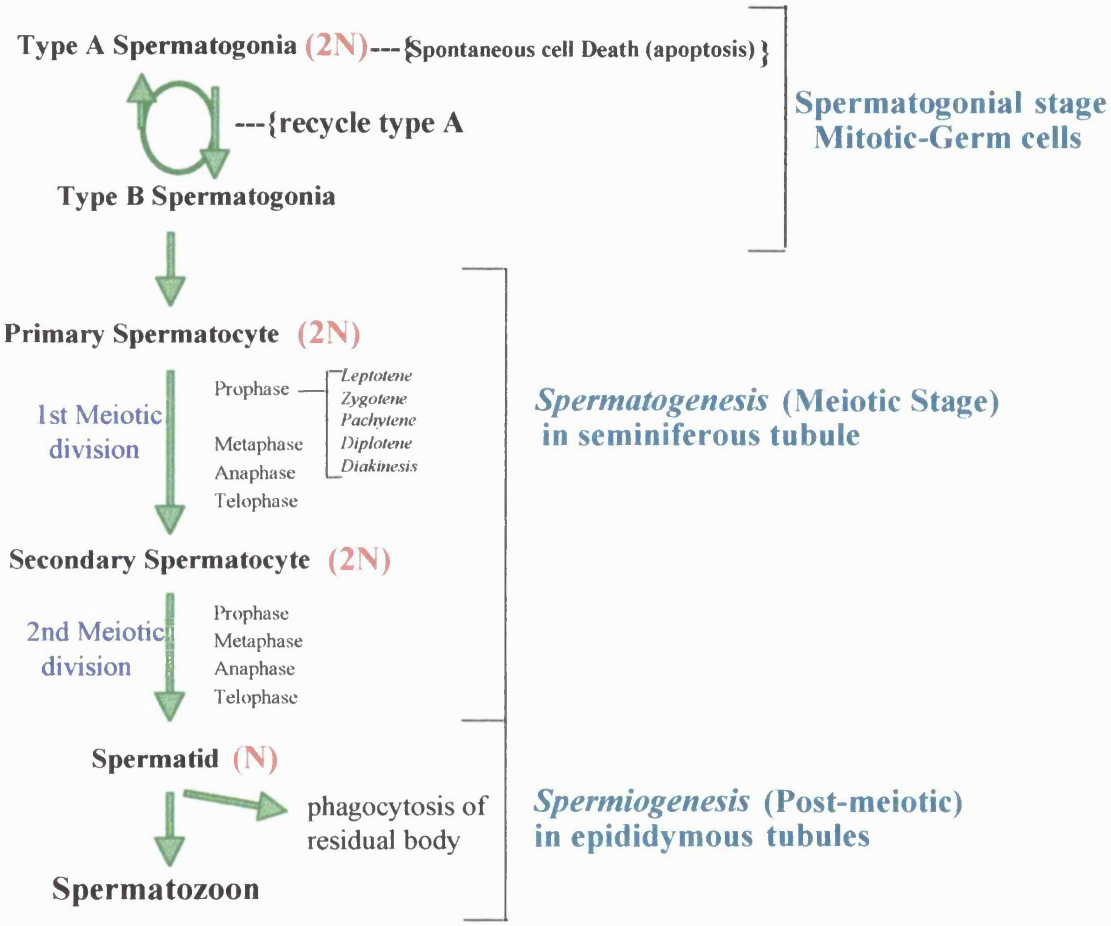


Figure I.4: Schematic showing the stages of mitotic and meiotic division during the differentiation of germ cells. The ploidy of each cell type is indicated. The whole process of sperm maturation from spermatogonia to presence of sperm in an ejaculate will take at least 80 days and depends on hormonal signals combined with a mechanical process (Jequier , 2000).

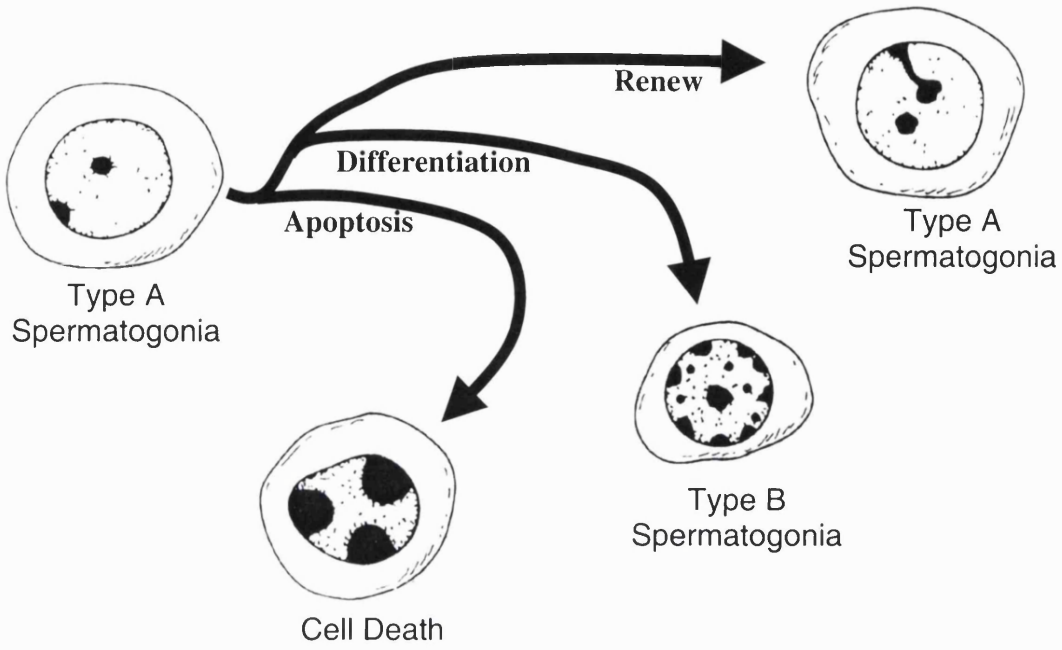


Figure I.5: Fates of type A spermatogonial stem cells.

1.3 Physiology and function of testis

The testis can be thought of as a gland and as having two roles; the first, carried out by the somatic cells, is its function in exocrine and endocrine secretion producing semen and male sex hormones (androgens) respectively in mature testis. The second role is the production of germ cells and their maturation to fertile spermatozoa. These processes of spermatogenesis and spermiogenesis also require a contribution from the somatic cells (Hecht, 1995). The two functions of the testis are closely related, since androgens are necessary for the production of spermatozoa, and also for their successful delivery, which requires normal sexual behavior, and the development of the secondary sexual characteristics, which are also under the control of androgens.

1.3.1 Signals are involved in regulating spermatogenesis and spermiogenesis.

The whole process of sperm production depends on a complex network of endocrine, paracrine and autocrine communication (figure I.6). It requires a system of extracellular signals to control cellular survival and proliferation, patterning and promotion of cell differentiation. Signalling mechanisms require a signal source (hormones), a signal reception system (such as G protein-coupled receptors) and an intracellular signal response system (which can be G proteins discussed in chapter IV). Hormonal signals that originate from a distance and are distributed via the circulation are endocrine signals.

Paracrine signals originate from nearby cells. There are many different paracrine interactions between the germ cells and Sertoli cells due to their intimate location. Sertoli cells provide physical support, junctional complexes or barriers (testis- blood-barrier) in the seminiferous tubules. For instance, it has been reported by Grandijean and co-workers (1997) that Sertoli cell-mediated phagocytosis to eliminate apoptotic spermatic cells in testis is regulated by a complex set of positive and negative signals emitted by the germ cells at distinct maturation stages. Cells can also signal themselves; these kind of signals are autocrine signals and this occurs during germ cell apoptosis (Callard *et al.*, 1995; Review Vaux and Strasser, 1996).

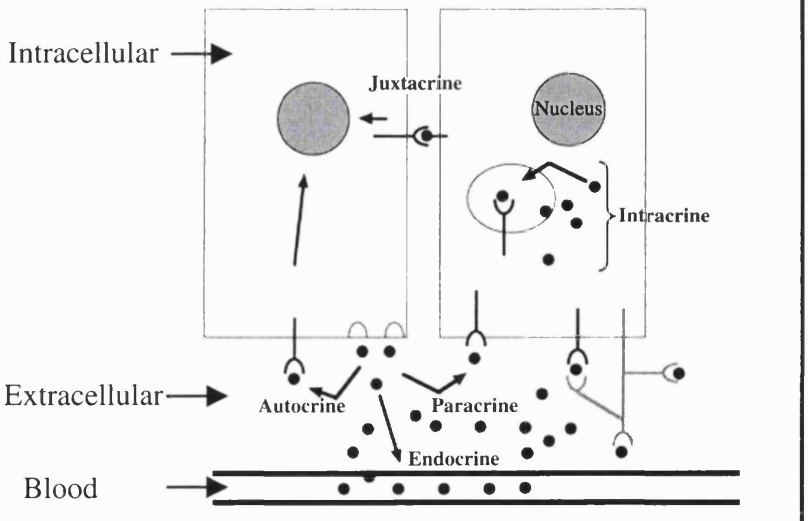


Figure I.6: Different methods of action for signalling mechanisms between sources and targets of signals. These different kinds of transduction (such as growth factors) are seen between cells in testis (see the text for specification).

In the classic pathway of activation, steroid hormones, such as androgens and female sex hormones, pass through the cell membrane and bind to intracellular receptors. These receptors are hormone-regulated transcription factors and after binding to hormone, activated components migrate into the nucleus and regulate gene expression. On the other hand, it has been published that steroids can cause effect by another way as a "non genomic action" (review, Wehling, 1997). The signal is transferred by the interaction of steroid with a non-specific or steroid-specific membrane receptor. In addition, it has been shown that when esteradiol makes a complex with an intracellular receptor it activates the tyrosine kinase and MAP-kinase pathway in human mammary cancer MCF-7 cells and acts like a peptide mitogen (Migliaccio *et al.*, 1996). In *Nature* (1998) Didier Picard warned us not to " underestimate nature's wizardry- because steroids can reach specific receptors inside cells and elicit changes in gene expression, but this does not mean that they always have to".

In the testis, migration, proliferation and differentiation of the most immature germ cells (spermatogonia) are regulated by the c-kit ligand (Bhasin *et al.*, 1994; Griswold, 1993; Niederberger, 1993; Sharpe, 1994). The c-kit proto-oncogene encodes a transmembrane tyrosine-kinase receptor with homology to the receptors for platelet-derived growth factor (PDGF) and CSF-1 (Chabot, 1988). The CSF-1 growth factor (also known as colony stimulating factor-1 or macrophage colony stimulating factor) promotes the growth and survival of cells of the myelomonocytic lineage such as macrophages. Stem-cell factor (SCF), also designated kit ligand, mast cell growth factor, steel factor, or SLF, is a hematopoietic and tissue growth factor, which is produced by the Sertoli cells, that binds to the receptor encoded by the c-*kit* proto-oncogene (Witte, 1990; Rossi *et al.*, 1993). The c-Kit receptor, is present in type A spermatogonial population (Sorrentino *et al.*, 1991). The Sertoli cell hormone binds to the c-kit tyrosine kinase receptor on spermatogonia and produces these profound events. The above-mentioned growth factors belong to the PDGF family which is the major protein growth factor in human serum. Growth factors can be placed into several groups (table I.1).

Table I.1: Classification of Growth Factors

1- PLATELET-DERIVED GROWTH FACTOR (PDGF) FAMILY
2- EPIDERMAL GROWTH FACTOR FAMILY
3- FIBROBLAST GROWTH FACTOR FAMILY
4- INSULIN FAMILY
5- HEPATOCYTE GROWTH FACTOR
6- NERVE GROWTH FACTOR FAMILY
7- LIGANDS FOR AXL/UFO FAMILY
8- LIGANDS FOR EPH-LIKE RECEPTORS

1.3.2 Signal transduction cascades in mitotic cells.

The first part of spermatogenesis is mitosis. The best known pathway, the Ras → Raf → MAPK cascade [mitogen-activated protein (MAP) kinase], is typically strongly stimulated by mitogens, extracellular factors such as growth hormones and proto-oncogene products (Cobb *et al.,* 1995). This pathway is implicated in both regulated and deregulated cell proliferation (induced by growth factors and Ras transformation, etc.) as well as the control of differentiation (Schonwasser *et al.,* 1998). MAP kinases phosphorylate and activate nuclear transcription factors that in turn activate genes controlling the cell cycle. This will be discussed at more length in chapter III and IV.

1.3.3 Endocrine signals related to spermatogenesis/spermiogenesis.

The overall co-ordination of spermatogenesis is orchestrated by endocrine interactions between the hypothalamus, the pituitary gland, the adrenal gland and the somatic cells of the testis.

1.3.3.1 Androgens. Testosterone and androstendione, are secreted by the Leydig cells at two stages, the first during embryogenesis (Section 1.1.2) and then again after puberty. Most cells in the body have receptors for androgens. There are two different targets for testosterone; somatic cells in a variety of tissues and organs such as skeleto-muscular or central nervous system and the testis itself. In somatic tissues testosterone stimulates the development of secondary sexual characteristics such as the growth of facial hair in males. It maintains sexual drive and is absolutely required for successful sexual intercourse or copulation by the male. In

the testis the targets are Sertoli and germ cells but androgen receptors are expressed more strongly in Sertoli cells (Sar *et al.*, 1993). Testosterone diffuses into the seminiferous tubules, where it is converted into dihydrotestosterone (figure I.7).

1.3.3.2 Luteinizing hormone (LH) and **follicle stimulating hormone (FSH).** Both are released from the pituitary gland under control of **gonadotropin-releasing hormone (GnRH)** which is released from the hypothalamus (Review by Chieffi *et al.*, 1991). There is a classical feed-back mechanism to control the proper amount of LH and FSH in the circulation system by hypothalamic stimulation (figure I.7). The combined action of FSH and testosterone secreted by the Leydig cells under LH stimulation causes Sertoli cells to secrete particular peptides (described later) which cause germ cell development (paracrine regulation). One effect of FSH on Sertoli cells is to cause them to secrete androgen-binding protein, which binds to androgens and may facilitate their direct effects on germ cell differentiation. Gonadotropin receptors belong to a subgroup of the super G protein-coupled receptors. These receptors are characterised by a large N-terminal extracellular domain containing leucine-rich repeats that is responsible for binding of the hormone.

LH and FSH interact with specific G-protein-coupled membrane receptors and stimulate adenylate cyclase activity. The consequent activation of the cAMP-dependent protein kinase A results in phosphorylation of a family of transcription factors containing the basic domain/leucine zipper motifs which bind as dimers to cAMP-responsive elements(CREs)(see section 3.1.2, CREM).

1.3.3.3 Female hormones. It has been reported by several authors that female hormones such as estrogen and progesterone are made by both male and female and have individual roles in spermatogenesis and/or spermiogenesis. The adrenal gland is the only other gland beside testis (and ovary in female) in the body that is known to secrete both male and female steroid sex hormones. However, both sexes make both hormones; it must be noted that circulating estrogens originate by the aromatization of androgens catalysed by the cytochrome P450 aromatase. So the

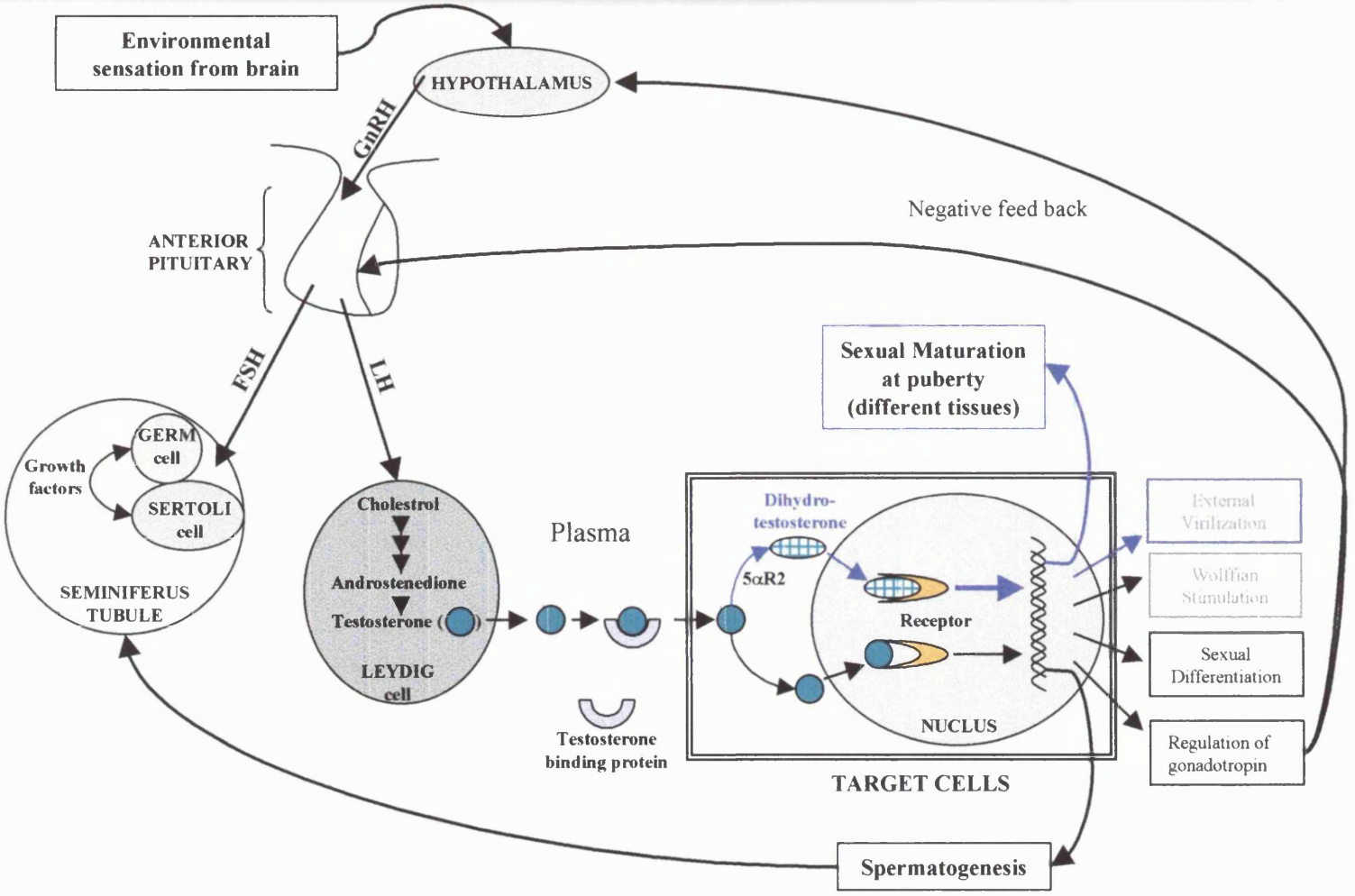


Figure I.7: This schematic diagram shows the regulation and transduction of pituitary and testis hormones. 5a-reductase (5aR2) converts testosterone to dihydro-testosterone. This changes a weak hormone to more potent hormone and is essential for many androgen actions. Deficiency of this enzyme causes defects in blue boxes. In the seminiferous tubules of the testis Sertoli cells are one of it's targets.

critical points in the determination of the sexual distinctions are the quantitative divergence in sex hormone concentrations and the differential expression pattern of steroid hormone receptors.

1.3.3.3.1 Estrogen. Circulating estrogens originate by the aromatization of androgen catalysed by cytochrome P450 aromatase. This enzyme is present and active in germ cells of adult mouse and rat testis (Janulis *et al.*, 1996 and 1998) and its deficiency causes spermatogenic impairment. (Robertson *et al.*, 1999).

Estrogen is also produced by the somatic cell component of the testis (Cook B.A, 1996; McLachlan *et al.*, 1996). Estrogen may even have an essential role for male fertility because male mice lacking the alpha form of the estrogen receptor are infertile (Eddy *et al.*, 1996).

1.3.3.3.2 Progesterone. Progesterone has no known role in testis function. The main action of progesterone in female genital tract before fertilisation is progesterone-induced acrosome reaction in human sperm to stimulate release of the acrosomal lytic content (Osman *et al.*,1989; Baldi *et al.*, 1995&1998). In high concentrations progesterone inhibits the secretion of FSH and LH (review, De Kretser and Phillips, 1998).

1.3.3.4 Gonadal peptides and growth factors as mediators of development and functional control of the testis. Growth factors are proteins that bind to receptors (G protein-coupled receptors) in the surface of target cells, and either stimulate cell division or alter cell fate. The activation process and function of G proteins will be discussed more in chapter IV.

Sertoli cells produce a number of growth factors. In the foetal testis, one is **seminiferous growth factor** (SGF), which stimulates somatic cell proliferation and blood vessel production and causes foetal and postnatal testis development (Feig *et al.*, 1980). In the adult, Sertoli cells respond to their own production of SGF by producing **sulfated glycoprotein-2** (SGP-2). SGP-2 is the major secretory product

of adult Sertoli cells and acts as a paracrine signal with a classical negative-feedback control and binds to the membranes of spermatozoa.

Another secretion of Sertoli cells are **transforming growth factors β** (TGF-β), a superfamily of growth factors, whose receptors are also expressed in the mammalian testis. TGF-ß members regulate a variety of developmental processes including spermatogenesis (review, Massague, 1998).

The Sertoli cells secrete **inhibin**, one member of the TGF-β family, a peptide hormone that inhibits FSH secretion (by a classical negative-feedback mechanism) and androgen-binding protein (ABP) that helps testosterone bind within the seminiferous tubule and stimulates Leydig function (Bremner, 1989; Lin *et al.*, 1989).

Another category of TGF-ß molecules that plays a role in spermatogenesis is the **bone morphogenetic protein (BMP)** sub-family. At least one member of this sub-family, BMP8b, which is produced within the germ cells themselves, is essential for spermatogenesis in mice. It is suggested that BMP8 is required for the resumption of male germ-cell proliferation at puberty and the maintenance of the germ cells in the adult (Zhao *et al.*, 1996).

Sertoli cells secrete other factors that have both negative and positive effects on Leydig cell functions; activin, arginine vasopressin, the cytokine interleukin-1 (IL-1) and tumour necrosis factor (TNF) have negative effects while endothelin is a stimulator of Leydig cell function. It should be noted that the production of these factors by Sertoli cells may depend on the developmental stage of the associated germ cells thereby stimulating Leydig cells along the length of the seminiferous tubule to periodically produce hormones and growth factors appropriate to the developing germ cells in the adjacent tubule (Saez, 1994; Gnessi *et al.*, 1997).

Apart from androgen hormones, Leydig cells produce small amounts of many hormones the majority of which act directly on the cells that secrete them (autocrine factors). The autocrine factors that inhibit steroidogenesis may include angiotensin II, corticotrophin-releasing factor (CRF), arginine vasopressin (AVP), renin, oxytocin, endorphin, and ACTH. The role of most of these agents in the regulation of testicular function remains undetermined, but it has been reported that endorphin

influences Sertoli cell division and inhibin production (review; Gnessi, *et al.*, 1997). Gnessi *et al.* have reviewed many proteins and growth factors which are secreted in the testis by either Sertoli cells or Leydig cells.

In summary, signalling mechanisms in spermatogenesis are associated with different pathways containing a variety of growth factors and proteins. They are encoded by different genes and controlled by cascades of transcription factors some of which are discussed in later sections. Many of these gene products such as growth factors and hormones are not only present in testis but also are associated with other pathways in other tissues. Therefore, to identify specific gene expression in a complex tissue like testis it is necessary to compare it to tissues containing similar pathways and signals.

1.3.4 Exocrine products and semen components.

Sperm is not semen. Spermatozoa mature and are stored in the epididymis. Transport of spermatozoa to the epididymis depends on age and sexual activity. During passage of spermatozoa through the epididymis, the sperm develop an increased capacity for progressive motility and also acquire the ability to penetrate the oocyte during fertilization (review, Moore, 1996). This process of sperm maturation in the epididymis is not well understood, but it has been reported in several publications that the epididymis secretes proteins such as specific androgen-dependent protein and glycoproteins that make changes on the surfaces of spermatozoa. The modifications of sperm membrane are coordinated carefully at different zones of the epididymis and presumably are important in post-ejaculatory survival and function of the spermatozoa (review, Jones, 1998; review, Moore. 1998).

Secretions from the testis, epididymis, bulbourethral glans (Cowper's glands), glands of Littre (periurethral glands), prostate, and seminal vesicles make up the normal seminal fluid. The fluid is released from the glands in a specific sequence during ejaculation. Final composition of semen is 5% from testes and epididymis, 60% from seminal vesicles, 30% from prostate, 5% from bulbourethral glands (Jequier, 2000). Although the seminal vesicles are responsible for the formation of a

coagulum, proteinase secreted by the prostate is responsible for semen liquefaction. Several proteases, including prostate-specific antigen and plasminogen activators, play a role in semen liquefaction.

Part two

2.1 Infertility

Infertility is defined by most authorities as the inability of a couple to achieve a pregnancy after one year of unprotected intercourse. The World Health Organization (WHO) estimates that approximately10% of couples experience some form of infertility problem (Okabe *et al.*, 1998). However, the incidence of infertility may vary from region to region. Male infertility plays a significant role in about 50% of infertile couples. An estimated six percent of adult males are thought to be infertile (Purvis and Christiansen 1992). A significant proportion of these are infertile either due to not producing enough sperm, (hypospermatogenesis/ oligozoospermia), or not producing sperm at all (azoospermia). A typical male produces 20-60 million sperm/ml of semen; if there are less than 20 million motile sperm/ml this person is considered infertile due to oligozoospermia; 20-40% of motility is termed as asthenozoospermia and more than 40% of abnormal morphology is designated teratozoospermia (Moosani *et al.*, 1995).

2.1.1 Why male infertility needs to be considered.

Over the past twenty years, a number of investigators have reported a significant decline in the sperm quality and quantity and also a concomitant increase in reproductive problems of humans (Carlsen *et al.*, 1992; Irvine *et al.*, 1996; review, Irvine, 1997). Although there are significant difficulties of interpretation because studies have been conducted in different countries, at different times and have applied different methods of subject selection and laboratory methodologies (Irvine, 1997; Fisch and Goluboff, 1996; Paulsen, 1996; Fisch *et al.*,1996). The main message is, "there has been a genuine decline in semen quality over the past 50 years." reported by Skakkebaek (1992), and as a result, there has been a decrease in the standard lower limit of human sperm count from $60x10^6$/ml to $20x10^6$/ml in the WHO report in 1992.

Recent epidemiological data in humans and animals suggest that this reduction in sperm quantity may be due to the adverse effects of industrial and

environmental toxins, and some authors believe in the adverse effects of 'environmental estrogens'. This has been published as the 'hormone disrupter hypothesis' (Soto *et al.*, 1998), and these chemicals are believed to act in various ways since they might: a) antagonize the effect of endogenous hormones; b) mimic the effect of endogenous hormones; c) disrupt the synthesis and/or metabolism of hormone receptors (review, Berruti, 1998).

2.2 Infertility classification

There are different classifications of infertility in the literature; these are mainly based on causes of infertility. However, the basic cause of male infertility remains obscure in ~30% of the cases (Nieschlag, 1997). For lack of better understanding these patients are assigned diagnoses such as idiopathic male infertility. I have categorised the principle causes of male infertility in table I.2 in a scheme based on those of Krausz and Forti (2000) and Shaban (1999). Each subgroup has been discussed later, with more focus on genetic disorders and chromosomal defects.

Infertility classification

Pre-testicular causes of infertility

- Congenital diseases
  - Isolated gonadotropin deficiency
  - Isolated LH deficiency
  - Isolated FSH deficiency
  - GnRH deficiency
  - Single-gene defects
- Secondary diseases
  - Tumors
  - Infections
  - Iatrogenic
  - Hemochromatosis
  - Hormonal irregularity

Testicular causes of infertility

- Chromosomal abnormalities
  - Klinefelter's syndrome
  - 47,XYY syndrome
  - Noonan's syndrome
  - Sex chromosome anomalies
  - Structural changes of autosomal chromosomes
- Single-gene defects
- Infertility as a symptom of other inherited diseases
- Sertoli Cell Only Syndrome
- Congenital anatomical abnormalities
  - Cryptorchidism
  - Scrotal varicocele
- Gonadotoxins and secondary infertility

Post-testicular causes of infertility

- Disorder of sperm transport
  - Congenital defects
  - Secondary oligospermia or azoospermia
- Disorder of sperm motility or function
  - Asthenozoospermia
  - Epididymis function impairment
  - Immunological defects and anti-sperm antibodies
- Single-gene defects
- Infections
- Sexual dysfunctions

Table I.2: The classification of infertility. Single-gene defects will not be discussed in this part. However, part III focuses more on infertilities which are caused by defects in genes which are expressed in germ cells and are directly involved in the process of spermatogenesis.

2.2.1 Category one: Pre-testicular causes of infertility.

These consist of hypothalamic and pituitary diseases (primary and secondary).

2.2.1.1 Congenital diseases:

2.2.1.1.1 Isolated gonadotropin (LH and FSH) deficiency: Kallmann's syndrome or olfactogenital dysplasia occurs in both sporadic and familial forms, at a frequency of 1 in 10000 men (OMIM no. 308700). The syndrome causes hypogonadism and is often associated in addition with facial abnormalities (reviewed by Layman 1999). It appears to be inherited in both autosomal recessive and dominant forms and shows incomplete penetrance.

2.2.1.1.2 Partial gonadotropin, LH, deficiency results in fertile eunuch syndrome (Faiman 1968)(OMIM no. 228300) with variable degrees of virilisation and very small numbers of sperm in semen.

2.2.1.1.3 Isolated FSH deficiency: This sort of deficiency with normal virilisation and reduced number of sperm is rare (OMIM no. 229070).

2.2.1.1.4 There are other congenital syndromes with different symptoms in addition to hypogonadism, such as Prader-Willi syndrome and Lourance-Moon-Bardet-Biedel syndrome in which patients have defects in the hypothalamus and GnRH deficiency.

2.2.1.2 Secondary diseases:

Hormonal control in the whole body is very important for proper spermatogenesis. For example, Tumours within or close to the pituitary and infarctions of the pituitary can both cause hypogonadism resulting in impotence and infertility. These symptoms may occur years before other tumoural symptoms such as headache, visual problems and thyroid/adrenal deficiency are noticed. Hyperprolactinemia caused by a prolactin-secreting tumour, a microadenoma or a macroadenoma, can result in an inadequate pituitary response leading to depression of testosterone and altered spermatogenesis. Iatrogenic causes such as poor surgery or radiation treatment in the pituitary region may produce infertility. The majority

of patients who suffer from hemochromatosis also have hypogonadism. This problem can be the result of iron deposition in the pituitary and/or testis. If pituitary insufficiency occurs before puberty, the major clinical symptom is growth retardation associated with adrenal and thyroid deficiency prior to infertility.

Estrogens act primarily by suppressing pituitary gonadotropin secretion, resulting indirectly in testicular failure. Over-exposure to estrogen reduces the number of Sertoli cells. A knowledge of the role of estrogens and estrogen receptors (see section 1.3.3.3.1) in the male can provide better understanding of this sort of sub-fertility or infertility.

Both hyper and hypothyroidism can affect spermatogenesis. Exogenous hormones from pathogenic sources such as adrenocortical tumours, Sertoli cell tumours and interstitial cell tumours or administered during the therapy of some chronic diseases, for example ulcerative colitis, asthma or rheumatoid arthritis, can all be causes of infertility.

2.2.2 Category two: Testicular causes of infertility

2.2.2.1 Chromosomal abnormalities:

Several autosomal and sex chromosome abnormalities are associated with severe oligospermia or azoospermia. Chromosome abnormalities can be numerical anomalies such as trisomies or monosomies, or large structural changes such as translocations, inversion or deletions, or more subtle such as microdeletions of particular regions of chromosomes containing genes involved in the process of spermatogenesis (review, Wieacker and Jakubiczka, 1997).

2.2.2.1.1 Klinefelter's syndrome (47, XXY) is a sex chromosome aneuploidy with an incidence of about 1:500 males. About 10% of the patients have chromosome mosaicism. Female secondary sexual characteristics with hypotrophic gonads is a feature of this disorder.

There is a variant of Klinefelter's syndrome, Sex Reversal Syndrome, in which male patients have a 46,XX karyotype (OMIM no. 278850) but have a portion of a Y chromosome somewhere in their genome. This can be the result of translocation of the testis determining factor (TDF) (see section 1.1.2) to the X

chromosome, or undetected 46,XX/47,XXY mosaicism (review, Van Assche *et al.*, 1996; Burgoyne, 1998; Wieacker and Jakubiczka, 1997).

2.2.2.1.2 47,XYY syndrome has about the same incidence as Klinefelter's syndrome and is sometimes associated with male infertility due to abnormal pairing during meiosis (Speed *et al.*, 1991). These patients are tall and it has been claimed that they exhibit anti-social behaviour, but a recent carefully controlled study showed that any slightly anti-social behaviour principally results from their lowered IQ (Gotz, *et al.*, 1999).

Also, a case report has been published of a patient with karyotype 46,XY (18.9%)/48,XYYY (81.1%) mosaicism in lymphocytes who has behavioural disturbances and in whom a testicular biopsy revealed fibrohyalinization in about 10% of the tubules and subnormal spermatogenesis in the others (Teyssier and Pousset, 1994).

2.2.2.1.3 Noonan's syndrome (OMIM no.163950) is an autosomal dominant condition characterised by short stature; broad or webbed neck; sometimes different congenital heart defects and a characteristic faces that changes with age (Allanson, 1987). It is the phenotypic equivalent of Turner's syndrome with a male phenotype. Patients often have cryptorchdism (60%) with other abnormalities of the genital tract and are infertile. Linkage analysis data has suggested a locus for dominantly inherited Noonan syndrome on chromosome band 12q22-4 (Jamieson, *et al.*, 1994; Legius *et al.*, 1998). No candidate gene in this region has yet been identified. Many cases of Noonan's syndrome are sporadic. It is sometimes possible that a disorder, such as XO/XY mosaicism may mimic features of the Noonan's syndrome's (Elsawi, *et al.*, 1994).

2.2.2.1.4 Sex chromosome anomalies and rearrangements can cause spermatogenic impairment and meiotic arrest. They can be the result of pairing failure. XYY mice are usually sterile with signs of meiotic impairment and it has been suggested by Rodriguez and Burgoyne (2000) that sex chromosome asynapsis (or some consequence thereof), rather than Y gene dosage, is the major factor leading to the meiotic impairment of XYY mice. Other examples of rearranged Y-chromosomes are either acrocentric, ring, inverted or dicentric (Maeda *et al.*, 1976;

Chandley, 1979; Delobel *et al.*, 1998; review, Johnson, 1998). In addition, another Y-chromosome anomaly is a deletion, the loss of sequences including vital genes for spermatogenesis (review, Burgoyne, 1998).

It has been shown that deletion of a region of the Y-chromosome long arm can give rise to azoospermia and this led to the definition of one or more azoospermia factors (AZF) thought to be located there (Tiepolo and Zuffardi, 1976). More recently, several gene families have been identified within this region, some of which will be discussed in a later section. Nakahori and colleagues (1996) estimate that around 10% of men with primary azoospermia have deletions or microdeletions within the AZF region.

Tiepolo and Zuffardi (1976) found six cases out of 1170 azoospermia patients with large, cytogenetically visible, deletions of the distal fluorescent heterochromatic part of the long arm of the Y and including part of the non-fluorescent euchromatic region proximal to it (figure I.8). In subsequent studies, AZF was initially mapped to interval 6 within the band Yq11.23 of the Y chromosome (Ma *et al.*, 1992, Vogt *et al*,1996). Further studies identified three non-overlapping regions on the long arm of the Y chromosome. They have been termed AZFa, AZFb and AZFc. However, recent investigations based on the extensive screening of a proven fertile male population in tandem with 514 infertile males, derived from three different patient selection protocols with some specific STSs (Sequence Tagged Sites), suggest the existence of a fourth region, AZFd located between AZFb and AZFc (Kent-First *et al.*, 1999) (figure I.8). The AZFa and AZFb regions are estimated to be between 1-3Mb in size and the AZFc region is 1.4Mb. A number of genes, both Y-specific and X-Y homologous, have been identified in these regions (Review Krausz and McElreavey, 1999). Detailed histopathology analysis of 13 patients revealed that those with deletions within the AZFa region have no germ cells at any stages of spermatogenesis and patients with the AZFb deletion have no post-meiotic germ cells. Deletions within the AZFc region can cause infertility or sub-fertility and usually patients have a small number of germ cells in the seminiferous tubules. Therefore, it might be suggested that

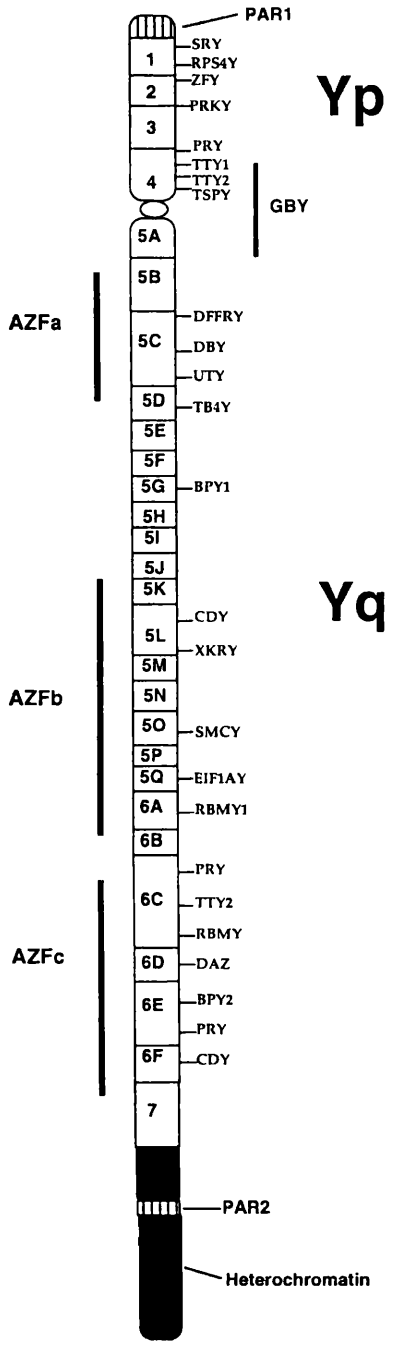


Figure I.8: Schematic representation of the human Y chromosome showing genes and gene families in the non-recombining region. The deletion intervals (AZFa, AZFb and AZFc) are as defined by Vogt et al., *(1996). The aproximate positions of the AZFc deletion (Kent-First* et al., *1999) and the loci associated with gonadoblastoma GBY, and pseudoautosomal regions 1 and 2 are indicated (figure redrawn from McElreavey* et al., *2000).*

AZFc gene products are involved in the maturation process of the post-meiotic germ cells (Vogt *et al.*, 1996).

2.2.2.1.5 Structural changes of autosomal chromosomes, both reciprocal and Robertsonian translocations, have been reported that can give rise to spermatogenic impairment. For examples a Robertsonian translocation of t(13q;14q) or reciprocal translocation t(3;4)(p12;q21) might be associated with male infertility and subfertility (Sasagawa, *et al.*, 1992). Another example is 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). Moreover, it has been reported that chromosome analysis of a patient with azoospermia, mild mental retardation, and minor physical anomalies demonstrated the presence of additional material on the long arm of one chromosome 13 (q13.3), leading to specific trisomy of 5q12 to (Nordgren, *et al.* 1997).

2.2.2.2 Single gene defects:

Single gene defects as causes of spermatogenesis will be discussed in detail in part III of this chapter.

2.2.2.3 Infertility as a symptom of other inherited diseases:

About 80 inherited disorders with a symptom of sub-fertility or infertility were obtained by searching some disease databases, for example Online Mendelian Inheritance in Man (OMIM). Many of them have been explained or mentioned in different parts of this chapter. Here, other examples including types of muscular dystrophy:

Myotonic dystrophy (OMIM no.160900) is an autosomal dominant inherited disease and one of the major clinical features in the patient is male infertility (Bundey 1982). Fu *et al.* (1992) suggested that the gene be referred to as Dystrophia Myotonica protein kinase (MDPK or DMK-protein). It has been demonstrated that the brain and heart transcripts of the *MDPK* gene are subject to alternative RNA splicing in both human and mouse (Jansen *et al.* 1992). In both species another

active gene, called *DMR-N9*, was found in close proximity to the *MDPK* gene. The DMR-N9 transcript is mainly expressed in brain and testis. Clinical symptoms, mental and testicular, of myotonic dystrophy may be caused by the expanded CTG-repeat, an unstable (CTG)5-30 motif, which results in the production of alternative forms of the DM-kinase protein (Jansen *et al.*, 1995).

Another example, a family with adult spinal and bulbar muscular atrophy with X-linked recessive inheritance (Kennedy's disease (OMIM 313200)) associated with testicular atrophy and severe oligospermia has been described (Arbizu, *et al.*, 1983).

2.2.2.4 Sertoli Cell Only Syndrome:

Sertoli Cell Only Syndrome (SCOS) is characterised by azoospermia; slightly smaller than normal testes; absence of germ cells in the testes; and elevated FSH but normal LH and testosterone concentrations (hence, it is associated with normal virilization). SCOS may be the result of congenital absence of the germ cells, gene defects, or androgen resistance, and may be of varying severity. For example, deletions occurring in AZFa result in type I Sertoli Cell Only Syndrome (containing no germ cells at all) and deletions in AZFb and AZFc can be associated with type II Sertoli Cell Only Syndrome (with limited spermatogenesis)(review Krausz and McElreavey, 1999). A large percentage of idiopathic SCOS may be genetically determined, possibly by genes located in the regions of the Y chromosome indicated above (Foresta *et al.*, 1998). Deletions in specific autosomal regions might also result in SCOS. For instance, a case where testicular biopsy showed Sertoli Cell Only Syndrome and the karyotype of the patient was del(1)(q44) may illustrate a link between the development of germinal cell maturation and genes in the 1q44 area (Hathout *et al.*, 1998).

2.2.2.5 Congenital anatomical abnormalities:

2.2.2.5.1 Cryptorchidism is the absence of at least one testis in the scrotum and is a common developmental defect with an incidence of almost 3% of boys in western countries (review, Cortes, 1998). Testicular descent occurs in two

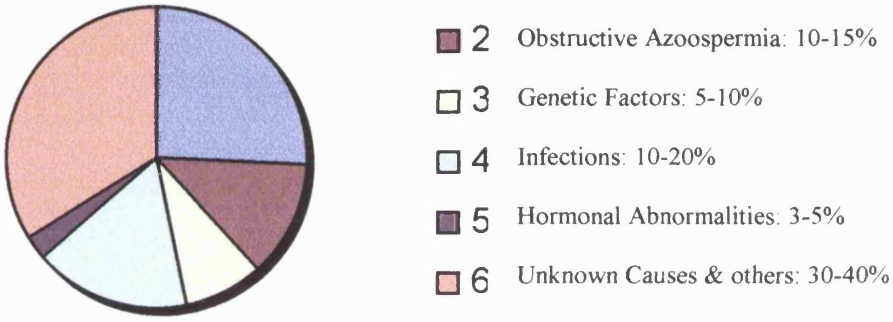
morphologically and hormonally distinct phases (section 1.1.3). The first phase, descending, is controlled by Mullerian inhibiting substance but the second phase, inguinoscrotal migration, is androgen dependent and is possibly mediated indirectly through the release from the genitofemoral nerve (GFN) of the neuropeptide calcitonin gene-related peptide (CGRP) (Hutson *et al.*, 1994). Cryptorchidism might be a feature of abnormalities in the hypothalamo-pituitary-testicular axis and of an abnormal differentiation of the male sexual organs (review, Cortes, 1998). The undescended testes become morphologically abnormal after age two and, without corrective surgery are susceptible to cancer in adulthood. Semen quality is poor in patients with bilateral undescended testicles even after surgical treatment.

2.2.2.5.2 Scrotal varicocele is the most common first finding in infertile men (figure I.9). The incidence of varicocele in adult male and the infertile population is approximately 15% and 30% respectively (Jequir, 2000). About one in two of those who suffer from varicocele can have insufficient spermatogenesis especially in boys as young as 17 to 19 years, by the decrease in motility, vitality and number of normal forms of spermatozoons (Paduch and Niedzielski, 1996). The anatomical problem is valvular deficiency and backflow of blood in the spermatic veins. Because of the anatomical situation of the internal somatic vein most (90%) varicocele formation is on the left side. To date, no definitive reason has been suggested for this association between infertility and scrotal varicocele. Jequier (2000) has reviewed other theories which have been suggested for infertility caused by varicocele. Treatment of varicocele stops the progress of testicular damage and improves spermatogenesis and semen parameters.

Elevated intratesticular temperature has been suggested as a cause of defective spermatogenesis in undescended testes (Crew, 1921) but some authors argue against the theory stating that raising scrotal temperature by 0.8–1.0°C does not affect spermatogenesis or sperm function (Lund and Nielsen, 1996; Wang *et al.*, 1997). There are some molecular data to support this theory, however. An elevated temperature in mouse male germ cells results in decreased expression of cold-inducible RNA-binding protein (CIRP) which is structurally highly similar to RBM1 (section 3.1.5), a candidate for the human azoospermia factor (Nishiyama *et*

al., 1998). In addition, in wild-type mice, a single heat exposure caused the activation of heat shock transcription factor 1(HSF1) which can induce the primary spermatocytes to undergo apoptosis (Nakai *et al.*, 2000).

1- *The Male Factor Breakdown*



2- *Causes of male infertility*

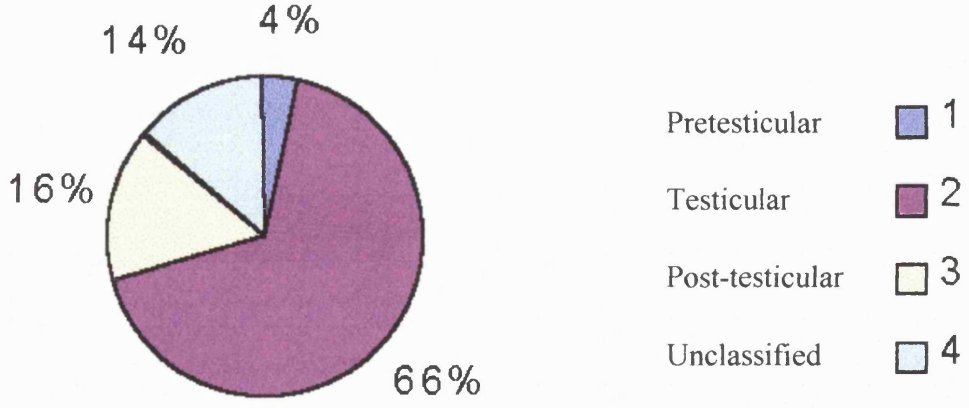


Figure I.9: Pie diagrams show percentages of causes of male infertility, sources of the first pie diagram are different papers which have been mentioned in the text and the second pie diagram has been drown from data in a recent statistical study of causes of male infertility (Devoto et al., 2000).

2.2.2.6 Gonadotoxins and secondary infertility:

Germinal cells divide rapidly and so are very susceptible to agents which interfere with of cell division. Secondary infertility and impairment of spermatogenesis can result directly or indirectly from exogenous or endogenous gonadotoxins. Drugs such as those used in cancer chemotherapy like Cyclophosphamide or war gases such as nitrogen mustard are particularly toxic to testis (review, Hess, 1998). Some other drugs like ketoconazole, spirolactone, cimetidine and even alcohol indirectly affect spermatogenesis by interfering with testosterone synthesis or by acting as testosterone antagonists. Uremia as a result of chronic renal failure or accumulation of some metabolic products like estradiol due to cirrhosis of the liver can both cause infertility (review, Baker 1998; Schmidt and Holley, 1998).

Radiation exposure of more than 600 rads causes irreversible germ cell damage (de Rooij and Vergouwen, 1991). Testicular atrophy can develop after testis torsion or trauma, iatrogenic injury during inguinal surgery or post infections such as mumps and subsequent inflammatory (orchitis) damage. Certain testicular tumours, germ cell tumour, teratoma and seminoma impair sperm production, often severely.

2.2.3 Category three: Post-testicular causes of infertility

2.2.3.1 Disorders of sperm transport:

Obstruction of the ductal system of the male genital tract is a common cause of infertility and may occur in around 15% of all infertility cases (Jequier, 1986). It can be congenital or acquired.

2.2.3.1.1 Congenital defects can be intratesticular obstruction such as Hypercurvature syndrome, which is still of unknown aetiology. It is not common and was first described by Averback and Wight (1979). Rete testis obstruction that might be a result of some form of immune response was demonstrated histologically by Guerin and colleagues in 1981 (Guerin *et al.* 1981). Epididymal

obstruction is the most common site of this obstructive azoospermic group. The aetiology of this has changed over the last 60 years from being mostly the result of gonococcal infection to congenital and idiopathic causes. For example congenital absence of portions of the epididymis and Young's syndrome (OMIM no. 279000) both result in obstructive azoospermia (Wollin *et al.*, 1987). This syndrome consists of bronchiectasis, sinusitis and obstructive azoospermia (Young, 1970). In addition, there are some infiltrative diseases that can cause epididymal obstruction like Sarcoidosis, Polyarteritis nodosa and Fabry's disease (which is an X-linked metabolic disorder resulting from defective activity of the enzyme galactosidase, Nistal *et al.*, 1983). Congenital absence of the vas deferens and Cystic fibrosis (OMIM 219700) are also well-recognised causes of obstructive azoospermia. This is important since, with the use of antibiotics and gene therapy, men with Cystic fibrosis are today surviving well into adult life. Mullerian duct cysts are one of the common causes of ejaculatory duct and vas deferens obstruction.

2.2.3.1.2 Secondary oligospermia or azoospermia may be caused when autonomic neuropathy damage results in lack of peristalsis of the vas deferens and/or failure of the bladder neck to close at the time of ejaculation leading to retrograde ejaculation. Such damage can be seen in diabetes or after pelvic surgery or trauma. In addition, bacterial, viral and filarial infections in the epididymis and vas deferens can all cause this kind of infertility.

2.2.3.2 Disorders of sperm motility or function:

2.2.3.2.1 Asthenozoospermia (low sperm motility) is not rare. It may either contribute to or be the primary cause of infertility in 30% of infertile men (Moore and Reijo-Pera, 2000). Sperm motility problems may be a sign of infection (especially when it is accompanied by a high number of leukocytes in the semen) or exposure to toxic substances. Asthenozoospermia also results from congenital defects of the sperm tail. Primary ciliary dyskinesia (PCD) or immotile cilia syndrome (OMIM 242650) refers to a group of heterogeneous disorders usually inherited as an autosomal recessive trait. The phenotype is characterised by axonemal abnormalities of respiratory cilia and sperm tails leading to

bronchiectasis, sinusitis and male sterility. PCD is approximately 50% associated with situs inversus and this disorder, situs inversus and PCD together, is called Kartagener syndrome(OMIM 244400). The main ciliary defect found in PCD is an absence of dynein arms in axonema of cilia, affecting almost all cilia (Afzelius 1985). Dynein arm structure is made by a large family of dynein proteins involved in many types of microtubule-dependent cell motility in both lower and higher eukaryotes. In recent studies, a mammalian gene, named DNAI1 (which is predicted to encode a protein related to a C. *reinhardtii* dynein protein) has been isolated and its involvement in human PCD has been demonstrated (Pennarun *et al.*, 1999). This gene, which is highly expressed in trachea and testis, is located in the p13-p21 region of chromosome 9.

In general, mammalian sperm motility is regulated by a cascade of cAMP-dependent protein phosphorylation events mediated by protein kinases and any defects in this cascade can cause sperm motility problem.

2.2.3.2.2 Epididymis function is important for sperm maturation and for reabsorption of testicular fluid to concentrate and store spermatozoa. It is debatable whether sperm maturation is dependent on specific regional functions of the epididymis (review, Temple-Smith *et al.*, 1998; Aitken and Vernet, 1998) and the process by which sperm mature is unknown. In the animal model, however, it is well-known that the epididymis secretes proteins which correspond to a change in spermatozoa surface protein, but no epididymal proteins have been identified that appear to be directly involved in modifying the sperm membrane (review, Dacheux *et al*, 1998). Surprisingly, in human, as a result of clinical work using corrective surgery (epididymovasostomy) on patients with congenital absence of the vas deferens or long term blockage of the epididymal duct, the view has arisen that passage of spermatozoa through the epididymis is not necessary for attainment of their fertilization capacity. This is opposite to all established and current information on sperm maturation in animal species (Jones, 1999).

2.2.3.2.3 Immunological defects and anti-sperm antibodies have been shown experimentally to play roles in states of male factor infertility in about 3-7% of infertile males. Autoimmunity is a condition in which the antibodies of the immune

system attack specific cells in the body, mistaking them for foreign antigens. If this process targets the sperm male infertility can result. The most common example of this is the development of antibodies following vasectomy. Antibodies, which can be from male or the partner, bind to particular parts of the sperm (e.g., the head or tail) and cause immunological reaction depending on the site of attachment. Sperm may stick together (agglutinate), fail to interact with cervical mucous, or be unable to penetrate the egg.

2.2.3.3 Infections:

The pattern of infections during past decades has altered and declined due to changes in sexually transmitted diseases and improvements of antibiotic-therapy. Sexually transmitted diseases, particularly repeated Chlamydia trachomatis infections, can trigger events leading to male infertility. Mycoplasma is an infectious organism that appears to bind itself to sperm cells and render them less motile. Another common infection in the male genital tract is chronic prostatitis.

Most of the components of semen are made by the accessory glands, the prostate, seminal vesicles and bulbouretral glands. Seminal vesicles and prostate gland secretions have high fructose, prostaglandins, zinc (which stabilises the cell membrane and nuclear chromatin of spermatozoa), citric acid and cholin which are supposed to have antibacterial effects (Sitaram and Nagaraj, 1995; Lin *et al*, 2000). Apart from causing direct scar tissue damage (see section 2.2.3.2), infections can indirectly cause changes in semen components and in its pH which can affect the longevity of sperm.

2.2.3.4 Sexual dysfunctions:

These disorders should be considered at the first step of evaluation of male fertility problems. Ejaculatory disturbance, impotence and sexual problems with a variety causes may prevent conception. Sexual dysfunction also increases with age.

2.3 Treatment of infertility

In the last decade, the treatment of male infertility for patients with post or almost post-testicular problems has become more promising. First of all, the infertile couple must receive appropriate counselling about the problem before treatment. Medical and surgical treatments concerned with the cause of infertility are the second step of treatment. For example, the performance of a microsurgical epididymovasostomy can sometimes result in success, but this procedure seems to be less successful in Young's syndrome than in other causes of epididymal duct obstruction (Jequier, 2000). The technique of intercytoplasmic sperm injection (ICSI) needs only one spermatozoon in order to fertilise an oocyte; most subfertile or infertile, i.e. men with obstructive azoospermia or extreme oligospermia can now be fathers. Combination of this powerful technique and microsurgical epididymal sperm aspiration (MESA-ICSI) or testicular sperm extraction (TESE-ICSI) has shown that it can be extended even for some of the patients with testicular infertility. Using TESE-ICSI sperm recovery in azoospermic patients with aplasia, arrest and hypoplasia in testis resulted in a pregnancy rate about half that using ICSI with a normal sperm (Tournaye, 1999). However, broadly speaking, the chance for birth is only about 10% in each ICSI. In addition, pregnancies and live births were achieved in a few reports with ICSI using devitalised spermatozoa (Hoshi *et al.* 1994), testicular elongated spermatids injection (ELSI)(Fishel *et al.*, 1995; Vanderzwalmen *et al.*, 1995) and round spermatid injection (ROSI)(Tesarik and Mendoza, 1996; Vanderzwalmen *et al.*, 1997; Choavaratana, et at., 1999). Implantation rates especially after ROSI (Vanderzwalmen *et al.*, 1997; Ghazzawi *et al.*, 1999) or testicular round spermatids nuclei injection (ROSNI)(banned for clinical use in many countries including the UK) has been extremely low. In fact, no pregnancy in human has been achieved using ROSNI (Yamamoto *et al.*, 1999).

In humans, procedures such as ROSNI, ROSI, and ELSI are used for patients with spermatogenic arrest, and therefore the results are not comparable to the animal studies when normal fertile males are used. Thus, the differences in pregnancy and live birth rates between animal models and human cases may reflect abnormalities caused by disturbed spermatogenesis. This hypothesis can be further

supported by the recent findings of Vanderzwalmen *et al.* (1997) and Amer *et al.* (1997) who reported an inverse correlation between ROSI fertilisation and pregnancy rates, and severity of spermatogenesis failure.

Furthermore, DNA sperm fragmentation is negatively correlated with fertilisation rates after ICSI (Lopes *et al.*, 1998). It has been speculated that round spermatids from azoospermic males may also demonstrate increased susceptibility to DNA damage due to their spermatogenic arrest status. Spermatogenic arrest, which may result in abnormal DNA packaging, may contribute to the failure to fertilise (Tesarik *et al.*,1998; Jurisicova *et al.*, 1999). This hypothesis has been demonstrated in that one of the features of the pathology associated with azoospermia is fragmented DNA in haploid germ cells (Jurisicova *et al.*, 1999). This raises questions about the suitability and morality of using these cells for fertility treatment.

In addition, the use of sperm from oligospermic or azoospermic men for ICSI presents a risk because of the possible defects in nuclear genes required for cellular function. For example, Nudell at el. (2000) showed that the mutation rate is higher in the germline DNA, compared with the somatic (blood) DNA, of men with meiotic arrest or incomplete meiotic arrest (oligospermic).

Finally, it is known that a great deal of causes of infertility are directly or indirectly related to genetical defects. By identifying more of these genes it will be possible to screen for mutations that will be inherited by the offspring and will allow any drawbacks of this procedure to be taken into account.

Part three

3.1 Genes involved in spermatogenesis.

The expression of a wide variety of genes is developmentally regulated during human meiosis. The experimental study of this process is more difficult in human than in animals, so the detection of correlated genetic defects and their phenotypes depends on the occurrence of natural mutations and infertile patients within the human population. Most information about the genes involved in human spermatogenesis has been found by the study of infertile patients. Because of ethical and technical restrictions for human studies, spermatogenesis in other mammals has been used as a model of human spermatogenesis. Very many genes are expressed in germ cells, estimates suggest 15,000 to 20,000 genes (Zhang *et al.*, 1997). Many of these are not confined in their expression solely to germ cells but have roles in other cell types also. However some genes are especially involved in spermatogenesis.

The genes which have been described to date to be involved in mammalian spermatogenesis can be categorised into ~~three~~ four groups:

1- Genes expressed in the testis and in spermatogenic cells which are housekeeping genes and are therefore also expressed in almost all cells of the body. These genes will not be described further.

2- Genes which produce a testis specific form of a somatic-cell protein and those genes where expression is altered during spermatogenesis (see section 3.1.2).

3- Genes that are expressed only in the testis and solely or predominantly during spermatogenesis (see section 3.1.3).

4- Spermatogenesis genes not exclusively expressed in testis (see section 3.1.4).

Before discussing examples from each class, it should be noted that it is sometimes difficult to place some spermatogenic genes in a single category. That is why different classifications (noted briefly later) have been suggested by many

authors. For example, several publications and reviews from different authors have focused on genes, vital for fertility, that are Y-linked (section 3.1.6) (Burgoyne, 1987). However, many genes which are important in spermatogenesis, are also located on autosomes.

3.1.1 Haploid gene expression and translation during spermatogenesis.

The analysis of the expression pattern of many genes during testicular development has shown diploid transcription and haploid translation (review, Nayernia *et al.*, 1996). However, another way in which genes might be classified is by whether they are expressed pre or post meiotically.

During spermatogenesis, spermatic cells lose many cellular apparati and enzymes. It is probable that the principle reason for this change is to reduce cell volume. A dramatic decrease is seen in the amount of cytoplasm of spermatid compared with spermatogonia (section 1.2.3). By the end of spermatogenesis at the start of spermiogenesis, when sperm are released into the lumen of the tubule, their ribosomes are nearly absent, and their endoplasmic reticulum (ER) has been lost from the cytoplasm (Clermont and Rambourg, 1978; review, Okabe *et al.*, 1998). This transformation in spermatic cell body during the process of meiotic division and the post-meiotic period is accompanied by changes in the expression pattern of genes which are necessary for post-meiotic functions in spermatid or sperm. Obviously, many proteins are still required during these stages of spermatogenesis and even after that for sperm function during travel in the female genital tract. In addition, it has been experimentally demonstrated that sperm can survive and remain motile for up to 17 days (Akhondi *et al.*, 1997). It was previously thought that a surge of RNA synthesis immediately before the meiotic reduction, along with RNA storage and translational regulation, provided sufficient resources during spermiogenesis to prevent the need for haploid gene transcription (review, Erickson, 1990). However, advances in cloning strategies have led to the identification of genes that are not transcribed until after the meiotic reduction (review, Hecht, 1990). Therefore, genes transcribed during spermatogenesis can be classified into two groups:

1- Genes required for pre-meiotic cell functions. These genes are necessary for cell life, division and differentiation before the second meiotic division.

2- Genes that are involved in spermatid and sperm function, and possibly even for the first divisions of the Zygote, termed post-meiotic genes.

The process of haploid gene expression and translation can also be subdivided into two types. Firstly, those genes that are transcribed during the diploid stage of meiosis and which are subsequently translated in the haploid spermatids and sperm. Secondly, those genes which are both transcribed and translated at the haploid genome stage. Examples of haploid expression are mentioned later in this section.

There are some important questions about post-meiotic translation whose answers are not yet clear. For example, does any translation occur in sperm right up until fertilisation and what is the exact process of translation in the tiny amount of cytoplasm which remains in sperm during spermatogenesis? If there is post-meiotic translation where is the origin of transcripts for protein production? And then, are these transcripts made post-meiotically or are they from mRNAs stored at pre-meiotic stages? The expression of proteins in the acrosome and in the Golgi apparatus of sperm together with the importance of the ER-chaperone protein calreticulin (which is remarkable for its effect on cellular adhesion in the fibroblast) are likely signs of translation in sperm (review, Okabe *et al.*, 1998). Repression and activation of translation presumably require proteins or antisense RNAs that bind to regulatory elements in the RNA and regulate the translatability of messages. Two points during translational control in mammalian spermatogenesis are more characterized from studies on protamine mRNAs. mRNA binding proteins (RBPs) bind to specific sequence sites of a transcript and prevent translation thus creating stored "paternal" mRNAs. For instance, after transcription of protamine 1 in round spermatids the mRNA is stored in an untranslatable form as messenger ribonucleoprotein (mRNP) particles for as long as a week before it is translated (section 3.1.3). Protamine 1 contains a "Y box" binding site for RBP attachment. Y box proteins are sequence-specific DNA binding proteins, and non-specific RNA binding proteins (Wolffe, 1994). Msy1, a mouse Y box protein, sediments in the 60-

80 S mRNP fraction of testis extracts suggesting it may be associated with untranslated RNAs (Tafuri *et al.*, 1993). The mouse p48 and p52 Y box proteins are highly enriched in the testis and have been shown to bind non-specifically to various RNAs *in vitro*, including *Prm1, Prm2, Tnp1, hGH* and *pGem-2* RNAs (Kwon *et al.*, 1993). Within the testis, p48 and p52 are firstly detected in the cytoplasm of mid-pachytene cells and remain through the elongating spermatid stage, consistent with a role in translational repression (Oko *et al.*, 1996).

A structure known as the chromatid body in germ cells has been presumed to be an mRNA storage organelle. There is evidence both to support but also to contradict this hypothesis (review, Hecht, 1998).

The second point at which translational control is exerted is via 3′ and 5′ UTR sites of mRNA. At least two separate regions of the protamin 1 gene 3′ UTR are capable of repressing the translation of reporter mRNA in transgenic mice (review, Braun, 1998). RNA electrophoretic mobility shift assays (EMSAs) using testis protein extracts revealed some new RNA binding sites in the protamine mRNAs (Kwon and Hecht, 1991). The 5′ UTR also seems to play a significant role in isoprotein switching in male germ cells. For example, the cytochrome c_s mRNA in testis is nonpolysomal, perhaps due to the extensive secondary structure of its 5′ UTR (Hecht, 1998).

Translational activation requires that the mRNA be mobilised from repressed mRNPs onto polysomes in elongated spermatids by some unknown mechanism. The repression apparatus must be modified to release the mRNA for translation. Synthesis of a translational activator such as a protein kinase or phosphatase that modifies the mRNP is a possibility. There is evidence for involvement of the cytoskeleton in translational control processes (review, Hesketh, 1994). Evidence that mRNPs and polysomes can associate with cytoskeletal elements supports the conception that this is a mechanism for organised protein synthesis (Braun, 2000).

Shortening of the poly (A) tail of mRNAs from about 100-200 nucleotides to around 30 nucleotides occurs in spermatogenesis and poly(A) tails may not be required for translational initiation. This is in contrast to other cell types where it has been suggested that translational activity in general is promoted by lengthening

of poly (A) tails (Brawerman, 1981; Richter, 1991; Wormington, 1993;) and de-adenylation may be a sign of the beginning of mRNA degradation (Schafer *et al.*, 1995).

3.1.2 Genes which produce a testis specific isoform of a somatic-cell protein

A number of genes undergo stage-specific splicing events during spermatogenesis, resulting in the production of novel transcripts (Willison and Ashworth 1987).The reason for the presence of tissue-specific expression of genes and of testis-specific isoforms of proteins in testis may be due to the specific physiology of the testis. For example, the functional temperature in testis is less than 37°C (see section 2.2.2.5.2) and, in this organ enzyme and proteins have to be optimised for reactions at 33-36°C in contrast to other parts of the body.

In addition, gene expression in actively dividing tissues with a variety of undifferentiated and differentiated cells (such as cancers or germ cells in testis) is more complicated than in a non-dividing and/or a well-differentiated tissue. It is no surprise to find unexpected active genes during this process, for instance male germ cells express the smooth muscle gamma actin gene even though they are not smooth muscle cells (Gu *et al.*, 1996). Moreover, many proto-oncogenes have altered expression during germ cell division; the reason for these cases is not known. For example, the mouse genes *c-abl* and *c-mos*, a protein tyrosine kinase and a protein serine/threonine kinase respectively, are both expressed in many tissues, but in the testis they are present with a different transcript size compared to those seen in somatic tissues. Each, however, contains the somatic cell open reading frame with an altered 5′ or 3′ untranslated region. In addition, a protein apparently binds to a negative regulatory element (NRE) of the *c-mos* promoter to repress expression in somatic tissues, but not in germ cells (Xu and Cooper, 1995). Sequences nearly identical to the NRE of *c-mos* are present in the promoter regions of other genes exclusively expressed in germ cells, e.g. PRM2, Pk2, cytochrome c_T and Hst70, suggesting this may be a common mechanism for suppressing their expression in somatic cells (review, Eddy, 1998).

There is some evidence that many of the novel testicular transcripts show differential capability of translation as a result of either secondary structure alterations or interaction with specific RNA-binding proteins (section 3.1.5) (review, Hecht, 1998).

Testis specific promoters have been described for a number of genes expressed during spermatogenesis, including the alternative promoters of cytochrome c_s and cytochrome c_T which contain similar binding sites (review, Hecht 1998). Some genes utilise a different promoter in testis that activates a unique transcription start site, either up-stream of the usual start site, as occurs for a housekeeping gene cytosolic aspartate aminotransferase (Toussaint *et al.*, 1994), or down-stream of the usual start site, as occurs for angiotensin converting enzyme (ACE) in which case the altered promoter site is somewhere within the gene (mentioned later in this section). They may also be produced by utilisation of alternative exons such as for hexokinase I in the sperm tail (Mori *et al.*, 1998). Alternative polyadenylation signals, as for β1,4-galactosyltransferase, is another example (Shaper *et al.*, 1990). The question arises whether the mechanism of alternative promoter activation is a result of promoter-region hypermethylation (section 3.2) as is used in other actively dividing tissues such as cancers (review, Herman and Baylin 2000; methylation section later in part III). However, the mechanisms responsible for generating most alternative transcripts in spermatogenic cells are still not known.

CREMτ

Genes which have a similar temporal expression pattern during germ cell differentiation tend to contain the same transcriptional regulatory motifs. One motif, the binding site for the cAMP response modulator element CREM, is present in the promoters of many testis-specific genes (reviews, Goldberg, 1996; Sassone-Corsi, 1997). Transcription factor CREM plays an important role in the nuclear response to the cAMP signal transduction pathway in neuroendocrine cells. This pathway begins with the interaction of gonadotropin hormones with specific G-protein-coupled membrane receptors (section 1.3.1)(Stehle *et al.*, 1993; Sassone-Corsi,

1995). In testis, the altered CREMτ, a germ cell-specific activator isoform, is generated by alternative splicing after the late pachytene spermatocyte stage and is highly expressed in post-meiotic germ cells of the testis (Foulkes *et al.*, 1993; review, Sassone-Corsi, 2000). The more ubiquitous CREM α, β and γ isoforms act as spermatogenic inhibitors and are weakly expressed in premeiotic germ cells (Walker *et al.*, 1994). This CREM switch has been reported to be regulated by FSH acting through the Sertoli cells in the mouse, rat and hamster testis. The extinction of *CREMτ* expression in testis has been observed after artificially removing pituitary gland in animals, while the direct administration of FSH restores the expression (Foulkes *et al.*, 1993). In addition, male mice lacking the CREM gene undergo a severe impairment of spermatogenesis, resulting in a post-meiotic arrest at the first step of spermiogenesis and germ cell apoptosis (Nantel *et al.*, 1996, Blendy *et al.*, 1996). A significant decrease in *CREMτ* expression in the semen from oligozoospermic patients has also been shown (Peri *et al.*, 1998).

Hsp70

A family of the heat shock proteins (Hsps), the 70 kDa protein (Hsp70) multigene family, is universally expressed in species ranging from yeast to insects and human, and is essential for growth both at high and low temperatures (Lindquist, 1988). Two Hsp70 isoforms, *Hsp70-2* in humans and *Hsc70t* in mice, are exclusively synthesised during spermatogenesis at high levels in spermatocytes and spermatids. These proteins are molecular chaperones and are thought to be linked to mechanisms that inhibit apoptosis of spermatocytes (Dix *et al.*, 1996; Tsunekawa *et al.*, 1999; review, Eddy, 1999). Hsp70-2 has been better characterised and in *Hsp70-2* knockout mice synaptonemal complexes fail to desynapse, metaphase spermatocytes are not observed, and development of spermatids does not occur (Dix *et al.*, 1997).

LDHc

The three isozymes of lactate dehydrogenase, each encoded by a separate gene, are developmentally regulated and differentially expressed in tissue-specific

patterns. Zinkham *et al.* (1964) found the lactate dehydrogenase-C (LDHc) isozyme in mature testes of many species including man. Its gene (*LDHC*, mouse *Ldh3*) is expressed exclusively in males and in the germ line during spermatogenesis, whereas lactate dehydrogenase-A (*LDHA*) and B (*LDHB*) genes with about 75% homology are active in somatic tissues. This gene is, in human, located to 11p15.5-p15.3 (Sakai *et al.*, 1987).

PGK-2

Phosphoglycerate kinase (PGK) enzyme is required in all metabolically active cell types. Two functional *PGK* genes are present in the mammalian genome, one, *PGK-1*, is X-linked and ubiquitously expressed in all somatic tissues, and a second, *PGK-2*, is an intronless autosomal gene and is expressed specifically during spermatogenesis. This gene has been mapped to chromosome 19 (OMIM no. 172270; Gartler *et al.*, 1986; McCarray and Thomas, 1987). It is likely that *PKG-2* is needed to compensate for inactivation during meiosis of the X- chromosome linked *PGK-1* gene. The expression of this testis-specific isozyme is initiated with the onset of male germ cell meiosis, and continues into the later stages of spermatocytes and in the round spermatids. mRNA transcribed before meiosis is stored for several days before being translated (Gebara and McCarrey, 1992; McCarrey *et al.*, 1996). Thus both transcription and translation of this gene are post-meiotic.

H1t

Histones are the basic nuclear proteins responsible for the nucleosome structure within the chromosomal fibre in eukaryotes and contribute to chromatin condensation. The mammalian testis-specific histone H1t gene is transcribed excessively in primary spermatocytes during spermatogenesis (Doenecke *et al.*, 1994). This protein binds more weakly to DNA than the other six known linker histones (review, Grimes *et al.*, 1997). The most remarkable changes in histone patterns occur at the stage of pachytene spermatocytes when most of the linker H1 histones are replaced by the H1t (review, Doenecke *et al.*, 1997). Therefore, it is

likely that the testis-specific histone is essential for the dramatic changes in chromatin structure and the changing patterns of gene transcription seen during spermatogenesis.

By study of somatic cell hybrids, in situ hybridisation and analysis of a YAC contig Albig *et al.* (1993, 1997) demonstrated that the H1t gene is located within a cluster of 35 histone genes mapped to 6p21.3.

G3PD

Glyceraldehyde 3 phosphate dehydrogenase (G3PD or GAPD) also has a post-meiotic isozyme (GAPD-S) which is expressed only in male germ cells (Welch *et al.*, 1995). Western blot analysis of isolated germ cells confirmed that GAPD-S is not detected in pachytene spermatocytes or round spermatids but indicated that GAPD-S is tightly associated with the fibrous sheath of the flagellum. Therefore, it has been suggested that it has a role in regulating glycolysis and the energy production that is required for sperm motility (Bunch *et al.*, 1998).

ACE

The angiotensin-converting enzyme (ACE) gene (*ACE* in human, *Ace* in mouse) encodes both a somatic isozyme which is present in blood and several other tissues, including the epididymis, and a testis-specific isozyme (testis ACE) found only in differentiating spermatids and mature sperm (Sibony *et al.*, 1993). The testis isozyme is encoded by the second half of the gene under the control of a testis-specific promoter located within the 12^{th} intron of the gene (Howard *et al.*, 1990). This gene provides an example of haploid expression and post-meiotic translation. The gene expression of testis ACE in mouse spermatogenic cells begins in late pachytene spermatocytes and continues after meiosis. The protein has been detected in haploid spermatids (Langford *et al.*, 1993; Sibony *et al.*, 1994). By using knockout mice, it has been demonstrated that ACE plays an important functional role only for male reproductive function during fertilisation because the absence of both ACE isozymes (testis ACE and somatic ACE) causes defects in sperm transport in the oviducts and in binding to zonae pellucidae. However, the absence

of somatic ACE does not impair male fertility. Subsequently, Hagaman *et al.*, (1998) have speculated that the testis ACE isozyme is important for reproductive function and the product of somatic ACE catalysing angiotensin I, is unlikely to be the substrate for testis ACE.

3.1.3 Testis specific genes are expressed solely or predominantly during spermatogenesis

Approximately twenty thousand different transcripts are present in a germ cell, relatively few of them have been characterised in detail. Here are some examples of genes which are exclusively or predominantly expressed in germ cells.

PRM

An example of genes expressed exclusively during spermatogenesis is the gene *PRM* coding for protamine, a main nuclear protein in the sperm. This gene is also an example of the haploid gene expression mentioned above. Histones present in somatic cells and premiotic male germ cells are replaced by transient proteins TNP1 and TNP2 and then by the smaller and more basic protamines, at the final stage of nuclear restructuring towards the end of mammalian spermatogenesis. The molecular characteristics of the protamines as well as the number of protamine types present in the spermatozoon vary from species to species. In mice and men, there are two protamine families *PRM1* and *PRM2*, members of which are encoded by genes clustered in human band 16p13.3 and on mouse chromosome 16 (OMIM no.182880; Viguie *et al.*, 1990). It has been suggested that TNP1, TNP2, PRM1 and PRM2 are associated with sperm nuclear shape and that they initiate condensation of the chromatin and packaging into the sperm head. 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). *PRM1* and *PRM2* are transcribed post-meiotically in the early spermatids, are transitionally repressed for several days and are then translated in elongating spermatids where they replace histones (Schafer *et al.*, 1995; Wykes *et al*; 1995). Numerous mRNAs are under post-transcriptional control during spermatogenesis,

but in only one case, that of the mouse PRM1 gene (*Prm1*), has a mechanism been demonstrated. It is sequences in the 3'UTR that control the translation of this gene (review, Braun, 1998).

Proacrosin

Proacrosin, the zymogen form of the serine protease acrosin, is located within the acrosomal vesicles of mammalian spermatozoa and it has been suggested that it is involved in the fertilisation process. In mouse and rat, expression of the proacrosin gene starts in pachytene spermatocytes and continues through the early stages of spermiogenesis (Nayernia *et al.*, 1994).

Sprm –1

Members of the POU-domain gene family encode proteins which are involved in transcriptional regulation. POU-domain family members are responsible for cell specification and differentiation of several organ systems, including anterior pituitary, sensory neurons, and B lymphocytes.

A member of this family, *Sprm-1* is exclusively expressed during spermatogenesis, predominantly in the haploid spermatid. It was thought that *Sprm-1* encoded a DNA-binding protein required for differentiation of the haploid stage of the male germ cell (Andersen *et al.*, 1993). But a few years later, experiments using "knockout" mice, suggested that the Sprm-1 protein has a particular regulatory role in the haploid spermatid, which the optimal function of the male germ cell. The mutant mice exhibit normal testicular morphology and produce normal numbers of sperm but display subnormal fertility (Pearse *et al.*, 1997).

SPAM1

Another example of a spermatogenesis specific gene is the sperm adhesion molecule 1 (SPAM1). The gene has been localised in the same region as the cystic fibrosis gene (*CFTR*), to 7q31.3. The gene 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 cell layer (Jones *et al.*, 1995).

3.1.4 Spermatogenesis genes not exclusively expressed in testis

A number of genes have been described in which defects cause infertility in both human and mouse strongly suggesting that they are directly or indirectly necessary for meiosis, but they are not solely testis-specific genes. Some genes are vital for spermatogenesis and are necessarily expressed in the testis but are also expressed in some other tissues. Below are some examples, including, in some cases, some that are important even in female meiosis.

AR and AIS

The gene encoding the androgen receptor, alternatively known as the dihydrotestosterone receptor (DHTR), has been mapped to Xq11-q12. Mutation of this gene can cause the androgen insensitivity syndrome (AIS) or the testicular feminisation syndrome (TFM) (OMIM no. 313700 and 300068). The clinical manifestations of androgen insensitivity vary from external genitalia that are completely female (complete AIS) to degrees of partial masculinisation (partial AIS) (review, Hiort *et al.*, 1998). The *AR* gene contains 2 polymorphic trinucleotide repeat segments that encode polyglutamine and polyglycine tracts in the N-terminal transactivation domain of the AR protein. Expansion of the polyglutamine tract causes spinal bulbar muscular atrophy, which is associated with low virilization, reduced sperm production, testicular atrophy, and male infertility associated with impaired spermatogenesis. Spermatogenesis is androgen-dependent, and there is a direct relation between the length of the AR polyglutamine tract and the risk of defective spermatogenesis that is attributable to the decreased AR functional competence that occurs with longer Gln tracts. This polymorphic trinucleotide repeat (CAG)*n* followed by a terminal CAA triplet is located in exon 1 (Lubahn *et al.*,1989; Jenster *et al.*, 1995). One study of 30 patients with idiopathic azoospermia or oligozoospermia concluded that the mean CAG repeat size increased with severity of spermatogenic defect (Dowsing *et al.*, 1999). However, other investigators using an infertile population of Swedish and German ethnic origin could not confirm a relationship between the size of the polyglutamine tract and

impaired sperm production (Giwercman *et al.*, 1998; Dadze *et al*, 2000). There was no significant association between the polyglycine tract and infertility (Tut *et al.*, 1997).

Pms2 and Mlh1

Pms2 and *Mlh1* genes are mismatch repair genes which are also involved in male meiosis (Baker *et al.*, 1995 and 1996; Edelmann *et al.*, 1996) where they have effects on chromosome pairing. Defects of these genes cause male but not female sterility (review, Cooke *et al.*, 1998).

ATM and ATR

ATM was originally isolated as the gene mutated in the inherited human disease ataxia-telangiectasia. It is a member of a family of kinases involved in DNA metabolism and cell-cycle checkpoint control. Meiosis in an infertile mouse (Atm-/-) is arrested at the zygotene/pachytene stage of prophase I as a result of abnormal chromosomal synapsis and subsequent chromosome fragmentation (Xu *et al.*, 1996). The human and mouse counterpart of the *Schizosaccharomyces pombe* Rad3 protein, named Atr (for ataxia-telangiectasia- and rad3-related) have been characterised. It has been demonstrated that *ATR* mRNA and protein are expressed in human and mouse testis in seminiferous tubules, in the process of meiosis I. This gene is also required for cell-cycle checkpoint pathways that respond to DNA damage by blocking replication in yeast (Keegan *et al.*, 1996). The Atr and Atm protein kinases associate with different sites along meiotically pairing chromosomes (Keegan *et al.*, 1996).

3.1.5 RNA Binding Protein (RBP).

Several genes have been cloned that encode putative RNA binding proteins expressed during spermatogenesis (table I.3). However, in most cases the RNA targets are unknown, as are the biochemical function of the RBPs.

Symbol	Name	Comment	X or autosomal homologue and location
RBMY	RNA-binding motif Y	Several subfamilies including *RBMY1* and *RBMY2*. *RBMY2* may be non-functional. Predicted to have RNA binding activity (section 3.1.5).	*RBMY* may be an ancestral *hnRNPG* gene, *RBMX*
DAZ	Deleted in azoospermia	Predicted to have RNA binding activity (section 3.1.5).	DAZL1 chromosome 3p25,
XKRY	XK related Y	Shows similarity to XK protein, a putative membrane transport protein (Ho *et al.*, 1994)	XK; Xp21.2-p21.1, Yq; interval 5L
PRY	PTP-BL Related Y	Shows similarity to the mouse PTP-BL a putative membrane transport protein tyrosine phosphatase (Hendriks *et al.*, 1995)	None found, Yp; interval 4A Yq; intervals 6C and 6E
TTY-1	Testis-transcript Y1	No ORF	Yp; intervals 3C and 4A
TTY-2	Testis-transcript Y2	No ORF	Yp; interval 4A Yq; interval 6C
CDY1	Chromodomain Y 1	Contains a chromodomain and a putative catalytic domain (James and Elgin 1986); intronless; arose by retroposition from CDYL gene (chr. 6); may be involved in chromatid modification during spermatogenesis	Yq; interval 6F
CDY2	Chromodomain Y 2	Shares 98% identity at the aa level with CDY1	None found, Yq; interval 5L
BPY1	Basic protein Y 1	Basic protein, l 125 aa, of unknown function	None found, Yq; interval 5G
BPY2	Basic protein Y 2	Basic protein, l 106 aa, of unknown function; same gene family as BPY1	Yq; interval 6E

Table I.3: Y linked genes (see figure I.8 for Y location) and gene families with expression restricted to the testis and which map to AZF regions (review, Ma et al., 2000; Kostiner et al., 1998).

RBMY

As explained in section 2.2.1.4, the analysis of infertile men with a deletion or microdeletion within the Yq AZF region, revealed at least two gene families that encode RNA binding proteins. The first of these referred to as *RBMY* (RNA Binding Motif, formerly RBM and YRRM) (Ma *et al.*, 1993; Chai, 1997&1998) 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). Men that carry an AZFb deletion have a reduced level of RBMY protein and problems of spermatogenesis (with type II Sertoli Cell Only Syndrome and limited spermatogenesis) at the meiotic to post-meiotic transition (Eliott and Cooke 1997; review Krausz and McElreavey, 1999). *RBMY* members code for proteins with a 90 amino acid 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). Researchers have indicated the presence of at least six RBMY subfamilies termed RBMY1 to RBMY6 and RBMY1, the largest family contains active members (Prosser *et al.*, 1996; Chai *et al.*, 1997). Mice also contain an Rbm gene family with 66% similarity to human *RBMY1* that is located on the short arm of the Y chromosome (Elliot *et al.*, 1997). *Rbm* genes are expressed from 4 days after birth, increasing transcription until 14 days after birth and then are expressed at lower levels in adult testis.

RBMY contains an RNA recognition motif and is highly (around 60%) similar to an autosomal gene on the short arm of human chromosome 6 that encodes heterogeneous nuclear ribonucleoproteins (hnRNP-G). RBMY may be a germ cell homologue of hnRNPs (Soulard *et al.*, 1993; Le Coniat *et al.*, 1992). hnRNPs are predominantly nuclear RNA-binding proteins that form complexes with RNA polymerase II transcripts. They are associated with primary nuclear polyadenylated RNAs and perhaps are involved in pre-mRNA processing and splicing (Soulard *et al.*, 1993; Weighardt *et al.*, 1996). Some hnRNPs shuttle between the nucleus and the cytoplasm and may function in the nuclear to cytoplasm transport of mRNAs. It has been demonstrated that RBM is a nuclear protein that is expressed exclusively

in human male germ cells of foetal, prepubertal and adult testis (Cooke and Elliot, 1997). The actual RNA target(s) of RBMY are still unknown in common with the targets of most hnRNPs. RBMY may interact with a general class of pre-mRNAs, or alternatively, there may be a few selected mRNAs for which it performs some important function (Braun, 2000).

Recently, Delbridge and her colleague (1999) found that RBMY and hnRNPG probes detect dosed bands on Southern blots containing male and female marsupial and human DNA suggesting that the RBMY family has an X-linked homologue in both species. The gene RBMX has been identified and it shares 99% homology at the 3′UTR with the published hnRNPG cDNA (GenBank accession no. Z23064), and is expressed in many tissues. The locations of RBMY-like sequences have been confirmed on the long arm of human X chromosome as well as chromosome 6 by FISH, also homologous sequences been have detected on chromosomes 1, 4, 9 and 11. Comparison of the genomic structure of RBMY1A and RBMY-like sequences on human chromosomes X and 6 shows there is an alternative splicing of RBMX such that three exons are omitted from the RBMX mRNA and that there are no introns present in the chromosome 6 copy, which is presumably a retroposed pseudogene. Like other gene pairs on the X and the Y chromosomes (*Zfx/Zfy*, *Sox3/Sry*), it is postulated that RBMX retains a widespread function and RBMY1 has evolved to become a testis-specific gene (Delbridge *et al.*, 1999).

DAZ

Another gene family located in the AZF region is referred to as *deleted in azoospermia* (*DAZ*) and 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. It has a single RBM with a single RNA recognition motif and a series of between 8 and 24 copies of a 24 amino acid unit termed the "DAZ repeat", its function is unknown (Reijo *et al.*, 1995; Yen *et al.*, 1997). *DAZ* is located in the AZFc region and appears to have arisen from transposition and then amplification of an autosomal gene (Vogt *et al.*, 1996; Saxena *et al.*, 1996). Multiple copies of *DAZ* are 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 which is expressed in human testis and foetal gonocytes, that encodes a 3.3kb transcript (Yen *et al.*, 1996; Seboun *et al.*, 1997; Nishi *et al.*, 1999; Braun 2000). The mouse protein, Dazl1, has been detected at low levels in the cytoplasm of B spermatogonia and leptotene and zygotene spermatocytes, and high levels in pachytene cells (Ruggiu *et al.*, 1997). A *DAZ* homologue, *boule*, has been described in *Drosophila.* Mutation in *boule* results in male sterility due to blockage of meiotic divisions, pachytene arrest and limited spermatid differentiation (Eberhart *et al.*, 1996). Disruption of the *Dazl1* gene in mouse causes loss of germ cells and complete absence of gamete production, in homozygotes both males and females are sterile (Ruggiu *et al.*, 1997). In addition, the localisation of Dazl1 and Boule proteins in the cytoplasm of germ cells at the end of meiotic prophase suggests that they may play a role in cytoplasmic post-transcriptional events (Cheng *et al.*, 1998). Although the *Dazl1* knockout male mouse (*Dazl*$^{-/-}$) remained infertile when the human DAZ transgene was introduced, histological examination revealed a partial and variable rescue of the mutant phenotype (Slee *et al.*, 1999). In *Xenopus*, *Xdazl* is expressed in premeiotic germ cells in adult testis and the protein has RNA binding properties *in vitro* exhibiting specificity for G or U rich RNA sequences (Houston *et al.*, 1998). Interestingly, the *Xenopus Xdazl* gene can rescue meiotic entry of spermatocytes in *Drosophila boule* mutants, suggesting that perhaps there can be functional conservation of the *Daz* family over evolutionary time (Houston *et al.*, 1998).

A recent survey of infertile patients with idiopathic azoospermia and severe oligozoospermia showed that deletions in the AZFc region involving the *DAZ* gene were the most frequent and that they were more often observed in severe hypospermatogenesis than in Sertoli cell-only syndrome (Rejio *et al.*, 1995; Vereb *et al.*, 1997). Therefore it has been suggested that deletions of this region are not sufficient to cause complete loss of the spermatogenic line (Ferlin *et al.*, 1999).

3.1.6 The Y chromosome and spermatogenesis

There are several genes on the Y chromosome that are thought to play roles in testis formation or testis function (spermatogenesis). Some of these, *SRY*, *RBMY* and *DAZ* have been mentioned above. Figure I.8 and tables I.3 and I.4 show most of these genes, indicating also the homologues on other chromosomes. Apart from the genes mentioned already the most studied gene is *TSPY*.

TSPY

The *TSPY* (testis-specific protein, Y-encoded) gene family consists of at least 20-40 tandemly repeated copies. The repetitive element was identified before the gene was described and was named *DYZ5*. On Yp, there are two clusters, TSPYA and TSPYB, within deletion map interval 3; two further clusters on Yq are located in intervals 4 and 5 (see figure I.8) (Arnemann *et al.*, 1991; Vogt *et al.*, 1997). TSPY has been reported to be expressed solely in both prenatal and adult testis. The result of immunostaining in the adult shows it is in a subset of spermatogonial cells which are developmentally immediately prior to the spermatogonia to spermatocyte transition and in particular, are located around the basal lamina of the seminiferous tubules (Arnemann *et al.*, 1987; Zhang *et al.*, 1992; Schnieders *et al.*, 1996). Based on the above result Schnieders *et al* (1996) have hypothesised that TSPY might regulate the normal proliferation of spermatogonia, and their passage into meiotic differentiation. TSPY transcript products have been found as splice variants, and one of them, a 1.3Kb 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). Gonadoblastoma arises within gonadal dysgenesis of individuals who possess Y chromosome material (Barbosa *et al.*, 1995. Using STSs Tsuchiya *et al.* (1995) have narrowed the region associated with this tumour to a 1-2 Mb fragment, surrounding the centromere. Therefore, the localisation may make *TSPY* a candidate gonadoblastoma gene. In addition, it has been suggested that aberrant expression of *TSPY* may also be involved in tumourogenesis of the prostate gland and an up-regulation of *TSPY* expression in a

Symbol	Name of gene	Comments	Locations of both X and Y homologues	X homologue and amino acid identity (McElreavey et al.,2000)
DFFRY (also known as USP9Y)	*Drosophila* fat facets (faf) related Y-linked	two isoforms by alternative splicing; may regulate ubiquitin precursors; Deubiquinating enzyme (Brown et al., 1998)	Xp11.4 / Yq11.2	DFFRX 91%
EIF1AY	Eucaryotic translation initiation factor 1A Y	X homologue implicated in translation initiation; Y copy, unknown function (Lahn and Page 1997)	Xp11.2 / Yq11.2	EIF1-1AX 98%
TB4Y	Thymosin β4 Y	May be involved in actin sequestration (Gondo et al., 1987)	Xq21.3-q22 / Yq11	TB4X 93%
SMCY	Selected mouse cDNA on Y	X and Y copies functional; encode human H-Y antigen epitope HLA-B7 (Agulnik *et al.*, 1994)	Xp11.1-p11.2 / Yq11.2	SMCX 84%
DBY	Dead Box Y	Putative RNA helicase; functional X and Y copies (Linder et al., 1989)	Xp11.3-p11.23 / Yq11	DBX 91%
UTY	Ubiquitous TPR motif Y	Functional X and Y copies; contains 10 tandem tetratricopeptide repeats (TPR) that may be involved in protein-protein intractions (Greenfield et al., 1998);	Xp11.2 / Yq11	UTX 85%

Table I.4. Ubiquitously expressed "housekeeping" genes that map to the AZF regions on Y chromosome that have been implicated in male infertility. These genes are present in a single copy on the Y chromosome, but they have X homologues that might escape X inactivation. The degree of sequence identity between both homologues is more than 84%.

prostatic cell line supports this hypothesis that functional *TSPY* is regulated by male hormone and its receptor in the prostate as well as testis (review, Lau, 1999).

Homologous gene families have been located on Y chromosomes of the great apes, pygmy chimpanzees, cattle and mammals (Schempp *et al.*, 1995; Conrad *et al.*, 1996; Jakubiczka *et al.*, 1993; Vogel *et al.*, 1997). In the cow, the TSPY cluster is located on Yp with 50-200 family members. By contrast in the rat there are only two *TSPY* gene homologues located in Yp, one is functional and expressed in testis and the other is truncated and possibly non-functional. The mouse *Tspy* gene has been identified as a Y-linked, single copy gene but is non-functional (Mazeyrat and Mitchell, 1998; Dechend *et al.*, 1998). It produces only low levels of aberrantly spliced transcripts, and is thought to be a pseudogene (Mazeyrat and Mitchell, 1998). It is not yet clear whether TSPY is present on the marsupial Y chromosome (Delbridge *et al.*, 1997).

Recent studies and accumulating sequences in databases have shown that other genes with reasonable homology to the TSPY-SET-NAP1 family exist. These are known as TSPY-Like (TSPYL) and occur in mouse and man. For instance, one TSPYL gene is located on human chromosome 6 and mouse chromosome 10 and *Tspyl* is expressed in at least eight different tissues (by Northern analysis)(Vogel *et al.*, 1998). However, both the human and mouse *Tspyl* homologues are intronless which may indicate creation by a retroposition event in the past and possibly imply that they are pseudogenes.

3.2 Methylation patterns in spermatogenesis

DNA methylation in eukaryotes involves addition of a methyl group to the carbon 5 position of the cytosine ring. This reaction is catalysed by DNA methyltransferase in the context of the sequence 5′-CG-3′, which is also referred to as a CpG dinucleotide (review, Singal and Ginder, 1999). The methylation of DNA has been indicated to be involved in many cell biological processes including the mechanisms of gene regulation, differentiation, genomic imprinting, X chromosome inactivation, genome defence and carcinogenesis. This methylation helps to defend

against proliferation of parasitic sequences such as retroposons and stabilises chromatin in an inactive configuration and inhibits gene transcription (Yoder *et al.*, 1997; review, Bestor, 1998). In the testis, differential gene expression during development of germ cells is regulated at the transcriptional level by several mechanisms such as specific protein-DNA interactions and changes in chromatin structure. One of the changes in chromatin structure that often proceeds activation of tissue specific gene expression is demethylation (review, Cedar, 1988; Singal and Ginder 1999). Many studies in the mouse have shown that DNA methylation undergoes a dynamic series of changes in the germ line and early embryo (Kafri *et al.*, 1992). Some of these changes during spermatogenesis are occasioned by the general maintenance of genomic material of the germ cell and others include gene specific changes. It is speculated that specific gene expression is controlled by and associated with methylation of special sites in the promoter region of mammalian genes. The DNA methylation status of only a few testis-specific genes has been reported. For example, methylation of the lactate dehydrogenase (LDH-C) gene was found to be low in testis compared to somatic tissues (Alcivar *et al.*, 1991). Similarly, the testis-specific histone H2B gene was found to be hypomethylated in testis but not in somatic tissues (Choi and Chae, 1991). Several male germ line-specific genes (MAGE-type genes) belong to a unique subset of germ line-specific genes that use DNA methylation as a primary silencing mechanism. Most MAGE-type genes have promoters that are highly methylated in somatic tissues, but they are largely unmethylated in male germ cells where they are transcribed (De Smet *et al.*, 1996 and 1999). Results from another study suggest that cytosine methylation may contribute to the transcriptional silencing of the testis-specific histone (H1t) gene in nonexpressing tissues such as liver (Singal *et al.*, 2000). However, the opposite of this up-regulation has been reported in other genes (Choi *et al.*, 1997). Some genes, *TP1*, *PK2*, apolipoprotein A1 and *Oct-3/4* are expressed during spermatogenesis and yet become methylated at some sites prior to expression in the testis (Trasler *et al.*, 1990; Ariel *et al*, 1994).

DNA methylation patterns during gametogenesis show some differences in the genomes of male and female. For example, a study by Rubin *et al.* (1994)

shows that Alu elements (CpG island, satellite, and L1 sequences) are methylated in oocytes and unmethylated in sperm DNA from rhesus monkey (review, Trasler 1998).

The active mammalian DNA methyltransferase, encoded by the *Dnmt1* gene, is able to methylate both unmethylated (de novo methylation) and hemimethylated DNA (maintenance methylation) (Bestor, 1992). This gene is expressed at higher levels in the testis and ovary than all other tissues tested in adult mice (Trasler *et al.*, 1992). The DNA methyltransferase gene is expressed in mitotic (spermatogonia), meiotic and post-meiotic germ cells, with the exception of pachytene spermatocytes, at the time of genetic recombination in male germ cells, where the enzyme has not been detected by Westerns and immunocytochemical techniques (Jue *et al.*, 1995). The homozygous gene knockout mouse dies at mid-gestation prior to germ cell development (Li *et al.*, 1992).

It is predicted that most (but not all) gametogenic methylation differences between alleles will be removed during the genome-wide demethylation that occurs during preimplantation development. However, imprinted loci retain features of methylation differences throughout development (Tremblay *et al.*, 1997). It has also been reported that DNA methylation of some testis-specific genes increases again as the sperm pass through the epididymis (Ariel *et al*, 1994). As shown in figure I.10, the level of methylation during germ cell development is altered. It is important to determine when the imprinting process is completed as this will be relevant in the treatment of infertility using early germ cells, including spermatids, for ICSI. Failure to consolidate a genomic imprint prior to fertilization could at least in theory cause problems in offspring such as tumour susceptibility and other imprinting disorders (section 2.3) (Tycko *et al.*, 1997; review, Trasler, 1998).

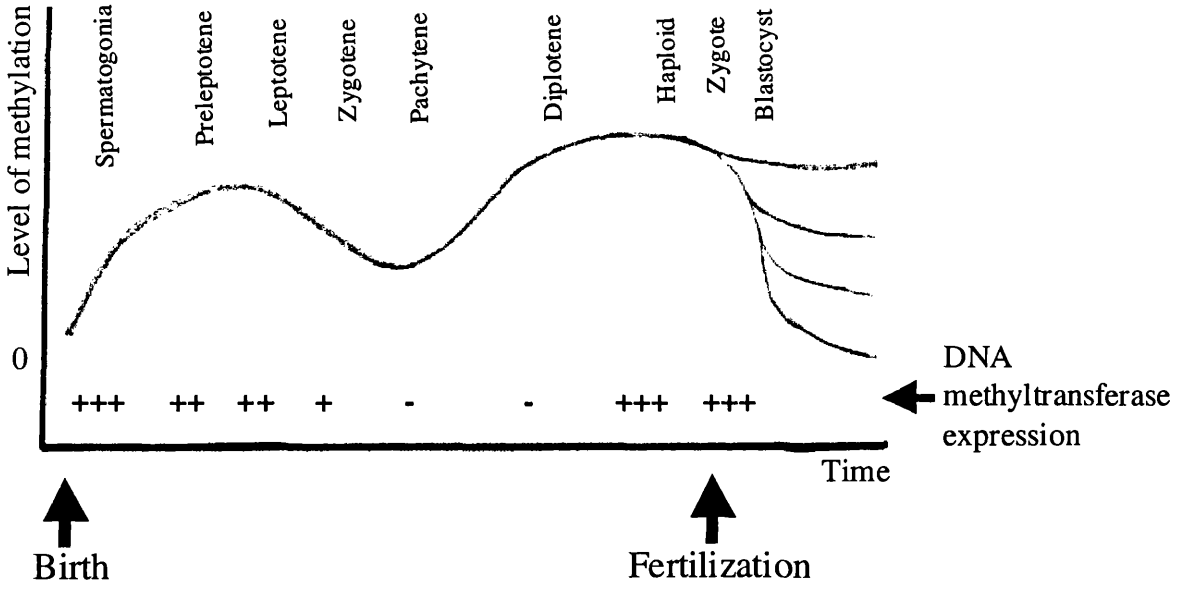


Figure I.10: DNA methylation and DNA methyltransferase expression during spermatogenesis. Spermatogonia and preleptotene, leptotene and zygotene spermatocytes all express varying levels of nuclear DNA methyltransferase protein, while the enzyme is undetectable in pachytene spermatocytes. The enzyme is re-expressed in haploid round spermatids. Relative levels of DNA methyltransferase expression (+ to +++) or absence (-) are depicted. The line shows the level of methylation during germ cell development. During preimplantation development the genomic material in general is demethylated but some genes retain significant levels of methylation, for example, the mouse H19 gene is hypermethylated on the paternal allele in blastocysts (Tremblay et al., 1997).

Part four

4.1 Tools for gene discovery and differential expression analysis

The study of genes and their proteins provides information about cellular growth, communication, and organisation and aids the understanding of the complex biological signals and pathways within each cell. In the past decade, the genomes of species including bacteria, yeast, *Drosophila, Arabidopsis thaliana* and the nematode *C. elegans* have been sequenced. Many more complete genome sequences are soon to be available and the goal of the Human Genome Project is to sequence the entire human genome. The exponential accumulation of human genome raw sequences has emphasised the need to be able to identify expressed genes within repetitive and non-coding sequences especially those genes related to clinical disorders.

Recently, several protocols have been developed to access genes with particular expression profiles thus identifying differentially expressed transcripts (DETs). In this section the principles, advantages and disadvantages of some of the commonly used techniques for identification of genes and DETs are reviewed.

4.1.1 RNA fingerprinting methods for DETs

The reasons for identifying DETs may be to compare transcripts expressed in different tissues and find those unique to one tissue, to compare transcripts expressed in normal and disease tissue, or to find sequences in common between specific tissues such as tumour tissues which are not in normal tissues. One way to do this is to generate a "fingerprint" of the transcripts by amplifying them by PCR and then separating the fragments by electrophoresis and comparing the band patterns that are generated.

There are a number of protocols which employ variations on this principle. Some use randomly generated (arbitrary) PCR primers whereas others first cleave

cDNA fragments with restriction enzymes and then anneal linkers to the overhangs (Vos *et al*, 1995; Ivanova *et al.*, 1995; Kato, 1996).

For example, a technique for DNA and cDNA fingerprinting named amplified fragment length polymorphisms (AFLP), which is used to detect polymorphism, should be mentioned for this application (Vos *et al.*, 1995; Money *et al.*, 1996). In this method, the same principle of fingerprint production is used: a pool of restriction fragments ligated to adapters is amplified with adapter-specific primers extended by several bases. This may reveal polymorphism when applied to genomic DNA. AFLP produces very good and reproducible fingerprints, but cannot be used for efficient display of cDNA fragments because it lacks the "one messenger-one displayable fragment" feature. In fact, in AFLP there is a mixture of cDNA fragments several times more complex than the original mRNA pool because each messenger is represented by several independent restriction fragments. In contrast, as explained later in other techniques a cDNA is digested but only the 3′end fragment is displayed.

Two other methods which rely on amplification with an adapter-oligo(dT) primer and adapter-specific primer after processes of digestion and ligation have been introduced. These methods are for comparison of gene expression profiles and schematics illustrating the processes of ordered differential display (ODD) (Matz *et al.*, 1997; Matz and Lukyanov 1998) and another similar technique suggested by Kato (1996) are both shown in figure I.11.

Two similar technique are Gene Expression Fingerprinting (GEF) and Restriction Landmark cDNA Scanning (RLCS)(Ivanova and Belyavsky, 1995; Suzuki *et al.*, 1996). As shown in figure I.12 the techniques are rather complicated and additionally RLCS employs two-dimensional gel electrophoresis which requires a lot of skill to obtain reproducible patterns. These techniques will not be described in detail.

Figure I.11: Outline of two different techniques which are based on using restriction enzyme digestion and ligation of primer-adaprors for identifying DETs. Schematic representation of the ODD protocol (right) and another similar method which have been introduced by Matz et al.(1997) and Kato (1996), respectively.

A) In the ODD method, after cDNA synthesis, using an oligo(dT) primer which has a small adapter sequence at 5´ end, the cDNA samples are digested by Rsa1. They are taken for ligation with pseudo-double-stranded adapter, a long oligo and a short one, complementary to the 3´ part of the longer one. PCR amplification is carried out using adapter-specific primer and the oligo(dT) primer. During the first PCR, fragments flanked by inverted terminal repeats can not be amplified with the single primer corresponding to the outer part of the repeat, based on PCR suppression effect (Siebert et al., 1995; Lukyanov et al., 1996). The second PCR at the high annealing temperature (67°C) using both the oligo(dT) and a labelled adapter-specific primer extended by two arbitrary, but defined, bases at their 3`ends produces a subset of PCR products corresponding 1/192 part of the total pool (as there exist 16 possible variants of adapter-specific primer extension and 12 of the oligo(dT) primer as shown in the figure). In addition, an intentional mismatch (with respect to the sequence of adapter-restriction site junction) has been made in the sequence of the extended adapter-specific primer in the fourth position from its 3´end. It has been found that this enhances the distinction substantially, because in such a situation a perfect match to all three 3´ distal bases (including the two `selector' bases corresponding to extension) is essential for efficient priming. As a result, the population of 3´fragments is subdivided into 192 subsets, which are displayed on an ordinary sequencing gel.
B) In the first step of Kato's protocol, cDNA is digested by EcoRI and ligated with an EcoRI adaptor and in the second stage cDNA is digested by a class IIS restriction enzyme (such as FokI) which produces four base overhangs of unknown sequence. Then, the cDNA is ligated separately to 64 adapters, each designed for one of the possible overhangs. The adapters have a 5´four nucleotids overhang of which the outmost is N (a mixture of A, C, G and T) and the other three are selective defined bases. The difference occurs at the ligation step, as the ligase attaches a specific adaptor almost exclusively to the complementary overhangs. The ligation products are recovered on streptavidin dynabeads using biotin attached to the adapter, and the resulting subset is amplified from the
bound sample with the adaptor-specific primer and one of the three 1 base anchored T primers. By repeating the experiment with three different class IIS restriction enzymes, the cDNA can be subdivided into 576 populations.

In the ODD method, in obtained samples each transcript is represented by a single fragment of characteristic length, between oligo(dT) primer and adapter sequence but not in the original Kato's protocol. In the Kato's protocol by using oligo(dT) primer with an adapter during cDNA synthesis and deletion of the first step digestion it can be achieved the "one messenger-one displayable fragment" feature.

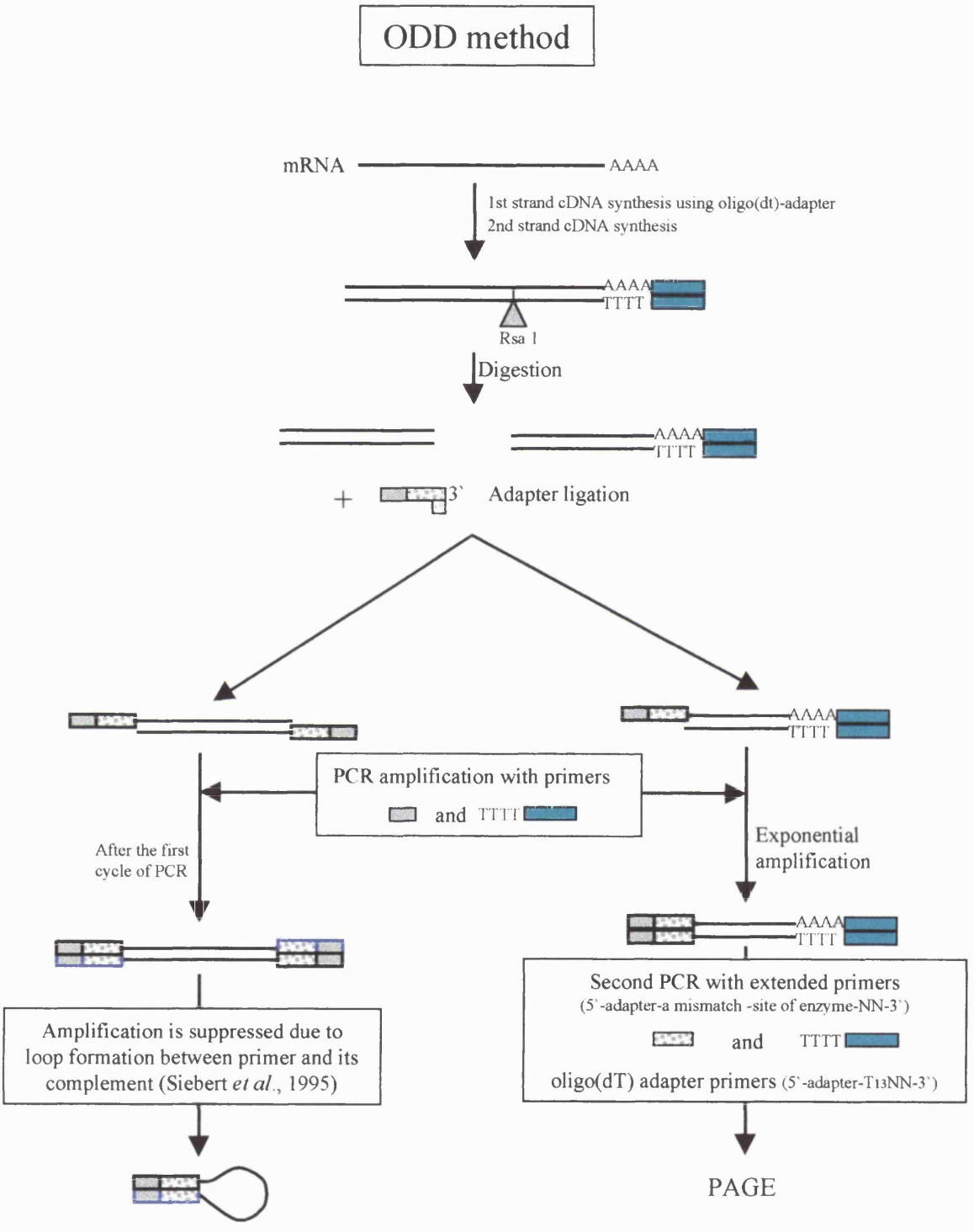

ODD method
mRNA
AAAA
1st strand cDNA synthesis using oligo(dt)-adapter
2nd strand cDNA synthesis
AAAA
TTTT
Rsa I
Digestion
AAAA
TTTT
3`
Adapter ligation
AAAA
TTTT
PCR amplification with primers
and TTTT
After the first cycle of PCR
Exponential amplification
AAAA
TTTT
Amplification is suppressed due to loop formation between primer and its complement (Siebert et al., 1995)
Second PCR with extended primers
(5`-adapter-a mismatch -site of enzyme-NN-3`)
and TTTT
oligo(dT) adapter primers (5`-adapter-T13NN-3`)
PAGE

Figure I.11-A

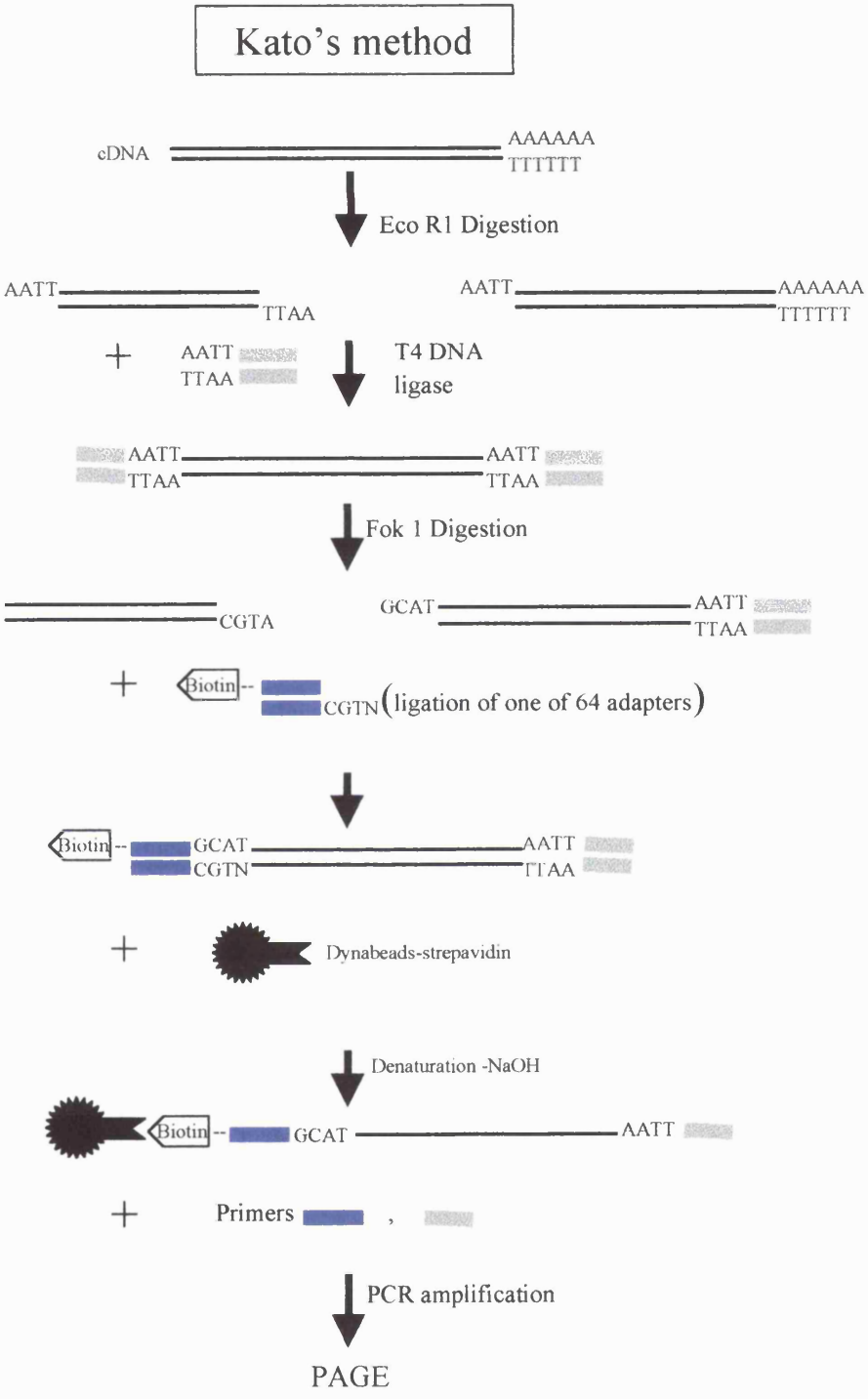

Kato's method
cDNA
AAAAAA
TTTTTT
Eco R1 Digestion
AATT
TTAA
AATT
AAAAAA
TTTTTT
AATT
TTAA
T4 DNA
ligase
AATT
TTAA
AATT
TTAA
Fok 1 Digestion
CGTA
GCAT
AATT
TTAA
Biotin
CGTN (ligation of one of 64 adapters)
Biotin
GCAT
AATT
CGTN
TTAA
Dynabeads-strepavidin
Denaturation -NaOH
Biotin
GCAT
AATT
Primers
PCR amplification
PAGE

Figure I.11-B

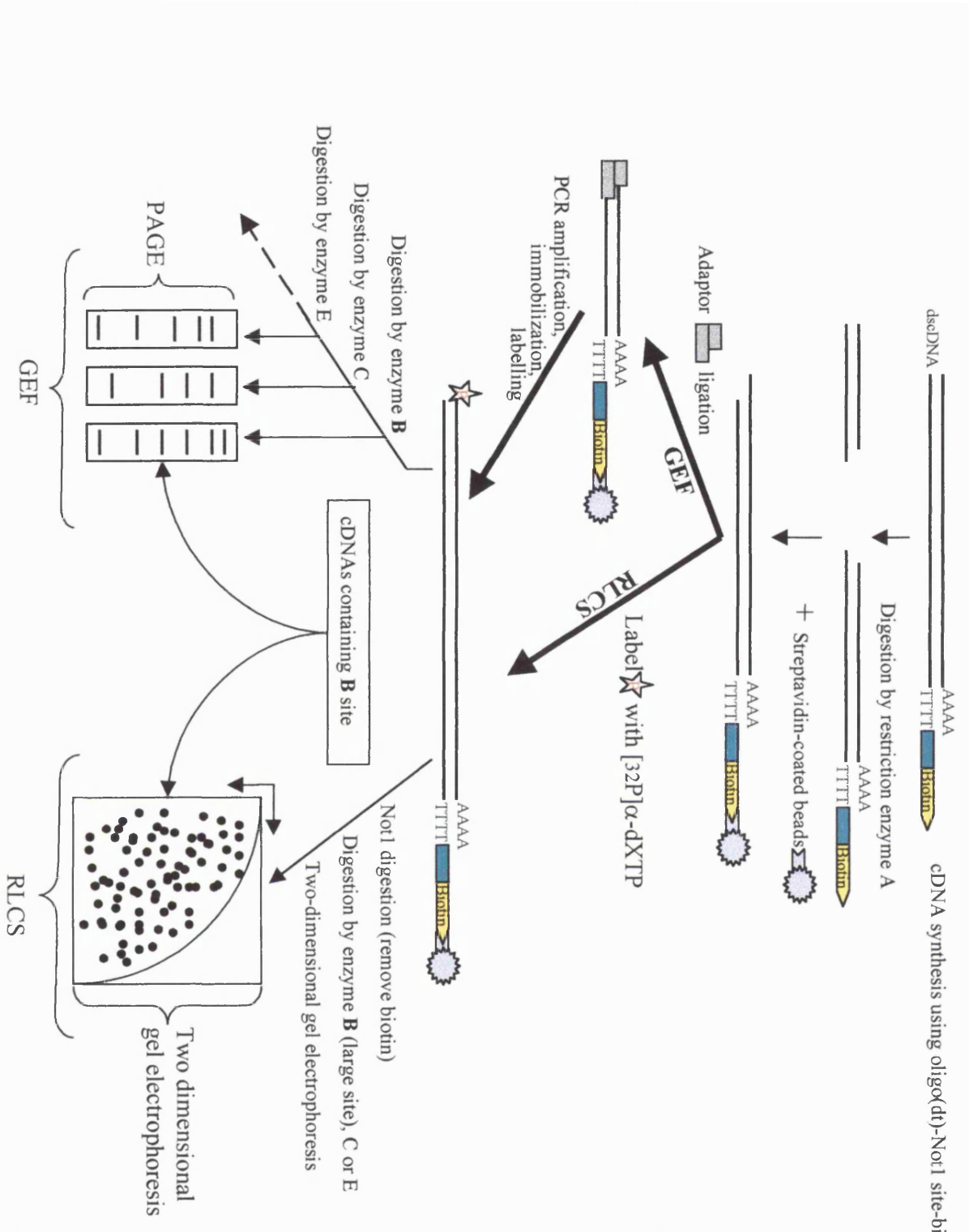


Figure I.12: Schematic representation of the principle of RLCS and GEF methods. In brief,the first stages, cDNA synthesis with a biotin-oligo(dT) primer and first digestion are the same in both (GEF and RLCS). In GEF method, cDNAs are ligated with an adapter, amplified and then labelled. cDNAs are immobilised using Avidin-Biotin reaction in both techniques. Different digestion stages for samples are carried out . Differences between samples can be displayed by running on a normal polyacrylamide gel (in GEF) or two dimensional gel electrophoresis (in RLCS). Both techniques represent one message-one displayble fragment feature because only the 3′end of each cDNA (mRNA) is immobilised and other parts of cDNAs resulted from digestion are washed away.

4.1.2 Subtractive hybridisation and suppression subtractive hybridisation (SSH)

An approach to the molecular analysis of differential gene expression is to construct cDNA libraries corresponding to different tissues or developmental stages, and then to enrich for genes expressed in a particular tissue or at a particular time by subtractive hybridisation. Screening involves checking sequences (clones), picked at random from two libraries, for presence in one of the original RNA samples in higher concentration than in another. Subtractive hybridisation is based on a procedure that increases the effective concentration of induced sequences expressed in an experimental RNA population (target) but not in a control RNA population (driver) (Sargent, 1987). This technique enriches samples for DETs prior to differential screening. Unhybridised cDNAs, which represent differentially expressed, tissue specific sequences can thus be isolated (Hara *et al.*, 1991). However, during subtractive hybridisation and cloning protocols for the isolation of activity-dependent genes, two main points have to be considered which are of general concern for a variety of other applications. This method requires a large amount of mRNA and is not particularly effective in identifying low abundance transcripts, and also following subtractive hybridisation an enrichment of induced sequences is frequently accompanied by an accumulation of sequences unable to form heteroduplexes for purely technical reasons (Li, *et al.*,1994; Konietzko and Kuhl, 1998). However, some contemporary techniques of subtractive hybridisation largely overcome the problem of bias toward abundant mRNAs (Gurskaya *et al.*, 1996; von Stein *et al.*, 1997). By using a combination of restriction enzyme digestion, hybridisation and PCR amplification a version of the subtractive hybridisation technique has been developed and is termed suppression subtractive hybridisation (SSH). This PCR-based cDNA subtraction method is used to amplify target cDNA fragments (differentially expressed) but simultaneously suppresses nontarget DNA amplification. The procedure, which has been explained by Diatchenko *et al.*, (1996), is shown in figure I.13. Here two cDNA samples are shown; it is possible

1- cDNA synthesis from two mRNA samples under comparison (tester and driver)

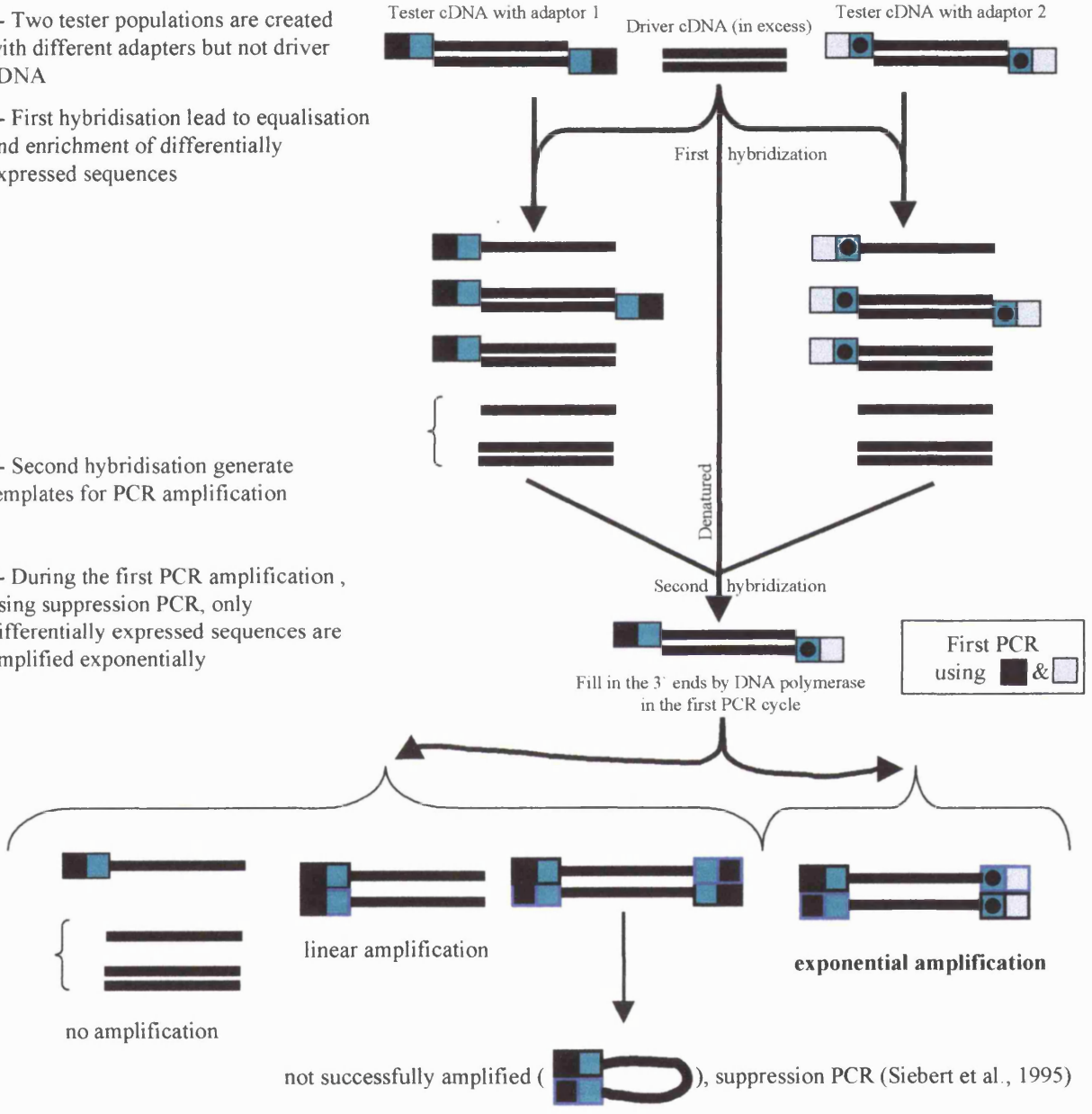
2- Both cDNAs are separately digested to obtain shorter, blunt-ended DNA

3- Two tester populations are created with different adapters but not driver cDNA

4- First hybridisation lead to equalisation and enrichment of differentially expressed sequences

5- Second hybridisation generate templates for PCR amplification

6- During the first PCR amplification , using suppression PCR, only differentially expressed sequences are amplified exponentially



7- Second PCR amplification using & , to enrich differentially expressed sequences

Figure I.13: Scheme of the SSH technique. Adapters are ligated to digested ds cDNAs made from tester and driver mRNAs. Green (with and without a black dot), black and blue boxes are adapters (which can be used as a site for nested primers) and two different primers. PCR products would be displayed on a gel and cloned in a set of vectors for further analysis.

to make one of these a mixture of RNA from several tissues, thus comparing more than one tissue at a time which can be an a significant advantage of this technique.

4.1.3 Representational difference analysis (RDA)

Representational difference analysis (RDA) is another approach to the identification of the differences between complex genomes and could be valuable as the first step for the positional cloning of genes of interest (Lisitsyn *et al.*, 1993; Hubank and Schatz 1994). RDA is basically a method of DNA subtraction, to investigate differences at the DNA level rather than RNA, similar to other methods such as chromosome painting, comparative genomic hybridisation (CGH), restriction landmark genome scanning (RLGS) and high-throughput analysis of LOH (Loss of heterozygosity). It can be used in situations such as the comparison of normal and tumour tissue (review, Gray and Collins 2000). Numerous studies have been done in cancers using RDA for the detection of genetic lesions, such as deletion (reviews, Baldocchi and Flaherty 1997; Wallrapp *et al.*, 1999; Frohme *et al.*, 2000). A possible application of RDA might be to display differences between normal men and patients with infertility caused by deletion with comparison of genomic DNA of fertile and infertile males. RDA has been successfully adapted to identify genes that are differentially expressed between two populations of cells. Figure I.14 shows the steps involved in the RDA protocol. Representative cDNA fragments from the two population to be compared are first generated by restriction endonuclease digestion of cDNAs followed by PCR amplification. The resulting mixtures, termed "amplicons", are then subject to successive rounds of subtractive cross-hybridisation followed by differential PCR amplification. This leads to progressive enrichment of cDNA fragments that are more abundant in one population than the other. However, absolute expression level has not been a crucial determinant for identifying genes, because fragments from both abundant and rare transcripts have been isolated (Braun *et al.*, 1995). As a disadvantage, this method does not resolve the problem of the wide differences in abundance of individual mRNA species because each RDA recovers typically only 6-12 differentially expressed cDNA fragments (Hubank and Schatz 1994; Braun *et al.*, 1995; Chu and

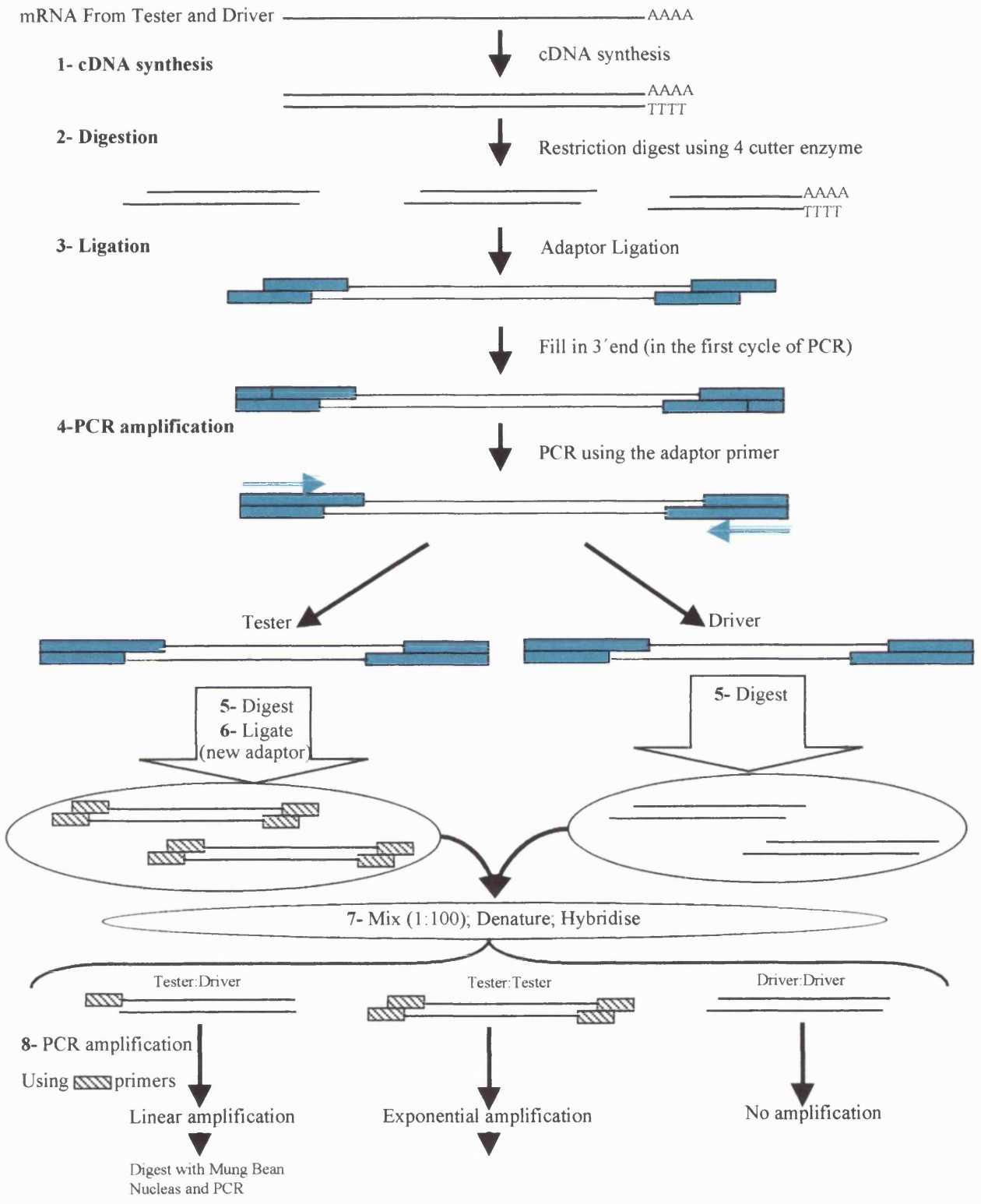


Figure I.14: Schematic diagram of cDNA RDA procedure (Lisistsyn et al 1993). Step 1 to 5 are the same in both tester and driver samples but tester cDNA (partial digests) are ligated with a new adaptor to make a new end for one of the samples which is used during the second PCR amplification after hybridisation stage.

In this method, one mRNA can be represented by many fragments.

Paul, 1997). Consequently, multiple rounds of subtraction are still required. In addition, like all subtractive methodologies, RDA can only compare two populations at a time and has generally requires large (~100 μg) quantities of starting mRNA.

4.1.4 DNA microarrays

DNA microarray technology is one of the most important recent breakthroughs in experimental molecular biology. It takes advantage of the preferential binding of complementary, single-stranded nucleic acid sequences (Schena *et al.*, 1995; DeRisi *et al.*, 1996&1997; Wodicka *et al.*, 1997). A microarray is usually a glass slide, to which single-stranded DNA molecules are attached at fixed locations (spots). There may be hundreds or even thousands of spots on an array, each representing a single gene. For display of DETs, RNA is extracted from the sample, which could be for instance a tissue or a cell line treated with a pharmaceutical product, and from control cells. The RNA is then labelled with two fluorescent labels: for instance, a red dye for RNA from the sample population and a green dye for that from the control population. Both labeled RNA extracts are washed over the microarray. Gene sequences from the samples and the control hybridise to their complementary sequences in the spots. The dyes make possible the amount of sample bound to a spot to be measured from the level of fluorescence emitted when it is excited by a light wave. If the RNA from the sample population is predominant, the spot will be red; if the RNA from the control population is more plentiful, it will be green; if the both RNAs bind equally, the spot will be yellow and when neither binds it will appear black. Thus, the relative expression levels of the genes in the sample and in control populations can be estimated.

This method provides a snapshot of gene activity for thousands of known genes and creates a considerable amount of valuable data. For example, it has been used in diagnosis of cancer by the pattern of gene expression (Alizadeh *et al.*, 2000; Berns, 2000) and for expression profiling in mammal development (Grimmond *et al.*, 2000). But it is still relatively expensive, needs specific equipment and is not

useful for the purposes of novel gene discovery. There is a rapid improvement in this technique during the last three years (review, Greenfield, 2000). For example, for a higher resolution view of gene structure, Shoemaker *et al.* (2001) have applied an 'exon array' approach to a detailed analysis of human chromosome 22. They have also used 'tiling arrays' to refine the structure of new genes discovered by exon analysis (these techniques will not be described). Recently, accumulation of results using this method has made bioinformatic scientists face the necessity of management and analysis of these huge and increasing volumes of data (Brazma *et al*, 2000; Lemkin *et al.*, 2000; see site: http://www.lecb.ncifcrf.gov/MAExplorer/hmaeHelp.html).

4.1.5 Library Screening

In addition to these techniques concerned with comparing patterns of expression, a number of screening strategies have been developed in order to identify potential genes. Some of these are explained below.

4.1.5.1 Random partial sequencing of clones from tissue specific cDNA libraries using vector primers can generate a very large number of expressed sequence tags (ESTs)(Okubo *et al.*, 1991; Boguski *et al.*, 1993). dbEST is a division of GenBank and isolated ESTs reported from all over the word as part of the Human Genome Project are available at http://www.ncbi.nlm.nih.gov/dbEST/dbEST_access.html (So far, more than 6.4 million ESTs from different cDNA libraries and organisms). Each EST has a datasheet which contains sequence data and other information such as tissue source, the name of organism, accession no. and clone name. Non-human ESTs are very helpful to compare with human homologues. ESTs can then be mapped to a chromosome by PCR amplification of somatic cell hybrids (see chapter III, table III.4)(Jones *et al* 1997) and can be used to retrieve entire genes by using either computational or experimental methods. Gene structure may be revealed from the genomic sequence data using computational programs like GRAIL (http://www.hgmp.mrc.ac.uk/Registered/Option/xgrail.html) which is used for gene prediction and can quickly identify potential exon/intron boundaries. NIX

(http://www.hgmp.mrc.ac.uk/NIX/index.html) is another computational program that can be used to analyse the obtained sequences by comparing results of many available programs at the same times. Large-scale cDNA sequencing projects have been applied to testicular cDNA libraries (Hoog, 1991; Starborg *et al.*, 1992; Yuan *et al.*, 1995). Following that screening Hoog *et al.*, (1995) have demonstrated some differences and similarities in expression patterns between brain and testis by comparing the functional categories of ESTs obtained from the two tissues.

Libraries from different mouse germ cell lines such as spermatocytes have been constructed to search for stage specific gene expression patterns of spermatogenesis (personal communication, Burgoyne, 1999 and see chapter III, section III.9).

4.1.5.2 Another screening technique is to screen human cDNA libraries at low stringency with non-human cDNA probes that are known (from databases) to be relatively conserved between species. For example, a yeast meiotic gene has been used (serine threonine kinase) as a probe to screen a mouse testis cDNA library (Shalom and Don 1999). In addition, this method could be very useful for finding other members of a gene family. The *DAZLA* gene was identified by screening a human testis cDNA library with a *DAZ* probe (Yen *et al.*, 1996). Screening with genomic clones is technically demanding and variable as some genomic clones, particularly YACs, give a very high background due to their complexity and repetitive sequence content. A converse 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).

4.1.5.3 Direct selection of cDNAs from human genomic DNA, is a technique which enables the identification of genes which are expressed in a particular tissue or cell type and which map to a specific chromosome(s) such as Y chromosome (or even to smaller regions of chromosomes) (Lovett *et al.*, 1991; Parimoo *et al.*, 1991, personal communication, Makrinou, 2000). The basis of this technique is the hybridisation of a pool of cDNAs to immobilised genomic clones. Non-hybridized sequences are eliminated and selected cDNAs ,i.e., those that

hybridise to ranges of the genomic clones, are recovered by elution and PCR amplification. The PCR products are either cloned or subjected to further selection/amplification cycles. This technique has also been combined with PCR amplification to facilitate the use of a small amount of mRNA. This method 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. Limitations of this technique may be contamination with mitochondrial cDNAs and repetitive DNA, the long procedure of creating the library and relatively time consuming sequencing of the selected PCR products (personal communication, Makrinou, 2000).

4.1.5.4 Differential screening of a subtracted cDNA library is another method which has been introduced using a combination of this and the array technology, to search for genes preferentially expressed in multiple tissues. For instance, Jin *et al.* (1997) followed this strategy to identify genes expressed in both testis and ovary. First they generated testis-specific cDNAs using the suppression subtractive hybridisation technique, which were then cloned and arrayed in microplates. The inserts from the subtracted testis-specific library were amplified by PCR and spotted onto filters. The dot blots from 2000 putative clones were screened by hybridisation using either testis- or ovary-specific subtracted cDNA mixtures as probes. The results showed that only three clones were preferentially or exclusively expressed in both testis and ovary. This technique is not simple or without biases and involved a number of stages.

4.1.6 DDRT-PCR

Differential display RT-PCR (DDRT) is a method that has been developed for the identification of genes differentially expressed between different cell or tissue types by creation of an RNA fingerprint for each sample (Liang and Pardee, 1992). Arbitrarily primed PCR fingerprinting of RNA is another very similar method for fingerprinting of RNA populations using an arbitrarily selected primer at low

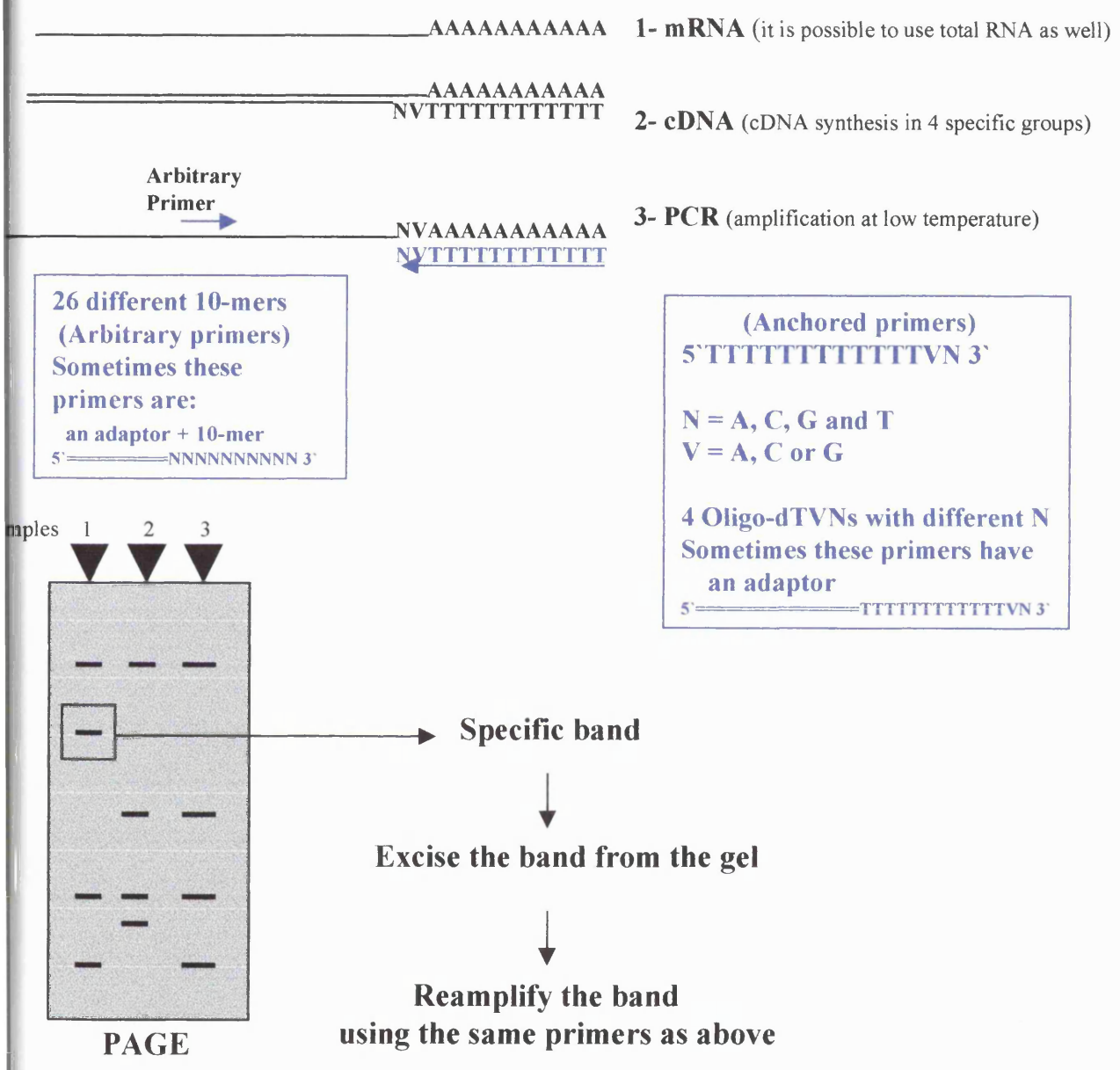


Figure I.15: The process of classic DDRT-PCR. In some of the improved techniques adapters at the 5′ end of the oligo(dT) and random primers result in more specific reamplification of the selected bands.

stringency for both the first and second strand cDNA synthesis. PCR is then used to amplify the products (Welsh *et al.*, 1992). DDRT-PCR is a technique that converts a subset of the mRNA present in the particular tissue or cell into short cDNA fragments using an anchored poly-(A) primer and an arbitrary 10-mer primer. A low annealing temperature (40°C) is required to achieve arbitrary priming. A simultaneous comparison of the cDNAs that are taken from various tissue samples can be made by running the samples side by side on a nondenaturing polyacrylamide gel, then individual cDNAs specific to a particular cell type or tissue can be isolated (figure I.14). The technique can be used to identify genes that are active at different stages of development or in different tissues. In addition, it may be useful to detect genes (activated) in response to pharmaceutical drugs or in abnormal tissues such as tumours in comparison to the normal tissue. DDRT-PCR has been cited extensively in the literature and has resulted in several hundred publications reporting its successful application. However, there are many more uncontrolled problems with DDRT-PCR technique than were apparent from the original publications (Debouck, 1995). For instance, false positives, the main disadvantage of DDRT-PCR, comprise a significant portion of isolated fragments. The proportion of false positives (not actually corresponding to DETs) often reaches 50% or more (Debouck,1995; Wan *et al.*, 1996). The most obvious cause of false positives is arbitrarily primed PCR with short (10-12 bases) oligos, in a low stringency PCR.

Therefore, a number of improvements were proposed to overcome its difficulties and fulfil the aims of the DDRT-PCR method and several promising techniques for displaying cDNA fragments were recently developed that differ from the original DDRT-PCR method (Matz and Lukyanov, 1998)(practical details of this method have been discussed in chapter 3). The original method, based on arbitrary primed PCR is usually termed classic DDRT-PCR and the techniques that involve DDRT-PCR combined with more complex methods, such as ODD, GEF and RLCS are named systematic differential display (DD) (Matz and Lukyanov, 1998). The scheme of systematic DD methods basically includes two stages: first, obtaining a sample of cDNA fragments in which each mRNA is represented by a

single fragment of characteristic length; and second, generating simplified non-overlapping subsets of these fragments consisting of ~70-100 members, short enough to be analyzed on a polyacrylamide gel. In this method, the same principle of fingerprint production as ODD, GEF, RLCS and Kato's method is used: a pool of restriction fragments ligated to adapters is amplified with adapter-specific primers extended by several bases.

4.1.7 Differential display computational tools

Recently specialised informatics tools have been provided to simplify the analysis of the huge accumulative complex data sets which is presented in databases. Some of them for gene expression analysis such as Digital Differential Display (DDD), cDNA Expression Profiler and Serial Analysis of Gene Expression Map (SAGE) are available at the site of NCBI (http://www.ncbi.nlm.nih.gov/CGAP/hTGI/). For instance, Digital Differential Display (DDD)(http://www.ncbi.nlm.nih.gov/CGAP/info/ddd.cgi) is a computer method for comparing many fingerprints using results of sequenced ESTs from specific cell, tissue or organ cDNA libraries. Using a statistical test, genes whose expression levels differ significantly from one tissue to the next can be identified and shown to the user.

4.2 Which method is better for displaying of DETs?

The advantage and disadvantage of each technique (the original technique not the improved methods) are shown in table I.5. Subsequently techniques have been improved to eliminate as many disadvantages as possible until the revised technique sometimes becomes completely different from the original method. The techniques used in a particular situation relate to their advantages and disadvantages. For example, those techniques that are based on PCR amplification are very sensitive and a small amount of starting mRNA can be used but false positive and non-specific results are seen. In contrast, those using a hybridisation strategy need more mRNA and are not sufficiently sensitive to detect low copy transcripts.

Microarray technique is a good method for DD but for genes which are already known, requires an expensive equipment and was very new at the beginning of this experiment.

At the time of starting this study, there was a lot of debate in the literature on the possible advantages and disadvantages of classical DDRT-PCR. It has also been indicated that classical DDRT-PCR reveals far less difference than is expected in some well-studied biological systems (Bertioli *et al.*, 1995; Graf *et al.*, 1997). This may be attributed to under-representation of the minor mRNA fraction to which most DETs belong.

The results in this thesis were produced using Differential Display RT-PCR. Despite all the problems mentioned, classical DDRT-PCR has a significant advantage over all other DET search methods in that it is the simplest technique for mRNA comparison and it does not require any special reagents, materials or instruments such as those needed for array technology. Modified DDRT-PCR techniques which in some publications are called systematic DDRT-PCRs are interesting and sometimes worth using. However, some of those protocols are too complicated to be used widely in the present form and require further development. Classical DDRT-PCR is a satisfactory method, at least in cases when a pronounced difference is expected and there is no need to detect all DETs, for example when searching for tissue markers in cancer tissues or, as is the case of this thesis, searching for novel transcripts unique to a particular cell type.

ADVANTAGE AND DISADVANTAGE (D)	DDRT-PCR	ARRAYS	RDA	SUBTRACTIVE HYBRIDISATION	LIBRARY SCREENING
Advantages					
Identification of new genes	*		*	*	*
Starting from total RNA	*	*			
To obtain full length cDNA and coding information		*			
Comparison of many different samples at a time	*				
To obtain 3` end information	*	*			
Disadvantages					
Starting from mRNA		*	*	*	*
Hybridisation technology necessary		*	*	*	*
Time consuming					*
Expensive equipment necessary		*			
False positive results	*			*	*

Table I.5: Advantage and disadvantage of the main strategies for the identification of DETs.

Chapter 2

Materials and Methods

II.1 Materials

II.1.1 Standard reagents

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

II.1.2 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 with the buffer IV (15mM $MgCl_2$, 1.25ml of 200mM $(NH_4)_2SO_4$, 750mM Tris-HCl, pH 8.8, 0.1% Tween). Advantage *Taq* polymerase was from Clontech. Bio-*Taq* polymerase was from Bioline. All other enzymes were from Boehringer Mannheim.

II.1.3 Primers

The majority of primers were supplied by Oswel and were already resuspended in 1ml of sterile water. Some of the primers were purchased lyophilised from Amersham and were resuspended in 1ml of 10mM Tris (PH 7.5). All primers were aliquoted and diluted to a final concentration of 50pmoles/100µl of PCR reaction.

II.1.4 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.

II.1.5 Kits and others

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. Amersham life science 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 PCR purification, mini plasmid purification and maxi plasmid/cosmid purification kits. mRNA QuickPrep Micro purification kit was from Pharmacia. PMOSBlue blunt ended cloning kit and TA PCR cloning kit were from Amersham and Invitrogen, respectively. Lymphoprep was from Nycomed, Oslo, Norway.

II.1.6 Standard and commonly used solutions and buffers

'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: 0.3% Diethyl pyrocarbonate added to distilled water, incubated for 2-16 hours 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 Denhardt's: 2% Ficoll, 2% polyvinylpyrolidone, 2% BSA (Bovine Serum Albumin)

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

20 x PBS: 23g Na_2HPO_4, 4g KCl, 160g NaCl, 4g KH_2PO_4, 11g Na Butyrate made up to 800mls with distilled water. Autoclave and add 200mls 0.5M EDTA.

Loading buffer: 30% glycerol in TE plus bromophenol blue

Alkaline lysis buffers:

Solutions for mini-preparation of DNA from plasmids:

Solution I: 50mM glucose; 25mM Tris-Cl (pH 8.0); 10mM EDTA (pH 8.0)

Autoclave at 10 lb/sq. in. for 15 minutes and store at 4°C.

Solution II: 0.2M NaOH, freshly diluted from 10M stock; 1% SDS

Solution III: 60ml 5M KAc; 11.5ml glacial acetic acid, made up to 100mls with dH_2O.

(3M with respect to potassium and 5M with respect to acetate)

Solutions for mini-preparation of DNA from PAC and BAC:

P1 (filter sterilized, 4°C): 15 mM Tris, 10 mM EDTA, 100 μg/ml RNAse

P2 (filter sterilized, room temp., freshly made): 2N NaOH and 1% SDS

P3 (autoclaved, 4°C): 3M KOAc, pH 5.5

II.1.7 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

NZY Broth (per litre): 2g of $MgSO_4.7H_20$ was added to L-Broth and for NZY agar, 20g of bacto-agar was included.

L top agarose: NZY broth with 0.7% agarose.

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

All growth mediums were autoclaved immediately when they were made up and antibiotics were added before usage.

Ampicillin: 100mg/ml in distilled water, 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 distilled water, filter sterilised through a 0.2μm filter. Stored at 4°C for a short term period used at a working concentration of 25μg/ml

Tetracycline: 5mg/ml in ethanol. Used at a working concentration of 5μg/ml. Stored at -20°C.

IPTG: Stock solution made up at 100mM, filter-sterilized through a 0.2μm filter. 4μl of the stock was spread on a 90mm agar plate (with the appropriate antibiotic) and left to absorb for 30 minutes. Stored at -20°C.

X-Gal: Stock made up to 20mg/ml in dimethylformamide. 40μl of the stock was spread on a 90mm agar plate (with the appropriate antibiotic) and left to absorb for 30 minutes. Stored at -20°C.

20% Maltose: 10g of maltose made up to 50mls with ddH_2O. Filter-sterilized through a 0.22μm filter. Stored at 4°C.

SM buffer: 5.8g NaCl, 2g $MgSO_4.7H_20$, 6.05g Tris base, 5mls 2% gelatin made up to 1l with dH_2O. Adjust to pH 7.5 with HCl. Autoclaved.

E.coli strain XL1-Blue from ZAP-cDNA Synthesis kit.

E.coli strain SOLR from ZAP-cDNA Synthesis kit.

II.1.8 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 Galton laboratory and Biology Dept., UCL.

II.1.9 Acknowledgements for materials, human and mouse samples

The human brain RNA was from Dr. A. Bennett, University of Nottingham. The human tissues samples were from Tehran University of Medical Sciences hospitals and RNAs were extracted in Biotechnology Dept. of Pasteur Institute of

Iran. Testis samples (as a biopsy) of infertile and Azoospermic patients were from Dr. Jabal-Amoli (Medical School of Tehran University of Medical sciences) and Mr. D. Ralph (University College London), respectively.

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 age was determined by either hand or foot measurements.

The somatic cell hybrids used were from the Galton Laboratory Somatic Cell Hybrid Panel, a gift from S. Jeremiah and Prof. S. Povey.

A human adult testis cDNA library has been made by Prof. Y.Edwards. A mouse testis cDNA library which has been made from spermatocyte were from P.S. Burgoyne, National Institute of Medical Research.

II.2 Methods

II.2.1 DNA related Methods

II.2.1.1 DNA extraction from tissues and blood for PCR

DNA was isolated from several individual (male and female) blood samples and a few human tissues using a modified protocol which is routinely used in Biotechnology Dept. of Pasteur Inst., Tehran. Approximately 1 to 2ml of blood or 100μg of tissue was placed in a 5ml plastic bijou with 1ml of lysis buffer, and homogenised using a mechanical homogeniser or by hand. The cells were centrifuged at about 5,000rpm for 5 minutes in a centrifuge. 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 times, or until the pellet became white (especially for blood to remove as much haemoglobin as possible). After the final wash, the pellet was dissolved in 100μl of 50mM NaOH and the suspension was boiled at 99°C for 20 minutes. 0.2ml 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, the pellet was discarded and supernatant was transferred to another 1.5ml Eppendorf tube, aliquoted and stored at -20°C. However, it is possible to keep this at 4°C until further use because there is no DNase present after boiling for 20mins.

II.2.1.2 DNA extraction from *E.coli*, mini-preparation of DNA from plasmids

Cloned cDNAs 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.

If the DNA extraction was intended solely for PCR, 0.5ml of overnight growth culture were centrifuged at about 5,000rpm for 5 minutes in a microcentrifuge. The supernatant was discarded and the pellet was resuspended in 0.1ml of solution I, mixed well and was heated at 95-99°C for 5 minutes in PCR

machine. The microtube was vortexed for 1-2 mins and centrifuged at 12000 rpm for 3-5mins. The supernatant was stored at -20°C and used for PCR.

II.2.1.3 DNA extraction from *E.coli*, mini-preparation of DNA from PAC or BAC

A single isolated bacterial colony was inoculated into 2ml TY (or LB) media supplemented with 25μg/ml kanamycin and grown overnight (up to 16h) shaking at 225-300 rpm at 37°C using a 12-15 ml snap-cap polypropylene tube. It was centrifuged (SM24 or similar rotor) at 3,000 rpm for 10 min. The temperature of the spin was not critical at that stage. Supernatant was discarded and the pellet was resuspended (vortex) in 0.3ml P1 solution. 0.3ml of P2 solution was added and the tube "gently" shaken to mix the contents and then left at room temperature for 5 min. The appearance of the suspension should have changed from very turbid to almost transparent. 0.3ml P3 solution was slowly added to the tube and gently shaken during addition. The tube was placed on ice for at least 5 min and then centrifuged at 10,000g for 10 min at room temperature. A thick white precipitate of protein and *E.coli* chromosomal DNA was formed. The supernatant was transfered to a 1.5ml Eppendorf tube and centrifuged again at 10,000g for 5 min and then transferred to another new 1.5ml Eppendorf tube containing 0.8ml ice-cold isopropanol. Tubes were mixed by gently inverting a few times, and were placed on ice for at least 5 min., the supernatants were removed and 0.5ml of 70% EtOH was added to each tube. The tube was again gently inverted several times and the DNA pellets were precipitated in a microfuge for 5 mins. This step was repeated one or two times and the pellet air dried at room temp. until the DNA pellet turned from white to translucent in appearance (when most of the ethanol has evaporated). The pellet was resuspended in 40 μl TE with a large pipette tip or, alternatively was allowed to redissolve by leaving the tubes to stand with occasional tapping of the bottom of the tube.

II.2.1.4 DNA extraction from *E.coli*, Maxi-preparation of DNA from PAC and BAC

500ml of 2x TY plus 25μg/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 clones.

II.2.1.5 DNA extraction from polyacrylamide gel

The bands were cut from the gel and dried on the Whatman paper using a scalpel, and the gel slice (still attached to the Whatman paper) was placed in an Eppendorf tube and eluted in 100μl of TE for 1 hour at room temperature. The sample was then heated at 95-99°C for 15 minutes and centrifuged briefly. Supernatant was used for PCR or cloning.

II.2.1.6 Agarose gel electrophoresis and DNA extraction from agarose gel

DNA fragments were size separated through 0.8-2.0% agarose gels (depending on the length of the samples being electrophoresed) which were 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 sizes were checked by comparison with 1Kb (Gibco) or 100bp (Gibco or Promega) ladder molecular weight markers.

The band was cut from the gel using a sterile scalpel in as small a piece as possible and was placed in the base of a drip column. The column was made using a 0.5ml tube with a tiny amount of glass wool which was siliconised and autoclaved according to Maniatis *et al.*, 1993. A small hole (made by an autoclaved tooth-pick or a sterile needle) had previously been made at the bottom. The drip column was

placed in another 1.5ml tube and was centrifuged at 12000g for 5mins. The inner tube was discarded and the DNA in TBE was used for PCR, cloning, sequencing or labelling. Concentration of the DNA was estimated by electrophoresis and comparison to molecular weight standards of known concentration or by spectrophotometry.

II.2.1.7 Purification of PCR product

For the purification of PCR products, to remove contaminants such as primer-dimers and leftover primers, the QIAquick PCR purification kit was used, according to the manufacturer's instructions (QIAGEN).

II.2.1.8 DNA precipitation and assay of concentration

DNA precipitation was routinely performed by the addition of 3M sodium acetate, pH5.5 to a final concentration of 0.3M and 2 volumes of absolute ethanol before chilling at -20°C overnight or -70°C for >1h. Precipitated DNA was recovered by centrifugation at 13000g and 4°C and washed in 70% (v/v) ethanol before air-drying and resuspending in double distilled water or TE. DNA was quantified by measurement of absorbence at 260nm; 1OD unit is 50μgml^{-1} (the ratio of OD_{260}/OD_{280} should be 1.8 for pure DNA).

II.2.1.9 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 incubator. The DNA fragments were then separated on a 0.8% agarose gel.

II.2.1.10 Cloning

PCR amplified DDRT-PCR products were cloned into either pMOSBlue blunt ended vector or pTAg vector following the kit protocols.

II.2.1.11 Preparation of glycerol stocks of plasmids

Glycerol stocks were prepared from all the selected cDNA plasmid clones by using 70% L broth containing the appropriate antibiotic, 15% glycerol and 15% of a single colony cultured medium.

II.2.1.12 Sequencing

Sequencing was carried out using the dideoxynucleotide chain termination method (Sanger et al., 1977). Radioactive sequencing was performed using the Thermosequenase radiolabelled terminator cycle sequencing kit (Amersham Pharmacia Biotech) with ^{33}P labelled Redivue terminators.

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_2O. 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.

II.2.1.13 Polyacrylamide gel electrophoresis

The sequencing apparatus used was 21 x 40cm in size (Biorad), and the gel thickness was 0.4mm. A 24-well sharkstooth comb was used to create the wells for loading.

Before assembly, both the front and back plates were cleaned with detergent and then 70% ethanol. A minimal amount of bonding solution was applied to the

front plate, and then rubbed with 70% ethanol, and a small amount of repelcote was applied to the back plate and then washed with water.

For a 6% denaturing polyacrylamide gel: 29g of urea, 23mls H_20, 6mls 10 x TBE and 9mls 40% (19:1) acrylamide:bisacrylamide solution were combined to a total volume of 60mls. 80µl of freshly prepared 25% ammonium persulphate (made up in ddH_20) and 60µl of TEMED was added just before pouring the gel.

After polymerization, the gel was pre-run at 50 watts for 20-30 minutes in 1 x TBE till it reached a temperature of 45-50°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 35W (45-50°C) for about 90 mins (for short run) to 5 hours (for long run) depending on the length of the samples being sequenced 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 minutes and then exposed to Kodak biomax MR 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.

II.2.1.14 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 plugged 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.

II.2.1.15 Radiolabelling DNA probes by random priming

DNA to be used as probes was labelled by random priming (Feinberg and Vogelstein; 1984) 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 (3000 Ci.mmol^{-1}). After incubation at 37°C for 20-60 minutes, unincorporated [α-^{32}P]dCTP and primers were removed by passing the labelling reaction either through a Sephadex G-50 column made as explained above or through a commercial Nick column. The labelling reaction volume (48µl) was applied onto the column and eluted with 400µl 1X TE. Percentage incorporation was estimated by comparing the cpm retained on the G-50 Sephadex column, which represents unincorporated [α-^{32}P]dCTP, with the cpm in the eluate. Reactions which failed to achieve more than 60% incorporation were discarded.

II.2.1.16 Radiolabelling DNA probes with T4 polynucleotide kinase

25-100ng of double-stranded DNA PCR product which had been extracted from an agarose gel, cleaned using QIAquick PCR purification kit and redissolved in ddH2O was 5`-end labelled using T4 polynucleotide kinase (Gibco) and [γ-^{32}P] ATP. All reagents were kept on ice. Labelling was carried out in a 25µl reaction volume containing 5µl of 5x forward reaction buffer (Gibco) with 2µl (20 units) T4 DNA kinase enzyme and about 10µ Ci of [γ- ^{32}P] ATP (3-4µl at>3000 Ci.mmol^{-1}). This mix was incubated at 37°C for 15 min followed by 10 min at 65°C for inactivation of the enzyme. The unincorporated [γ-^{32}P] ATP was removed by passing the labelling reaction through a commercial Nick column. The level of labelling was usually more than 80% incorporation.

II.2.1.17 Radiolabelling DNA probes by PCR

Polymerase chain reaction (PCR) products were labelled by incorporating [α-^{32}P]dCTP into the product during the amplification reaction. Reactions utilised Advanced Biotechnologies *Taq* enzyme with 10x buffer IV. Amplifications of 2-5ng of template DNA were carried out in 50µl reaction volumes with 1x buffer IV, 3µl each of dATP,dGTP and dTTP (5mM of each), 1.5µl of dCTP (0.5mM), 50pmoles of both forward and reverse primers, 1µl [α-^{32}P]dCTP (3000 Ci.mmol^{-1}) and 0.25µl *Taq* (1.25 units). Each reaction was covered with 0.1ml of mineral oil to prevent evaporation during thermal cycling. The reaction was heated at 94°C for 5 minutes and then cycled for 30 cycles, using conditions established for each primer pair, on a Hybaid thermocycler machine as follows:

Denaturation step	94°C	20 seconds
Annealing step	T_m-5°C	30 seconds
Extension step	72°C	30 seconds
With a final extension step	68°C	7 minutes

II.2.2 RNA related methods

II.2.2.1 Equipment preparation

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_2O before use. Tips were autoclaved two times and a special set of clean micropipettes was kept solely for RNA work. Liquid nitrogen was available in a small flask during experiments with RNA to keep tissues and extracted RNA in the liquid N_2 and when experiments were over all RNA and tissue samples were transferred and stored at -70°C.

II.2.2.2 Isolation of total 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, following the manufacturer instruction, frozen tissue was homogenised in RNAzol B in a mechanical homogeniser (2ml per 100mg of tissue) and 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. After centrifugation at 12000 rpm for 10-15 minutes the pellets were 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/μl. Human RNAs (except samples from liver, testis, a testis biopsy from an azoospermic patient and brain) were extracted by me from tissues at the Pasteur Institute in Tehran. After checking their concentrations, 2.5 volume 100% ethanol was added to each RNA sample and carried in liquid N_2 to London and then RNAs were precipitated at 13000rpm (4°C) for 15min. The RNA pellet was drained well and resuspended usually in 50μl DEPC-treated water.

For RNA extraction from sperm, the first quality of semen was checked according to WHO protocol (chapter I, section 2.1) and then after liqinification 2ml of the semen was washed three times with 10ml 1x PBS and centrifuged at 800g. After the final centrifugation step 2ml RNAzol B was added and the solution was passed through a needle (gauge no.18) using a sterile syringe and then the above procedure was carried out.

II.2.2.3 Total RNA preparation from white blood cells

White blood cells were prepared using Lymphoprep (Nycomed): 10mls of human male blood was mixed with 10mls of 1x PBS, and the mixture was then layered on top of 0.5 volumes of lymphoprep (10mls in 2 sterilin tubes). These were

then centrifuged in a Mistral 2000 bench centrifuge at 1500rpm for 15 minutes. The top layer was removed and discarded and the second, white blood cell layer, was removed into a new sterilin. The sterilin was then filled with 1x PBS to wash the cells and centrifuged at 2000rpm for 15 minutes. The PBS was then tipped off and discarded.

An alternative method for separation of WBCs from RBCs based on soft lysis of RBCs was also performed as described (Michel *et al.*, 1997).

200µl of RNAzol B is usually added for every 10^6 cells and 10mls of blood contains about $2x10^7$ cells; so 4mls of RNAzol B was added to the white blood cells in two centrifuge tubes. After extraction with 200µl chloroform, the aqueous phase was divided again into two eppendorfs and the RNA precipitated with an equal volume of isopropanol. The RNA was washed with 100µl 75% ethanol and centrifuged at 5600g in a microfuge for 2 minutes. At this stage the RNA pellet was usually resuspended in 30µl DEPC-treated water.

A second ethanol precipitation step was then carried out with 0.2M NaCl and 1 volume of isopropanol. The RNA was precipitated at -20°C for 1 hour and then centrifuged for 10 minutes. The RNA was washed with 50µl 70% ethanol and centrifuged for a further 2 minutes. The RNA was incubated at 65°C for 5 minutes in 20µl DEPC-treated water to dissolve it.

II.2.2.4 Poly $(A)^+$ RNA preparation from tissue

The Microprep mRNA kit was used to prepare poly $(A)^+$ RNA, following the kit protocol.

II.2.2.5 Control of the RNA samples using RNA gel agarose electrophoresis and spectrophotometer

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 to

the samples and UV transillumination. 0.24-9.5kb RNA ladder was electrophoresed alongside the samples for sizing.

All RNA samples were checked using the spectrophotometer. RNA was quantified by measurement of absorbence at 260nm; 1OD unit is 40μgm/l (the ratio of OD_{260}/OD_{280} should be 2.0 for pure RNA).

II.2.2.6 DNAse treatment of RNA

1μg of RNA was combined with 10 units of DNAse enzyme (10 units/μg RNA), RNAse Inhibitor enzyme (1units/μl total reaction volume) and 1 x reverse transcription buffer (Advanced Biotechnologies). The reaction volume was made up with DEPC-H_20. This was incubated at 37°C for 15 minutes before the DNAse enzyme was inactivated by heating at 75°C for 8 minutes.

II.2.2.7 cDNA synthesis

When the purpose of making cDNA was to amplify using primers which were located in 3`UTRs or with no intron between them, RNAs were checked for genomic DNA contamination by two successive PCR amplifications (without a reverse transcription step) using primers designed from the gene G3PDH (Clontech, table II.1) which has processed pseudogene copies. The PCR reaction was set up using standard material for PCR (explained later in this chapter). The negative control from the first PCR was always used as a template in the second PCR. If the RNA had genomic contamination it was treated with DNase enzyme and retested.

The reaction was heated at 94°C for 3 minutes and then cycled for 35 cycles on a Hybaid thermocycler machine as follows:

Denaturation step	94°C	20 seconds
Annealing step	56°C	30 seconds
Extension step	72°C	30 seconds

With a final extension step at 68°C 7 minutes

The total DNAsed reaction or RNA (about 1μg) was combined with 100-200 pmoles random hexamers. Before adding RT-buffer, dNTPs and enzymes the reaction was heated at 70°C for 5 minutes to denature the secondary structure of the RNA, and then cooled on ice for 1-2 minutes for annealing of random primers to the RNA. 200μM dNTPs (final concentration), 10 units of MMLV reverse transcriptase enzyme (Advanced Biotechnologies), RNAse Inhibitor enzyme (1unit/μl total reaction volume) and 5μl of 5 x reverse transcription buffer were added and the reaction volume made up with DEPC-H_20 to 25μl. This was then centrifuged briefly and incubated at 37°C for 60 minutes before the enzyme was inactivated by heating at 75°C for 15 minutes. 1/20th of the RT reaction was used in subsequent PCR reactions using specific primers.

Single-stranded cDNA was also prepared from total RNA of various tissues using MMLV Reverse Transcriptase (Advance Bio.) 10u/μg of RNA, RNase Inhibitor enzyme (Pharmacia) 1u/μl reaction volume, 1x RT buffer (with DTT, Advance Bio.), 10mmol dNTP and 500μg oligo-$dT_{(12-18)}$ primer in a total volume of 20μl. RNA and oligo-dT primers were incubated at 70°C for 5 minutes before adding enzymes, dNTPs and RT-buffer and the mix was then incubated at 37°C for 1h followed by 70°C for 15 min to inactivate the enzyme.

When a longer cDNA was required this procedure was carried out using SuperScript II reverse transcription enzyme (Gibco) following the enzyme manufacturer's instructions using various primers (oligo-$dT_{(12-18)}$, oligo-$dT_{(12-18)}$ with adapter (table II.1) or gene specific primers (GSPs)). SuperScript II lacks RNase H activity.

RT-PCR amplification with the same G3PDH gene primers as above was also used to confirm the success of reverse transcriptase reactions using cDNA as a template.

In addition, when the reason for making cDNA was to carry out PCR using primers which were known or were likely to be separated by an intron, cDNAs were made by random priming and the integrity of the cDNA was checked using primers

designed from exons 10 and 11 of the housekeeping gene *PGM1* (see table II.1 for primer sequences).

II.2.2.8 Reverse transcription of RNA with DDRT anchor primer

1μg of total RNA was denatured by heating at 70°C for 5 minutes. This was then combined in a 30μl reaction volume with 10 units MMLV reverse transcriptase (Advanced Biotechnologies), 30 units RNAse Inhibitor, 2.5 μM anchor primer (see Appendix, section A1.1.2 for primer sequences), 20 μM dNTPs and 5 x reverse transcriptase buffer made up to 30μl with DEPC-H_2O. The reverse transcription reaction was carried out for 60 minutes at 37°C, and the enzyme was inactivated by heating the reaction to 75°C for 15 minutes.

II.2.3 PCR and RT-PCR

PCRs in this study have been carried out using Red Hot *Taq* (routinely), and two other *Taq* DNA polymerase enzymes from Bioline (more sensitive, often giving additional PCR products and more of the desired product) and Promega (less sensitive, giving fewer PCR products and less of the desired product).

When the prevention of Taq mistakes was important for the PCR product sequence, a DNA polymerase (Clontech Advantage cDNA polymerase) with proof reading capability was used for the PCR amplification.

PCRs were routinely carried out using the following templates; 0.5-1μl of cDNA (from 20μl RT-PCR), 1 to 500ng of extracted DNA (such as plasmid or genomic DNA), different amounts of diluted PCR product, a PCR band (using a needle touch to the band or DNA extracted from an excised band) or, a colony from picked up from culture by tip of autoclaved tooth-pick. A standard PCR reaction incubated the following in a 50μl reaction volume (or half of this): 200μM dNTPs, 25 pmoles each of forward and reverse primers, 1 unit of *Taq* polymerase enzyme and 1 x enzyme buffer (containing 1.5mM $MgCl_2$). The amount of DNA included in the PCR reaction varied depending on the template; the reaction was made up to 50μl with ddH_20 and overlaid with mineral oil (two to three drops). Positive and

negative PCR controls were routinely set up; the negative control contained ddH_20 instead of DNA. The standard cycling conditions used were: initial denaturation at 94°C for 3 minutes (5min when template was a bacterial colony) followed by various numbers of cycles (20-35) of: denaturation (94°C for 20 seconds), annealing (varied) and elongation (72°C). A final elongation step was then carried out for 7 minutes. 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 was calculated from the primer sequence using the formula (Sambrook *et al.*, 1989):

T_m=69.3+0.41(%G+C content of primer)-(650/length of primer)

However, annealing temperatures were usually modified in the light of experience to achieve the best conditions and result for each specific PCR amplification. The reactions were carried out in a Hybaid thermocycler machine.

Primers which were used in these experiments are mentioned in figure III.22, figure IV.10, figure V.4 and table II.1.

II.2.4 DDRT-PCR

Routinely, half a microlitre from 20µl of the RT reaction was used for the PCR. The components of the PCR reaction were: 0.5µl reverse transcribed cDNA, 2.5µM anchor primer, 0.5 µM oligo 10-mer (see Appendix, section A1.1.2 for primer sequences), 2 µM dNTPs, 0.3 µM α^{35}S dATP, 10 x Taq Polymerase buffer (15 mM Mg^{2+}), 1 unit Taq polymerase (Promega) made up to 10µl with ddH_20.

The cDNAs were amplified using the following conditions: 94°C for 30 seconds, 42°C for 60 seconds, 72°C for 30 seconds, cycled 40 times. A final extension was carried out for 5 minutes at 72°C.

The PCR reaction was carried out in a total reaction volume of 10µl and about 2µl from the PCR reaction was routinely loaded onto a polyacrylamide gel for electrophoresis.

II.2.4.1 Non-denaturing acrylamide gels

6% non-denaturing gels were prepared in the same way as denaturing gels for sequencing (mentioned above, section II.2.1.12) except that urea was not included in the gel mixture and no bonding solution was applied to the front glass plate; the gel mixture therefore consisted of 45mls H_20, 6mls 10 x TBE and 9mls 40% (19:1) acrylamide:bisacrylamide solution to make a total volume of 60mls.

2µl of the radioactive PCR product was combined with 0.5µl of formamide loading dye and loaded onto the gel. After electrophoresis for an hour at 50W, the gel was removed onto a piece of Whatman 3MM paper, covered in cling film and dried in an 80°C oven for 20 minutes. The gel and Kodak Biomax MR X-ray film were taped together and clamped together between two glass plates. Hole punches or scissor cuts were used to mark the gel and X-ray film so they could be re-aligned later. The gel was exposed to the X-ray film at room temperature overnight in a light-tight black bag. After exposure, the gel was re-aligned with the X-ray film and differentially expressed bands were cut from the gel using a scalpel. For DNA extraction from the gel see section II.2.1.5.

II.2.4.2 Radioactive marker for non-denaturing acrylamide gels

10µg of Lambda DNA was digested with 20 units of *Hin*fI enzyme, 1 x enzyme buffer in a total volume of 20µl. The reaction was incubated at 37°C for 3 hours. The digested DNA was then radiolabelled with α^{35}S d-ATP by end-filling (to distinguish between this reaction and end-labelling using T4 Kinase) with 20 units of Klenow enzyme. This reaction was incubated at room temperature for 15 minutes.

200µl of TE buffer and 28µl of formamide stop solution were then added. 1-2µl of the radioactive ladder was loaded per gel. This produced a radioactive ladder with a known size range of 10 - >300bp which could be visualised clearly upon exposure to X-ray film.

II.2.5 Northern Blotting

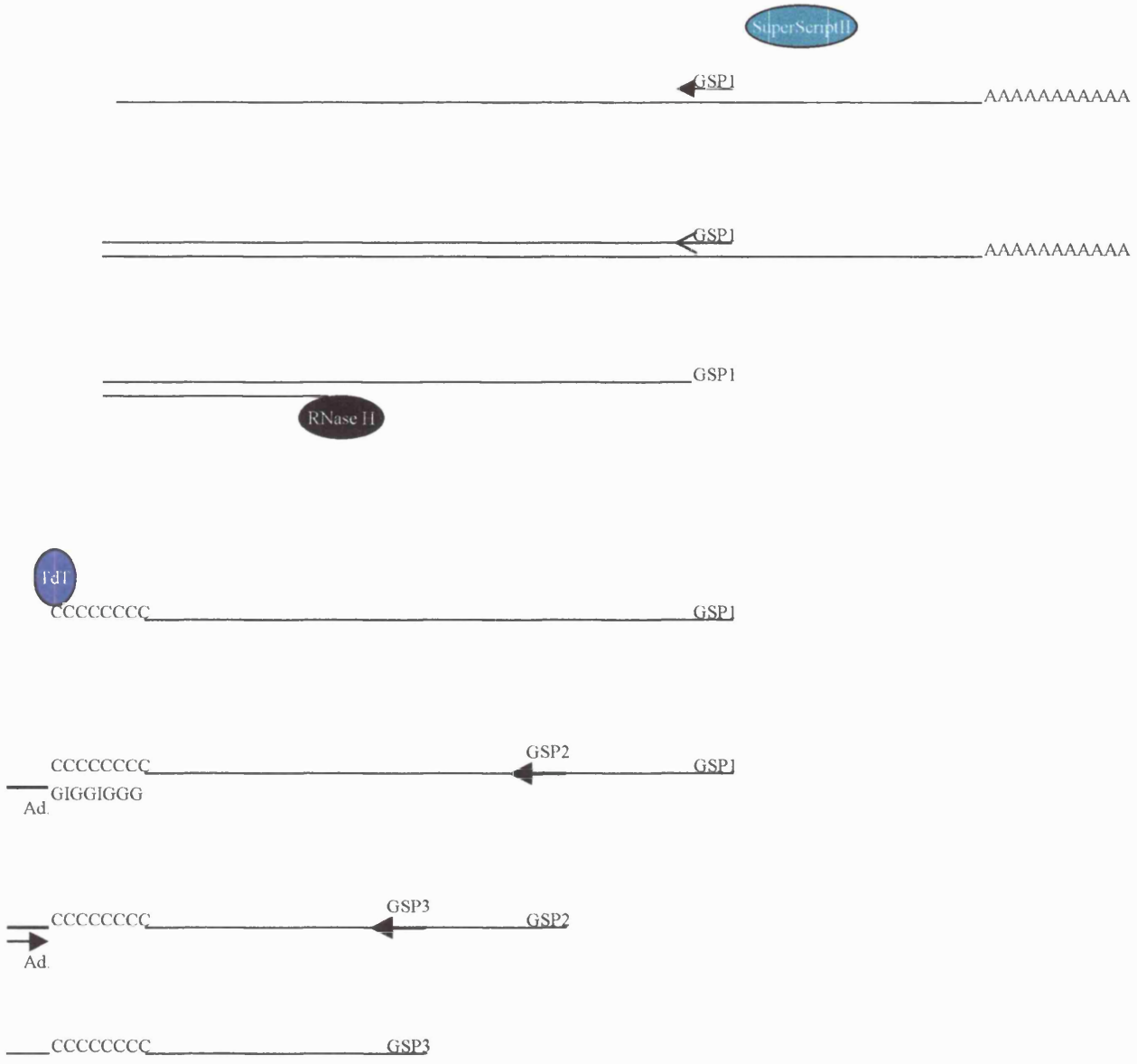
A multiple human tissue Northern blot (Human MTN Blot2, Clontech) was probed with a radioactively labelled RT-PCR product generated using primers GSP2 and R6 (figure III.21). Pre-hybridization and hybridization were carried out for 2 hours at 60° C and 20 hours at 65°C respectively in solutions made according to the manufacturer's recommended protocol. The blot was washed twice in 2xSSC and 0.05% SDS for 30 min at room temperature and once in 0.1xSSC and 0.1% SDS for 40 min at 55°C. The blot was exposed to X-ray film (Kodak-Biomax-MR) for 36 hours.

II.2.6 5' RACE

For all 5' RACE experiments the 5' RACE system for the rapid amplification of cDNA ends, version 2.0 (Life technologies, GIBCO BRL) was used following the manufacturer's instructions. In brief, first strand cDNA is synthesised from total RNA using a gene-specific primer and SuperScript^{tm} II, an RNase H- derivative of Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLVRT). The primer, 3μg of total RNA and DEPC treated water were heated to 75°C (for two minutes) and added to a reaction mix containing 20mM Tris-HCI (pH 8.4), 50mM KCI, 2.5mM $MgCI_2$, 10mM DTT, 100mM cDNA primer, and 400μM each of dATP, dCTP, dGTP, dTTP (final concentration). The reaction mix was incubated at 42°C for one minute. 1μl (200 units) of the SuperScript^{TM} II RT was added, mixed and incubated at 42°C for 60 minutes.

The reverse transcription reaction was stopped by incubating at 70°C for 10 minutes, the RNA template was digested by treatment with an RNase mix (containing RNase H which is specific for RNA: DNA heteroduplex molecules, and RNase T1) at 37°C for 30 minutes. Following brief centrifugation the reaction mix was passed through a Glassmax column. Prior to application to the column 120μl of

binding solution (6M Nal) was added to the newly synthesised cDNA. The column was centrifuged at 13000g for 20 seconds, the flow-through collected and kept at-20°C. The column was washed four times with the wash buffer (supplied by manufacturers), and twice with 70% ethanol to remove any unincorporated dNTPs,



1- Specific cDNA synthesis using mRNA or total RNA and Gene Specific Primer 1(GSP1).

2- Single strand cDNA was prepared and mRNA was removed by using RNase H.

3- Single strand cDNA was cleaned and all dNTP and primers were removed from the reaction.

4- A polyC tail using dCTP and TdT enzyme was made.

5- The first RACE-PCR using GSP2 (or GSP1) and PolyG with an adapter primer (abridged anchor primer)

6- The second RACE-PCR using GSP3 (or GSP2) and the adapter primer

7- The PCR result band from an agarose gel was cut out for sequencing

Figure II.1; The procedure of 5`RACE

Name	Sequence (5' to 3')	Note
5′RACE abridged anchor primer	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	For 5′RACE
Abridged universal amplification primer (AUAP)	GGCCACGCGTCGACTAGTAC	For 5′RACE
Adaptor-poly(dT)$_{18}$-VN primer	TTGCAGTGGTAACAACGCAGAGTACTTTTTTTTTTTTTTTTV (A,C or G) N (A,C,G or T)	For 3′RACE
Adaptor primer	TTGCAGTGGTAACAACGCAGAGT	For 3′RACE
5′GAPDH primer	ACCACAGTCCATGCCATCAC	
3′GAPDH primer	TCCACCACCCTGTTGCTGTA	
PGM1-forward primer	TCCGACTGAGCGGCACTGGGAGTGC	Exon 9
PGM1-reverse primer	GCCCGCAGGTCCTCTTTCCCTCACA	Exon 11
T7 primer	TAATACGACTCACTATAGGG	Vector primer
T3 primer	ACAGCTATGACCATG	Vector primer
U19 primer	TTTTCCCAGTCACGACGT	Vector primer

Table II.1: Primers were used in 5′RACE and 3′RACE experiments and RT-PCRs (except gene specific primers).

gene-specific primer and proteins. The cDNA was eluted from the column with 50µl of sterilised, distilled H_2O at 65°C by centrifugation at 13000g for 20 seconds. A homopolymeric poly -C tail was added to the 3' end of the cDNA using terminal dexynucleotidyl transferase (TdT). TdT tailing creates the binding site for the abridged anchor primer on the 3' end of the cDNA, which is the primer used for the creation of the second cDNA strand. A reaction mix containing the cDNA in 10mM Tris-HCl (pH 8.4). 25mM KCl, 1.5mM $MgCl_2$, 200µM dCTP was prepared and incubated at 94°C for three minutes. The mix was placed on ice and 1µl of TDT added. After incubation at 37°C for ten minutes, the reaction was stopped by heat inactivation at 65°C for ten minutes.

The tailing buffer is PCR-compatible. This means that the reaction can be directly amplified without intermediate organic extractions, ethanol precipitation, or dilutions. PCR amplification was carried out using *Taq* polymerase, a nested gene-specific primer corresponding to a site located within the cDNA molecule, and a novel deoxyinosine-containing anchor primer provided with the kit. The thermal cycler was preheated to 94°C prior to use and reaction mixes were prepared on ice. The first amplification used the abridged anchor primer in a reaction mix of final composition, 20mM Tris-HCl, 50mM KCl, 1.5mM $MgCl_2$, 400mM gene-specific primer (termed either GSP1 or GSP2), 400mM abridged anchor primer, 200µM each of dATP, dCTP, dGTP and dTTP, tailed and 2.5 units of *Taq* polymerase. Several different *Taq* polymerase enzymes were tried for the best result. Each reaction was covered with three drops of mineral oil to prevent evaporation during thermal cycling. The reaction was heated at 94°C for 3 minutes and then cycled for 30 cycles using conditions established for each primer pair on a thermocycler machine as follows:

Denaturation step	94°C	20 seconds,
Annealing of primer	(varied depending on GSPs)	45 seconds,
Primer extention	72°C	1.5 to 6 minutes,

followed by a final extension at 72°C for 7 minutes. 10µl of the first 5' RACE product was analysed on a 2 % agarose gel. In general, no product was visible at this stage.

The second amplification was carried out using second nested gene-specific primer (GSP2) and various dilutions of the first amplification product as template. Generally the first amplification products were diluted in double distilled sterilised water so that the following dilutions, 1:5, 1:100 and 1:500 were obtained. The PCR mix for each sample contained the following, 20mM Tris-HCl (pH8.4), 50mM KCl, 1.5mM $MgCl_2$, 200nM GSP2, 200nm AUAP (universal primer), 200µM each of dATP, dTTP, dGTP and dCTP (final concentration), diluted first round PCR product and 2.5 units of *Taq* DNA polymerase. Each reaction mix was covered with mineral oil and the reactions heated at 94°C for 3 minutes and then cycled for 30 cycles using conditions established for the primer pair.

Denaturation	94°C	45 seconds,
Annealing step	(varied)	45 seconds
Extension step	72°C	90 seconds

A final extension period of 7 minutes at 72°C. 10µl of each reaction mix was analysed by electrophoresis on a 1.2% agarose gel.

Details of the primers used for each 5'RACE (figure II.1) reaction are listed in table II.1, figure III.22, figure IV.10 and figure V.4.

II.2.7 3'RACE

3'RACE is the acronym for Rapid Amplification of 3'cDNA Ends. In this procedure mRNA (about 0.5-1µg) was converted into cDNA using reverse transcriptase enzyme (SuperScript II) and an oligo(dT) adapter primer consisting of 25 nucleotides of arbitrary sequence with a T18-VN 3'tail (see table II.1). The reverse transcription reaction was set up using the protocol explained in section II.2.2.7 for making cDNA using an oligo(dt) primer. Gene specific cDNA was then amplified by PCR using a specific primer that anneals to an internal region of the cDNA sequence and an adapter primer which is identical to the arbitrary 5'end of the RT-PCR primer (see table II.1 and figure IV.10). To generate a specific

amplification product, I found it advantageous to design a second semi-nested, gene specific primer at sequence within the cDNA further downstream and for use in a second amplification reaction with the adapter primer. 0.2μl of the first 3′RACE-PCR product without further purification or 0.5-1μl of DNA extracted from an excised gel was used for the second 3′RACE-PCR. This gave a single band on an agarose gel. The band was cut from the gel and, after DNA extraction was used for sequencing using either the second gene specific primer or a third downstream primer. Genomic DNA and water (ddH_2O) were always used as controls during amplifications. In the second 3′RACE-PCR, a control reaction was also set up using the first PCR product from genomic DNA.

A schematic of the procedure of the 3′RACE is shown in figure IV.5 in chapter IV.

II.2.8 Screening human adult testis and mouse spermatocyte cDNA libraries

A human adult testis cDNA library in lambda ZAP was titred and screened according to the conditions advised by Stratagene, the suppliers of lambda ZAP, using XL1-Blue as the host strain. Another cDNA library which had been made from mouse spermatocyte in Uni-ZAP (Stratagene) was also screened according to the manufacturer's instruction (nearly the same for both libraries). The titre of the human and mouse libraries were estimated at 1×10^8 and 2.5×10^8 p.f.u./ml, respectively.

II.2.8.1 Preparation of competent cells

E.coli strain XL1-Blue was streaked onto L agar plates supplemented with 100μg/ml ampicillin and incubated, inverted, at 37°C overnight. A single colony was then isolated and used to inoculate 10mls of L broth supplemented with 10mM $MgSO_4$ and 2% maltose, and shaken at 250rpm overnight at 37°C.

1ml of this overnight culture was added to 50mls of pre-warmed L broth (containing the same supplements as before) and shaken at 37°C until the cells grew to an OD_{600} of 0.5 (approx. 2.5×10^8 cells/ml). The culture was then cooled on ice and centrifuged at 1500rpm in a Mistral 2000 bench centrifuge for 10 minutes. The

cells were resuspended in 15mls of ice-cold sterile 10mM $MgSO_4$ which was made up using filtered double distilled water. The cells were usually used immediately, but could be stored at 4°C for 2 weeks with little loss in plating efficiency (so 5-10% more aged cells were used).

II.2.8.2 Plating out the cDNA library

Approximately 3 x 10^5 pfu from the human adult testis cDNA library and 2 x 10^5 pfu from the mouse spermatocyte cDNA library were each plated as follows onto four 22 x 22cm megaplates containing NZY agar (pre-warmed to 37°C). An appropriate dilution of the cDNA library was added to 2mls of competent cells and incubated at 37°C for 15 minutes. 50-60mls L top agarose was melted and cooled to 45°C and 1ml of 20% maltose was added before the cells were added and the top agarose was poured on top of an NZY agar plates. After setting, the plates were incubated at 37°C inverted, overnight.

Filter lifts were taken from the plates by placing a wet Hybond N^+ membrane after soaking in 2x SSC or ddH2O on each plate (orientated with pin-pricks and ink) for 2 minutes, and then a second, duplicate, membrane for 4 minutes. These filters were placed (DNA side up) on Whatman 3MM paper soaked with denaturing solution for 2 minutes, and then submerged in 1M Tris base for 2 minutes and neutralizing solution for 5 minutes, twice. Before the filters were air dried and then baked in an 80°C oven for 2 hours they were soaked in 2x SSC for 2 minutes. 1µl of the unlabelled probes were used to make a dot on top of the filters to control the hybridisation.

II.2.8.3 Pre-hybridisation and hybridisation of DNA fixed onto filters

Pre-hybridisation buffer which was used for hybridisation was 5x SSPE or SSC, 1-3xDenhardt's, 0.1-0.5% SDS, 100-200 µg/ml denatured, sheared, Herring sperm DNA and 5M saturated tetra-pyrophosphate sodium salt in ddH2O. 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 (60-65°C) in a rotating hybridisation oven (Hybaid).

The radiolabelled DNA probe was then denatured by boiling for 5 minutes and then added to the hybridisation buffer after removing about 2/3 of the buffer. Hybridisation was usually carried out overnight at 63-65°C in a shaking water bath or rotating hybridisation oven.

II.2.8.4 Washes after hybridisation and signal detection

The membranes were primarily washed with 2X SSC and 0.1% SDS twice at room temperature and then at low stringency with the same washing buffer for 20-40 minutes at 60-65°C temperature and thereafter at increasingly higher stringency up to 0.1x SSC and 0.1% SDS at room temperature for 20-40 minutes. The washing conditions were varied depending on the probe and checking radioactivity of filters by a Geiger. 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 (Kodak biomax MR) at -70^0C in a light-proof cassette with intensifying screens. Exposures ranged from 16 hours to 3 days.

II.2.8.5 Isolation of positive plaques and preparation of amplified phage stocks

Positive plaques, present in both first and duplicated filters, were isolated from the original cDNA library plates and stored in 0.5ml of SM buffer containing 20µl of chloroform. Plaque plugs as small as possible were isolated in order to have less irrelevant plaques in the secondary screening. The plaques were allowed to diffuse for 1-2 hours at room temperature and then stored at 4°C. Further screenings were carried out by plating dilutions of these primary plaques onto 90mm agar plates so that secondary plaques could be isolated.

An amplified stock was made from single positive plaques and also both human and mouse libraries by plating about 10^5 pfu onto 90mm agar plates and growing at 37°C overnight. Each plate was overlaid with 5mls of SM buffer and left at 4°C for 2 hours with intermittent shaking. The SM buffer was harvested and a

further 1ml was added to the plate for 15 minutes; this was then combined with the first harvest. 0.1ml of chloroform was added to the amplified stock, vortexed and centrifuged at 3000rpm in a Mistral 2000 bench centrifuge for 10 minutes. The supernatant was then recovered, one drop of chloroform was added and the stock stored at 4°C. Amplified stocks had titers ranging from 10^{10}-10^{11} pfu/ml.

II.2.8.6 Isolation of cDNA inserts

The lamdaZAP and Uni-ZAP XR vector has been designed to allow *in vivo* excision and recircularisation of any cloned insert contained into the lambda vector to form a phagemid (pBluescript) containing the cloned insert. This system provides efficient excision of the pBluescript phagemid from the vector, while eliminating problems associated with helper phage coinfection. The ExAssist helper phage contains an amber mutation that prevents replication of the phage genome in a nonsuppressing *E.coli* strain such as SOLR cells. This allows only the excised phagemid to replicate in the host, removing the possibility of productive coinfection from the ExAssist helper phage. A liquid lysate was prepared by incubation of 10^8 XL1-Blue cells, 10^9 of ExAssist helper phage and 10^7 pfu of the isolated phage. The phage were allowed to adhere to the cells at 37°C for 15 minutes without shaking, and then by adding 3ml of LB broth were grown at 37°C for about 3 hours, shaking at ~250rpm. During this time, both phages grow, recombination excision of the phagemid takes place and cells lyse. The culture was centrifuged for 15 minutes at 3000 x g to remove cellular debris and intact cells. The culture supernatant was heated at 70°C for 15 minutes and centrifuged for 15 minutes at 3000 x g. A mixture of 200µl of SOLR cells (OD_{600}=1) and 100µl of the heated phagemid solution after incubation at 37°C for 15 minutes was plated onto two or three plates containing LB agar and 50µg /ml ampicilin. After overnight incubation at 37°C colonies containing the pBluescript double-stranded phagemid with cloned DNA insert were checked by PCR.

II.2.9 Genomic and cDNA Library screening (human and mouse)

6 commercial filters (Research Genetics Inc.) containing a gridded BAC library of 125,000 clones (library RPCI-11 prepared by P. de Jong at Roswell Park Cancer Institute) were screened as recommended by the manufacturer. Two RT-PCR products from different regions of *TSGA10* gene (see chapter three) were electrophoresed through a 1.6% agarose gel and bands were cut out and purified by centrifugation through glass wool and the "QIAquick" PCR purification kit (Qiagen). Purified DNA was labelled using random prime labelling system (explained above), and radioactive ^{32}PdCTP. Pharmacia NICK columns were used to remove unincorporated ^{32}P-labelled nucleotides. Buffer containing 6× SSC, 0.5% SDS and 5× Denhardt's solution was used for pre-hybridization and hybridization after adding 12μg/ml of denatured sheared Herring sperm DNA.

The position of positive clones were given reference numbers by aligning the autoradiographs to a numbered grid provided by the manufacturer. The corresponding BACs were obtained from Research Genetic Inc. in the form of agar stabs which required overnight incubation at 37°C. Cells from each BAC stab were streaked onto L agar plates with kanamycin at 25μg/ml and grown overnight at 37°C. A single colony was picked from these plates for each BAC and grown in 3ml L-broth (with kanamycin) at 37°C overnight. Glycerol stocks were made and plasmid DNA was isolated.

Selected BAC clones were checked by PCR (using different pairs of primers located in different regions of the gene) for the presence of the gene and were grown at 37°C on L agar.

40 filters derived from a gridded PAC human genomic library of 125,000 clones were a gift of M. Ross (Sanger Centre) and were screened as previously described (Hornigold et al. 1997) using radioactive ^{32}P dCTP labelled probes made from TG2 and T2G3 gene cDNAs (see chapter 4 and 5, respectively). The major difference with the above screening was the Pre-hybridisation and hybridisation mix which was 6x SSC, 1%Sarcosyl, 10% Denhardt's solution, 50mM Tris-HCl and 10% Dextran Sulfate in ddH_2O.

A mouse genomic PAC library (Osoegawa *et al.*, 2000) on 7 gridded filters was obtained from the UK HGMP Resource Centre (http://www.hgmp.mrc.ac.uk/Biology/descriptions/mouse_pac.html). The library consists of approximately 128,000 clones and has an average insert size of 147kb. Further information about the library and the vector are available from the WWW site: http://www.chori.org/bacpac/. Pre hybridisation and hybridisation with a probe made by RT-PCR using human primers and mouse cDNA (explained in chapter III, section III.9) was carried out as explained above. The positions of positive signals from both above PAC genomic libraries were identified according to the HGMP resourse centre instructions. The corresponding PACs were requested from the centre and were studied using the same procedures as for BAC clones.

Single isolated bacterial colonies were checked with PCR using mouse TSGA10 R12 and G12 primers (see chapter three) which were designed from a cDNA clone and are located in the 5′ end of the gene.

II.2.9 Fluorescence In situ Hybridisation

BAC or PACs containing both ends of genes were labelled (Nick translated or Random Prime Kit, Gibco BRL) with Avidin- conjugated Fluorescein isothiocyanate (FITC-Avidin) following the manufacturer's instructions. They were hybridised to male human or mouse metaphase spreads prepared for FISH analysis using standard procedures. Hybridization and detection conditions on metaphase chromosomes were performed as described (Gillett et al., 1993). FISH signals and the DAPI (4,6-diamidine-2-phenylindole dihydrochloride) banding pattern were merged for figure preparation. Images of metaphase preparations were captured by a cooled CCD camera using the Cyto Vision Ultra image collection and enhancement system (Applied Imaging Int. Ltd.).

Chapter III

III.1 Detection of genes involved in spermatogenesis using DDRT-PCR

This chapter describes the optimisation of the published DDRT-PCR and 5`RACE techniques. The gene *TSGA10* which was discovered using these techniques is described as an example of the methodology and of the characterisation of a gene.

III.2 Background

I have attempted to isolate genes expressed in spermatogenesis using DDRT-PCR. The products of such genes may be involved in the cell cycle or be important for sperm structure or function.

Recently, several protocols have been developed to access genes with particular expression profiles. Differential Display Reverse transcription-PCR (DDRT-PCR) is a powerful method used widely for the identification of genes that are differentially expressed between different cell or tissue types by creation of an RNA fingerprint for each sample (chapter I, part four). Briefly, the scheme of this method, classic DDRT-PCR, and other procedures for characterisation and localisation of the DDRT-PCR products are shown in figure III.1. The steps are explained separately and the results obtained using these methods are also presented.

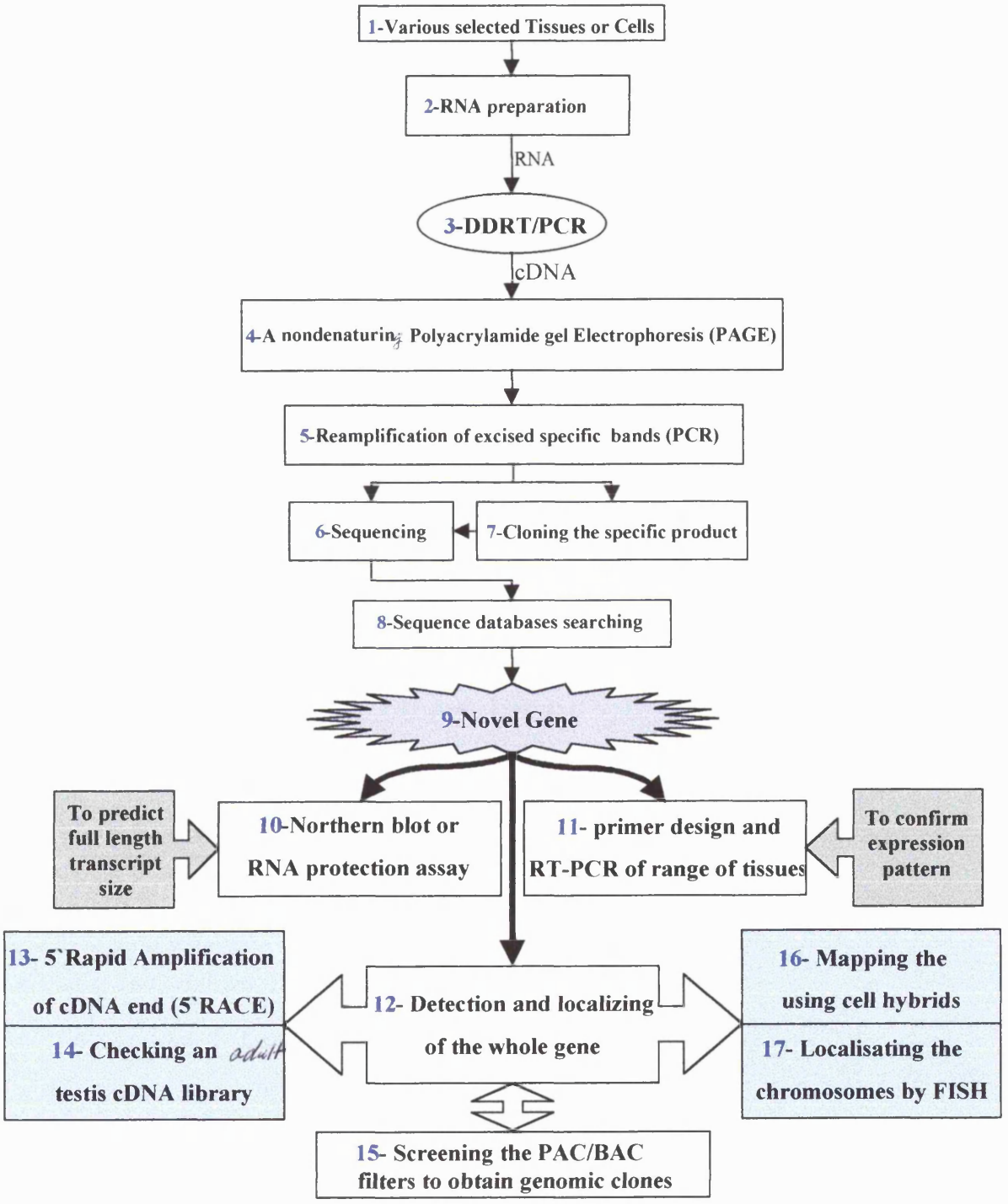


Figure III.1 : Process of DDRT-PCR and the characterisation of genes obtained by this method.

Step one: Tissue selection

The goal of this study is the identification of genes involved in the processes of spermatogenesis and spermiogenesis, so the best samples for carrying out DDRT-PCR, in comparison with control tissues, would be the individual germinal cells from testis and sperm. However, due to technical problems the selection of human germ cells at in each stage of spermatogenesis for DDRT-PCR is currently too difficult.

A major advantage of this technique is that it requires very small amounts of cellular RNA which are obtainable from small pieces of human tissue biopsy. A second important feature of DDRT-PCR is the ability to compare more than two RNA samples at once. This feature is nearly impossible in differential screening if subtractive hybridisation is involved (Matz and Lukyanov, 1998) and other techniques which have been discussed in chapter I, part IV.

In our experiments tissues were chosen which have similarities to the functions of testis except for spermatogenesis, so DDRT-PCR would be more likely to reveal differences in gene expression related to the involvement of the testis in spermatogenesis. In particular, the adrenal gland was selected because it secretes testosterone and the prostate because it is a site of production of components of semen. Some genes expressed in the testis are essential housekeeping genes and, since testis also includes a proportion of general somatic cells as well as blood vessels, neurones and connective tissue, tissues such as male liver, brain and muscle tissues were also selected. The brain RNA especially helps in this case as this tissue is known to express a wide variety of genes and therefore eliminates a large number of cDNAs (Adams *et al.*, 1993). Lastly but most obviously, gene products involved in spermatogenesis might be absent from the testis of an azoospermic patient with no active tubules. In summary, DDRT-PCR was carried out using eight tissues (see table III.1).

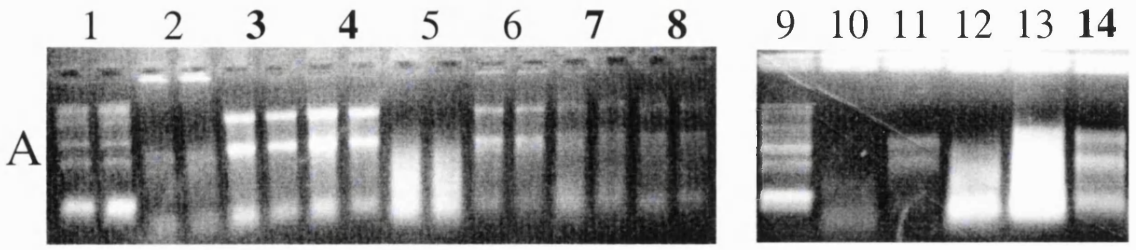
1- Normal testis 2- Testis tissue from Azoospermic patient 3- Semen (containing predominantly sperm and also small numbers of WBC and Epithelial cells) 4- Adrenal gland tissue (adrenal gland is the second source gland for the secretion of testosterone after testis) 5- Prostate gland tissue (prostate gland has exocrine secretion such as Prostaglandins like testis) 6- Liver tissue (as a general tissue with high level of gene expression) 7- Muscle tissue (as a general tissue, also there is small amount of muscle in testis) 8- Brain tissue (as a general tissue with high level of gene expression and also testis has some neurones)

Table III.1: Tissues have been used in DDRT-PCR and the reason for selection.

The patient sample was from somebody who referred to a infertility clinic about two years after marriage and suffered from non-obstructive infertility. Testing his semen, no sperm has been found and he is therefore classified as an azoospermic patient. The histopathology of his testis showed seminiferous tubules and contained germinal cells at various stages of maturation but obviously in lower numbers than normal testis. Semen was obtained from a fertile man according to WHO protocol (chapter I, part II). It should be noted that there are three different sorts of cells present in a normal semen sample. The majority are sperm cells but white blood cells (WBCs) and epithelial cells are also present. Obviously the most differentially displayed cDNAs, and thereby the most testis-specific cDNAs, will be identified when using the greatest range of RNAs in the initial display reactions. It should also be pointed out that having tissue samples from normal prepubertal testis and a foetal testis would be the best selection to display genes involved in spermatogenesis during DDRT-PCR but to obtain them would be invasive and difficult.

Step two: RNA preparation for DDRT-PCR

One of the advantages of DDRT-PCR is the ability to carry out the experiment using either total RNA or mRNA, using a quantity of about 100 ng of total RNA.



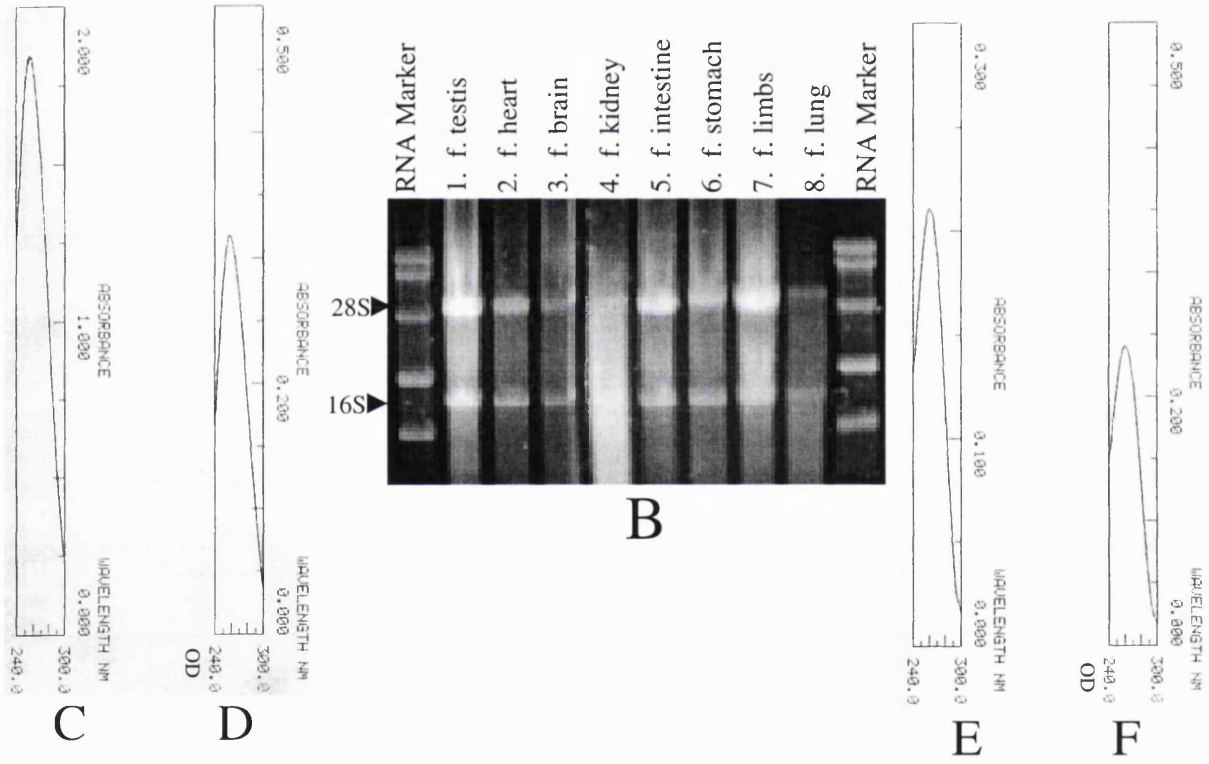


Figure III.2: A) RNA isolated from human tissues; lanes 1-8 RNA gel electrophoresis and lanes 9-14 a normal 1.2% agarose electrophoresis: 1-RNA ladder size marker, 2-Lymph node (not used in the experiment), 3-Normal adult testis, 4-Adrenal gland, 5- Benign polyp (not used in the experiment), 6-Squamus cell carcinoma, 7-Infertile adult testis, 8- Muscle, 9-RNA ladder, 10-Colon (not used in this experiment), 11-Thyroid gland tumor, 12-Parotid tumor 13-Granulation tissue, 14-Prostate. Only RNA in the lanes marked with bold numbers was used for DDRT-PCR.

B)RNA isolated from human fetal tissues (used for RT-PCRs).

Spectrophotometer results from some of the RNA samples: C-Adrenal gland, D-Liver, E-Normal adult testis, F-Sperm.

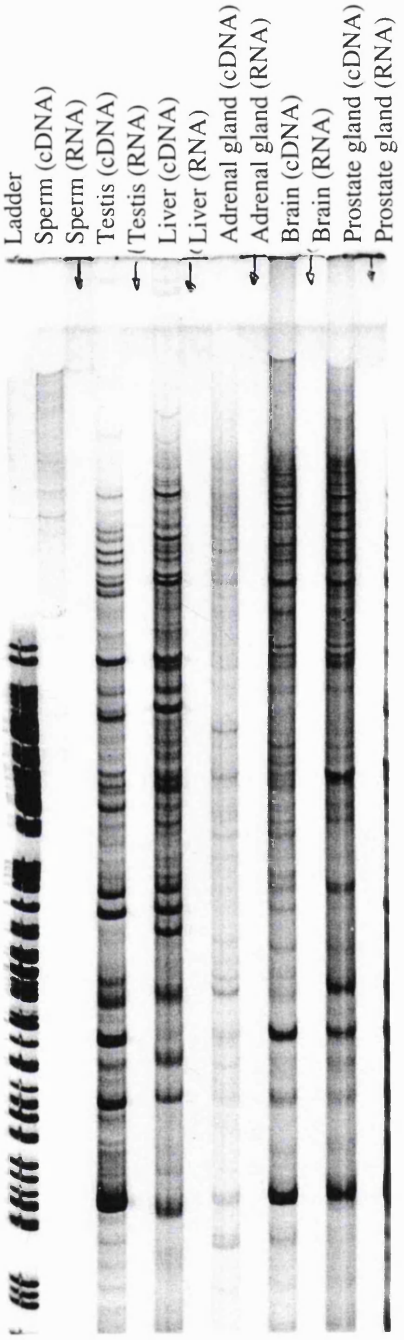


Figure III.3: DDRT-PCR results using only RNA as a negative control comparing to results from cDNA as a positive control.

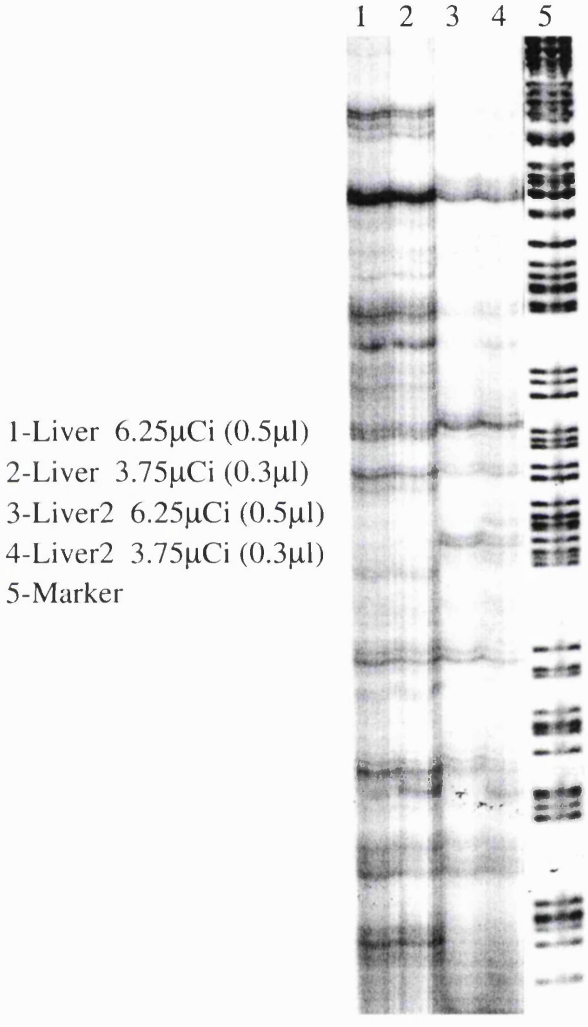


Figure III.4: DDRT-PCR using different amounts of radioactive α^{35}S-dATP.

This experiment was carried out using total RNAs from eight different tissues. RNA samples were prepared according to the Materials and Methods. The quality and quantity of the RNA was checked by spectrophotometer and electrophoresis on an RNA MOPS-formamide gel (figure III.2).
A set of RT-PCR reactions were also carried out to show that the RNA used for the Differential Display reactions was initially free from DNA. The PCR reactions shown in figure III.3 demonstrate that when RNA was included in a PCR reaction using primers which would normally amplify products from DNA, no products were seen.

Step three: DDRT-PCR

The first stage of the DDRT-PCR reaction is the conversion of RNA to cDNA using four different Anchored oligo (dT) primers. The reactions for RNA from each tissue being studied are carried out in parallel. The anchored primers are 14 bases long, containing 12 Ts which will bind to the poly (A) tail present at 3′ end of most eukaryotic mRNAs, and two additional nucleotides at their 3′end which limit the number of mRNA species transcribed into cDNA. Without these two nucleotides providing specificity, a smear would result in the subsequent PCR reaction (Liang *et al.*, 1993) as too many products are generated for separation through a polyacrylamide gel.

In this experiment, reverse transcription reactions for all above-mentioned tissues were carried out as outlined in the Methods, section II.2.2.8 using $(dT)_{12}VG$ primer and 0.5μg of total RNA. cDNAs primed with $(dT)_{12}VC$ were provided by J.Cameron (see below).

At the second stage, following cDNA synthesis, anchored oligo(dT) primers and arbitrary 5′ oligonucleotide 10-mers are used to carry out polymerase chain reaction on reverse transcribed RNA. The arbitrary primers anneal to the reverse transcribed cDNA fragments resulting in many PCR products of different sizes. In a classic DDRT-PCR, using 10-mer primer, 50-100 products may be obtained. A radionucleotide, such as α^{35}S-dATP, is included in the PCR reaction so that the products can be visualized by autoradiography after electrophoresis.

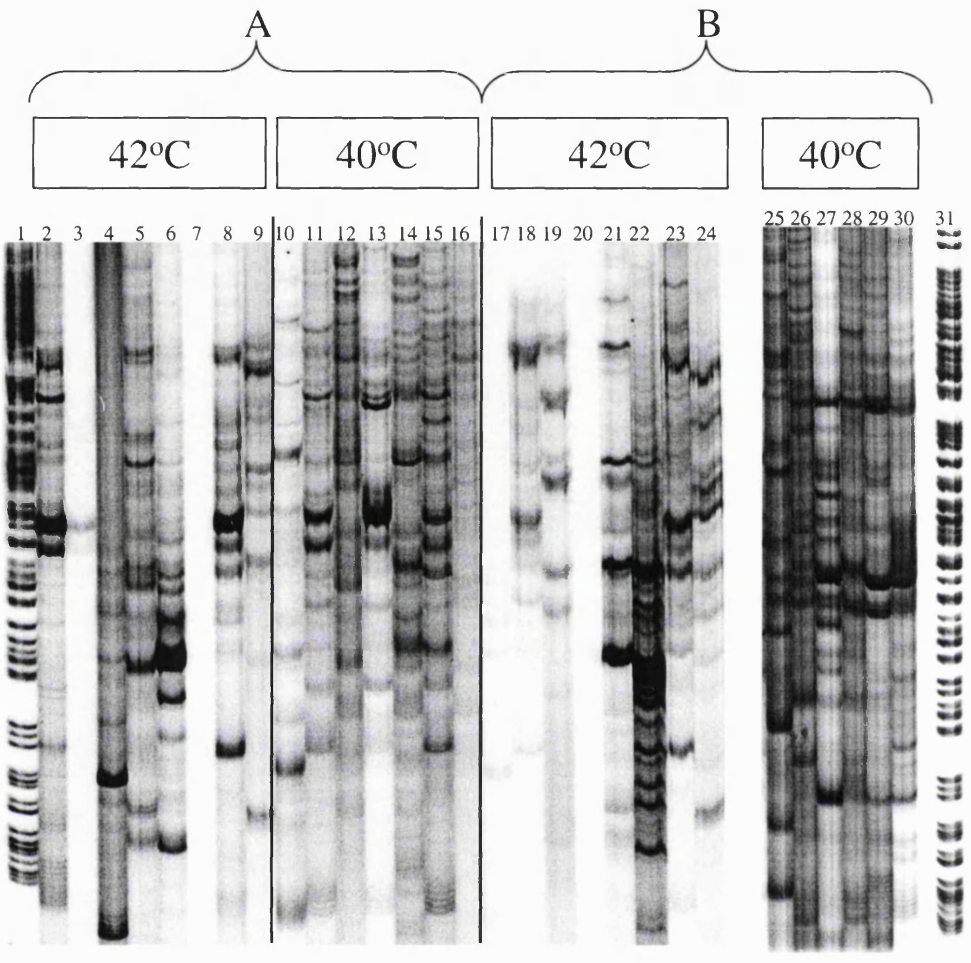


Figure III.5: DDRT-PCR at different temperatures. VG or VC refer to the anchored primer used. A & B experiments were carried out by two persons separately.

1-Marker
2-Adrenal gland VC
3-Liver VC
4-Testis VG
5-Adrenal gland VG
6-Liver VG
7-Liver2 VG
8-Muscle VC
9-Muscle VG
10-Testis VG
11-Adrenal gland VC
12-Adrenal gland VG
13-Liver VC
14-Liver VG
15-Muscle VC
16-Muscle VG
17-Testis VG
18-Adrenal gland VC
19-Adrenal gland VG
20-Liver VC
21-Liver VG
22-Liver2 VG
23-Muscle VC
24-Muscle VG
25-Adrenal gland VG
26-Muscle VC
27-Muscle2 VC
28-Adrenal gland VC
29-Adrenal gland2 VC
30-Liver VC
31-Marker

The protocol outlined in the Methods, section II.2.4 had been found to be optimal for DDRT-PCR by J. Cameron (1998), a previous PhD student in our lab, but I carried out some control experiments during the experiment which helped me to set up DDRT-PCR more easily. For example, a reduced amount of α^{35}S-dATP was used in the PCR reaction: 3.75μCi (0.3μl) instead of 6.25μCi (0.5μl). This gave visible and distinguishable bands after overnight exposure of polyacrylamide gel to X-ray film, (figure III.4) and was both safer and cheaper. 0.1μl of α^{35}S-dATP was not enough to give an accurate result (J.Cameron, personal communication). In another experiment the annealing temperature of the PCR reaction was varied, comparing 40°C and 42°C (figure III.5). An annealing temperature of 42°C was used in all subsequent reactions because usually when a lower annealing temperature is used it is more likely that non-specific amplification products are produced.

It is interesting to note from this figure and also in figure III.4 that, although similarities can be seen, there was a significant difference in the band patterns produced from the same tissue used in different experiments. For this reason only band patterns generated in the same experiment were compared to each other.

DDRT-PCR required only a small amount of RNA. Each RT reaction gave sufficient material for about 40 PCR reactions (ie. with different arbitrary primers) and each of those gave sufficient material for about 4 gel electrophoresis comparisons.

Step four: polyacrylamide gel electrophoresis (PAGE)

The products of DDRT-PCR were separated by electrophoresis through a 6% non-denaturing polyacrylamide gels using a sharkstooth comb such as that used for sequencing gels for about an hour. Non-denaturing gels, which are easier and quicker to prepare than denaturing gels (containing urea), were used preferentially (Bauer *et al.*, 1993). The DDRT-PCR products were loaded onto the gel with a formamide loading dye which helps to give better (sharper) bands.

To control the electrophoresis of the DDRT-PCR fragments, a radioactive size marker (bacteriophage Lambda DNA digested with *HinfI* enzyme and end-labelled

with α^{35}S-dATP) was electrophoresed adjacent to the samples on each gel. It has been reported that polyacrylamide gels sometimes do not provide adequate resolution, resulting in the isolation of several different cDNAs from what appeared to be a single band. Each band may therefore represent different cDNAs of the same or similar molecular weight (Callard *et al.*, 1994).

Step five: Re-amplification

Using the above method it was possible to identify some specific bands that were differentially present just in the testis or in sperm (table III.2). Wherever differences were noted between lanes the bands were excised and eluted in 0.1ml TE (Methods, section II.2.1.5). After elution of the DNA, a fraction of the eluted sample was re-amplified by PCR. I carried out two separate PCR reactions using differing amounts of eluted DNA as a substrate (0.5µl and 3µl)(figure III.6-A), because it was thought that the variable amounts of cDNA present in the excised bands might not always be amplified by a standard PCR reaction. The effect of varying the conditions of the PCR reaction was tested. Altering the denaturation time (20-40 seconds), the annealing temperature (40°C or 42°C), the extension time (20-40 seconds) and the number of cycles (35 or 40) had no effect on the band patterns produced (comparative results not shown). As shown in table III.2, almost all of excised bands in this experiment were amplified successfully. However, sometimes the result of re-amplification PCR was two bands or a fuzzy rather than a sharp band which may be due to a number of smaller fragments trapped in the acrylamide gel during migration. This problem was seen especially when PCR was carried out with more than 35 cycles (figure III.6-B). Using non-denaturing polyacrylamide gels is helpful for the re-amplification procedure because urea in denaturing polyacrylamide gels has been shown to inhibit the PCR (Konecy and Redinbaugh, 1996).

Step six: sequencing

Short sequencing data from excised DDRT-PCR fragments should indicate whether the cDNA is already known, or represents repeats or other contaminating

Bands No.	OP primer	*Tissue RT-PCR results (VG-OP)								Re-amplified specific band	Estimated Length (bp)	Direct sequencing results	Cloning results
		NT	PT	SP	L	A	M	B	P				
1	1	+	+	+	-	+	+	+	+	TG1	220	>1products	+
2	1	+	+	+	-	+	+	+	+	SG1	290	N.D.	+
3	2	+	+	+	+	+	+	+	+	T1G2	160	>1products	+
4	2	+	+	+	+	+	+	+	+	T2G2	140	N.D.	+
5	2	+	+	+	+	+	+	+	+	S1G2	200	>1products	+
6	3	+	+	+	+	+	-	+	+	T1G3	240	N.D.	N.D.
7	3	+	+	+	+	+	-	+	+	T2G3	250	270bp	N.D.
8	3	+	+	+	+	+	-	+	+	S1G3	120	N.D.	+
9	3	+	+	+	+	+	-	+	+	S2G3	110	N.D.	N.D.
10	4	+	+	+	+	+	+	+	+	T1G4	210	196bp	N.D.
11	4	+	+	+	+	+	+	+	+	T2G4	200	194bp	N.D.
12	5	+	+	+	-	-	+	+	-	T1G5	230	N.D.	N.D.
13	5	+	+	+	-	-	+	+	-	T2G5	210	N.D.	N.D.
14	5	+	+	+	-	-	+	+	-	T3G5	200	164bp	N.D.
15	5	+	+	+	-	-	+	+	-	T4G5	180	N.D.	N.D.
16	7	+	+	?	+	+	+	+	+	T1G7	400	N.D.	N.D.
17	7	+	+	?	+	+	+	+	+	T2G7	370	N.D.	N.D.
18	7	+	+	?	+	+	+	+	+	T3G7	360	N.D.	N.D.
19	7	+	+	?	+	+	+	+	+	T4G7	280	N.D.	N.D.
20	8	+	+	?	+	+	+	+	?	T1G8	250	>1products	+
21	8	+	+	?	+	+	+	+	?	T2G8	230	>1products	+
22	8	+	+	?	+	+	+	+	?	T3G8	170	>1products	+
23	9	+	+	+	+	?	+	?	+	TG6	250	N.D.	N.D.
24	9	+	+	+	+	?	+	?	+	S1G9	220	>1products	+
25	9	+	+	+	+	?	+	?	+	S2G9	200	205bp	N.D.
26	9	+	+	+	+	?	+	?	+	S3G9	130	>1products	N.D.
27	6	-	-	-	-	-	-	-	-	-	-	N.D.	N.D.
28	11	-	+	+	+	+	+	+	+	SG11	300	N.D.	N.D.
29	17	-	-	-	-	-	-	-	-	-	-	-	-
30	25	+	?	?	+	+	-	+	?	-	-	-	-
31	26	+	+	?	+	+	?	-	-	TG26	200	N.D.	N.D.
The tissue RT-PCR results columns show whether it was possible to include the tissue in the comparison. NT: Normal testis, PT: Patient's testis, Sp: Sperm, L: Liver, A: Adrenal, M: Muscle, B: Brain, P: Prostate Isolated bands were present only in normal testis (eg. TG1)or sperm (eg. SG1).													
OP: Arbitrary primer from OPERON Inc. VG: Anchored oligo(dT) primer (dT12VG)													
N.D.: Not done													
* + sign means the PCR has worked, – indicates that the PCR for that particular tissue has not worked and ? shows where the PCR result are not convincing.													

Table III.2: The results of DDRT-PCR. The table shows the estimated size of these DDRT-PCR products and whether the product was cloned. Complete sequence information appears in the Appendix, section A.2.2.

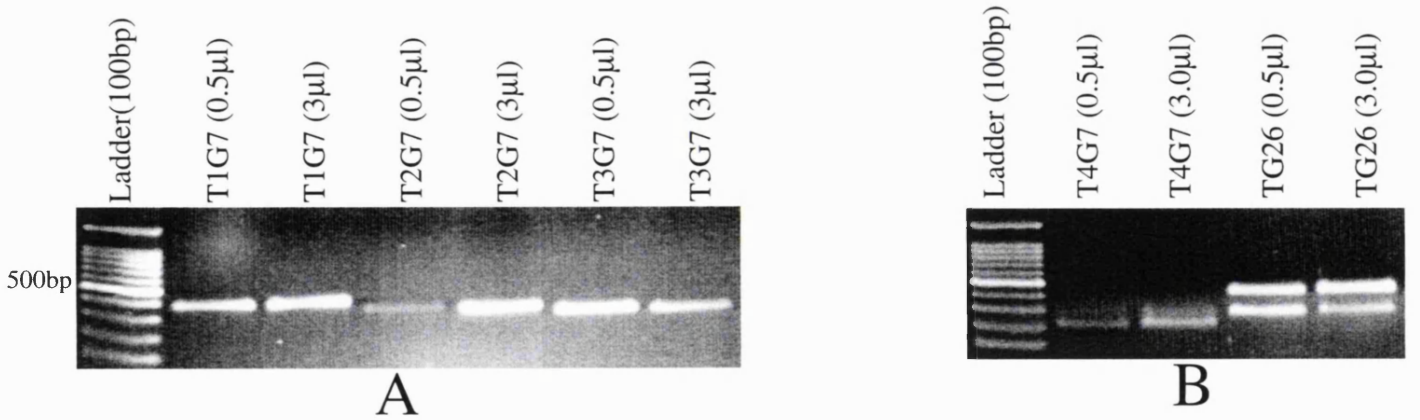


Figure III.6: A)Re-amplification of DDRT-PCR excised fragments using 0.5 and 3µl eluted bands. B) re-amplification reactions that resulted in more than one PCR product.

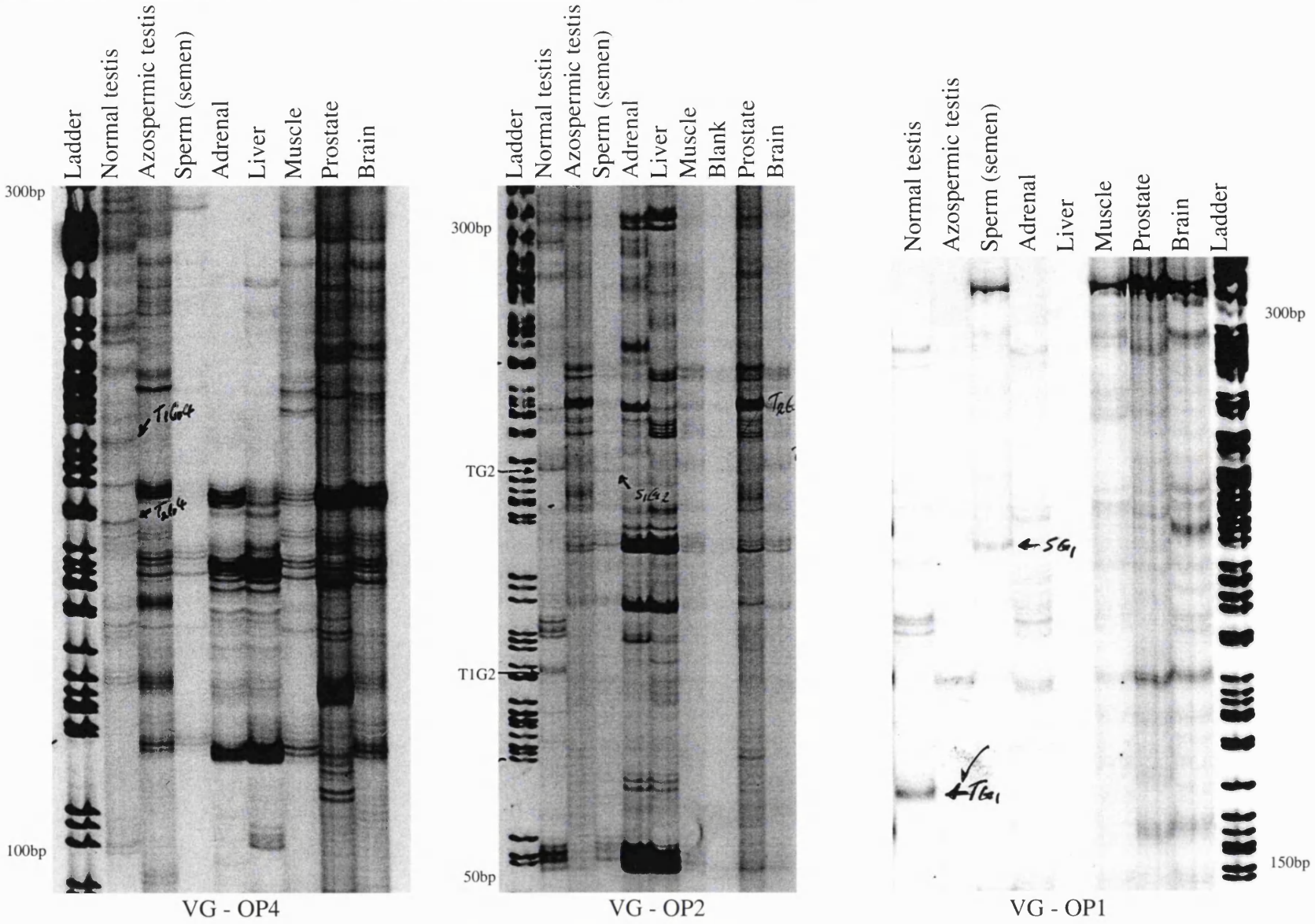


Figure III.7: DDRT-PCR products after PAGE. Various bands which were investigated are labelled on the autoradiograph. T2Gin brain lane of the middle panel refers to a band which is in a lane not shown.

genomic sequences. Sequencing can be carried out directly from the PCR amplified DNA using either the original anchor or arbitrary primer. Direct sequencing of the re-amplified band is the easiest way unless there is more than one cDNA product in the excised band (as a result of co-migration of DDRT-PCR fragments, or contamination from smaller products as described above, or error in the excising procedure of a single band). Table III.2 shows which of the DDRT-PCR fragments were sequenced directly from the excised band.

Step seven: cloning of the specific DDRT-PCR fragment

One of the easiest but time consuming methods for separating co-migrated bands is cloning. The eluted band (DDRT-PCR fragment) can be cloned into a bacterial vector and then either vector primers flanking the cloning site or the original anchor or arbitrary primers can be used for sequencing. It is, of course, necessary to sequence several of those clones to be certain of including the differentially expressed band.

III.3 Summary of the DDRT-PCR results:

The actual primer combinations used and sequence of bands excised from each differential display gel can be seen in the Appendix, section A.1. The nomenclature used for naming the DDRT-PCR fragments can also be found in section A.2.1.

Many DDRT-PCR reactions were carried out using VC and VG anchored primers with different combinations of arbitrary primers, 10-mers (OPs). Templates in these DDRT-PCR reactions were from the range of tissues explained earlier (table III.1). The main panel of this experiment was testis displayed against seven other tissues chosen as described earlier. 13 sets of reactions were carried out using VG anchored primer combined with different arbitrary primers (OPs) and the results are shown in table III.2. In addition, some differential display reactions were carried out using either VG or VC anchored primers for two other panels of RNAs: testis displayed against either a single other tissue (5 reactions) or against two other tissues (6 reactions). Some testis bands were isolated from these experiments (not

shown) but they were later found not to represent testis-specific cDNAs. The experiments using eight tissues, were more successful in generating testis specific cDNAs. Table III.2 shows that from the 13 set of DDRT-PCR reactions using eight different RNAs, 31 cDNA fragments were excised. All of those fragments appeared to be expressed solely or highly either in the testis or sperm (except one of them which was present in both tissues, S1G2 in figure III.7) and no similar sized bands were present in the lanes which contained DDRT-PCR products from other tissues. Some typical DDRT-PCR gels with testis displayed against seven other tissues, are shown in figure III.7. The bands are listed in table III.2. Almost all of the 31 excised bands in table III.2 could be re-amplified by PCR at the first or the second attempt with optimisation of PCR conditions. Of these, four were sequenced directly and 10 were sequenced indirectly after cloning. Nucleotide sequences were compared with those in the genome databases.

Step eight: Database searching results for DDRT-PCR products

The DDRT-PCR fragment sequences were compared with sequences in the Genbank database (http://www.ncbi.nlm.nih.gov/blast/blast.cgi), with specific searches targeting EST, High-Throughput Genome Sequences (HTGS) and non-redundant sequence databases (NR)(table III.3). Only the most significant matches with almost 100% identity have been mentioned.

Step nine: Novel gene

In this experiment, my aim was not to focus on the technique of DDRT-PCR. I was looking for novel DDRT-PCR fragment sequences to choose for further analysis and to find full-length cDNA and the complete gene as explained in figure III.1, steps eight to twelve. In addition, there were some DDRT-PCR fragments from a previous Ph.D. experiment (Cameron 1998) using testis, liver, lung, white blood cells, placenta, muscle and brain still left for further investigation. Six DDRT-PCR fragments were selected for confirmation of their testis-specificity.

DDRT-PCR fragment	bp used for database search	Databases searched	Homologies	Accession No.	Identity (similarity)	Notes
T1G4	196	NR EST htgs	No significant matches No significant matches No significant matches			
SG9	128	NR EST htgs	No significant matches No significant matches No significant matches			
SG2 (Colony 3&5)	197	EST NR htgs	No significant matches Human DNA sequence from clone RP3-347E1 on chromosome 6p11.2 21.1, complete sequence [Homo sapiens] Homo sapiens chromosome 6 clone RP11-124N8, WORKING DRAFT SEQUENCE, 16 unordered pieces	AL035670 AC073888	99% 99%	Gene: CYP39A1 Gene product: oxysterol 7alpha-hydroxylase
SG2 (colony1)	205	NR EST htgs	No significant matches No significant matches Homo sapiens chromosome 9 clone RP11-336P12, SEQUENCING IN PROGRESS	AL445186	99%	
T3G5	164	NR EST htgs	No significant matches No significant matches Homo sapiens chromosome 5 clone RP11-730N24 map 5, WORKING DRAFT SEQUENCE, 21 unordered pieces	AC025113	99%	It has a polyadenylation signal

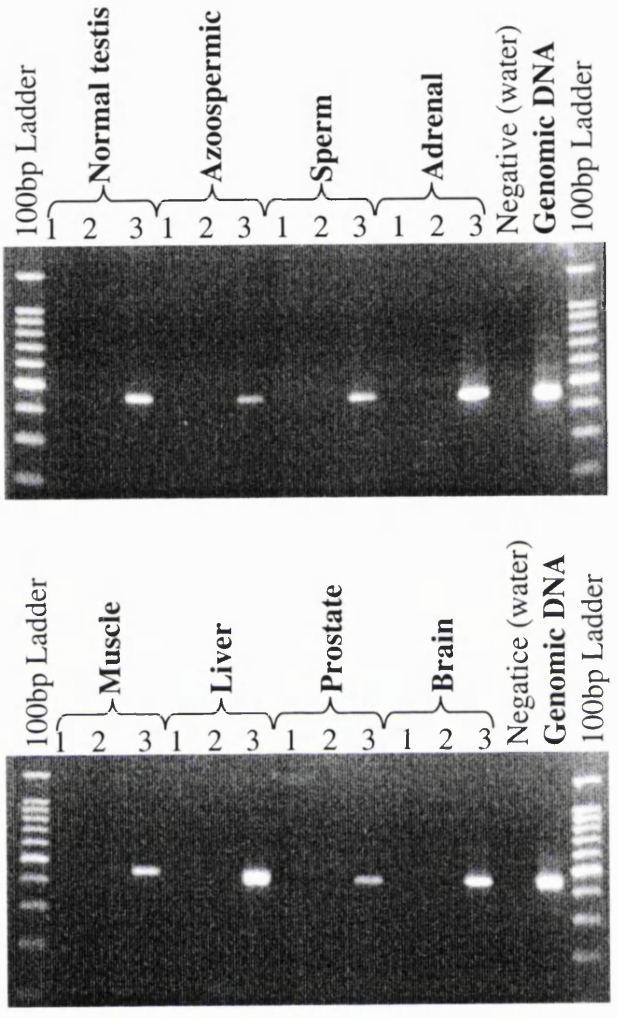
DDRT-PCR fragment	bp used for database search	Databases searched	Homologies	Accession No.	Identity (similarity)	Notes
TG2 (colony 26)	200	NR EST htgs	hereditary prostate cancer 1 (HPC1) No significant matches	AF172077		
TG2 (colony 23)	198	NR EST htgs	Many ESTs from testis, germ cell tumors and fetal tissues		100%	All matches from NR, ESTs and htgs databases discussed and shown in chapter 4
T2G4	194	NR EST htgs	No significant matches No significant matches No significant matches			30 bp of 3`end matched with many sequences
S2G9	205	NR EST htgs	No significant matches No significant matches Homo sapiens chromosome 9 clone RP11-336P12,SEQUENCING IN PROGRESS Homo sapiens chromosome 9 clone RP11-282P20 map q33.2-34.2, SEQUENCING IN PROGRESS	AL445186 AL162387	99%	
T2G3	270	NR EST htgs	No significant matches			All matches from NR, ESTs and htgs batabases discussed and shown in chapter 5

Table III.3: Results of Genbank database search for DDRT-PCR fragments.
****No significant matches means there is no match with identity more than 90% (to human) with high score (bits) (date of search January 2001).***

Step eleven: Confirmation of testis-specificity

The cDNA fragments generated by this technique must be tested by additional methods to ascertain whether they are expressed from other tissues in addition to the testis. It is usual to initially confirm expression patterns of the DDRT cDNAs using RT-PCR. This was performed on RNA from two normal testes, two different infertile patients' testes, sperm, adrenal gland, liver, muscle, prostate gland and brain. Following either cloning and sequencing or direct sequencing, specific primers were designed internal to the original set used for differential display. The usual criteria for primer selection were followed, utilising primer design software (such as the program at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi) to give primers that were generally 18-22 nucleotides long with 50-60% GC content. However, often DDRT-PCR fragments were too small and were very T/A rich, due to their location in the 3′ UTR region, and so it was not possible to choose primers with standard criteria. It is important to ensure that the primer regions selected do not have any identity to other sequences (at least known sequences) especially at 3′ end of the primer, therefore, the primer sequences selected were checked against databases during designing. RT-PCRs were then carried out using equal amounts of total RNA from the samples being compared under identical reaction conditions. The 3′UTR region is often intronless and so the PCR products from cDNA and genomic DNA are the same size. Therefore, RNAs or DNased RNAs were checked for genomic DNA contamination by two successive PCR amplifications (without a reverse transcription step) using primers designed from the gene *G3PDH* (sequence shown in Materials and Methods) which has processed pseudogene copies. PCR using the same *G3PDH* gene primers was also used to confirm the success of reverse transcriptase reactions (figure III.8).

Specific primers were designed for seven DDRT-PCR fragments that were thought to be novel or worthy of further study: TC10, T2G3, TG2, T1G4, T2G4, S2G9, TC13. The RT-PCR results from four cDNAs are shown in figure III.9. These RT-PCR results confirm that all of these cDNAs are expressed in the testis.



Templates:

1 = RNA or DNAsed RNA

2 = Re-amplification PCR from reaction (1)

3 = cDNA from RT reaction

Figure III.8: RNAs or DNAsed RNAs were checked for genomic DNA contamination by two successive amplifications using primers designed from the housekeeping gene G3PDH.

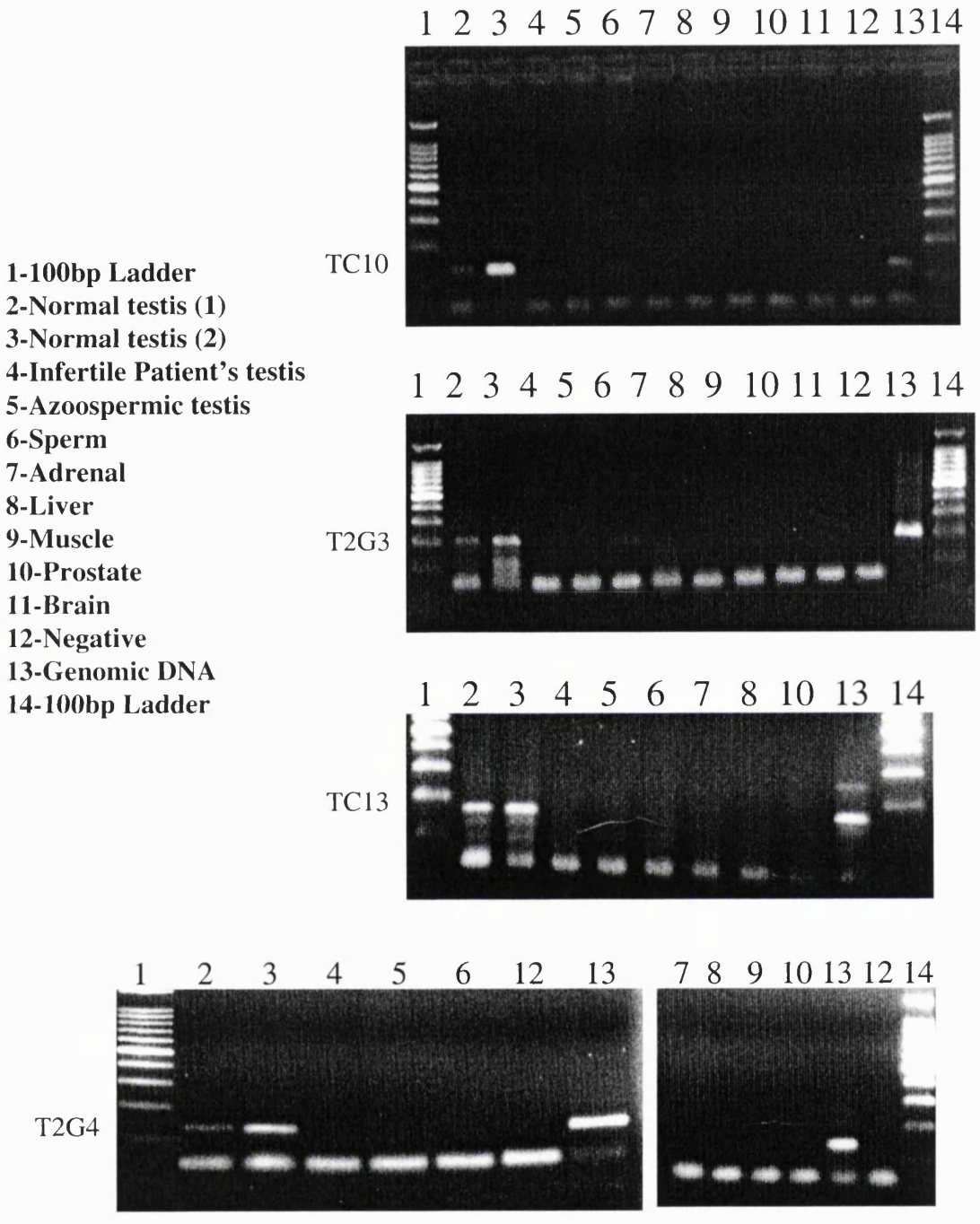


Figure III.9: RT-PCR results using specific primers designed from sequence located in DDRT-PCR fragments. All cDNAs used in this experiment were checked for genomic DNA contamination as shown in figure III.8.

No positive results were obtained for S2G9 and T1G4 in the RT-PCR experiments. Therefore these were believed to be false positive results. It is most likely that the original DDRT-PCR bands were amplification products from genomic DNA contaminating the testis RNA.

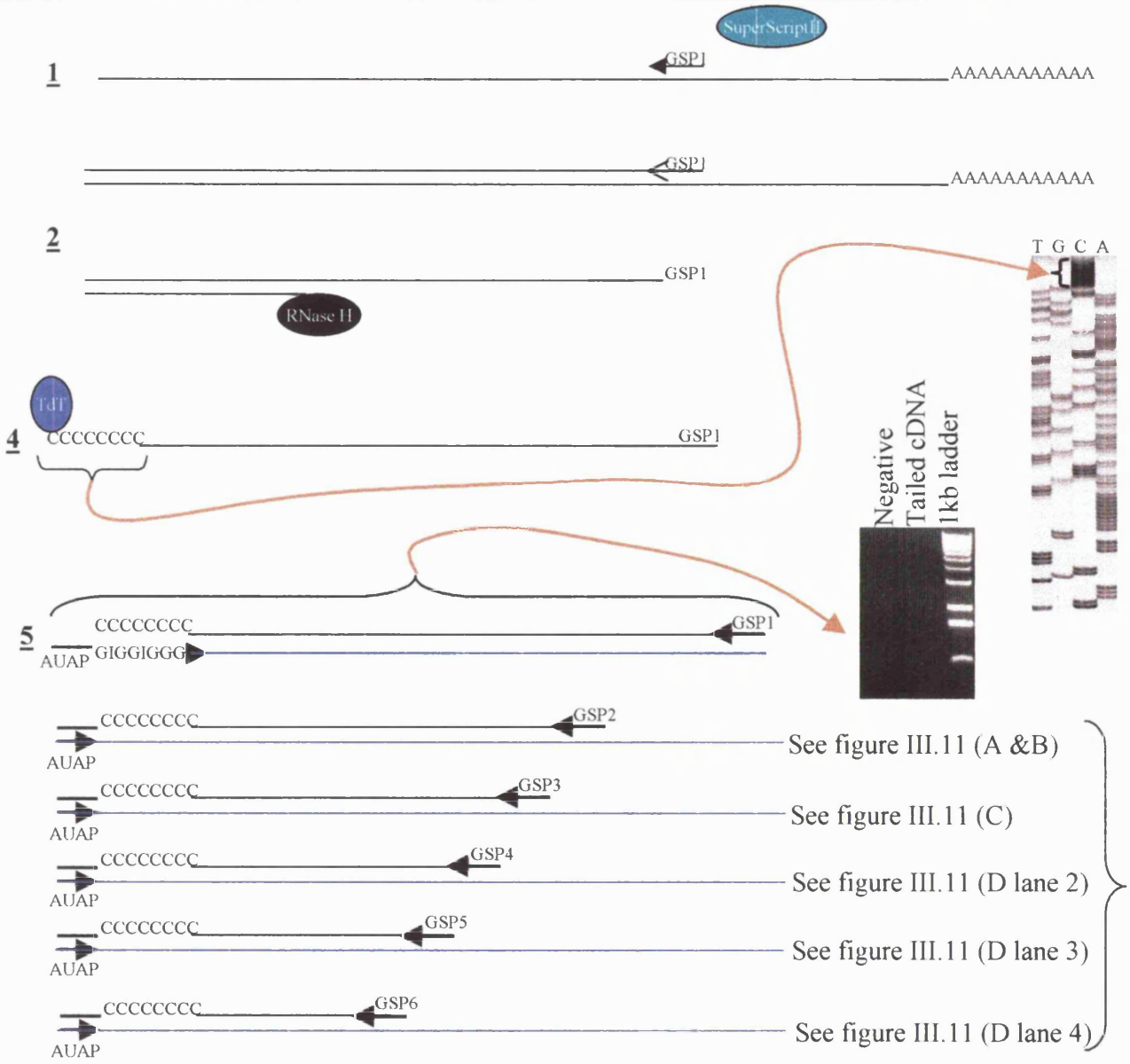
Northern blot analysis is another method which can be used to examine the expression pattern of the novel DDRT-PCR products. Using this method, not only would it be possible to confirm which tissues the DDRT-PCR cDNAs were expressed in, but the size of the transcript would also be known. However, I did not use this method at this stage to check the DDRT-PCR fragments because of problems in hybridisation of DDRT cDNAs to Northern blots. Firstly, DDRT-PCR fragments are always small, between 100-200bp, which is not sufficient to be a good probe; random-prime labelling of these cDNAs may generate even shorter probes. Secondly, it is possible that Northern blots are not sensitive enough for the detection of low abundance transcripts (Sambrook *et al.*, 1987) especially using DDRT-PCR products. Finally, unsuccessful attempts by a previous PhD student (Cameron 1998) made me postpone **step ten**, use of Northern, at this stage until I had obtained full-length cDNA. This is described after step 13.

Step twelve: Detection and localisation of the whole gene

This was carried out using the techniques in steps 13-17. Most of the results described in this chapter from here onward are from cDNA fragment TC10.

Step thirteen: 5′ Rapid Amplification of cDNA Ends (5′RACE)

To obtain full-length transcripts the strategy of 5′RACE was used (figure III.10). This technique facilitates the isolation and characterisation of 5′ ends from low copy messages (Frohman, 1993; and Loh, 1991). In general, 5′RACE procedures are efficient and effective. This is reflected in the number of times this technique appears in the literature; in the last year, 2000, there have been more than a thousand publications citing this technique. There are many applications of the technique: for example, it can be used to clone rare mRNAs that may escape conventional methodologies such as construction plus screening of cDNA libraries.



1- Specific cDNA synthesis using mRNA or total RNA and Gene Specific Primer 1(GSP1).

2- Single strand cDNA was prepared and mRNA was removed by using RNase H.

3- Single strand cDNA was cleaned and all dNTP and primers were removed from the reaction.

4- A polyC tail using dCTP and TdT enzyme was made (it is seen in the sequencing result of the full-length cDNA).

5- The first RACE-PCR using GSP1 and PolyG with an adapter primer (5′RACE abridged primer). No band was seen.

6- The second RACE-PCR using GSPs and the adapter primer (AUAP).

Figure III.10: 5′RACE procedure for obtaining full-length cDNA of TC10. Primers are shown as arrows.

Additionally, the RACE system may be applied to existing reliable cDNA libraries (rather than to mRNA preparations) so that the desired result is produced more easily and in less time (Bertchold, 1989). The RACE procedure can produce full-length transcripts from small fragments of sequence such as DDRT-PCR fragments or ESTs recovered from public databases by using both 3′ and 5′ RACE in combination. cDNA prepared by priming mRNA reverse transcription with random hexamers has been adapted to 5′RACE for amplification and cloning of multiple genes from a single reverse transcription reaction. An advantage of 5′RACE is that products can be sequenced directly without the need for intervening cloning steps (Frohman and Martin, 1989).

A problem can be associated with 5′RACE is the presence of incompletely spliced mRNAs in the pool of RNAs extracted from the tissue. For example, when the gene contains a small intron it can be present in 5′RACE amplified products.

The results obtained in this study during amplification at the 5′end of *TC10* mRNA highlight another problem. When 5′RACE was carried out using the testis RNA a smear of products ranging in size from very small to larger than 2kb was produced in the second PCR reaction (figure III.11). Sequencing of different parts of the smear demonstrated that the products were derived from the *TC10* transcript and were presumably the result of incomplete extension during the cDNA synthesis stage of the procedure.

Results of database searching showed that the TC10 DDRT-PCR fragment had homology with almost 100% identity to four ESTs which were all obtained from libraries in which human adult testis was either the sole tissue source or was one of the tissues used. The sequences of these ESTs were about 100 to 200bp longer than that obtained from DDRT-PCR. This confirmation that TC10 was expressed in testis was the reason why it was chosen for further analysis.

In an attempt to obtain full length cDNA of the DDRT 3′UTR fragment TC10, a 5′RACE experiment was carried out using human adult testis RNA and primers designed from the TC10 fragment. PCR results from the second amplification (second RACE-PCR) using different dilutions of the first RT-PCR reaction showed a smear with three bands in the background (figure III.11). Bands and smears were

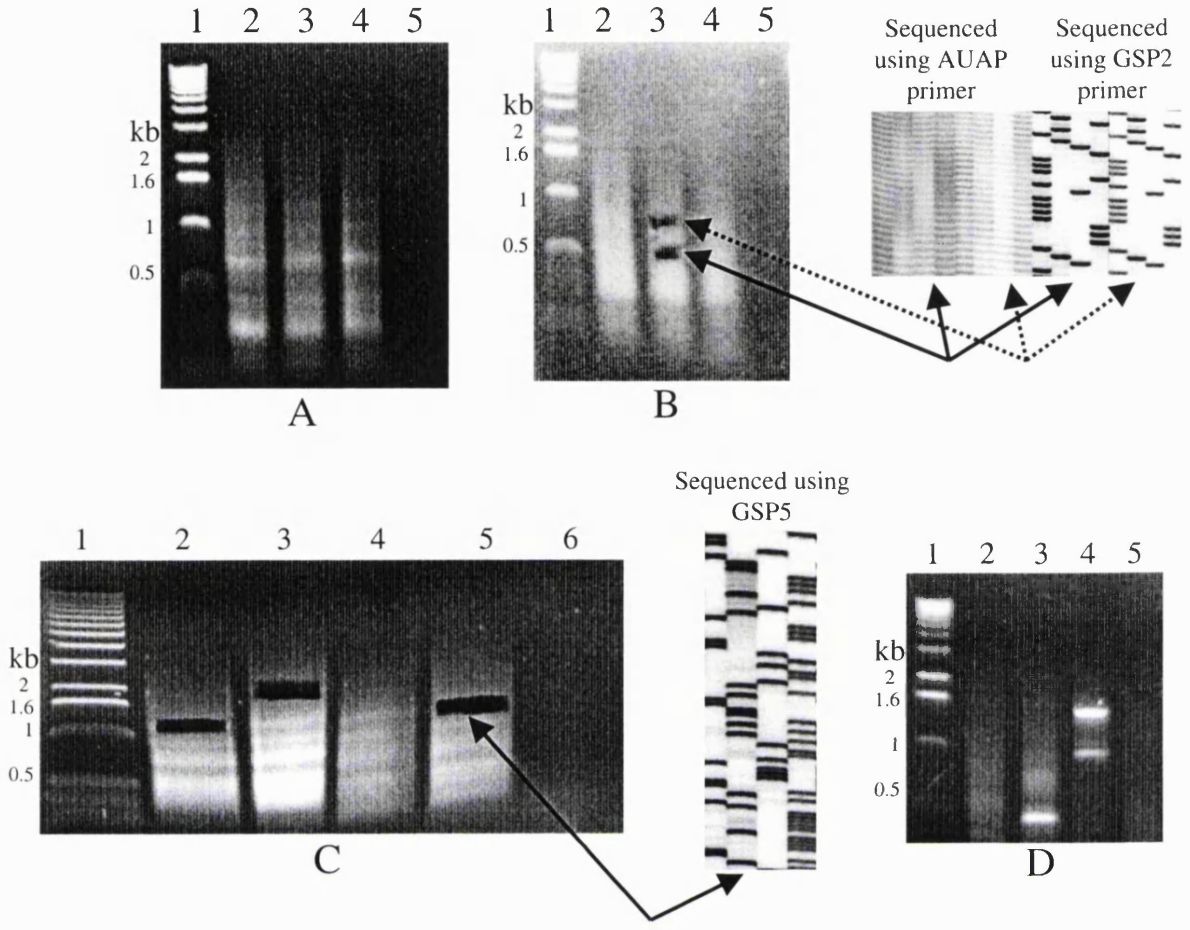


Figure III.11: TC10 5´RACE. (A) Second amplifications of 5´RACE were carried out using three different dilutions of the first PCR (with no visible band), 1/20 (2), 1/100 (3) and 1/500 (4) with a negative control (5). Two parts of the smear were cut out out from the 1.5% agarose gel and show as holes in the picture (B). Both, after DNA extraction and sequencing using GSP2 primer, showed that they contained only TC10 fragment and both gave the same sequence result. This experiment was carried out with five degree higher annealing temperature and two dilutions of the first PCR , 1/20 and 1/100 but results (C- 2 & 3) were very similar. Second RACE PCRs were also performed using other primers, GSP3 (C- 4 & 5), GSP4 (D- 2), GSP5 (D- 3) and GSP6 (D- 4). Last lanes are negative controls.

The band in (D- 3) is smaller than expected size but the large band from the second RACE PCR using GSP6 (lane D-4)was the expected size compared to the Rattus norvegicus homologue cDNA from databases.There was no result of the Northern at the time of this experiment but retrospectively it was also confirmed by the Northern. The band was excised for further sequencing.

In both the first and the second RACE PCRs the PCR primer at the 5´end of cDNA was AUAP primer.

excised and sequenced and the results showed the same sequences (figure III.11) but with ambiguities caused by a polyT region in the 3′UTR (figure III.25-B).

To sequence accurately the region 5′ to the poly T polymorphism a clone could have been used instead of the RT-PCR product. However, the easiest way was to design a primer from EST sequences with 100% identity in the sequenced region. Sequencing could not be carried out using the AUAP primer because of the many truncated cDNAs in the 5′RACE product. A typical sequence obtained using the AUAP primer is shown in figure III.11-B. The GSP primers were designed successively from the sequence data and used for second 5′RACE PCR. All gave smears in combination with the AUAP primer except GSP6 which gave a band of 1.6kb, consistent with later evidence concerning the full length of the mRNA. The entire *TC10* transcript was sequenced by primer walking along the cDNA template generated by 5′ RACE (using the primers GSP1-12 shown in figure III.22). It consists of 2994 nucleotides and contains an open reading frame of 698 amino acids (figure III.22).

During the 5′RACE experiments reported here it was noted for TC10 that the first round of amplification generated no visible product. The primer used was the same as that used in the RT reaction for cDNA generation. In contrast, when using a nested primer in the second round of amplification a visible product was always generated. However, it would be better to use three different nested primers for the three steps. This overcomes the problem that the first gene-specific primer is used in the reverse transcriptase step at a relatively low temperature of 42°C. This primer binds to a wide number of sites with the result that many mRNAs are amplified at low yield, very few of which correspond the correct product. The second nested primer finds and amplifies only the correct template whereas using the first gene-specific primer again simply leads to the amplification of many cDNA templates again at lower yields.

During sequencing of 5′RACE products one of the results had 144 nucleotides inserted (figure III.22) between nucleotides number 1192 and 1193, later found to be the boundary between exons 8 and 9. RT-PCR with primers GSP6 and R8 which span this boundary did not reveal any evidence of the longer product (figure III.12,

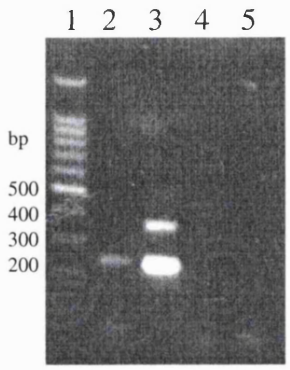


Figure III.12: RT-PCR using primers GSP6 and R8. Lane 1, marker, 100bp ladder. In lane 2 only one band was seen from testis cDNA but in lane 3 two bands using RACE PCR product as a template for PCR were seen. Lanes 4 and 5 are PCR products from genomic DNA and water (negative control), respectively. Both bands in lane 3 were excised and sequenced. The higher molecular weight band contains an insertion of 144bp at a position later shown to correspond to the boundary between exons 8 and 9 (exons and introns shown in figure III.22)

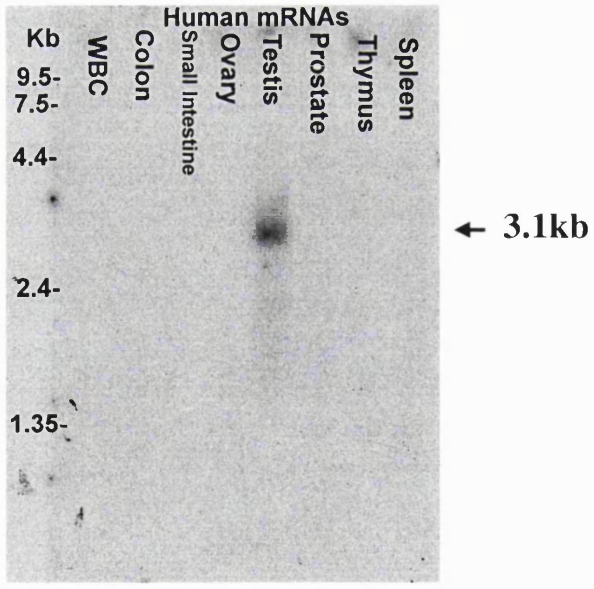


Figure III.14: The Northern confirmed the specific expression in testis.

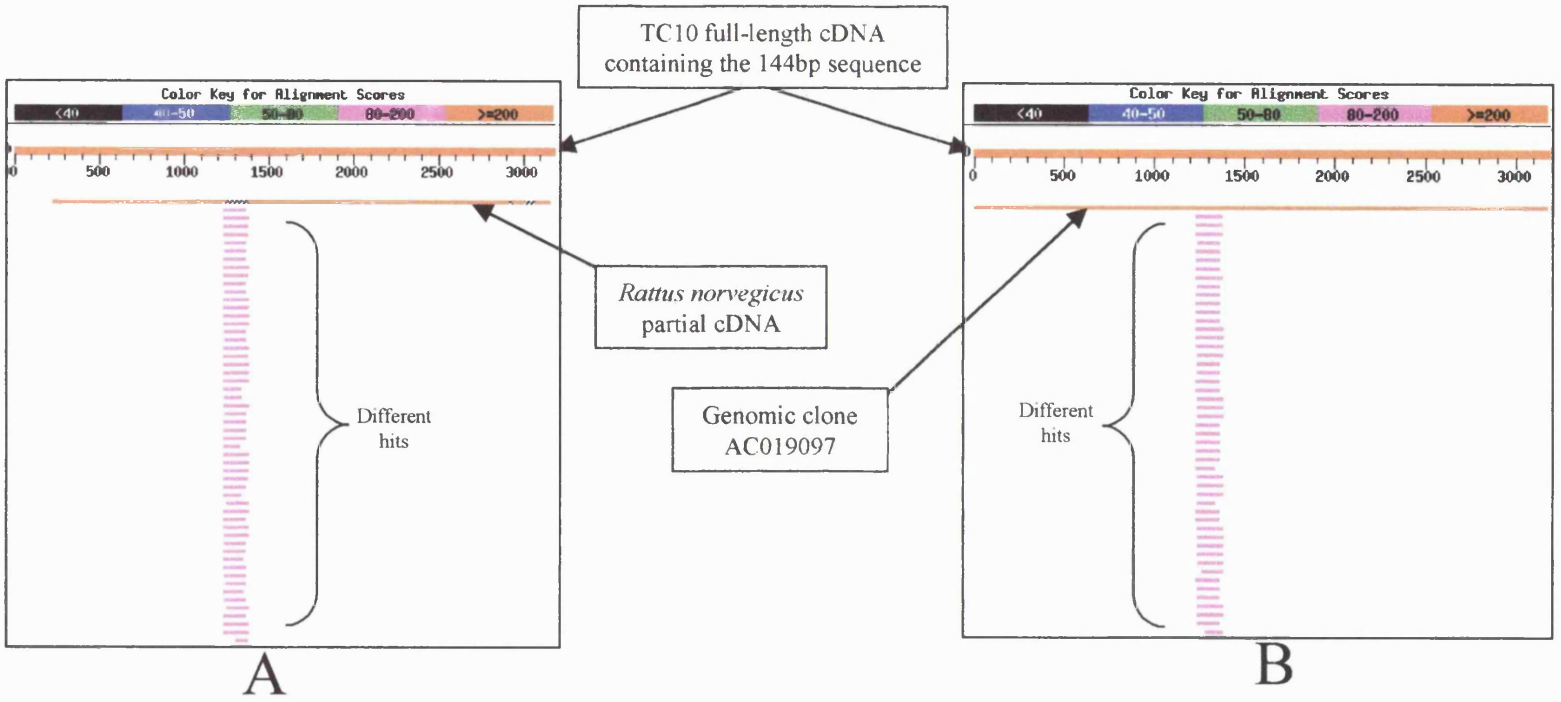


Figure III.13: Database search results. (A) from non-redundant db with many matches and (B) from htgs database with matches from a genomic clone, shows many matches, but only to the 144bp fragment which is inserted in 8th exon/intron boundary. The clone in the result (B) was not present in the databases at the time of the experiment.

lane 2). Retrospective database searches (figure III.13) showed that the 144bp insert is part of a repetitive element and is present in intron 8. The presence of the extra 144bp in a 5′RACE product was probably result of a rare incorrectly spliced mRNA. It is discussed more fully in section III.13.5.

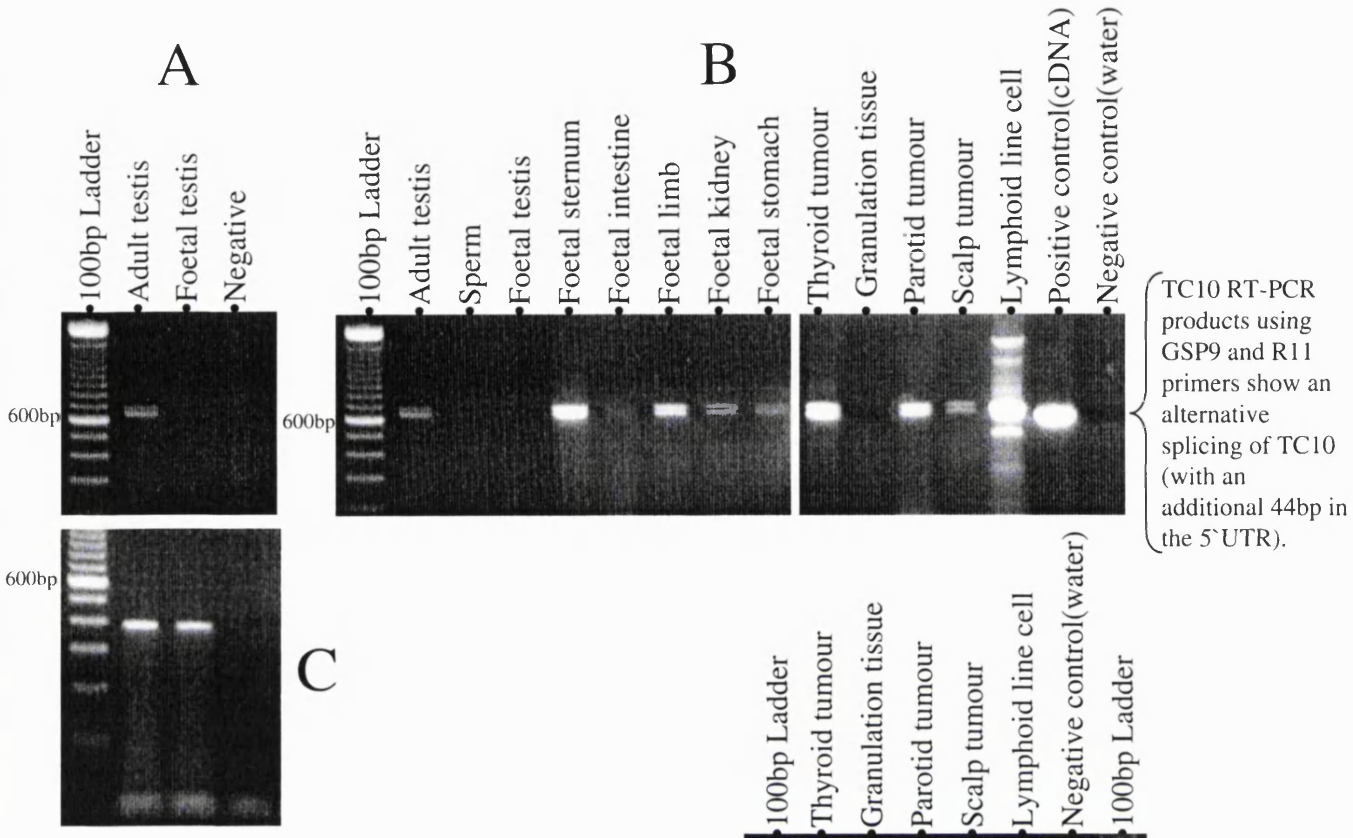
Step ten: Northern blot analysis to confirm the size and testis-specificity of *TC10 transcript.*

Northern blotting using the GSP3-R6 RT-PCR product as a probe revealed a single band of about 3.2 kb in testis which was not present in RNA from white blood cells, ovary, spleen, small intestine, colon, prostate gland or thymus (figure III.14). This mRNA size correlates with that predicted from the cDNA (assuming a polyA tail of about 100 nucleotides (Brawerman, 1981)). The Northern blot was separately hybridized with a β-actin control probe to confirm that all tracks contained intact mRNA (result not shown).

Step eleven: Expression of *TC10* in other tissues and alternative splicing

Using a combination of different forward and reverse primers, (shown in figure III.22), RT-PCR demonstrated expression of the gene and of an alternatively spliced mRNA in adult testis but not in foetal testis (figure III.15). The site of alternative splicing was located between GSP9 and R11 primers. Both RT-PCR product bands were cut out from an agarose gel and sequenced. Results revealed the incorporation of an additional 44 nucleotides (5′-ataattttagatactgaaaaagcacaaaataaatctccttctag-3′). This sequence also corresponded to the sequences obtained from the 5′RACE procedure and was later also found in the expected position in the genomic clone AC019097.

RT-PCR amplification of cDNA made from some foetal (16-20 weeks) and tumour tissues using GSP9 and R11 primers shows differing levels of expression with both alternative splice products present in almost all of the actively dividing tissues which were available; sternum, intestine, limb, kidney stomach, and also thyroid, parotid and scalp (squamous cell carcinoma) tumours (figure III.15). Neither strong band was present in foetal testis or granulation tissue.



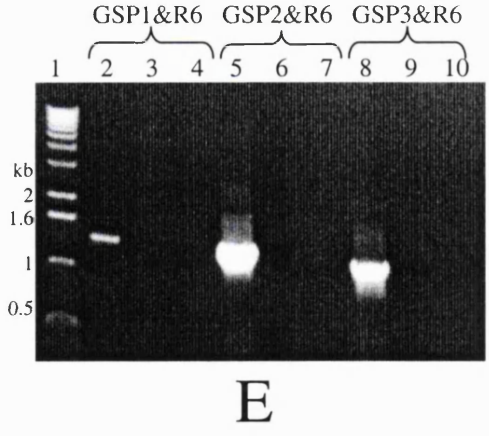


Figure III.15: (A) RT-PCR demonstrated expression of the gene and of the alternatively spliced mRNA only in adult testis. GSP9 and R11 primers (which span three introns and do not amplify from genomic DNA) were used to amplify cDNA from (lane 2) adult testis and (lane 3) foetal testis. Lane 1 is 100 bp markers and lane 4 is the negative (water) control. Expression of this gene has also been shown in some foetal and abnormal tissues as well as in a lymphoid cell line(B).
(C & D) Control PCR reactions were carried out with PGM1 primers to demonstrate the presence of cDNA in the samples prepared from adult testis (C-lane 2), foetal testis (C-lane 3) and other adult tissues (D). These primers span a large intron and do not amplify from genomic DNA. RNA from thyroid, Parotid and Scalp tumours and also granulation tissue have been checked for DNA and shown in figure III.2.

The gel (E) is adult testis RT-PCR results using R6 and GSP1 (2, 3 & 4), GSP2 (5,6 & 7) and GSP3 (8,9 & 10). Lanes 2, 5 & 8 were from cDNA, lanes 3, 6 & 9 were from genomic DNA and lanes 4, 7 & 10 were negative control.

In addition, expression was seen in RNA extracted from a lymphoid line cell culture using the same primers as above (figure III.15). About five additional (weaker) bands were seen in the products from the lymphoid cell line. These have not yet been further investigated.

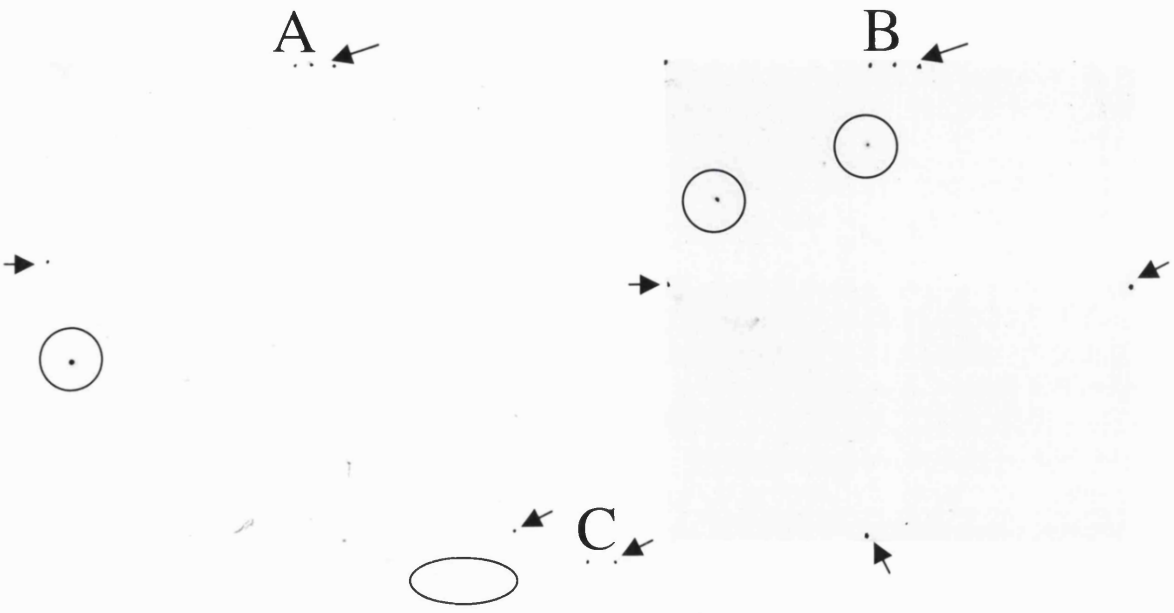
Step fourteen: Isolation of cDNA clones from an adult human testis cDNA library

Approximately 2-3 x 10^5 p.f.u of an adult testis cDNA library were plated out in duplicate and screened in an attempt to obtain full-length cDNA. Three cDNA clones were identified by hybridisation of the GSP5-R8 and GSP9-R12 PCR products to the cDNA library filters (made as described in Methods section II.2.8). Figure III.16 shows plaques from primary, secondary and tertiary screening of the cDNA library. Those cDNA clones were isolated and sequenced. None was full-length but together they covered almost all the sequence derived by 5'RACE.

Step fifteen: Screening PAC/BAC genomic filters to obtain genomic clones

Three BACs were identified by hybridisation of the GSP5-R8 and GSP9-R12 PCR products (676bp and 817bp, respectively, see figure III.22) to commercial BAC library filters, containing a gridded BAC library of 125,000 clones (figure III.17). All three BAC clones were obtained from Research Genetics Inc. and were confirmed by PCR which also showed that BAC R380D5 contains both ends of the gene. BAC 225M7 and BAC 120I13 contain only the 3´ region of the gene (figure III.18).

Searching the BAC end sequence database at the Telethon Institute for Genome Research (Tigr) (http://www.tigr.org/tdb/humgen/humgen.html) with the complete cDNA sequence, revealed a BAC end (CIT-HSP-2166P19.TF) with 100% homology over 108bp before interruption by an intron. Clone 2166P19 was also obtained from Research Genetics Inc..



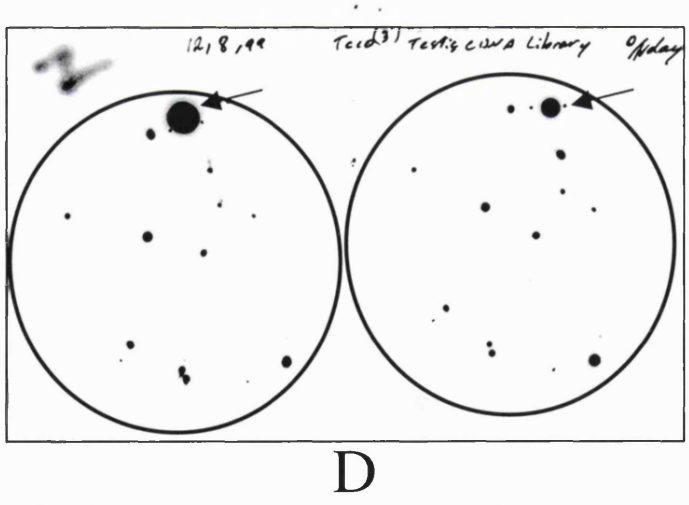


Figure III.16: Adult testis cDNA library screening results, primary (A, B & C) and secondary (D) screening. The secondary screening results in duplicate are from the signal in the filter A. Real signals (comparing to their duplicate) in the first screening are indicated by circles and oval. Radioactive dots (shown by arrows) were marked to evaluate the hybridisation procedure and are not real signals.

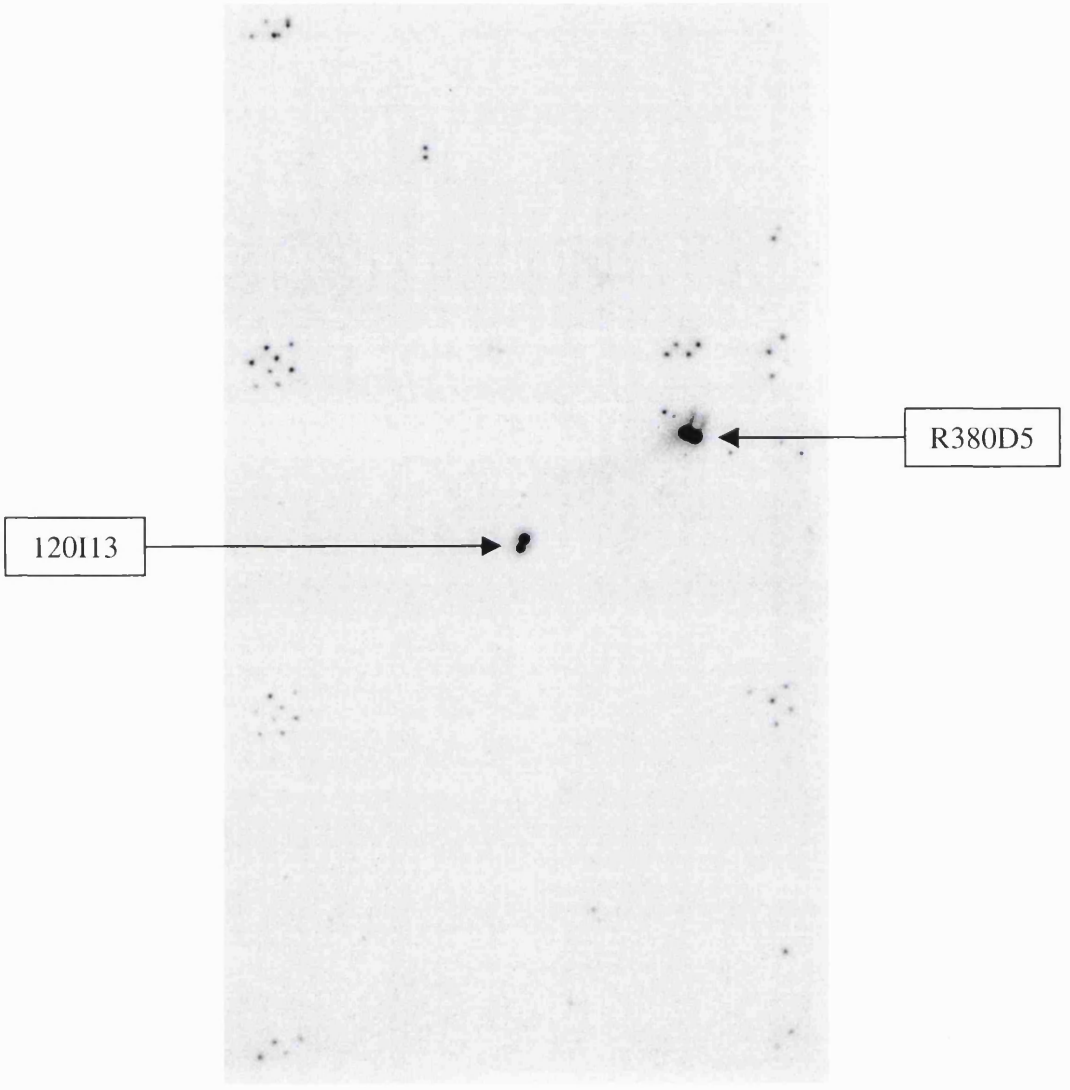


Figure III.17: One out of 6 genomic (BAC) filters screened with radioactive labelled probes. This filter contains two out of three positive colonies. Each clone on the filter is represented by two (duplicate) signals and other weaker signals are from markers and another unrelated probe which was hybridised to this filter at the same time. BAC R380D5 contains both ends of the gene and BAC 120I13 contains part of the gene. The intensity of the signals reflect the proportion of the gene presented in the clones.

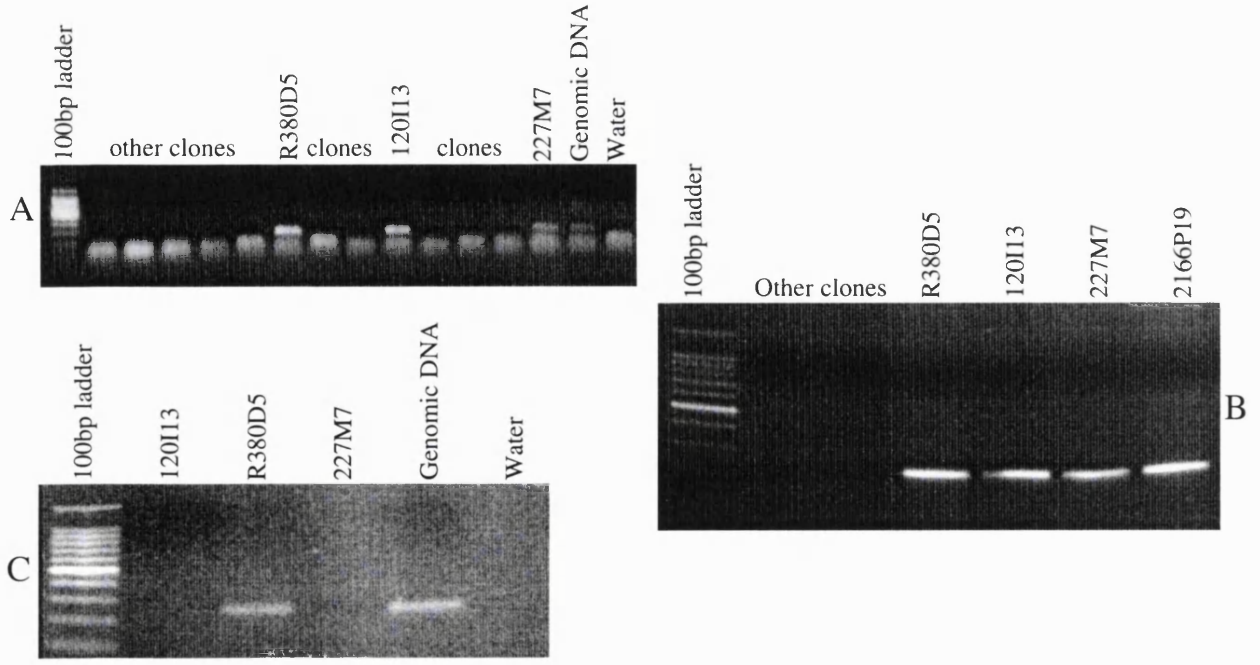


Figure III.18: PCR results show four BAC clones (three from genomic library screening and one from database search) contain some or all of the* TC10 *gene, R380D5, 120I13, 227M7 and 2166P19. PCR primers are (A) GSP1&R1 located in the 3´UTR, (B) TC10-IF4&R4 located within the coding region and (C) GSP11&R12 located in the 5´UTR.

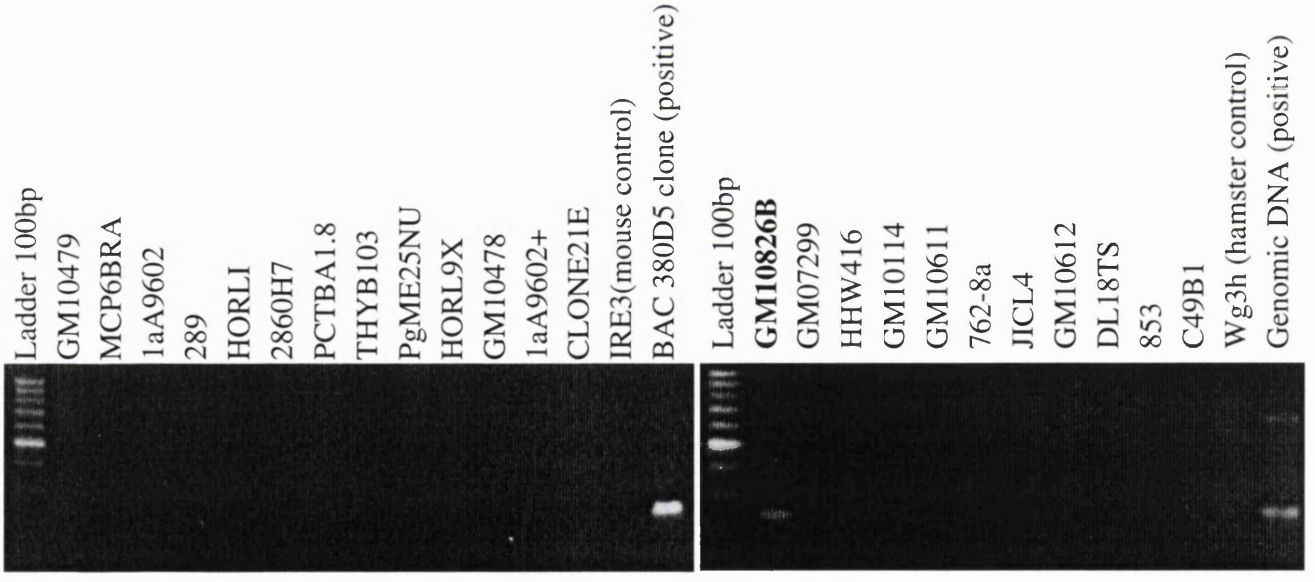


Figure III.19: The chromosomal location of* TC10 *was determined by PCR amplification using gene specific primers, IF4 (located in an intron) and R4 (present in an exon) on a panel of hamster or mouse cell hybrids that each contain a single human chromosome.
The PCR product (253bp) is present in the cell hybrid containing only human chromosome 2, GM10826B, and no product was obtained from hybrids containing other human chromosomes.

Steps sixteen and seventeen: Chromosomal localisation of the gene, *TC10*

The chromosomal location of the *TC10* gene was determined by two methods. Firstly by PCR amplification using gene specific primers, (IF4, located in an intron and designed from the BAC published clone 2166P19 end sequence, and R4 present in an exon) of a panel of rodent somatic cell hybrids that each contain a single human chromosome (table III.4). The primers were selected in order to avoid amplifying products from mouse and hamster genomic DNA or from any intronless pseudogene. I obtained a specific 253bp PCR product from the cell hybrid containing only human chromosome 2 and no product was obtained from hybrids containing other human chromosomes (figure III.19). Secondly, to refine the localisation of the *TC10* gene, BAC R380D5 which contained the whole gene was hybridised in situ (FISH) to human metaphase chromosomes from a fertile man. A specific fluorescent signal was observed at chromosome 2q11.2 (figure III.20).

III.4 Database searching results for the human testis-specific gene, *TC10*

To confirm the novelty, to find out about homologues and more about the expression pattern of the transcript, the whole cDNA sequence of *TC10* obtained from 5`RACE, 3038bp (with 44bp, the alternative splicing sequence) was used for searching EST databases containing ESTs from human and other species, non-redundant databases and even Unfinished High Throughput Genomic Sequences (htgs) which contains the most recent human genome project genomic sequences. A summary of the results is shown in table III.5.

III.4.1 EST database searching results

dbEST (Database of GenBank+EMBL+DDBJ sequences) (Boguski *et al.*, 1993) is a division of GenBank that contains sequence data and other information on "single-pass" cDNA sequences, or Expressed Sequence Tags, from a number of organisms. A search of EST databases (NBLAST at the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/blast/blast.cgi)) with the *TC10* sequence identified 13 ESTs with 95-99% homology from human and 11 ESTs with 84-95% homology from other species including mouse, rat,

	Chromosome																							TC10	TG2	T2G3
Hybrid DNA	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**X**	**Y**			
GM07299	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
GM10A26B	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-
GM10253	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HHW416	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
GM10114	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
MCP6BRA	-	-	-	-	-	/	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	/	-	-	-	-
CLONE21E	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
C4A	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-*	-*	-*
GM10611	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
762-8a	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
JICL4	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1Aa9602+	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-
289	-	-	-	-	-	-	-	/	-	-	/	/	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
GM10479	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	/	-	-	-	-	-	-	-	-	-	+	-
HORLI	-	-	-	-	-	-	-	-	-	-	/	-	-	-	+	-	-	-	-	-	-	-	/	-	-	-	-
2860H7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
PCTBA1.8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-
DL18TS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-
GM10612	-	-	-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
GM10478	-	-	-	/	-	-	-	/	-	-	-	-	-	-	-	-	-	-	-	+	-	/	+	-	-	-	-
THYB1.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-
PgME25NU	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	/	-	-	-	-
HORL9X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
853	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-
IRE3	Mouse control																								-	-	-
Wg3h	Hamster control																								-	-	-

Table III.4: *Somatic cell hybrid PCR results for the 3 transcripts TC10,TG2 and T2G3 are shown in the last 3columns. a + or - (in colour) denotes whether a PCR product of the correct size was seen. TG2 and T2G3 are discussed in chapter 4 and 5, respectively. The presence or absence of a human chromosome in a cell hybrid is denoted by a + or -. /, indicates chromosome translocation, extra chromosomes or other modifications. GM10478 20, 4(PART), 8(PART), 22q, X*

** Not used in this experiment.*

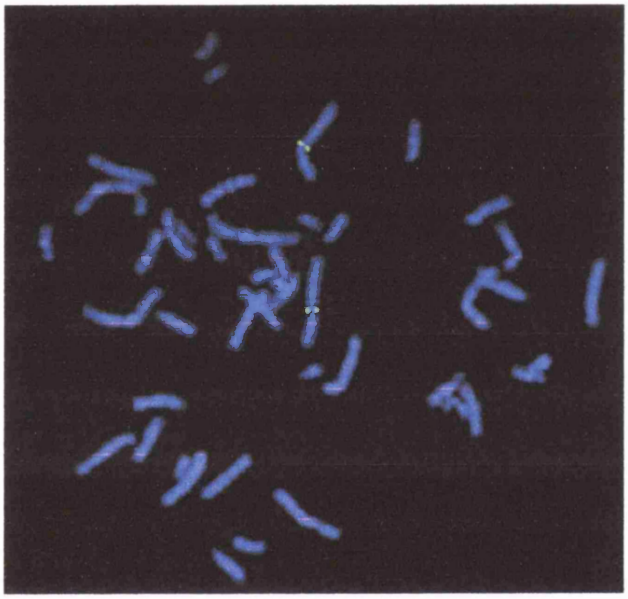


Figure III.20: Localisation of BAC R380D5 containing the entire gene TC10 to band 2q11.2 by FISH.

bovine and pig. Three ESTs were from human testis (Accession nos. AI024201, AA625993 and AA868661) and five ESTs from three normalised libraries all partially derived from human testis (Accession nos. AW269043, AI807982, AI240099 and AI221135). The sequences all matched the 3`UTR region of *TC10* with almost 100% identity. In addition, one testis EST (Accession no. AA393655) matched to the region of 1310bp to 1550bp but this EST appeared to be chimeric. It was not confirmed by RT-PCR amplifications and sequencing using a primer, R7 (data not shown), designed from an end of the EST which had no similarity to *TC10,* and either GSP2 or GSP3. Three more human ESTs (Accession no. AW057728, BE047007 and BF243403) isolated from poorly differentiated adenocarcinoma, acute myelogenous leukaemia and germ cell tumours were identical to the *TC10* transcript between nucleotides 512 & 1041, 229 & 788 and 2378 & 2981, respectively (table III.5 and figure III.21).

There are eleven ESTs from other species with approximately 90% identity to *TC10* (table III.5 and figure III.21).

III.4.2 Non-redundant database searching results

This database (NR) in NCBI contains all GenBank+EMBL (European Molecular Biology Laboratory) +DDBJ (DNA Data Bank of Japan) + PDB (Sequences derived from the 3-dimensional structure from Protein Data Bank) sequences but no EST (Expressed Sequence Tags), STS (Sequence Tagged Sites), GSS (Genome Survey Sequence, includes single-pass genomic data, exon-trapped sequences, and Alu PCR sequences), or phase 0, 1 or 2 HTGS (High-Throughput Genome Sequences) sequences. It seems all gene sequences and completed clones which have been reported to date are present in this database. Searching the non-redundant nucleic acid sequence databases at NCBI gave a match of 90% identity to a partial cDNA from spermatocyte of *Rattus norvegicus* (Accession no. AF092091) to *TC10* (figure III.21). This sequence was unpublished sequence and no more information could be obtained from the database.

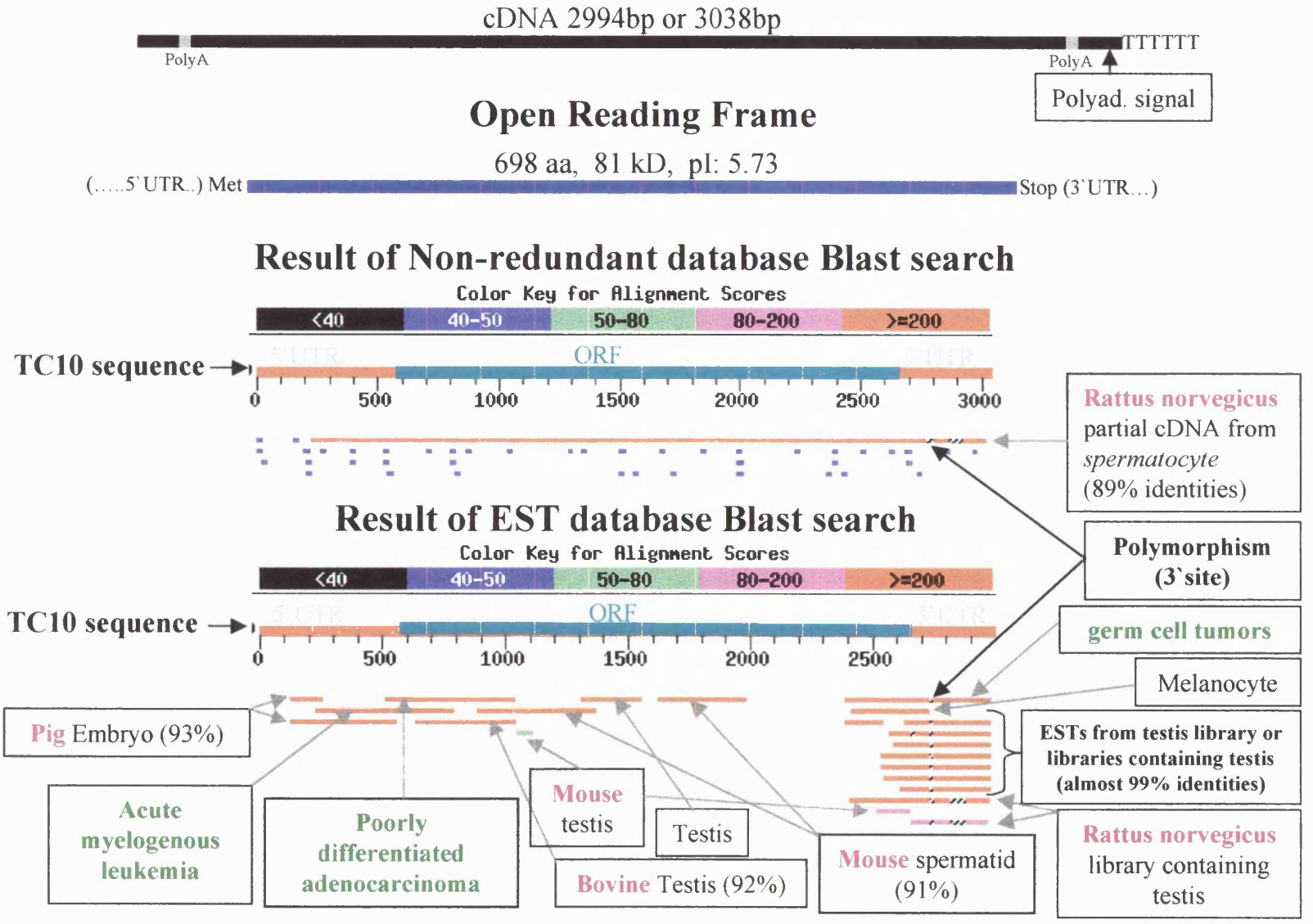


Figure III.21: schematic of **TC10** *gene and search results.*

Accession No.	Database searched	bp matched	Species	Source	Identity (similarity)	Notes
BE047007	EST	600/605	Human	Pooled germ cell tumors	99%	
N33757	EST	325/326	Human	Melanocyte	99%	
AI024201	EST	250/358	Human	**Testis**	97%	
AI799092	EST	156/157	Human	Three normalized libraries (fetal lung NbHL19W, **testis** NHT, and B-cell NCI_CGAP_GCB1)	99%	3' similar to TR:O75130 O75130 KIAA0635 PROTEIN
AW269043	EST	382/401	Human	Three normalized libraries (fetal lung NbHL19W, **testis** NHT, and B-cell NCI_CGAP_GCB1)	97%	
AA625993	EST	392/400	Human	**Testis**	98%	
AI807982	EST	443/454	Human	Three normalized libraries (fetal lung NbHL19W, **testis** NHT, and B-cell NCI_CGAP_GCB1)	97%	
AI240099	EST	437/445	Human	Three normalized libraries (fetal lung NbHL19W, **testis** NHT, and B-cell NCI_CGAP_GCB1)	98%	
AA868661	EST	427/436	Human	**Testis**	97%	

Accession No.	Database searched	bp matched	Species	Source	Identity (similarity)	Notes
AI221135	EST	363/371	Human	Three normalized libraries (fetal lung NbHL19W, **testis** NHT, and B-cell NCI_CGAP_GCB1)	97%	
AA393655	EST	241/241	Human	**Testis**	100%	478bp chimerical
BE102667	EST	452/524	Rattus norvegicus	A subtracted library derived from a mixture of the following tissues: hippocampus, thalamus, mid-brain, medulla, corpus striatum, cerebral cortex and **testis**	86%	
BF319562	EST	327/366	Mouse	**Testis**, round spermatids, pooled from multiple mice	89%	3' similar to TR:Q9Z220 Q9Z220 CP431
BF018459	EST	446/482	Mouse	**Testis**, round spermatids, pooled from multiple mice	92%	5' similar to TR:Q9Z220 Q9Z220 CP431
AW057728	EST	539/531	Human	Stomach, poorly differentiated adenocarcinoma	99%	3' similar to TR:O75130 O75130 KIAA0635 PROTEIN
BE233763	EST	127/133	Pig	Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos	95%	
BF243403	EST	541/571	Human	Bone marrow, from acute myelogenous leukemia	95%	

Accession No.	Database searched	bp matched	Species	Source	Identity (similarity)	Notes
BE233756	EST	366/388	Pig	Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos	94%	
BE751489	EST	379/410	Bovine	Library made from pooled tissue from **testis**, thymus, semitendonosus muscle, longissimus muscle, pancreas, adrenal , and endometrium	92%	
AW528471	EST	233/275	Rattus norvegicus	A subtracted library derived from a mixture of the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum and hippocampus	85%	
AV283000	EST	108/128	Mouse	RIKEN mouse **testis**	84%	
BB249130	EST	57/68	Mouse	RIKEN mouse Cerebellum, 7 days neonate	85%	
4933432N21	EST	108/128	Mouse	RIKEN mouse	84%	*
AC019097	htgs	2981/2981	Human	Genomic clone RP11-425I16, WORKING DRAFT SEQUENCE, 4 unordered pieces, Length = 185850	99%	Genomic clone localised on chromosome 2
AF092091	nr		Rattus norvegicus	**Testis**, spermatocyte	90%	

Table III.5: Result of databases Blast searches from the site: http://www.ncbi.nlm.nih.gov/blast/blast.cgi.
****This result is from another local database.***

III.4.3 Unfinished High-Throughput Genome Sequences (HTGS) and recently released database searching results

Unfinished High Throughput Genomic Sequences are phases 0, 1 and 2 (finished, phase 3 HTG sequences are in NR). Searching the recently released sequence database at NCBI revealed a working draft sequence, an incomplete clone with 4 unsorted contigs (Accession no. AC019097) which contains the *TC10* gene. The clone at the time of this experiment was unlocated but recently it has been localised to chromosome two (GenBank). Together with the information from the BAC end sequence, the gene can be seen to be composed of 19 exons (figures III.22) and to contain an alternative splice site at the 5′ end of exon 4.

III.5 Mapping of the gene

The draft browser program at http://genome.ucsc.edu/goldenPath/septTracks.html was used to examine the BAC genomic clone with accession no. AC019097 (containing the entire *TC10* gene). It places the clone at chromosome position 2q11.2. In addition, this program can search all other databases which are available for public use (in a similar way to the NIX program, see chapter 1, part IV) and display the results (figure III.23). Another program, with a few differences, at http://www.ensembl.org/ can also be used for this purpose.

III.6 TC10 protein prediction and similarity analysis

The predicted protein sequence of *TC10* (estimated molecular mass of 81 kDa and pI of 5.73) was used to search databases using the PBLAST program at NCBI. The search showed 92% identity to the predicted protein translated from the *Rattus norvegicus* cDNA mentioned above (section III.7.2). It also gave 40% identity to a human protein of unknown function (KIAA0635) predicted from a brain cDNA sequence (Accession no. BAA31610) which is located on chromosome 4 (q12). The search result is shown in figure III.24.

A)

```
 1-caggtatctaacctatctttgtatttttggaatttaatatagttcctgacataatcattg     -R12
 61-ttcaatgggtctaaTTTTTTTTTcctttaaaggcatttggataccttttgtgatagaaaag
121-gccctttttcttcatccattgatggtttgatagtgggctgggaaggaaagctgtgttcctc     -GSP12
181-cacattaggcagcaaatacttgattgatttgatatttagcttagtcaattgaagatctc      -GSP11
241-aaacagacaaatcatggcttggaagaatgtgttaggaaactcttggatagtaaggaggtg     -R11
301-gtaagcagtcaagtagatgatttaaccagccacaatgagcatctttgtaaagaattgatt
361-aaaattgaccaactagcagagcaactcgaaaaagagaaaaattttgtggtggattccgcc
421-aacaaggaacttgaagaagccaagattgatctcatttgccagcaaaataatattatagta     -R11/SPL
481-ttggaagatacaataaaaaggcttaaatcttacttgattcctttgtcaagactttggaagc
541-agacaaagatcactataagagtgaagctcaacatttgagaaagatgatgcgaagtaggtc     -GSP10
                                                    M  M  R  S  R  S
601-taaaagtccaagacgcccatcaccaactgcccgggtgcaaactgtgatgtagaactttt     -R10
     K  S  P  R  R  P  S  P  T  A  R  G  A  N  C  D  V  E  L  L
661-gaagacaacaacaagagatcgtgaagaacttaaatgcatgctggaaaaatatgagcgcca
     K  T  T  T  R  D  R  E  E  L  K  C  M  L  E  K  Y  E  R  H
721-tttggcagaaattcagggtaatgtcaaggttcttaaatctgagagagacaagatcttcct
     L  A  E  I  Q  G  N  V  K  V  L  K  S  E  R  D  K  I  F  L
781-tctttatgaacaggcacaggaagaaattacccgacttcgacgagaaatgatgaaaagctg
     L  Y  E  Q  A  Q  E  E  I  T  R  L  R  R  E  M  M  K  S  C
841-taagagtcctaaatcaacaacggcacatgctattctccggcgagtggagactgaaagaga     -GSP9
     K  S  P  K  S  T  T  A  H  A  I  L  R  R  V  E  T  E  R  D
901-tgtagcctttactgatttacgaagaatgaccacagaacgagatagtctaagggagaggct
     V  A  F  T  D  L  R  R  M  T  T  E  R  D  S  L  R  E  R  L
961-aaagattgctcaagagacagcatttaatgagaaggctcacctggaacaaaggatagagga
     K  I  A  Q  E  T  A  F  N  E  K  A  H  L  E  Q  R  I  E  E
1021-gctggagtgtacagttcataatcttgatgatgaacgtatggagcaaatgtcaaatatgac
     L  E  C  T  V  H  N  L  D  D  E  R  M  E  Q  M  S  N  M  T
1081-tttgatgaaggaaaccataagcactgtggaaaaagaaatgaaatcactagcaagaaaggc
     L  M  K  E  T  I  S  T  V  E  K  E  M  K  S  L  A  R  K  A
1141-aatggataccgaaagtgaacttggcagacaaaaagcagagaataattctttgagactttt     -R8
     M  D  T  E  S  E  L  G  R  Q  K  A  E  N  N  S  L  R  L  L
1201-gtatgaaaacacagaaaaagatctttctgatactcagcgacaccttgctaagaaaaaata
     Y  E  N  T  E  K  D  L  S  D  T  Q  R  H  L  A  K  K  K  Y
1261-tgagctacagcttactcaggagaaaattatgtgcttggatgaaaaaattgataactttac
     E  L  Q  L  T  Q  E  K  I  M  C  L  D  E  K  I  D  N  F  T
1321-aaggcaaaatattgcacagcgagaagaaatcagcattcttggtggaaccctcaatgatct     -GSP6
     R  Q  N  I  A  Q  R  E  E  I  S  I  L  G  G  T  L  N  D  L
1381-ggctaaagaaaaggaatgcctgcaagcatgtttggataaaaaatctgagaatattgcatc
     A  K  E  K  E  C  L  Q  A  C  L  D  K  K  S  E  N  I  A  S
1441-ccttggagagagtttggcaatgaaagaaaagaccatttcaggcatgaagaatatcattgc     -R6(RatH1)
     L  G  E  S  L  A  M  K  E  K  T  I  S  G  M  K  N  I  I  A
1501-tgagatggaacaggcatcaagacagtgtactgaggccctaattgtgtgtgaacaagacgt
     E  M  E  Q  A  S  R  Q  C  T  E  A  L  I  V  C  E  Q  D  V
1561-ttccagaatgcgtcggcaattggatgagacaaatgatgagctggcccagatcgccaggga
     S  R  M  R  R  Q  L  D  E  T  N  D  E  L  A  Q  I  A  R  E
1621-aagagatatcttggctcatgacaatgacaatctccaggaacagtttgctaaagctaaaca
     R  D  I  L  A  H  D  N  D  N  L  Q  E  Q  F  A  K  A  K  Q
1681-agaaaaccaggcactgtccaaaaaattgaatgacactcataatgaacttaatgacataaa
     E  N  Q  A  L  S  K  K  L  N  D  T  H  N  E  L  N  D  I  K
1741-acagaaggttcaagatactaatttggaggttaacaagctgaagaatatattaaagtctga
     Q  K  V  Q  D  T  N  L  E  V  N  K  L  K  N  I  L  K  S  E
1801-aaaatctgagaaccggcaaatgatggaacaacttcgaaaagccaatgaagatgctgaaaa     -R5
     E  S  E  N  R  Q  M  M  E  Q  L  R  K  A  N  E  D  A  E  N
1861-ctgggaaaataaagcccgtcaatcagaggcagataacaataccctcaaactggaacttat
     W  E  N  K  A  R  Q  S  E  A  D  N  N  T  L  K  L  E  L  I
1921-cactgctgaggcagagggtaacagattaaaagaaaaagtagattccctcaacagagaggt
     T  A  E  A  E  G  N  R  L  K  E  K  V  D  S  L  N  R  E  V
1981-tgagcaacacttaaatgcagaaaggtcttacaagtcccagatttctaccttacataaatc
     E  Q  H  L  N  A  E  R  S  Y  K  S  Q  I  S  T  L  H  K  S
2041-tgttgtaaaaatggaagaggagcttcagaaggttcagtttgaaaaagtgtccgctcttgc
     V  V  K  M  E  E  E  L  Q  K  V  Q  F  E  K  V  S  A  L  A
2101-agatttgtcttctactagggaactctgtattaaacttgactcaagcaaagaacttcttaa     -GSP5
     D  L  S  S  T  R  E  L  C  I  K  L  D  S  S  K  E  L  L  N
2161-tcgacagctggttgctaaagatcaagaaatagaaatgagggagaatgagttagattctgc
     R  Q  L  V  A  K  D  Q  E  I  E  M  R  E  N  E  L  D  S  A
2221-tcattctgaaattgaactcctgaggagtcagatggcaaatgagagaatctccatgcagaa
     H  S  E  I  E  L  L  R  S  Q  M  A  N  E  R  I  S  M  Q  N
2281-tctagaagctttgctggtggccaatcgagacaaagaatatcagtctcagatagcacttca     -R4
```

```
     L  E  A  L  L  V  A  N  R  D  K  E  Y  Q  S  Q  I  A  L  Q
2341-agaaaaagaatctgaaattcagcttcttaaagaacacctttgtttggcagaaaataaaat
     E  K  E  S  E  I  Q  L  L  K  E  H  L  C  L  A  E  N  K  M
2401-/ggccatccagagtagagatgtggcccagttcagaaatgttgtcacacaattggaagctga
     A  I  Q  S  R  D  V  A  Q  F  R  N  V  V  T  Q  L  E  A  D
2461-tttagacattaccaaaagacaactaggaacagagcgctttgaaag/ggagagggccgtaca    -GSP4
     L  D  I  T  K  R  Q  L  G  T  E  R  F  E  R  E  R  A  V  Q
2521-agaacttcgccgccaaaattattcaagtaatgcttatcatatgagttctacaatgaagcc
     E  L  R  R  Q  N  Y  S  S  N  A  Y  H  M  S  S  T  M  K  P
2581-aaatacaaaatgtcattcaccagaacgtgctcaccatcgatctcctgaccgaggcctaga
     N  T  K  C  H  S  P  E  R  A  H  H  R  S  P  D  R  G  L  D
2641-tcgatcattagaaga/gaatctttgctacagagatttctgacacctgaaatgattcttcac    -G3/Stop
     R  S  L  E  E  N  L  C  Y  R  D  F STOP
2701-atccctgagaaaggtcaaagttacaaactgaTTTTTTTTTTTTTTgctacatgagtgcat    -GSP3
2761-ttatcttttaaatgcttggcaatgttaaatgtatttattaactttgtgtctctgaatctc    -R1
2821-tgttctaatgtgccatgttgcagtgatctgagatgacttataaaaacaaaaatgtatatg    -GSP2
2881-gctctttctatccatgcaatgatagtgagtgtaaaatctgcttacttcactattgaacac    -GSP1
2941-tattctgttcactatctggagtaaataaagaagcttattaaaacagggaaaaaaaaaaaa
```

Figure III.22: **A)** ***The sequence of the*** **TC10** ***gene including complete coding sequence (amino acid sequence in capitals). Primers used in the paper are underlined and mentioned at the right, as is the polyadenylation signal. Polymorphic regions in the 5´ and 3´ UTRs are also indicated in upper case letters.***

There is an alternative splice acceptor site at the beginning of exon 4 (/) leading to the incorporation of an additional 44 nucleotides (5`-ataattttagatactgaaaaagcacaaaataaatctccttctag-3`).

B) ***The*** **TC10** ***gene contains 19 exons (solid black boxes are coding regions and solid gray boxes are untranslated regions, the alternatively spliced region in exon 4 is shown as a white box). Exons two to nineteen start at positions (/) 229, 445, 511, 635, 794, 965, 1043, 1193, 1310, 1466, 1522, 1691, 1802, 1988, 2198, 2401, 2506 and 2656 respectively. The length of introns are 791, >6380, 1337 or 1293, 397, 3135, 90, 1240, 22568, 2466, 5586, >1163, 2707, 3763, >18793, >1192, >1625, >427 and 19976 base pairs respectively. The length of the entire gene is more than 120kb.***

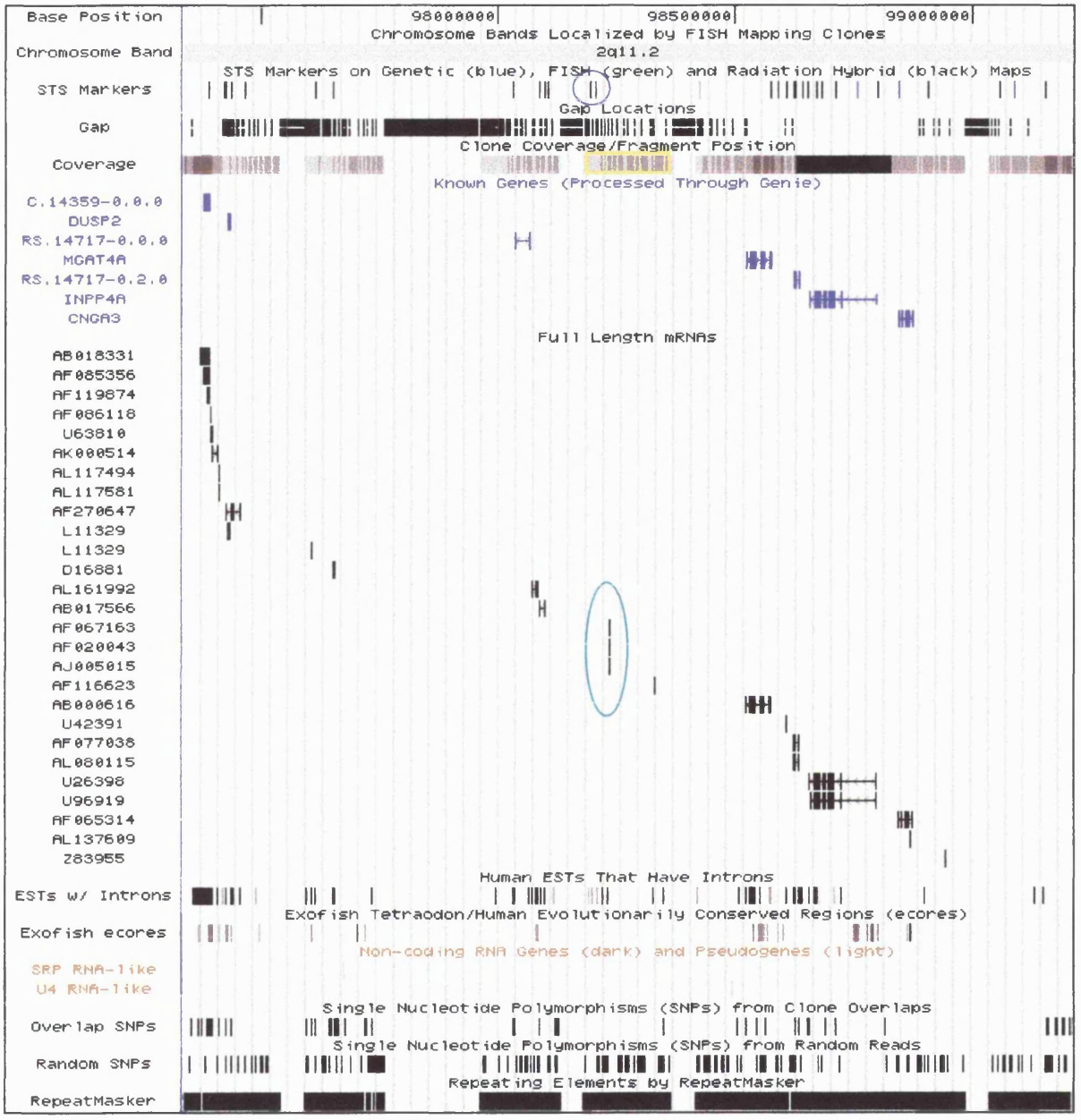


Figure III.23: The draft browsing result. More information in detail such as ESTs are obtainable by clicking on the main browser. The working draft clone AC019097, two STS markers (D2S1309 & stSG50234)within the clone and three human cDNA sequences (bamacan homologue- complete cDNA, chromosome-associated polypeptide (HCAP)-complete cDNA and partial cDNA for putative SMC-like protein) (see section III.16.1) has been shown in a yellow box, a blue circle and a green oval, respectively. This browser displays the draft assembly of the Sept. 5th, 2000 version of the human genome. For example at that time the unfinished clone consisted of 24 contigs but its sequence has now been improved and has only 4 contigs (in Jan. 2001).

Alignments

>gb|AAC72234.1| (AF092091) **cp431 [Rattus norvegicus]**
Length = 898

Score = 1234 bits (3193), Expect = 0.0
Identities = 647/701 **(92%)**, Positives = 662/701 (94%), Gaps = 20/701 (2%)

```
Query: 15  TARGANCDVELLKTTTRDREELKCMLEKYERHLAEIQGNVKVLKSERDKIFLLYEQAQEE 74
           T+R ANCDV+LLK+T RDREELKCMLEKYERHLAEIQGNVKVL S RDK FLLYEQAQEE
Sbjct: 201 TSRAANCDVDLLKSTARDREELKCMLEKYERHLAEIQGNVKVLTSVRDKTFLLYEQAQEE 260

Query: 75  ITRLRREMMKSCKSPKSTTAHAILRRVETERDVAFTDLRRMTTERDSLRERLKIAQETAF 134
           I RLRREMMKSC+SPKSTTAHAILRRVETERDVAFTDLRRMTT RDSLRERLKIAQ  AF
Sbjct: 261 IARLRREMMKSCQSPKSTTAHAILRRVETERDVAFTDLRRMTTARDSLRERLKIAQ--AF 318

Query: 135 NEKAHLEQRIEELECTVHNLDDERMEQMSNMTLMKETISTVEKEMKSLARKAMDTESELG 194
           NEKAHLEQRIEELECTVHNLDDERMEQMSNMTLMKETI+ VEKEMKSLARKAMDTESELG
Sbjct: 319 NEKAHLEQRIEELECTVHNLDDERMEQMSNMTLMKETITIVEKEMKSLARKAMDTESELG 378

Query: 195 RQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQRE 254
           RQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQRE
Sbjct: 379 RQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQRE 438

Query: 255 EISILGGTLNDLAKEKECLQACLDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQ 314
           EISILG TLNDLAKEKECLQ CLDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQ
Sbjct: 439 EISILGATLNDLAKEKECLQTCLDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQ 498

Query: 315 CTEALIVCEQDVSRMRRQLDETNDELAQIARERDILAHDNDNLQEQFAKAKQENQALSKK 374
            TEALI+CEQD+SRMRRQLDETNDEL QIARERDILAH+NDNLQEQFAK KQENQALSKK
Sbjct: 499 STEALIMCEQDISRMRRQLDETNDELGQIARERDILAHENDNLQEQFAKVKQENQALSKK 558

Query: 375 LNDTHNELNDIKQKVQD-----------------TNLEVNKLKNILKSEESENRQMMEQL 417
           LNDTHNEL+DIKQKVQD                 TNLEVNKLKNILKSEESENRQ+MEQL
Sbjct: 559 LNDTHNELSDIKQKVQDTNLEVNKLKNILKSEESTNLEVNKLKNILKSEESENRQIMEQL 618

Query: 418 RKANEDAENWENKARQSEADNNTLKLELITAEAEGNRLKEKVDSLNREVEQHLNAERSYK 477
           RKANEDAENWENKARQ EA+NNTLKLELITAEAEGNRLKEKVD+LNREVEQHLNAERSYK
Sbjct: 619 RKANEDAENWENKARQLEAENNTLKLELITAEAEGNRLKEKVDALNREVEQHLNAERSYK 678

Query: 478 SQISTLHKSVVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIE 537
           SQI+TLHKS+VKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIE
Sbjct: 679 SQIATLHKSLVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIE 738

Query: 538 MRENELDSAHSEIELLRSQMANERISMQNLEALLVANRDKEYQSQIALQEKESEIQLLKE 597
           M ENELDSA SEIELLRSQM NERISMQNLEALLVANRDKEYQSQIALQEKESEIQLLKE
Sbjct: 739 MMENELDSARSEIELLRSQMTNERISMQNLEALLVANRDKEYQSQIALQEKESEIQLLKE 798

Query: 598 HLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQELRRQNYSSNA 657
           HLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQELRRQNYSSNA
Sbjct: 799 HLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQELRRQNYSSNA 858

Query: 658 YHMSSTMKPNTKCHSPERAHHRSPDRGLDRSLEENLCYRDF 698
           YH+ S MKPNTKCHSPERAHHRSPDR LDRSLEENLCYRDF
Sbjct: 859 YHLGS-MKPNTKCHSPERAHHRSPDRDLDRSLEENLCYRDF 898
```

```
>ref|NP_055460.1| KIAA0635 gene product [Homo sapiens]
 dbj|BAA31610.1| (AB014535) KIAA0635 protein [Homo sapiens]
          Length = 846

 Score =  499 bits (1285), Expect = e-140
 Identities = 260/645 (40%), Positives = 431/645 (66%), Gaps = 10/645 (1%)

Query: 30  TRDREELKCMLEKYERHLAEIQGNVKVLKSERDKIFLLYEQAQEEITRLRREMMKSCKSP 89
           TR+R+EL+ MLE++E+++ +IQ NVK+L +ERDK+ +LY +AQEE++ LR+E      S
Sbjct: 208 TRERDELQRMLERFEKYMEDIQSNVKLLTAERDKLSVLYNEAQEELSALRKE------ST 261

Query: 90  KSTTAHAILRRVETERDVAFTDLRRMTTERDSLRERLKIAQETAFNEKAHLEQRIEELEC 149
           ++T  H I+  +E E+++A +DLRR+  E+++LRE+L+  +E +   K+ LE+ IE L C
Sbjct: 262 QTTAPHNIVSLMEKEKELALSDLRRIMAEKEALREKLEHIEEVSLFGKSELEKTIEHLTC 321

Query: 150 TVHNLDDERMEQMSNMTLMKETISTVEKEMKSLARKAMDTESELGRQKAENNSLRLLYEN 209
             H L+ E+ E  S + +MKETI ++E ++K  A+K      +   QK E NSLR++ E
Sbjct: 322 VNHQLESEKYELKSKVLIMKETIESLENKLKVQAQKFSHVAGDSSHQKTEVNSLRIVNEQ 381

Query: 210 TEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQREEISILGGTLNDLAKE 269
            ++ + D Q  L+ K+ EL+  Q +I  L+EKID    +  +Q EE  ++  T+  + KE
Sbjct: 382 LQRSVDDYQHRLSIKRGELESAQAQIKILEEKIDELNLKMTSQDEEAHVMKKTIGVIDKE 441

Query: 270 KECLQACLDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQCTEALIVCEQDVSRM 329
           K+ LQ  +D+K+E IA+L E+LA KEK ++ MK +I+E E +  Q  E L+  +++++ +
Sbjct: 442 KDFLQETVDEKTEKIANLQENLANKEKAVAQMKIMISECESSVNQLKETLVNRDREINSL 501

Query: 330 RRQLDETNDELAQIARERDILAHDNDNLQEQFAKAKQENQALSKKLNDTHNELNDIKQKV 389
           RRQLD  + EL ++ R R+I   +N  LQ+  A   +ENQ +S +L     E  ++K +V
Sbjct: 502 RRQLDAAHKELDEVGRSREIAFKENRRLQDDLATMARENQEISLELEAAVQEKEEMKSRV 561

Query: 390 QDTNLEVNKLKNILKSEESENRQMMEQLRKANEDAENWENKARQSEADNNTLKLELITAE 449
                EV++ ++++ ++E EN+ ++++ +  +  AE+WE KA Q+E ++++++LEL++ +
Sbjct: 562 HKYITEVSRWESLMAAKEKENQDLLDRFQMLHNRAEDWEVKAHQAEGESSSVRLELLSID 621

Query: 450 AEGNRLKEKVDSLNREVEQHLNAERSYKSQISTLHKSVVKMEEELQKVQFEKVSALADLS 509
            E   L+E+V+ L +E+++H+NA  +Y+SQIS++ K++ ++EEEL+  + EK + L DLS
Sbjct: 622 TERRHLRERVELLEKEIQEHINAHHAYESQISSMAKAMSRLEEELRHQEDEKATVLNDLS 681

Query: 510 STRELCIKLDSSKELLNRQLVAKDQEIEMRENELDSAHSEIELLRSQMANERISMQNLEA 569
           S RELCIKLDS K+++ +QL +K+ E E    EL++  SE +LL+ Q++NER +++NLE+
Sbjct: 682 SLRELCIKLDSGKDIMTQQLNSKNLEFERVVVELENVKSESDLLKKQLSNERHTVKNLES 741

Query: 570 LLVANRDKEYQSQIALQEKESEIQLLKEHLCLAENKMAIQSRDVAQFRNVVTQLEADLDI 629
           LL  NRDKE+ S +   EK++EIQLLKE L L+E+K+  QSR+    R  V QL+ D D
Sbjct: 742 LLATNRDKEFHSHLTSHEKDTEIQLLKEKLTLSESKLTSQSRENTMLRAKVAQLQTDYDA 801

Query: 630 TKRQLGTERFERERAVQELRRQNYSSNAYHMSSTMKPNTKCHSPE 674
            KRQ+ TER+ERERA+QE+RR   ++    +SST++  +  HSPE
Sbjct: 802 LKRQISTERYERERAIQEMRRHGLATPP--LSSTLR--SPSHSPE 842
```

Figure III.24: Result of PBLAST search at NCBI shows two best matches with the predicted TC10 Protein. The first match is predicted protein from the rat homologue of TC10. A duplicated region (17 aa) within the rat sequence is not present in the TC10 human predicted protein (Query). The second match is to a protein predicted from human cDNA sequence (see the text). Red and blue letters indicate identity or similarity between amino acids, respectively. The query sequence is the predicted TC10 protein.

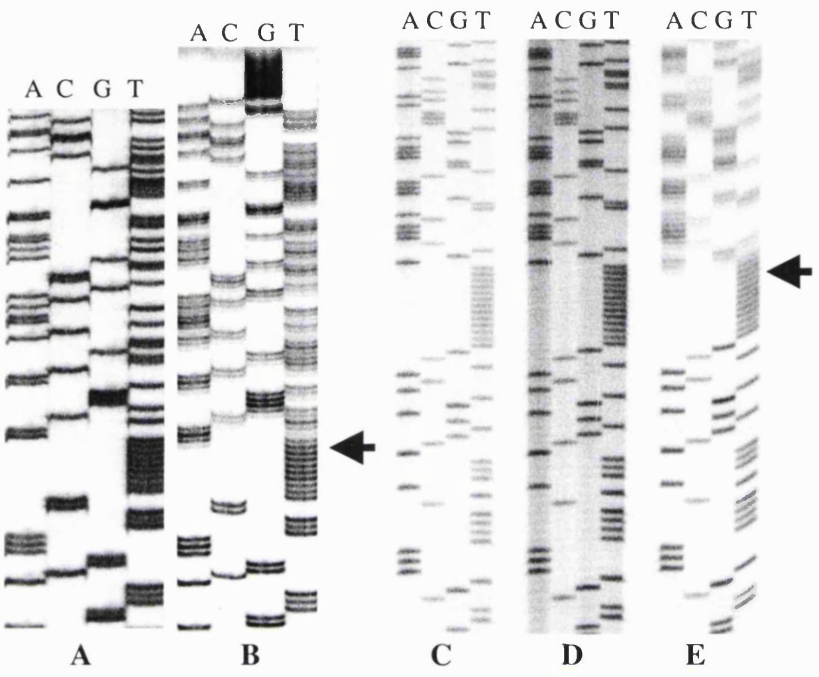


Figure III.25: Polymorphism in 5´UTR and 3´UTR. Sequencing 5´UTR (A & B) and 3´UTR (C, D and E) using GSP11 and GSP2 primers respectively. Templates from which the sequence was derived were RT-PCR products from testis (lanes B and E), a cDNA clone (lanes A and D) and a genomic clone, BAC 138D5 (lane C). Arrows show where the sequences become indistinct in the RT-PCR derived template lanes. The polyG tract visible in the 5`UTR (B)was an experimental construct.

III.7 Polymorphism

The 5′UTR and 3′UTR regions of the *TC10* gene were sequenced from cloned templates (both genomic and cDNA) and from the RT-PCR products derived from testis RNA. Poly T sequences occur at positions 74 (in the 5′UTR) and 2731 (in the 3′UTR) (figure III.25). The sequences of the clones were easy to read at these points (figure III.25, panels A, C and D) but the RT-PCR templates gave perfect sequence up to the poly T tracts but thereafter showed the co-existence of two templates possibly derived from transcription of mRNA from two alleles (figure III.25, panels B and E). Results from EST database searching also showed variation in the number of T nucleotides in the 3′UTR region.

III.8 Nomenclature and the symbol of the gene, *TC10*

The system used for naming the DDRT-PCR fragments at the beginning of this study is mentioned in the Appendix, section A.2.1. This DDRT-PCR fragment was originally named TC10, but as the evidence accumulated suggesting that the sequence could represent a novel gene, the gene symbol, ***TSGA10*** (*Testis Specific Gene A10*) was approved by the HUGO Nomenclature Committee. TC10, the name of the DDRT-PCR fragment that has been used in this study, is substituted by *TSGA10* for future use in this thesis.

III.9 Mouse homologue of the gene

The *Rattus norvegicus* and *TSGA10* sequences were aligned (shown in the Appendix, section A.5) and the similarity at the site of primers checked. GSP5, R5 and R10 were most similar with one nucleotide difference, thus they were selected for RT-PCR using mouse testis cDNA. RT-PCR amplifications were carried out with an annealing temperature 10°C lower than usual. RT-PCR products were electrophoresed onto an agarose gel, and then, after re-amplification as shown in figure III.26, bands were cut out from the gel. They were sequenced using GSP5, R6 and R10 primers and comparison with *TSGA10* sequence revealed around 90% similarity between mouse and human sequences. More than 1kb sequence of the

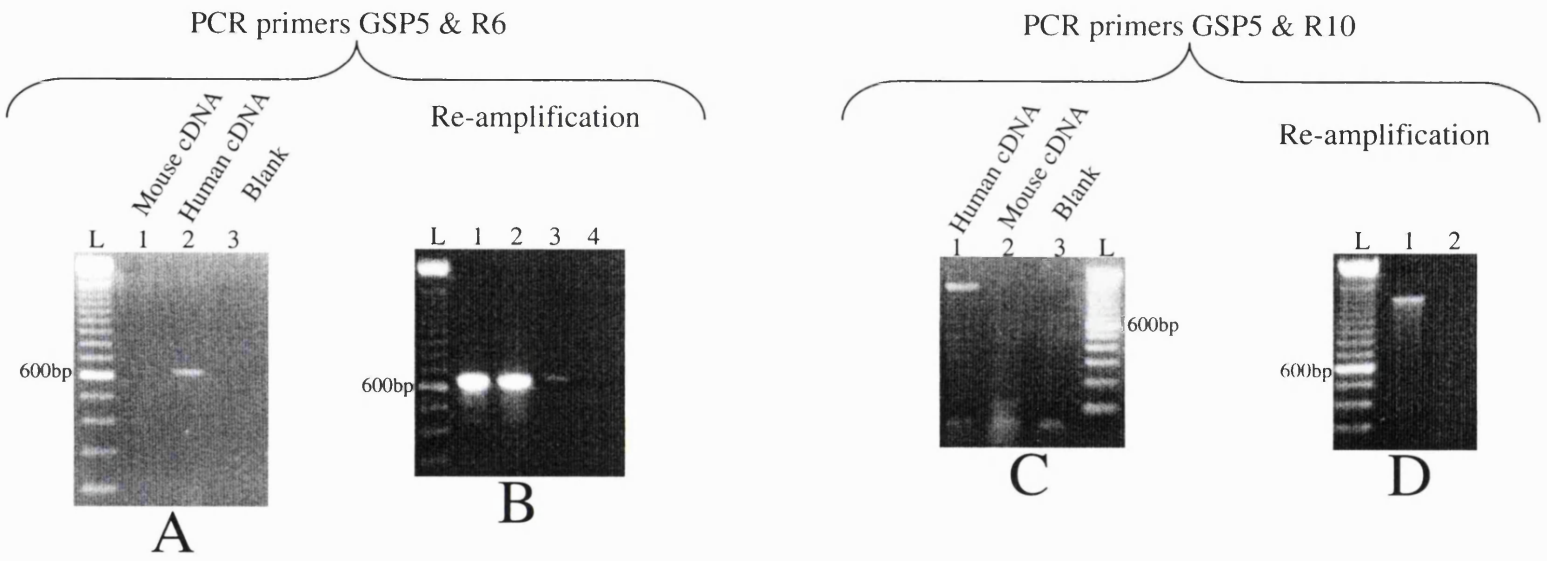


Figure III.26: RT-PCR using mouse cDNA and primers designed from the human TSGA10 gene. GSP5 & R6 and GSP5 & R10, respectively, at ten degrees lower than the calculated annealing temperature. In gels (A & C) a band from human cDNA are seen (lane A-2 & C-1). Lanes A-1 and C-2 were from mouse cDNA, and the third lane in both A & C are PCR negative control (water). A very faint band was visible in the lane A-1 the same size as human PCR product. (B & D) were re-amplification of PCR (A & C) products with 26 PCR cycle at calculated annealing temperatures. Lane B-1 was from the faint band (A-1) by tipping in a needle in the gel. Template in lane B-2 was 0.2 μl of the mouse RT-PCR product and lane B-3 was from human cDNA. Lane B-4 and D-2 were negative controls (water). A small part of lane C-2 at the expected size comparing to the band from human cDNA was cut out from the gel and the extracted DNA from that used for re-amplification (D-1). The bands in lane B-1, B-2 & D-1 were excised for sequencing. Lane L is 100bp ladder in all panels.

mouse homologue of the gene was retrieved by RT-PCR using human primers and mouse testis cDNA. The mouse sequence is given in appendix, section A.4.

Approximately $2x10^5$ p.f.u. of a mouse testis cDNA library, made specifically from spermatocytes (Material, section II.2.8) were plated out on two 22x22 cm plates and duplicate filter lifts were taken. Replica filters were screened with a labelled PCR product derived from the above mentioned mouse RT-PCR using GSP5 and R6 primers. Many positive plaques were seen in both plates (figure III.27). The strongest signal plaque plug from each plate was isolated from this primary screening. These plaques were then plated out onto two 90mm petri dishes and duplicate lifts were screened with the same probe and single plaques were isolated. Amplified phage stocks were prepared from these plaques and sequenced using vector primers. The largest insert size after PCR using two universal vector primers, T3 and T7, was estimated to be about 3.3kb.

The full-length of the mouse *TSGA10* gene was sequenced using templates from mouse RT-PCR products and the mouse spermatocyte cDNA clone (figure III.27 and section A.4).

III.10 *TSGA10* in different species

Comparing partial or complete sequences of *TSGA10* from different species (human, mouse, rat and pig) revealed that exon one in human and pig is different from that in rat and mouse (figureIII.28 and section A.5 in the Appendix). The mouse sequence has been aligned to the human genomic clone AC019097 to compare human and mouse exon/intron junctions. All mouse exons matched with about 90% identity to the same region of the human exons in the clone except exon one and exon nineteen. Human exon 19 is very similar to exon 19 of the rat sequence but there was no homology between this exon and mouse exon 19. An EST (Accession no. BF319562) from mouse adult testis had about 100% identity to exon 18 and 19 of mouse sequence. Interestingly, a match was found between mouse exon one and a region located 8260bp upstream of human exon one. The region was appropriately orientated and showed 89% identity between mouse exon one (base 1 to 130) and the human clone AC019097. The rat and pig cDNA

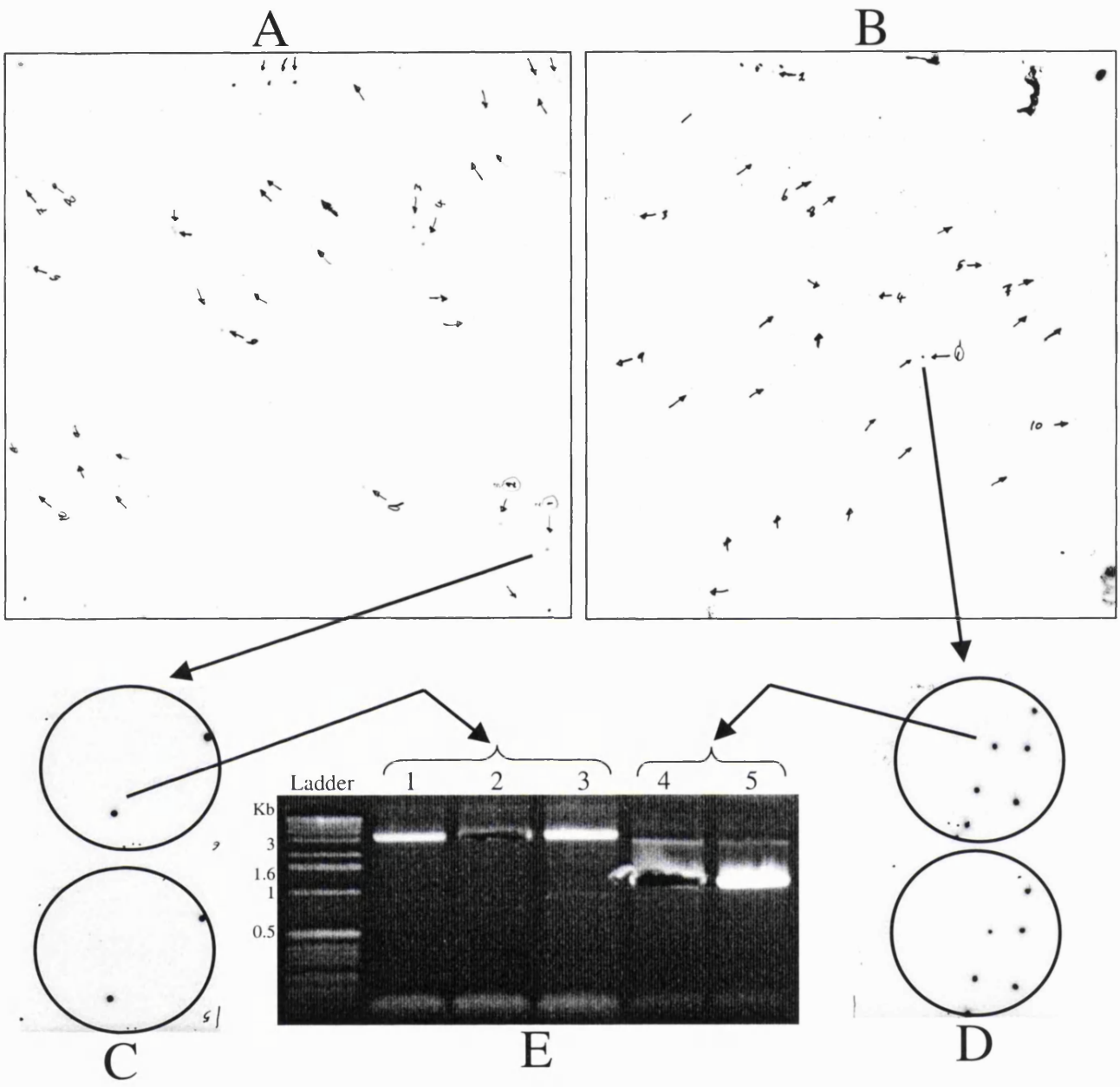
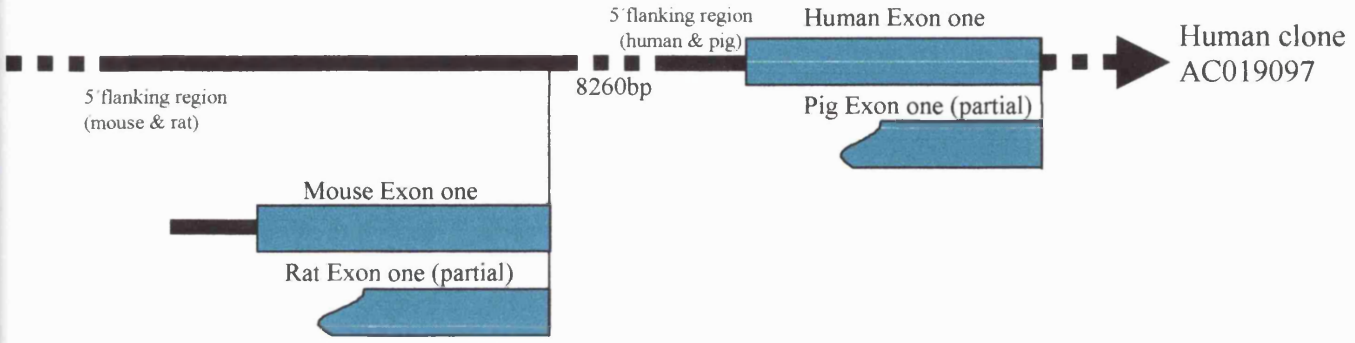


Figure III.27: Results of primary (A & B) and secondary (C & D) mouse spermatocyte cDNA library screening. Two signals(plaques) have been rescued and the size of cDNA inserts were examined by PCR amplification of three clones from the first plaque and two clones from the second plaque(E). PCR products using vector primers T3 + T7. Products from clone C were longer than from clone D. Bands (in lane 2 and 4) were cut out from the gel and sequenced. Clone from the first plaque contained full-length cDNA of mouse TSGA10.

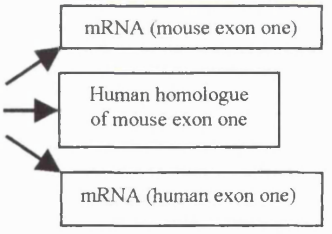


A

```
1 Mouse 5`flanking ------------------------------AG CTTGGTAACTCTGCT TTACTGATGGAAAGT TAAGAACAATAGAAA
2 H&M   5`flanking CACACAGAAAAAGTT TTATAAATTTTA-AG AAGGTAAAGTTTGAT GAAATCCTCTTAGCT GGTGGGAAATGGTTA
3 Human 5`flanking AGGGTACTAACACAG TATCTTAGACACAAT AGGTGCTCAATAAAT ATTTGTTAAATACAC AATGAGAGACAAGTA
_

1 Mouse 5`flanking ATATGAACGAAATCC ATTGTTACTAGTTTT CAAAACTTACTGATG TGAGAAGTCACAGCT
2 H&M   5`flanking AAAAAAATCTAA--G AATATTATCAATTTT CAAAATGTATTGATG TAAGAAATCACAGAG
3 Human 5`flanking CTATGGAAAAGAATT TGTACCATGACTTTA CCTCGTAGTTATTTT TGTTATTGCACTTA-
_

1 Mouse 5`flanking CTGTTTTAAATAGAG ---CTCATTCTTATA TCTCATTCTTTTTGT TAG
2 H&M   5`flanking CTGTTTTAAACTGAA TGCCTCTGTATTATT TCCCATTCTTTGTAC TAG
3 Human 5`flanking ACATGTTGTCTTTGT ATTTCCTCAAGTAAA CAAAGAGTTCCCTGA GGG
```



B

Figure III.28: A) Schematic position of exon one of each species (green) and 5'flanking regions from different species (black) have been aligned to human genomic clone AC019097. 5'flanking region of the mouse gene was sequenced from mouse PAC clone 618H14. B) Multiple sequence alignment results and comparison of (1) the 5'flanking regions of the mouse* TSGA10 *which was sequenced from a PAC mouse genomic clone, (2 or M&H) the 5'upstream region of the matching region of mouse exon one with the human genomic clone AC019097, 8.2kb upstream of human* TSGA10*, and (3) the 5'flanking region of human* TSGA10 *exon one. Similarities of human and mouse sequences adjacent to mouse exon one (1)&(2) have been shown in blue and the putative TATA boxes are in red.

sequences are partial but both contain most of exon one (enough to be compared to other sequences).

The mouse exon one has been used to search sequence database (HTGS) and two matches are noteworthy (figure III.29). The first one was the above mentioned match in human genomic clone AC019097 and the second match with 88% identity was from a human clone (accession no. AC007403) that was also located on chromosome 2, but on the short arm at 2p14. Figure III.30 displays various types of information about the AC007403 clone. Six human ESTs (containing introns) from testis or libraries containing testis match with the clone, none is related to *TSGA10*. The location of *TSGA10* as determined by FISH is 2q11. This homology to a sequence on the short arm of chromosome 2 may suggest an evolutionary duplication.

III.11 Genomic clones and localisation of the mouse homologue of *TSGA10*

Seven mouse PAC gridded filters (Osoegawa *et al.*, 2000) were provided by the UK HGMP Resource Center (http://www.hgmp.mrc.ac.uk/Biology/descriptions/mouse_pac.html), and contain DNA from approximately 128,000 clones. The average insert size was 147kb and further information about the library and the vector can be obtained from the WWW site: http://www.chori.org/bacpac/. They were screened with the same probe as was used for the mouse cDNA library screening. Eight positive signals resulted from hybridisation (figure III.31) and the eight single isolated bacterial colonies were checked by PCR using mouse TSGA10 R12 and G12 primers which were designed from the mouse testis cDNA clone sequence and are located in the 5′end of the gene. PCR results confirmed that colonies 635E6, 552E13, 652L5, 413O2 and 618H14 contained at least the 5′ end of the mouse gene. DNA extraction from colony 618H14 was carried out and it was labelled and used to hybridise to male mouse metaphase spreads. The results of fluorescence in situ hybridisation (FISH) localised the gene to mouse chromosome one, band B-C1. This location would be expected given the mouse-human chromosome homologies (figure III.32).

III.12 5′flanking region of the mouse sequence

The 5′flanking region of mouse exon one was sequenced from PAC 618H14. It was aligned with the 5′upstream region of the match site of mouse exon one from the human genomic clone AC019097 and the 5′flanking region of human *TSGA10* exon also from AC1019097 (figure III.28). 65% identity was seen between the mouse 5′flanking region and its homologue from the human clone. Putative TATA boxes are present in both sequences (figure III.28 and III.37). There does not seem to be much of a match to the human exon one 5′ region.

```
>AC019097 Homo sapiens chromosome 2 clone RP11-425I16, WORKING DRAFT
SEQUENCE, 4 unordered pieces
          Length = 185850

 Score =  145 bits (73), Expect = 9e-33
 Identities = 112/125 (89%)
 Strand = Plus / Minus

Query: 1     gttaacaccaaagaacatcagtttggtcttggggctccagtttctgaagctggaagaaac 60
              |||| |||||||||| || | |||||||| ||||||||||||||||||||||  ||   ||
Sbjct: 14117 gttaccaccaaagaatatgattttggtctaggggctccagtttctgaagctgaaaactac 14058

Query: 61    cagaatactttccaactagaacaagaagtgagaacccaagatagattcatctctacactg 120
              ||||||||| |||| |||||||||||||||||| |||||||||||||||||| ||||||
Sbjct: 14057 cagaatactctccagctagaacaagaagtgagaaaccaagatagattcatctcgacactg 13998

Query: 121   aaatt 125
              |||||
Sbjct: 13997 aaatt 13993


>AC007403 Homo sapiens clone RP11-547F18, complete sequence
          Length = 176321

 Score =  111 bits (56), Expect = 8e-23
 Identities = 92/104 (88%)
 Strand = Plus / Minus

Query: 22    tttggtcttggggctccagtttctgaagctggaagaaaccagaatactttccaactagaa 81
              |||| ||| |||| |||||||||||  ||| || | |||||||||| |||| ||||||
Sbjct: 55266 tttgttctgggggatccagtttctgacactgaaaaataccagaatactctccagctagaa 55207

Query: 82    caagaagtgagaacccaagatagattcatctctacactgaaatt 125
              |||||| |||||| |||||||||||||||||||||||||||||
Sbjct: 55206 caagaaatgagaaaccaagatagattcatctctacactgaaatt 55163
```

Figure III.29: Significant results of database search using mouse exon one (Query).

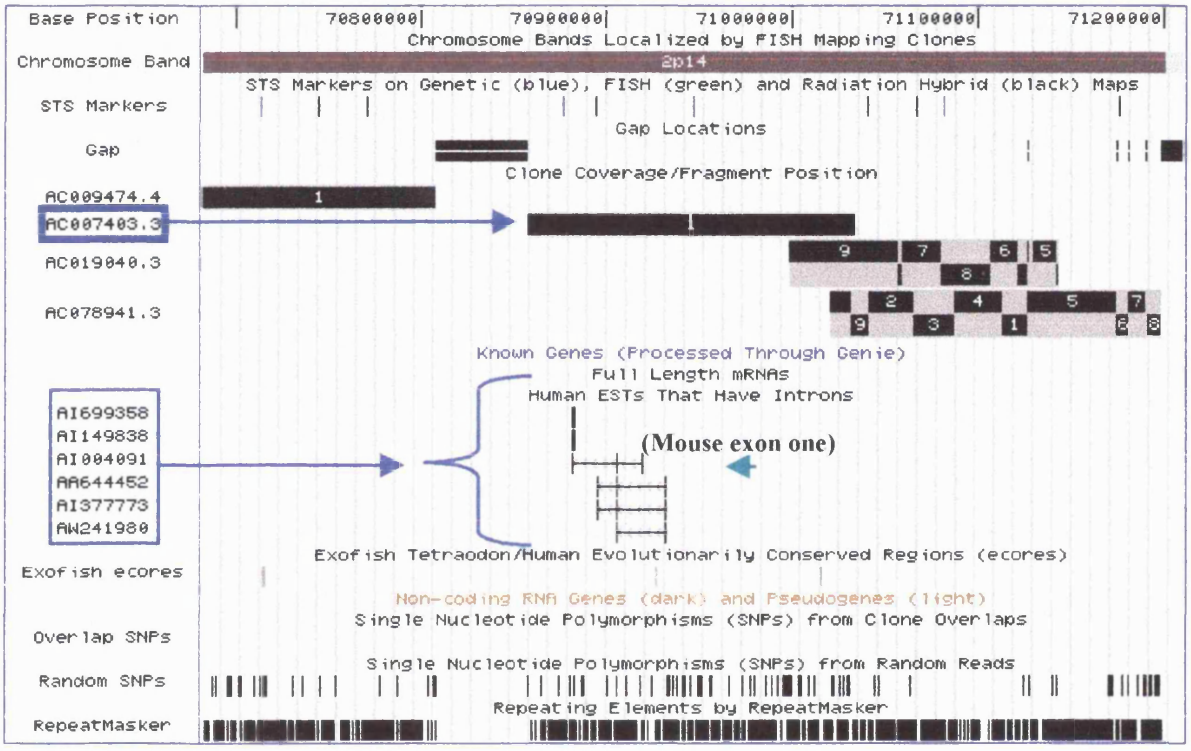


Figure III.30: The genomic position of the human clone AC007403 is shown in this draft compared to other sequences at this location in gene bank (Dec. 2000). Six human ESTs, which have introns, have been matched with the clone and all of them (in the box) originate from testis or libraries containing testis.

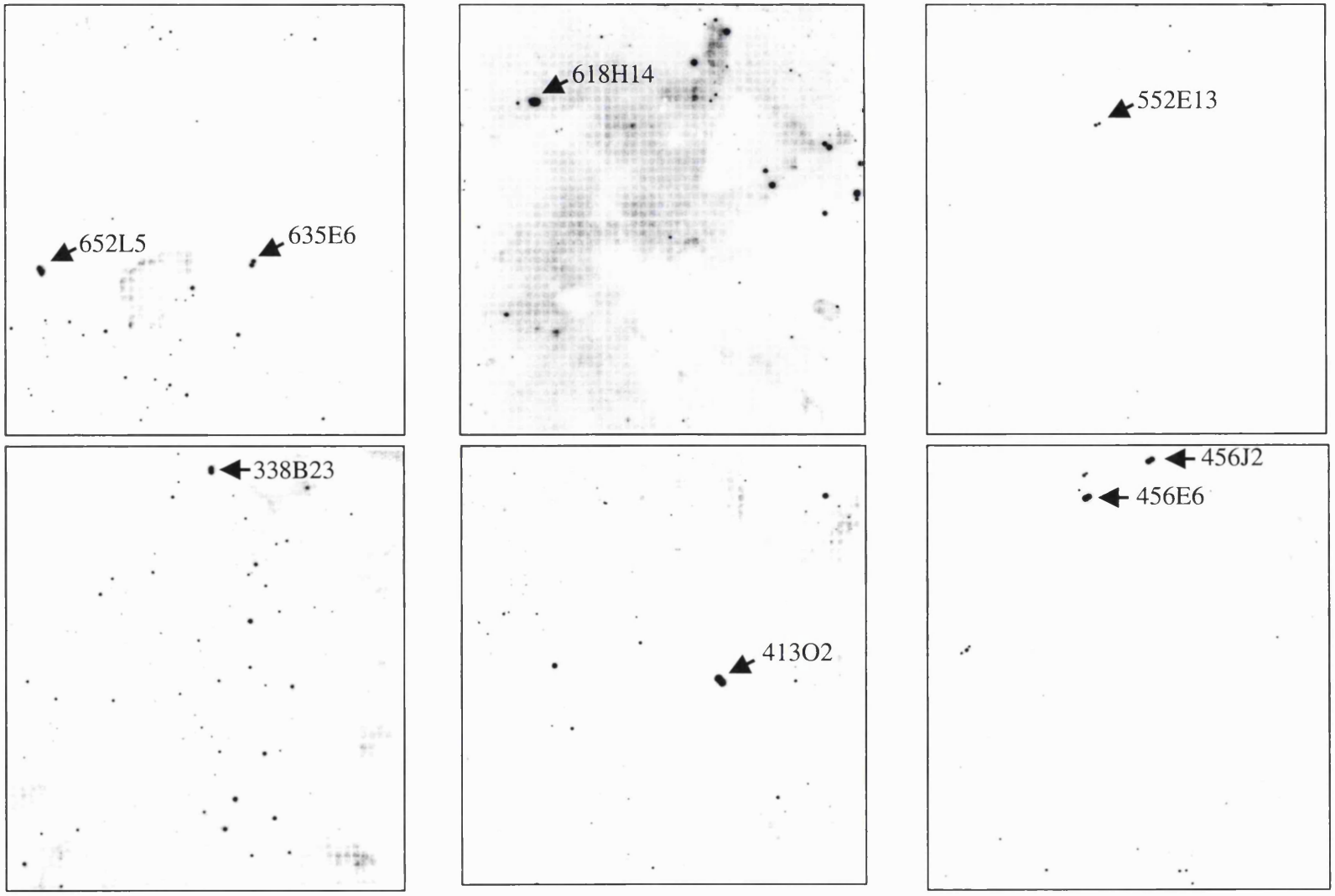
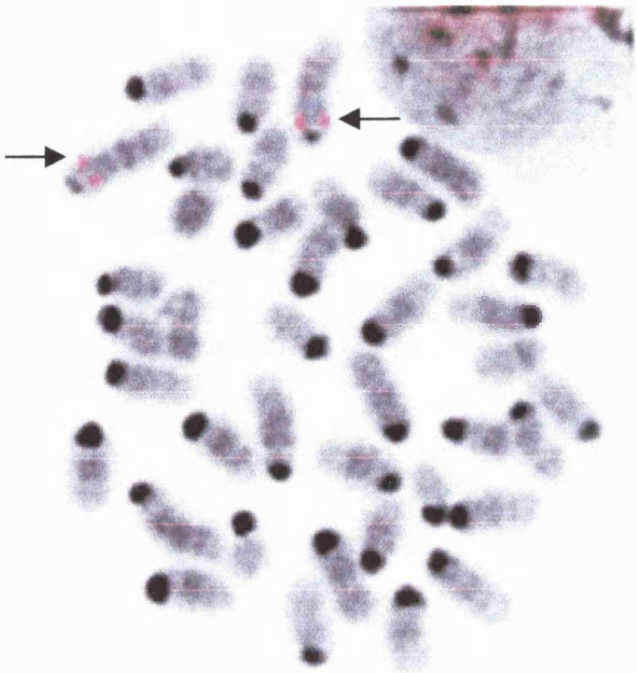


Figure III.31: PAC filters made from mouse genomic library was screened for clones which contain entire or part of the mouse TSGA10 gene. Positive, duplicate clones are represented as double signal (real signals have been noted by arrows).

A)



B)

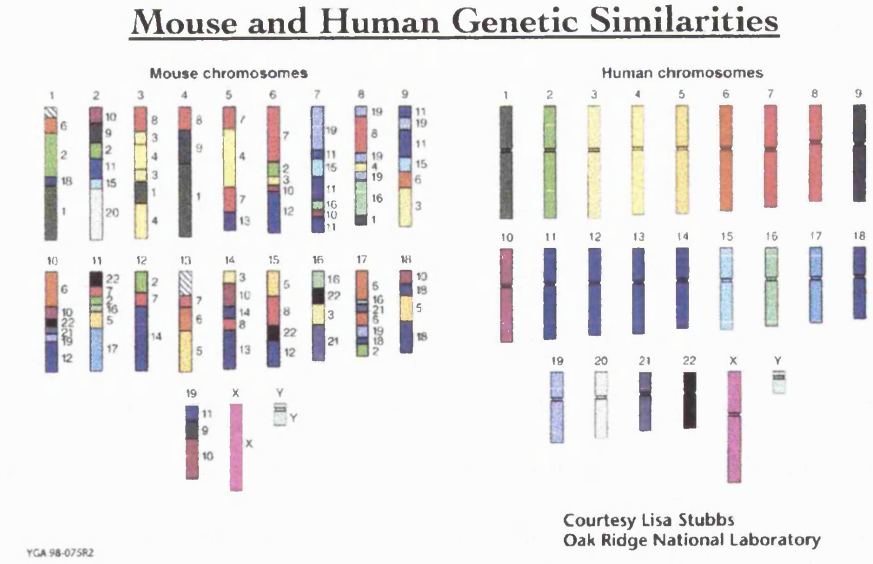


Figure III.32: A) Localisation of the mouse TSGA10 on mouse chromosome 1 (red dots) to the band B or C1. B) Schematic showing the synteny between human and mouse chromosomes.

III.13 THE GENE, *TSGA10* DISCUSSION

III.13.1 Is TSGA10 a gene?

Previously it has been reported that false positive results can arise in DDRT-PCR experiments with a frequency greater than 70% (Sun *et al.*, 1994). This ratio has been reduced using experimental improvements (Sompayrac *et al.*, 1995; Miele *et al.*, 1998; see chapter I, section 4.1.6) but confirmation of the gene still needs some additional experiments to be carried out (see figure III.1). Plenty of evidence supports the hypothesis that *TSGA10* is a real gene. Results from DDRT-PCR, RT-PCRs, the northern blot and cDNA adult testis library all suggest that this is a genuine transcript. This is supported by evidence from the sequence databases, ESTs and cDNAs from human, mouse, rat and pig. In addition, there is the human genomic clone AC019097 sequence which shows 19 exons all with consensus splice boundaries.

III.13.2 Expression pattern

It has been demonstrated by DDRT-PCR, RT-PCR and by Northern blotting that the gene *TSGA10* is expressed in normal testis and also in other actively dividing tissues (either foetal or malignant). It is not expressed in a variety of tissues functionally related to testis nor is it expressed in the testes of two infertile patients nor foetal testis. It should be noted that testis development in a foetus at the second and third trimester of pregnancy has already been completed and descent is in process with little cell division occurring (see chapter I, sections 1.1.2 and 1.1.3). Thirteen human ESTs homologous to *TSGA10* have been reported. Nine are certainly or probably derived from testis and three of them, which represent more than two exons of the gene, are derived from a poorly differentiated adenocarcinoma, an acute myelogenous leukemia and a germ cell tumour. At the present time there are about 6.85 million EST sequences present in databases (January 2001) and table III.5 shows the results of the Blast search of these with *TSGA10*. Most of the matches seem to be real ESTs derived from mRNA and not artefacts which sometimes arise from genomic contamination in the original

libraries. The observation that the sources of the ESTs, both human and other species, which show highly significant matches to human *TSGA10* are almost all germinal cells in testis, or tumours or embryonic tissues, may suggest that there is a link between expression of this gene and cell division.

Mammalian spermatogenesis provides an excellent model system for studying cellular division and differentiation. First, the mitotic division of germ cells is similar to many mitotically active undifferentiated tissues such as those in tumours or foetus. The formation of different organs and tissues is based primarily on differential gene expression. While some ' housekeeping' genes that contribute to basic structural or metabolic cellular functions are expressed ubiquitously throughout the body, other tissue-specific genes that contribute to specialised function in differentiated cell types are expressed in a regulated fashion. It is likely that, due to the structural and functional changes that take place in male germ cells during spermatogenesis, many germ cell-specific gene products are involved in this differentiation process. This process is seen to occur in parallel with changes in the expression of total RNA and stage specific proteins (Hoog, 1995, Monesi, 1965, Boitani *et al.,* 1980).

There may be a parallel in the expression of *TSGA10* and *c-kit*. In melanogenesis, *c-kit* (see chapter I, section I.3.1) is expressed in melanoblasts from the time they leave the neural crest and expression continues during embryonic development and in the melanocytes of postnatal animals (Besmer *et al.*, 1993). The c-kit receptor tyrosine kinase belongs to the PDGF/CSF-1/c-kit receptor subfamily (see chapter I, section I.3.1). The kit-ligand, KL, also called steel factor, is synthesised from two alternatively spliced mRNAs as transmembrane proteins that can either be proteolytically cleaved to produce soluble forms of KL or can function as cell-associated molecules. In gametogenesis c-kit is expressed in primordial germ cells and in spermatogonia (see chapter I, section I.3.1). Studies in white spotting and steel mice have shown that functional SCF and *c-Kit* are critical in the survival and development of stem cells involved in hematopoiesis, pigmentation and reproduction (review, Linnekin, 1999). Downstream of c-Kit, multiple signal transduction components are activated, such as the Ras-Raf-MAP kinase cascade

which is one of the known cascades in tumours and cell division (see chapter I, section I.3.2). There is some evidences that may support the links between *TSGA10* and *c-Kit*. Firstly, the site for phosphorylation by the tyrosine kinase of *c-kit* is similar to one of the sites on the predicted TSGA10 protein (explained later in the section). Secondly, one of the matches with 99% identity to the 3′ end of *TSGA10* (figure III.21 and table III.5) is an EST (accession no. N33757) from melanocyte and c-kit is also involved in pigmentation. Thirdly, the MAP kinase pathway and many components downstream of c-kit are related to cell division in which *TSGA10* is probably involved. Fourthly, The proto-oncogene c-kit is expressed in a broad spectrum of human cancers (Matsuda *et al.*, 1993) and *TSGA10* expression was also shown in at least six different tumours (three from EST database and others by using RT-PCR). Finally, the role of *c-kit* during reproduction and its expression in testicular germ cells has been demonstrated (see chapter I, section I.3.2), and experiments have proved the presence of the *TSGA10* transcript in testis and mouse and rat spermatocyte libraries.

In summary, the pattern of expression of *TSGA10* strongly suggests that has a role in spermatogenesis. This conclusion is strengthened by consideration of the *TSGA10* homologues in other species. Genbank contains a partial cDNA which is a rat homologue of *TSGA10*, and is claimed to be present in *Rattus norvegicus* spermatocytes. In addition, the mouse homologue of *TSGA10* was isolated in this study from a spermatocyte cDNA library. Making a cDNA library from only one kind of germinal cells of testis is difficult due to failure of purity of cell separation and contamination by other cells but it is encouraging that rat and mouse homologues of the gene were separately obtained from two cDNA libraries each made using spermatocyte cells. Two of the mice ESTs (BF319562, BF018459) which matched to *TSGA10* are derived from the later stage of spermatid (round spermatid).

III.13.3 Copy number of *TSGA10*

The human genomic BAC library which was screened consisted of 125,000 BAC colonies, each with an average size of 160,000-200,000 bp. This gives an

estimated ten fold coverage of the genome. The screen yielded three positive colonies. This indicates that the gene is likely to be single copy. In addition, the mouse genomic library which was used in this experiment consisted of 128,000 PAC clones with an average insert size 150,000bp equivalent to approximately six fold genome coverage and the hybridisation results (8 colonies) is also consistent with the single copy hypothesis.

III.13.4 Translation of mRNA and predicted protein

The 5′ and 3′ ends of mRNA are elements for translational regulation in mammals (review, Gray and Wickens, 1998). Regulation via the 5' UTR can be mediated by (1) modification of the cap structure, (2) secondary structure, (3) RNA-protein interactions, and (4) upstream open reading frames (uORFs). 3′UTR-mediated regulation can occur through (1) RNA-protein interactions, which may involve multi-protein complexes, (2) RNA-RNA interactions, (3) cytoplasmic polyadenylation elements, and (4) changes in poly (A) tail length.

The majority of eukaryotic mRNAs have 5′UTRs of 20–100 nucleotides (Kozak, 1987). Experimentally, shortening the 5′UTR of reporter mRNAs to less than 12 nucleotides impairs the efficiency of translation from the first AUG (Sedman *et al.*, 1990). Increasing the length of a 5′UTR can increase the efficiency of translation as additional 43S pre-initiation complexes can be loaded. This is sometimes termed pre-loading (Kozak, 1991). However, many cellular mRNAs with unusually long 5′UTRs are poorly translated as a result of the presence of upstream AUGs, uORFs, and/or secondary structure (Kozak 1987, 1991). Increasing the length of a 5'UTR appears to be especially common in mRNAs encoding proto-oncogenes, transcription factors, growth factors, and their receptors (Kozak 1987, 1991), which suggests that their translation is tightly controlled. The 5'UTR of the *TSGA10* gene is 626 nucleotides with (or 582 nucleotide without) the alternative spliced fragment (figure III.22) which is relatively long for a 5'UTR.

Short uORFs in the 5'UTR can modulate translation of the main ORF. An example is seen in mammalian β2 adrenergic receptor mRNA regulation (Parola and Kobilka, 1994). Human PDGF2 (platelet-derived growth factor 2) mRNA

contains three uORFs (Rao *et al.,* 1988), but it is the presence of extensive secondary structure within its 5'UTR that results in its poor translation, at least under the conditions tested. Interestingly, an additional T in the 5'UTR polymorphic site of *TSGA10* (see section III.7 and figure III.33) results in an uORF of 8aa in the same frame as the main predicted protein (figure III.33). Secondly, alternative splicing also revealed two uORFs with only 2aa in the main predicted protein frame. Both translated sequences are shown in the figure III.33 and may lead to different levels of expression in individuals with different alleles.

The use of an alternative initiation site is another point in translation regulation to produce more than one protein from a single mRNA. Many mRNAs contain AUGs within their 5'UTR that are not followed by in-frame stop codons prior to the start of the main ORF. Initiation at the downstream AUGs is the result of leaky scanning. Interestingly, several mRNAs produce full-length proteins from both upstream and downstream AUGs or sometimes CUGs (review, Gray and Wickens, 1998). An example of this bifunctional mRNA in human is *c-myc* (Hann *et al.*, 1988, 1992). The predicted TSGA10 protein shows that there are many "Met." or AUG sites within the ORF which could in theory be used after missing the upstream AUG by translational machinery to produce smaller proteins. The effects of uORFs and upstream AUGs can vary with cell type and during differentiation (review, Gray and Wickens, 1998).

Secondary structure of the mRNA can also cause an inhibitory effect on translation especially when it is G-C rich. A well-characterized example of mRNA that is regulated through secondary structure is the human PDGF2 mRNA (review, Gray and Wickens, 1998). PDGF2 (c-sis) mRNA encodes one of two proteins that form PDGF, a powerful mitogen, important in wound healing, embryogenesis, and development. Its role in wound healing involves its expression in bone marrow. Prior to megakaryocytic differentiation, PDGF2 translation is thought to be repressed by extensive secondary structure. (Bernstein *et al* 1995, 1997). The secondary structure of *TSGA10* mRNA (with and without the alternatively spliced site) and PDGF2 were predicted using computer program at site: http://www.genebee.msu.su/services/rna2_reduced.html and is shown in figure

1 Translated protein without a T at the 5`UTR polymorphic site

Q V S N L S L Y F W N L I **Stop** F L T **Stop** S L F N G S N F F F L **Stop** R H L D T F V I E K A L F F I H
Stop W F D S G L G R K A V F L H I R Q Q I L D **Stop** F D I **Stop** L S Q I E D L K Q T N H G L E E C V
R K L L D S K E V V S S Q V D D L T S H N E H L C K E L I K I D Q L A E Q L E K E K N F V V D S A
N K E L E E A K I D L I C Q Q N N I I V L E D T I K R L K S I I L D T E K A Q N K S P S R L D S F V K
T L E A D K D H Y K S E A Q H L R K **Met** **Met** R S R S K.....

2 Translated protein with a T the 5`UTR polymorphic site

R Y L T Y L C I F G I **Stop** Y S S **Stop** H N H C S **Met** G L I F F F L **Stop** R H L D T F V I E K A L F F I
H **Stop** W F D S G L G R K A V F L H I R Q Q I L D **Stop** F D I **Stop** L S Q I E D L K Q T N H G L E E C
V R K L L D S K E V V S S Q V D D L T S H N E H L C K E L I K I D Q L A E Q L E K E K N F V V D S
A N K E L E E A K I D L I C Q Q N N I I V L E D T I K R L K S I I L D T E K A Q N K S P S R L D S F V
K T L E A D K D H Y K S E A Q H L R K **Met** **Met** R S R S K.....

3 Translated protein with alternative splicing

Q V S N L S L Y F W N L I **Stop** F L T **Stop** S L F N G S N F F F L **Stop** R H L D T F V I E K A L F F I H
Stop W F D S G L G R K A V F L H I R Q Q I L D **Stop** F D I **Stop** L S Q I E D L K Q T N H G L E E C V
R K L L D S K E V V S S Q V D D L T S H N E H L C K E L I K I D Q L A E Q L E K E K N F V V D S A
N K E L E E A K I D L I C Q Q N N I I V L E D T I K R L K S T **Stop** F L C Q D F G S R Q R S L **Stop** E
Stop S S T L R K **Met** **Met** R S R S K .....

4 Translated protein without alternative splicing

G I **Stop** P I F V F L E F N I V P D I I I V Q W V **Stop** F F F P L K A F G Y L C D R K G P F L H P L **Met**
V **Stop** **Stop** W A G K E S C V P P H **Stop** A A N T **Stop** L I **Stop** Y L A **Stop** S N **Stop** R S Q T D K S W
L G R **Met** C **Stop** E T L G **Stop** **Stop** G G G K Q S S R **Stop** F N Q P Q **Stop** A S L **Stop** R I D **Stop** N
Stop P T S R A T R K R E K F C G G F R Q Q G T **Stop** R S Q D **Stop** S H L P A K **Stop** Y Y S I G R Y N
K K A **Stop** I Y N F R Y **Stop** K S T K **Stop** I S F **Stop** T **Stop** F L C Q D F G S R Q R S L **Stop** E **Stop** S S
T L R K **Met** **Met** R S R S K.....

Figure III.33: Translation of different alleles of* TSGA10 *gene (5'end sequences). Mets are starting codon of the predicted TSGA10 protein (pink). Mets are starting codon of uORFs which have been underlined. Blue shows the similarity of sequences before uORFs.

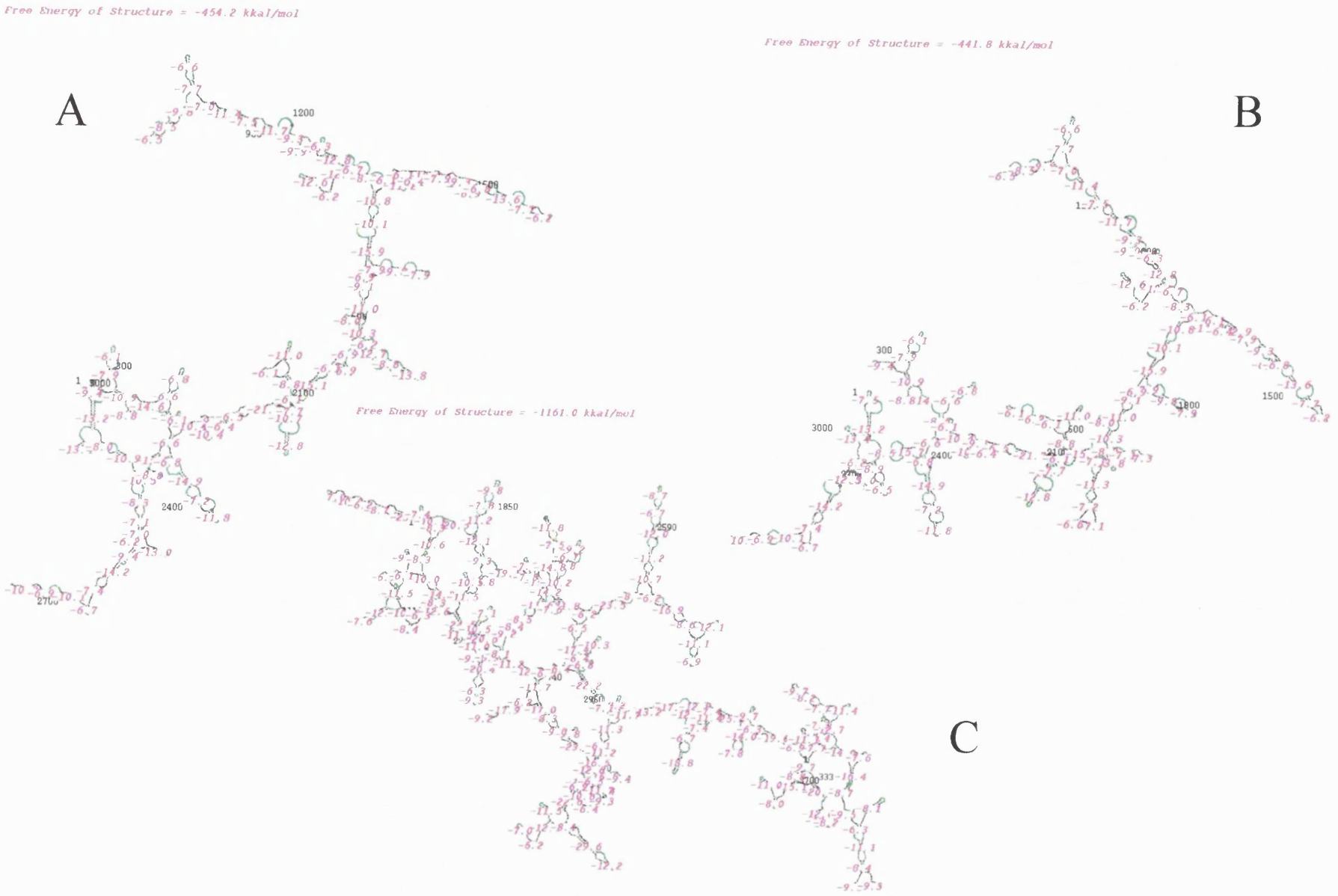


Figure III.34: Secondary structure of* TSGA10 *mRNA without (A) and with (B) alternatively spliced site. Secondary structure of* PDGF2 *mRNA(C) is also shown to compare with* TSGA10*. Nucleotides are indicated in black, purple numbers shows the energy between each complement site.

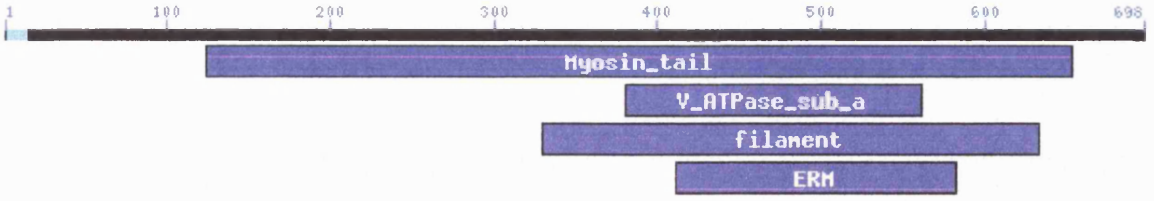
III.34. Comparing the predicted free energy of structure showed that there is a slight difference between the two *TSGA10* transcripts but a large difference with PDGF2. The *TSGA10* transcript is not very G-C rich and this can be one of the reasons for making less stable secondary structure than PDGF2. Knowledge of the secondary structure may be helpful for further investigation such as in situ hybridisation or anti-sense RNA studies.

Protein Blast searching revealed a 40% identity to a human protein of unknown function (KIAA0635) predicted from a brain cDNA sequence (Accession no. BAA31610, located on chromosome 4) but this identity is not localised at one or more sites as a motif (figure III.24). Brain tissue is known to express a very wide variety of genes (Adams *et al.*, 1993) and the complexity, i.e. the number of different RNA sequences transcribed in brain tissue, is considerably higher than many tissues and in uniform cell cultures. Many similarities have been observed between gene expression patterns in testis and brain. For instance, the relative frequency of expressed genes encoding cell surface proteins, or signal transduction, has been approximately the same (Hoog, 1995). Therefore, considering the 40% identity at the level of protein which is relatively significant, it is possible that both may be from the same family.

Computational biologists define conserved domains based on recurring sequence patterns or motifs. A Conserved Domain Database (CDD) and a search service at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) was used to search for a known domains in the predicted TSGA10 protein. The results shown in figure III.35 indicate very low homology to three domains but this homology (highest homology with 21% to myosin heavy chain) is not thought to be significant. Using another similar program, Pfam at Washington University (http://pfam.wustl.edu/) and Sanger center (http://www.sanger.ac.uk/Software/Pfam/dnasearch.shtml) also indicated a similar low level of homology to the ERM (Ezrin/radixin/moesin) family (figure III.35).

The protein product is predicted to be soluble and not a membrane protein (http://dot.imgen.bcm.tmc.edu:9331/seq-search/protein-search.html). It was examined using the PROSCAN program (http://pbil.ibcp.fr/cgi-

Producing significant alignments:



		Score (bits)	E value
pfam01576	Myosin_tail, Myosin tail	43.1	4e-05
pfam01496	V_ATPase_sub_a, V-type ATPase 116kDa subunit family	39.7	5e-04
pfam00038	filament, Intermediate filament proteins	39.3	6e-04
pfam00769	ERM, Ezrin/radixin/moesin family	35.4	0.009

pfam01576, Myosin_tail, **Myosin tail**

```
             Length = 860
             Score = 43.1 bits (100), Expect = 4e-05
Query:  124  ERLKIAQETAFNEKAHLEQRIEELECTVHNLDDERMEQMSNMTLMKETISTVEKEMKSLA  183
Sbjct:  141  DQLQKQKAKAEKDKSQLEAEVDDLLAQLDSIAKAKLNAEKKAKQLESQLSELQVKLDELQ  200

Query:  184  RKAMDTESELGRQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKID  243
Sbjct:  201  RQLNDLNSQKSRLQSENSEL--------------TRQLEEAEAQVSNLSKSKSSLESQLE  246

Query:  244  NFTRQNIAQREEISILGGTLNDLAKEKECLQACLDKKSENIASLGESLAMKEKTISGMKN  303
Sbjct:  247  EAKRSLEEETRERATLQAQLRQLEHDLDSLREQLEEESEAKAELERQLSKANAEIAQWRS  306

Query:  304  II-AEMEQASRQCTEALIVCEQDVSRMRRQLDETNDELAQIARERDILAHDNDNLQEQFA  362
Sbjct:  307  KFESEGAARAEELEELKKKLNQRISELEEQAEAANAKCDSLEKAKSRLQSELEDLQIELE  366

Query:  363  KAKQENQALSKKLNDTHNELNDIKQKVQDTNLEVNKLKNILKSEESENRQMMEQLRKANE  422
Sbjct:  367  RANAAASELEKKQKNFDKILAEWKKKVDELQQELETAQREARNLSTELFRLKNELEELKD  426

Query:  423  DAENWENKARQSEADNNTLKLELITAEAEGNRLKEKVDSLNREVEQHLNAERSYKSQIST  482
Sbjct:  427  QVEQLRR-------ENKNLQDEI-------HDLTEQLGEGGRNVHELEKARRRLEAEKEE  472

Query:  483  LHKSVVKMEEELQKVQFEKVSALADLSSTR-ELCIKLDSSKELLNRQLVAKDQEIEMREN  541
Sbjct:  473  LQAALEEAEAALELEEDKVLRAQVELQQIRSEIERRLAEKEEEFENTRKNHQRAIESLQA  532

Query:  542  ELDS-AHSEIELLRSQMANERISMQNLEALLVANRDKEYQS------QIALQEKESEIQL  594
Sbjct:  533  SLEAEAKGKAEASRIKKKLEGDINELEIALDHANKANADAQKNIKKYQQQVKDLQLQLEE  592

Query:  595  LKEHLCLAENKMAIQSRDVAQFRNVVTQLEADLDIT---KRQLGTERFERERAVQELRRQ  651
Sbjct:  593  EQRAREDAREQLAVAERRAAALEAELEELRSALEQAERARKAAETELAEASERVNELTAQ  652

Query:  652  NYS  654
Sbjct:  653  NSS  655
```

---

pfam01496, V_ATPase_sub_a, **V-type ATPase 116kDa subunit family**

```
             Length = 692
             Score = 39.7 bits (91), Expect = 5e-04
```

```
Query:  381  ELNDIKQKVQDTNLEVNKLKNILKSEESENRQMMEQLRKANEDAENWENKARQSEADNNT  440
Sbjct:  68   EFLDLEEKLLELEEEIKEVEESLESLEKELNELEEWLNVLDETKSFLEENLEELEELSNL  127

Query:  441  LKLELITAEAEGNRL--------KEKVDSLNREV-----EQHLNAERSYKSQISTLHKSV  487
Sbjct:  128  DDEFKYLRGGEMLRLGFVAGVINREKLESFERELWRVLRGQAEIEEPLEDPKKTVFIVFF  187

Query:  488  V------KMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIEMREN  541
Sbjct:  188  VGKEDLDKVKKILESFGFELYDVPETEGEPSELISKVNKRIEDLKRVL---EQTISHLEK  244

Query:  542  ELDSAHSEIELLRSQMANER  561
Sbjct:  245  VLVKIADELLAWDEQVSKEK  264
```

---

pfam00038, filament, **Intermediate filament proteins**

```
             Length = 312
             Score = 39.3 bits (90), Expect = 6e-04
Query:  330  RRQLDETNDELAQIARERDILAHDNDNLQEQFAKAKQENQALSKKLNDTH-NELNDIKQK  388
Sbjct:  3    KEQMQNLNDRLASYIDKVRFLEQQNKELEVKIEELRQKQAPSVSRLYELYETEIEELRRQ  62

Query:  389  VQDTNLEVNKLKNILKSEESENRQMMEQLRKANEDAENWENKARQSEADNNTLKLELITA  448
Sbjct:  63   IDQLTNERARLQ----LEIDNLREAAEDFRKKYEDEINLRQEA---ENDLVGLRKDLDEA  115

Query:  449  EAEGNRLKEKVDSLNREVE--QHLNAERSYKSQISTLHKSVVKMEEELQKVQFEKVSALA  506
Sbjct:  116  TLARVDLENKVESLQEELEFLKKNHEEEVKELQAQIQDTVNVEMDAARKL---DLTKALR  172

Query:  507  DLSSTRELCIK--LDSSKELLNRQLVAKDQEIEMRENELDSAHSEIELLRSQMANERISM  564
Sbjct:  173  EIRAQYEEIAKKNRQEAEEWYKSKLEELQTAAARNGEALRSAKEEITELRRQIQSLEIEL  232

Query:  565  QNL----EALLVANRDKEYQSQIALQEKESEIQLLKEHLCLAENKMAIQSRDVAQFRNVV  620
Sbjct:  233  QSLKAQNASLERQLAELEERYEEELRQYQALISQLEEELQQLREEMARQLREYQELLDVK  292

Query:  621  TQLEADLDITKRQL  634
Sbjct:  293  LALDIEIATYRKLL  306
```

---

pfam00769, ERM,

```
             Length = 365
             Score = 35.4 bits (80), Expect = 0.009
Query:  412  QMMEQLRKANEDAENWENKARQSEADNNTLKLELITAEAEGNRLKEKVDSLNREVEQHLN  471
Sbjct:  139  QMEEEMRKAQEELEEAEETADELEEKLKQEEEEAQLLEKKQSELEEENRRLEEEAMKSEE  198

Query:  472  AERSYKSQISTLHKSVVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVA  531
Sbjct:  199  ERERLEAEVDEATAEVAQLEEELERREEEARQLQTELRSAQEAHEEAREELLEALAAPTA  258

Query:  532  KDQEIEMRENELDSAHSEIELLRSQMANERISMQNLEALLVANRDKEYQSQI  583
Sbjct:  259  PPVSAPDNGSGLESGENGEEASADLETDGDMKQLEEERVTYAEKNERLQTQL  310
```

Figure III.35: Result of Pfam at (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) shows conserved domains matching with the predicted TSGA10 Protein. Red and blue letters indicate identity or similarity between amino acids, respectively. The query sequence is the predicted TSGA10 protein.

bin/npsa_automat.pl?page=/NPSA/npsa_proscan.html) which revealed many potential sites for N-glycosylation, for phosphorylation by many kinases (cAMP & cGMP dependent protein kinase, protein kinase C, Casein kinase II and Tyrosine kinase) and also for N-myristoylation. The site for Tyrosine kinase phosphorylation with nine amino acids, RSLEENLCY, is a perfect match to the consensus and therefore seems more likely to be of biological significance than do some of the other, shorter, matches. Tyrosine kinase receptors play some roles during signal transduction, cell division and differentiation (Schlessinger, 1994; Hsueh and Scheuermann, 2000) which are clearly seen in spermatogenesis.

III.13.5 Incorrectly spliced transcript

As mentioned above (step thirteen and figure III.12 & III.13) a partial sequence was found during 5′RACE between the eighth exon/intron boundary. RT-PCR experiments indicated that this is probably not another case of alternative splicing, and database searches indicate that it may perhaps be the result of errors in the splicing process. During RACE multiple amplifications can reveal even a small number of incorrectly spliced RNAs. The 144 bp repetitive sequence is present in the clone AC019097 almost in the middle of the eighth intron (figure III.36). In general, this position of nucleotides {5′ ...exon][GU...intron....AG][exon... 3′} are appropriate for splicing and it seems that the insert has appropriate splicing sites (see figure III.36). Therefore, it is likely that this addition is the result of a splicing error which has been amplified during the multiple amplification of the 5′RACE.

III.13.6 Transcription factor and promoter prediction

Transcriptional activity of a eukaryotic gene is governed by a complex interaction between transcription factors and DNA response elements (Johnson and McKnight, 1989; Mitchell, P.J. and Tjian, 1989).

300bp of 5′flanking region was sequenced from human BAC R380D5 using GSP12 primer and was used to search non-redundant databases. The search highlighted a region at about (-129 to -155) that is present in many sequences and

Figure III.36: The location of the insert (red) within intron 8. cDNA is in blue. / shows exon and intron boundary site. PCR products using primers GSP6 and R8 (underlined) shown in the figure III.13 were 231bp (small band = blue) and 375bp (large band = blue and red). Bold letters show nucleotides which are appropriate for splicing site.

-240 ACTGAAACTTCTGTGTGCGAGTGTTTTTTTTGTTGCTGTTGTTTTAATGCATCTTTACAG

-180 GGTACTAACACAGTATCTTAGACACAATAGGTGCTCAATAAATATTTGTTAAATACACAA

-120 TGAGAGACAAGTACTATGGAAAAGAATTTGTACCATGACTTTACCTCGTAGTTATTTTTG

-60 TTATTGCACTTAACATGTTGTCTTTGTATTTCCTCAAGTAAACAAAGAGTTCCCTGAGGG

1 caggtatctaacctatctttgtatttttggaatttaatatagttcctgacataatcattg.....**mRNA (Sense)**

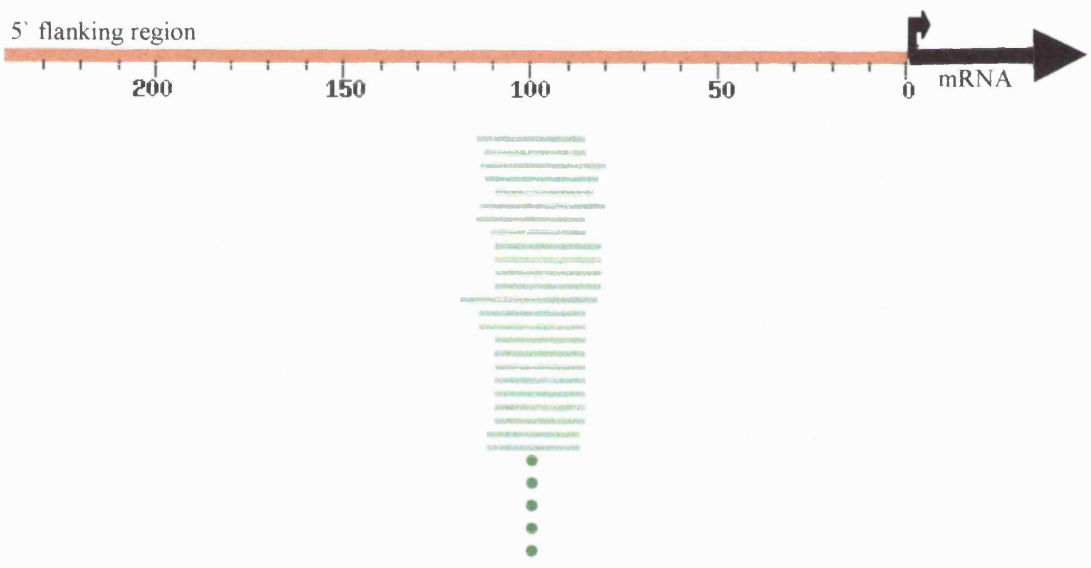


Figure III.37: Result of 240 bp 5 flanking region Blast search. A lot of hits to the area of TSGA10 between nucleotides -125 to -155 (blue sequence) are seen.

could be a potential promoter site (figure III.37). There was no conventional TATA box and no CCAAT box in the expected positions –30bp and –70bp respectively.

However, a CAAT/GT box is positioned at position –21 (figure III.37). It has been shown that a CAAT/GT box in the mouse lactoferrin gene termed the epidermal growth factor response element (EGFRE) mediates epidermal growth factor (EGF)-induced transcriptional activation of the promoter in transiently transfected human endometrial carcinoma RL95-2 (RL95) cells (Shi *et al.*, 1999). It has also been demonstrated that intestinal-enriched Kruppel-like factor (IKLF) which is a transcription factor with zinc finger domains binds to the CAAT/GT box of the mouse lactoferrin gene (Shi *et al.*, 1999). IKLF expression is limited to the epithelial lining of the intestine and is localised primarily to the base of the crypts in the adult intestine and also shows temporal changes in expression during embryogenesis (Conkright *et al.*, 1999). IKLF binds GC/GT-rich sequences (Liu *et al.*, 1996), however, the 5′flanking region of TSGA10 is not GC/GT-rich.

In addition, analysis of potential binding sites within 300bp of 5′flanking region of *TSGA10* for transcription factors using MatInspector V2.2 program at http://www.gsf.de/biodv/matinspector.html and TESS program (http://www.cbil.upenn.edu/cgi-bin/tess/tess33?RQ=SEA-FR-Query or http://searchlauncher.bcm.tmc.edu/seq-search/gene-search.html) identified many potential sites (shown in tables III.6 and III.7). The programs were set for highest similarity with transcription factor sequence. Results revealed that within 300bp upstream of the 5`end of the gene, there are five binding sequences with 100% similarity for; SRY (Harley *et al.*, 1992), PU.I (Faisst and Meyer, *et al.*, 1992), DEF (Jansen-Durr *et al.*, 1989), STE12 (Dolan *et al.*, 1989) and HiNF-A transcription factors (in order from 5′end). Recombinant SRY protein was able to bind to the same core sequence AACAAAG recognised by T-cell factor-1 (TCF-1) in a sequence dependent manner. The *PU.1* gene encodes an Ets family transcription factor which controls expression of many B cell- and macrophage-specific genes. Numerous expression studies have shown expression of *PU.1* to be restricted to hematopoietic cell lineages (reviewed by Fisher and Scott, 1998). *PU.1* is identical to the *Spi-1* proto-oncogene which is associated with the generation of spleen focus-

MatInspector Result

```
MatInspector Release public domain January 2000 Fri Mar 9 13:18:49 2001

Solution parameters:
   ~~~~~~~~~~~~~~~~~~~
   sequence file:   BCM_Search_Launcher_184.seq
   core sim:        0.80
   matrix sim:      0.85

   Explanation for column output:
   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
        ->      Matrix positions correspond to sense strand numbering, but all
                sequences are given in 5'-3' direction. + shows sense strand

                1>>>>>>>>>>>>>>>>>>>>300 1>>>>>mRNA>>>>>>
                   5'flanking region

        ->      n/a in column 'core simil.' indicates, that no core search
                was conducted.

        ->      Capital letters within the sequence indicate the core string.
```

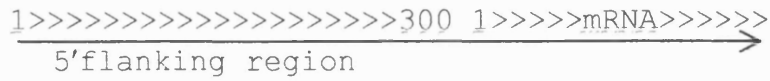
Matrix Name	Position(str) of Matrix	Core Simil.	Matrix Simil.	Sequence
V$CEBPB_01	32 (+)	1.000	0.975	gtatttaGCAAttt
V$NFAT_Q6	196 (+)	1.000	0.962	ctatgGAAAaga
V$FREAC4_01	275 (+)	1.000	0.952	ctcaagtaAACAaaga
V$FREAC2_01	275 (+)	1.000	0.951	ctcaagTAAAcaaaga
V$HFH2_01	80 (+)	1.000	0.937	gagTGTTttttt
V$SRY_02	280 (+)	1.000	0.936	gtaaACAAagag
V$HNF3B_01	78 (+)	1.000	0.931	gcgagTGTTttttttt
V$RORA1_01	19 (+)	1.000	0.930	actaagaGGTCag
V$XFD1_01	277 (+)	1.000	0.929	caagTAAAcaaaga
V$FREAC3_01	275 (+)	1.000	0.929	ctcaaGTAAacaaaga
V$HFH3_01	80 (+)	1.000	0.927	gagTGTTttttttt
V$XFD3_01	277 (+)	1.000	0.913	caagtaAACAaaga
V$HNF3B_01	160 (+)	1.000	0.913	ataaaTATTtgttaa
V$SRY_02	142 (+)	1.000	0.911	agacACAAtagg
V$FREAC7_01	155 (+)	1.000	0.909	gctcaaTAAAtatttg
V$HFH3_01	162 (+)	0.955	0.904	aaaTATTtgttaa
V$NKX25_01	276 (+)	1.000	0.900	tcAAGTa
V$LMO2COM_01	43 (+)	0.822	0.899	tttCAGAtgctc
V$HFH3_01	260 (+)	0.838	0.899	ttgTCTTtgtatt
V$CMYB_01	90 (+)	1.000	0.896	ttttgttgctGTTGtttt
V$SRY_02	175 (+)	1.000	0.894	atacACAAtgag
V$HFH2_01	162 (+)	0.903	0.889	aaaTATTtgtta
V$AP1FJ_Q2	216 (+)	1.000	0.888	caTGACtttac
V$TAL1BETAE47_01	41 (+)	1.000	0.883	aatttCAGAtgctcct
V$XFD2_01	157 (+)	1.000	0.882	tcaaTAAAtatttg
V$XFD2_01	277 (+)	1.000	0.882	caagTAAAcaaaga
V$HFH3_01	232 (+)	0.955	0.882	agtTATTtttgtt
V$HFH8_01	80 (+)	1.000	0.881	gagTGTTttttttt

V$OCT1_06	40	(+)	1.000	0.878	caatttcagATGCt
V$DELTAEF1_01	27	(-)	1.000	0.877	aaatACCTgac
V$BRN2_01	172	(+)	0.854	0.873	taaatacaCAATgaga
V$TCF11_01	27	(+)	1.000	0.873	GTCAggtatttag
V$HFH3_01	82	(+)	0.838	0.872	gtgTTTTttttg
V$XFD1_01	157	(+)	1.000	0.870	tcaaTAAAtatttg
V$TAL1ALPHAE47_01	41	(+)	1.000	0.867	aatttCAGAtgctcct
V$HNF3B_01	230	(+)	1.000	0.867	gtagtTATTtttgtt
V$SOX5_01	143	(+)	1.000	0.863	gacaCAATag
V$SOX5_01	176	(+)	1.000	0.861	tacaCAATga
V$AP1_Q2	216	(+)	1.000	0.860	caTGACtttac
V$HFH3_01	234	(+)	0.838	0.859	ttaTTTTtgttat
V$DELTAEF1_01	221	(+)	1.000	0.857	ctttACCTcgt
V$PADS_C	81	(+)	0.865	0.855	aGTGTTttt
V$HNF3B_01	264	(+)	1.000	0.855	ctttgTATTtcctca
V$HFH3_01	86	(+)	0.838	0.854	tttTTTTtgttgc
V$CEBPB_01	194	(+)	0.986	0.854	tactatgGAAAaga
V$AP1_Q4	216	(+)	1.000	0.852	caTGACtttac
V$HFH1_01	80	(+)	1.000	0.851	gagtGTTTtttt

Table III.6: Results of search of 300bp of 5'flanking region of TSGA10 for potential binding sites with MatInspector V2.2 program at http://www.gsf.de/biodv/matinspector.html.

Column Headings Results

1- Name and ref. No. of transcription factors
2- Start of the site in the query sequence (sense). Numbered from 1



3- Length of the site
4- Matching portion of the query sequence coloured or cased to indicate mismatches (black: 0.1 or less, best or perfect match; blue: 1.0 or less, pretty good match; red: others, mismatch)

1		2	3	4
T00029	AP-1	18	11	CTAAGAGGTCA
I00403	RAR-beta2	19	10	TAAGAGGTCA
I00405	RAR-gamma1	19	11	TAAGAGGTCAG
T00045	ARP-1	20	9	AAGAGGTCA
T00117	CF1	22	9	GAGGTCAGG
I00133	COUP	23	8	AGGTCAGG
I00201	NF-GMb	26	7	TCAGGTA
T01477	BR-C_Z1	28	18	AGGTATTTAGCAATTTCA
T00656	Oct-6	28	18	AGGTATTTAGCAATTTCA
T00459	C/EBPbeta	30	14	GTATTTAGCAATTT
T00360	**HiNF-A**	32	12	**ATTTAGCAATTT**
I00186	Ig/EBP1	32	9	ATTTAGCAA
Q00083	HNF-3	44	19	CAGATGCTCCTTACAGACT
I00334	**STE12**	62	6	TGAAAC
I00106	GHF-1	105	8	TAATGCAT
T02291	Croc	154	16	CTCAATAAATATTTGT
T01480	BR-C_Z4	155	13	TCAATAAATATTT
M00268	V$XFD2_01	155	14	TCAATAAATATTTG
T02290	HFH-2	160	12	AAATATTTGTTA
T00739	SBF-1	163	14	TATTTGTTAAATAC
M00268	V$XFD2_01	170	14	TAAATACACAATGA
T00906	XPF-1	197	22	TGGAAAAGAATTTGTACCATGA
I00147	ANF	219	10	CTTTACCTC
T00395	Hb	232	10	TTATTTTTGT
I00073	**DEF**	238	8	**TTGTTATT**
I00047	**PU.1**	270	6	**TTCCTC**
T01051	XFD-1	275	14	CAAGTAAACAAAGA
M00268	V$XFD2_01	275	14	CAAGTAAACAAAGA
T01479	BR-C_Z3	276	15	AAGTAAACAAAGAGT
T00371	HNF-3	277	9	AGTAAACAA
T00997	SRY	278	12	GTAAACAAAGAG

I00100	HNF-5	279	8	TAAACAAA
I00035	**SRY**	281	8	**AACAAAGA**
T00333	GR	282	12	ACAAAGAGTTCC
T00697	PR	283	12	CAAAGAGTTCCC

Table III.7: Results of search of 300bp of 5 flanking region of TSGA10 for potential binding sites with TSS program at http://www.cbil.upenn.edu/cgi-bin/tess/tess33?RQ=SEA-FR-Query.

forming virus-induced erythroleukemias. HiNF-A sequence is a recognition site for a nuclear protein with affinity for the 5′ flanking region of a cell cycle dependent human H4 histone gene (Van Wijnen *et al.*, 1987). Furthermore, within 300bp upstream of the 5′end of the gene, there are 5 core binding sequences for *SRY* or *SOX5* transcription factors (see table III.6). It remains to be seen whether any of these sites are of biological significance.

It was mentioned in this chapter that comparing partial or complete sequences of *TSGA10* from different species (human, mouse, rat and pig) revealed that exon one in human and pig is different from that in rat and mouse. Therefore, this result may suggest the presence of an alternative promoter in human and pig *TSGA10* genes compared to mouse and rat genes. In addition, the presence of different exon sequences in human and mouse implies that alternative use of exon one in different tissues is a possibility. The upstream areas of homology may indicate areas that are important recognition sites for transcription factors.

III.13.7 Polyadenylation signal

Some testis mRNAs do not have an obvious signal directing polyA tail addition (Sheets *et al.*, 1990, Nayernia *et al.*, 1999). However, the *TSGA10* transcript has both a polyA tail and polyadenylation signal (Proudfoot and Brownlee, 1976) (see figures III.20 and III.21).

III.13.8 Polymorphism in both 5` and 3` UTR

Two potentially polymorphic sites were recognised in the sequence. The first is present in the 5′UTR as different numbers of T residues in two different alleles (figure III.25). The second polymorphism is a variable polyT tract located in the 3′UTR region. This 3′ difference was also observed in cDNA from at least three EST sequences in databases. Several different 3′ alleles have been sequenced (11,12 and 14 T nucleotides) a cDNA clone sequenced in this study had 13 T nucleotides as did a genomic clone (figure III.25). However, the RT-PCR product showed the presence of both 12 and 13 T nucleotides. Interestingly, during the initial DDRT-PCR amplification, the oligoT12VC primer primed both from the polyA tail and

from this polymorphic site. The arbitrary primer did not take part in this amplification and therefore, the DDRT-PCR fragment of this experiment has been amplified using only one primer.

III.13.9 Neighbouring genes of *TSGA10* and location analysis

The chromosome region to which the *TSGA10* gene maps, 2q11.2, has been reported as the locus of at least two other genes with possible roles in mitosis and the regulation of the cell cycle. One is *BRRN1* (Cabello *et al.*, 1997) which is similar to the gene *barr* (*barren*) in *Drosophila melanogaster.* This gene is necessary for sister chromatid segregation and modulates topoisomerase II (Bhat *et al.*, 1996). A second gene in this region is *RANBP2* (Ran-binding protein 2) (Krebber *et al.* 1997) which is close to a related gene *RANBP2L1* (Nothwang *et al.*, 1998). RANBP2 plays an important role for nuclear import (Milchoir *et al.*, 1995) and it is involved in cell cycle progression, nucleocytoplasmic transport and pre-mRNA processing (Sazer, 1996). It is tempting to speculate that this region of chromosome 2 contains a functional cluster of genes involved in or subject to, cell cycle regulation.

There is a sequence within intron eight which is related to 3 human genes. These three genes have been shown previously to be involved in cell cycle regulation. They are three human cDNA sequences (bamacan homologue- complete cDNA, chromosome-associated polypeptide (hCAP)-complete cDNA and partial cDNA for putative SMC-like protein), however, given their location in genomic DNA (figure III.23), they seem to be pseudogenes and the main genes (hCAP, babacan and SMC like) are on chromosome 10 (the same region, q25.2) containing 22, 16 and 8 exons, respectively. The eukaryotic genome must compact and resolve into distinct chromosomes for proper segregation at mitosis. Chromosome condensation requires a family of highly conserved ATPases called structural maintenance of chromosome (SMC) proteins (Hirano and Mitchison, 1994 ; Hirano *et al.*, 1997). It has been suggested that bamacan (Perlecan and basement membrane-chondroitin sulfate proteoglycan) is a member of SMC family (Ghiselli *et al.*, 1999). Four SMC proteins have been identified in humans in the form of two

distinct complexes (Schmiesing *et al.*, 1998). The human chromosome-associated protein (HCAP)-C/HCAP-E complex associates with chromosomes at mitosis and is required for chromosome condensation (Steen *et al.*, 2000).

In summary, *TSGA10* is a single copy gene with homologues in other mammals that appears to be expressed predominantly in testis and other actively dividing tissues. The gene is about 120kb long with 19 exons. The cDNA spans 2994bp/3038bp, has an alternatively spliced transcript and encodes a predicted protein with 698 aa. Polymorphic sequences in both 5` and 3` UTR have been observed. A polyadenylation signal and potential sites for transcription factor binding are also predicted.

Chapter IV

Cloning, characterisation and mapping of the gene encoding the human G protein gamma 2 subunit

One of the DDRT-PCR fragments expressed in testis, called TG2, was cloned and sequenced. Initial sequencing indicated that this cDNA was the transcript from a member of the G protein gamma subfamily which had not yet been cloned in human. The characterisation of this gene is described here.

IV.1 Confirmation of testis-specificity

As mentioned in chapter III, the ESTs generated by the DDRT-PCR protocol must be tested by additional methods to ascertain whether they are expressed from other tissues in addition to the testis. DDRT-PCR fragment TG2 was 198bp long and presumed to be from 3'UTR. Firstly, GenBank databases were searched using the TG2 sequence to confirm novelty and expression pattern. The results of NR database search at that time confirmed that it was a novel gene. The result of searching human EST databases, which can be helpful to show pattern of the expression of a gene, revealed that this gene was expressed in many tissues, both related to testis such as germ cell tumour and seminoma and unrelated to testis, such as melanocyte (see 3'UTR part of the gene in table IV.1 and figure IV.6).

It is usual to initially confirm expression patterns of the DDRT cDNAs using RT-PCR. Specific primers (GSP1 and R1) were designed from the DDRT fragment sequence according to the usual criteria and were used to perform RT-PCR on RNA from two normal testes. The 3'UTR region is often intronless and so PCR products from cDNA and genomic DNA are the same size (as seen in figure IV.1-A). A second primer (R2) was designed from EST sequences 410bp from the 3'end. The location of the primers have been underlined in figure IV.10. RT-PCR using GSP1 and R2 showed a cDNA product in both normal testis samples (figure IV.1-B). These specific primers were also used to perform RT-PCR on RNA from two normal testes, two different infertile patients' testes, sperm, adrenal gland, liver, muscle, prostate gland and brain. All RNAs or DNased RNAs were checked for

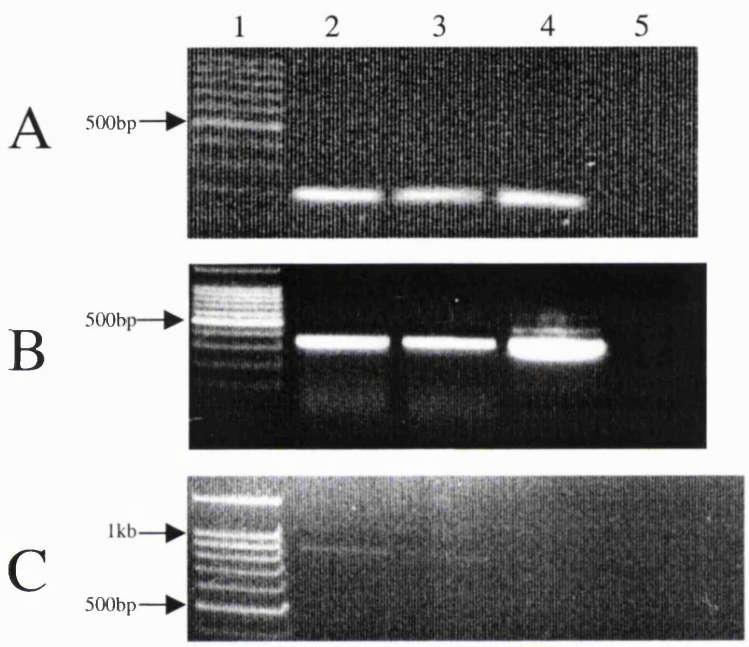


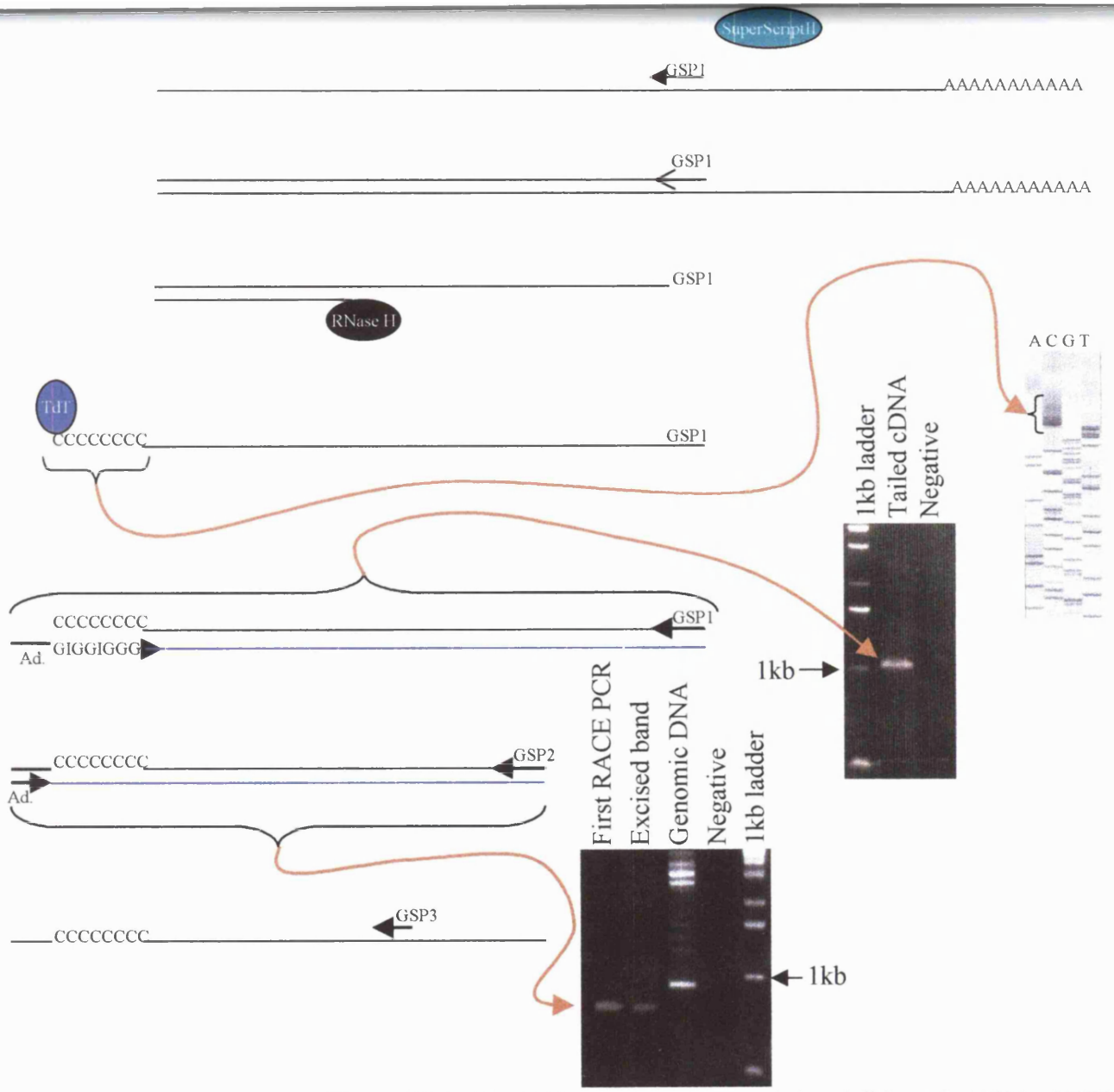
Figure IV.1: RT-PCR results using different primers, GSP1 & R1 (A), GSP1 & R2 (B) and GSP1 & R3 (C). Lane 1 is 100bp ladder. PCR templates in lanes 2 and 3 are cDNAs from two adult normal testis samples, in lane 4 is genomic DNA and in lane 5 is water (negative control).

genomic DNA contamination by two successive PCR amplifications (without a reverse transcription step) using primers designed from the gene G3PDH (sequence showed in Materials and Methods). PCR using the same G3PDH gene primers was also used to confirm the success of reverse transcriptase reactions (figure III.8). RT-PCR products were seen not only in the testis but also in the adrenal gland and the brain tissues (data not shown). Therefore, this DDRT-PCR fragment, TG2 (from colony 23), was a typical "false positive" result of DDRT-PCR. However, this fragment was characterised further since it was a novel human gene and secondly, the presence of many EST matches from databases were a helpful guideline in finding out about the expression and the full-length of the gene.

Using overlaps of ESTs (with 95 to 100% match) a sequence about 0.8kb was retrieved (figure IV.6). Primer R3 was designed from this sequence in a region that was present in ESTs from testis and contained an ATG codon and an open reading frame of 71 codons. PCR amplifications were carried out using primers R3 and GSP1 on cDNA and genomic DNA and a weak band of about 0.8-0.9 kb in cDNA and no band from genomic DNA (figure IV.1-C) supported the suggestion that there is an intron between the primers. The band was excised from the gel and the extracted DNA sequenced from both ends of the PCR product.

IV.2 5'RACE

5'RACE was applied to obtain full-length cDNA using human adult testis RNA and primers designed from the TG2 DDRT fragment. In brief, reverse transcription was carried out using a gene-specific primer, GSP1 and after adding a poly C tailing the same primer which was used for synthesis of cDNA (GSP1) and a complement primer to poly C with an adaptor, Abridged Anchor Primer, (see Methods, table II.1) were used for the first PCR amplification. The result of PCR was electrophoresed on a 1.2% agarose gel and one strong band about 1kb and another weak band of more than 2kb where visible (figure IV.2, number 5). Both bands were cut out from the gel and then DNAs were extracted and checked using a nested PCR reaction (primers GSP2 and R3) to confirm presence of the correct product. Only the strong band contained the full-length cDNA (figure IV.3).



1- Specific cDNA synthesis using mRNA or total RNA and Gene Specific Primer 1(GSP1).

2- Single strand cDNA was prepared and mRNA was removed by using RNase H.

3- Single strand cDNA was cleaned and all dNTP and primers were removed from the reaction.

4- A polyC tail using dCTP and TdT enzyme was made (it is seen in the sequencing result of the full-length cDNA).

5- The first RACE-PCR using GSP1 and PolyG with an adapter primer (5′RACE abridged primer).

6- The second RACE-PCR using GSP2 and the adapter primer (AUAP).

7-Cut out the PCR result band from an agarose gel for sequencing using GSP3.

Figure IV.2: 5`RACE procedure for obtaining full-length cDNA of TG2.

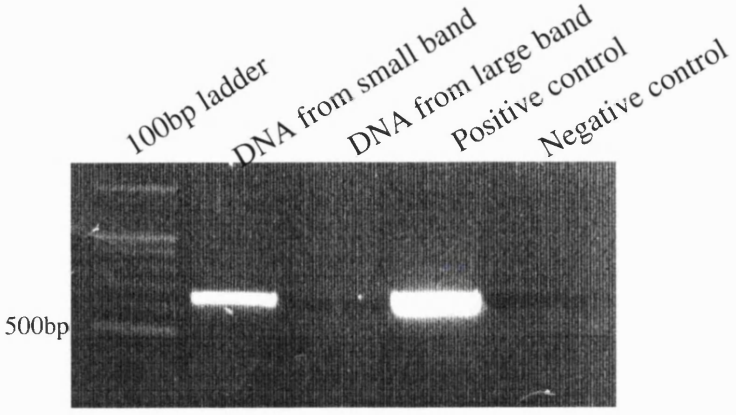


Figure IV.3: A nested PCR result using GSP2 and R3 primers and two bands (small, about 1kb and large, more than 2kb) from first 5´RACE-PCR. Templates in the first and second PCRs are extracted DNA from excised bands. The first 5´RACE was used for positive control and water was used for negative control.

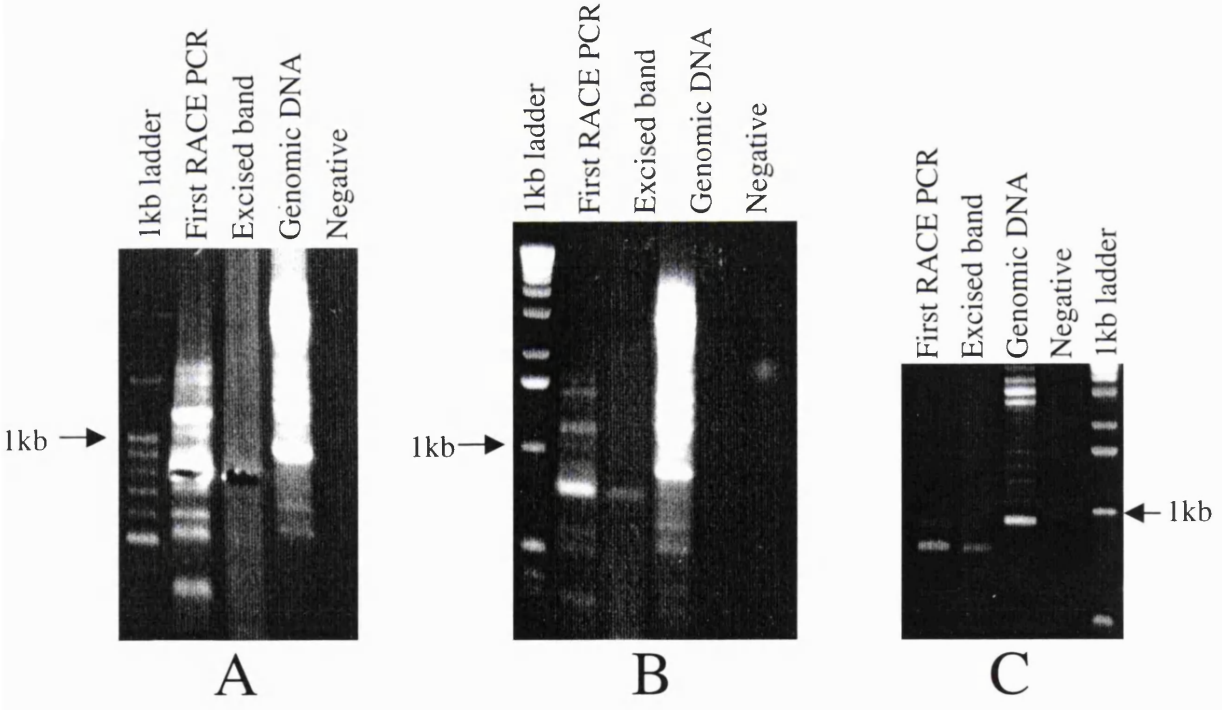
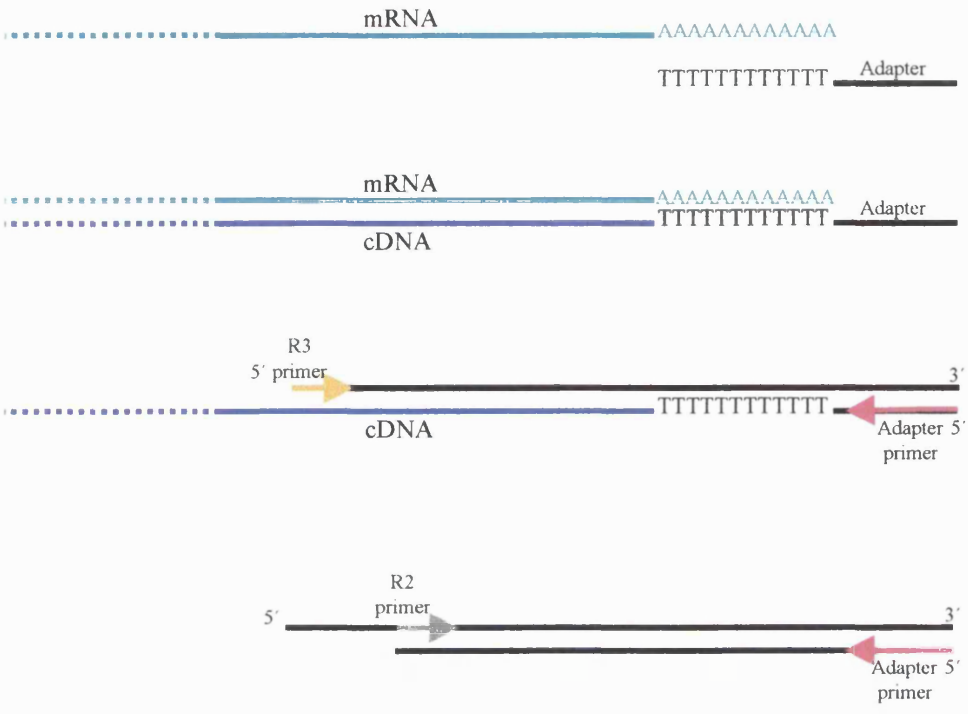


Figure IV.4: Optimization of second PCR of 5´RACE using GSP2 and AUAP primers with different annealing temperatures 61°C (A), 64°C (B) and 68°C (C). Bands were cut out all for sequencing (as shown in gel A).

However, sequence of the band (using primer GSP1) was not legible and indicated the presence of more than one product as a result of using the same primer, GSP1 for both cDNA synthesis and the first PCR. It was confirmed that one of those PCR products in the strong band was from the full-length cDNA. Therefore, a second, seminested PCR was carried out using as templates the first PCR product, the DNA excised from the gel and genomic DNA as a control with a second-specific primer (GSP2) and the AUAP primer. The second 5′RACE-PCR results showed an expected band (see figure IV.2, number 6) in both PCR products using the first 5′RACE-PCR and the DNA excised from the strong band. The procedure of 5′RACE was optimised to obtain the above result and results have been shown in figure IV.4. The second 5′RACE-PCR was carried out using different annealing temperatures 61°C, 64°C and 68°C to remove non-specific bands and results are seen in figure IV.4. The band from the second PCR was cut out to obtain a sequence of 1066 bp, the full-length cDNA (figure IV.10).

IV.3 3′RACE

3′Rapid Amplification of cDNA Ends (3'RACE) was carried out to get a complete picture of what is contained in the last exon of the mRNA (figure IV.5). At the first step mRNAs were converted into cDNA using reverse transcriptase enzyme and an adaptor-oligo$(dT)_{18}$VN primer (see table II.1). Specific cDNA was then amplified by PCR using R3 primer and an adapter primer (see table II.1) that was annealed to the polyA tail 3′ region. The first PCR product was run on 1.8% agarose gel and the expected band (about 950bp) was not visible but some small non-specific bands were seen (figure IV.5-A). There are two factors that may have contributed to this result. Firstly, the amplification was performed without intermediate organic extraction, ethanol precipitation or cleaning of cDNA. Secondly, the template for the first 3′RACE-PCR was a mixture of cDNAs made by the adapter-oligo$(dT)_{18}$VN primer. Therefore, the unexpected small bands were probably results of primer dimers, amplification of the adapter-oligo$(dT)_{18}$VN primer that remained in the reaction from the reverse transcription procedure and/or mispriming of the primers with cDNAs which where present in the reaction.



1- non-specific cDNAs were synthesized using mRNA or total RNA and oligo(dt)12 with an adapter

2- PCR amplification of single strand cDNA using the adapter primer and a gene specific primer (R3).

3- The second 3′RACE-PCR using another gene specific primer, R2 and the adapter primer (semi-nested PCR)

4-The PCR result band was cut out from an agarose gel for sequencing using either R2 or another internal primer.

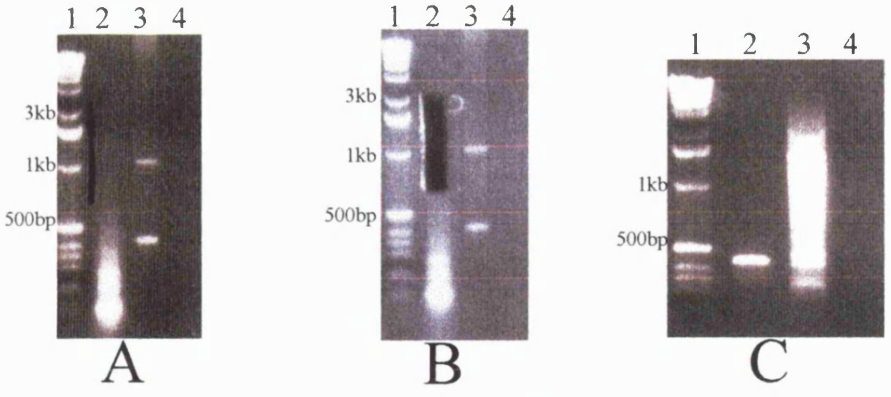


Figure IV.5: Schematic presentation of 3′RACE for GNG2. mRNA is coloured green, cDNA is blue, adapter primer is red and specific primers R2 & R3 are orange and grey, respectively.

The results of the first 3′RACE-PCR (A&B) and the second 3′RACE-PCR (C). The expected band is not visible in the lane A-2. The hole in the lane B-2 is an expected region for the non-visible band. The gel A and B are the same, before and after excision. Lane C-2 shows the expected band from the second 3′RACE-PCR. Lanes one are 1kb ladders (Gibco). Lanes 3 are 3′RACE amplification results using genomic DNA as a control in case of possibility of genomic DNA contamination of mRNA. Lanes 4 are from water as negative controls.

The second 3′RACE-PCR amplification, semi-nested PCR, was performed to generate a specific amplification product using R2 and the adapter primer. To avoid amplification of those non-specific bands, I found it advantageous to cut out a large area of the gel (between about 0.6 to 3kb, however, there was no visible band) (figure IV.5-B). DNA was extracted from the gel using the glass wool method (see the Methods, section II.2.1.6) and used as a template for the second 3′RACE-PCR to obtain a single band of 410bp (expected size) in a 1.8% agarose gel (figure IV.5-C). The band was cut out from the gel and sequenced using R2 and R1 primers.

Genomic DNA and ddH2O were used as controls for the first 3′RACE-PCR amplification and then the PCR products of these were also used as a template for controls in the second 3′RACE-PCR to eliminate unrelated bands which could be produced during repeated amplifications.

IV.4 The sequence analysis

The sequence contained an open reading frame (ORF) of 71 codons as shown in figure IV.9 and figure IV.10. The predicted amino acid sequence was used to search databases using the PBLAST program at NCBI (www.ncbi.nlm.nih.gov/blast/blast.cgi). This showed 100% homology to the bovine G protein γ6 subunit (Accession no. P16874, AAB34963) (whose name was later changed to γ2 (6)), and to the mouse and rat G protein γ2 subunit (Accession no.Q61013, D36204, AAB82554). Using the nucleic acid sequence, a search of EST databases using NBLAST provided many ESTs covering all regions of the transcript with almost 100% homology (figure IV.6 and table IV.1). Searching the nonredundant nucleic acid sequence database gave matches at about 90% identity to bovine (Accession no. J05071 and M37183), mouse (partial cDNA, Accession no. AF098489, AF098488 and U38496) and rat (partial cDNA, Accession no. AF022087) G protein γ2 subunits and a recent search also revealed another match at 100% identity to a human cDNA clone (Accession no. AK001024), 2124bp in length, derived from a 10 week old embryo.

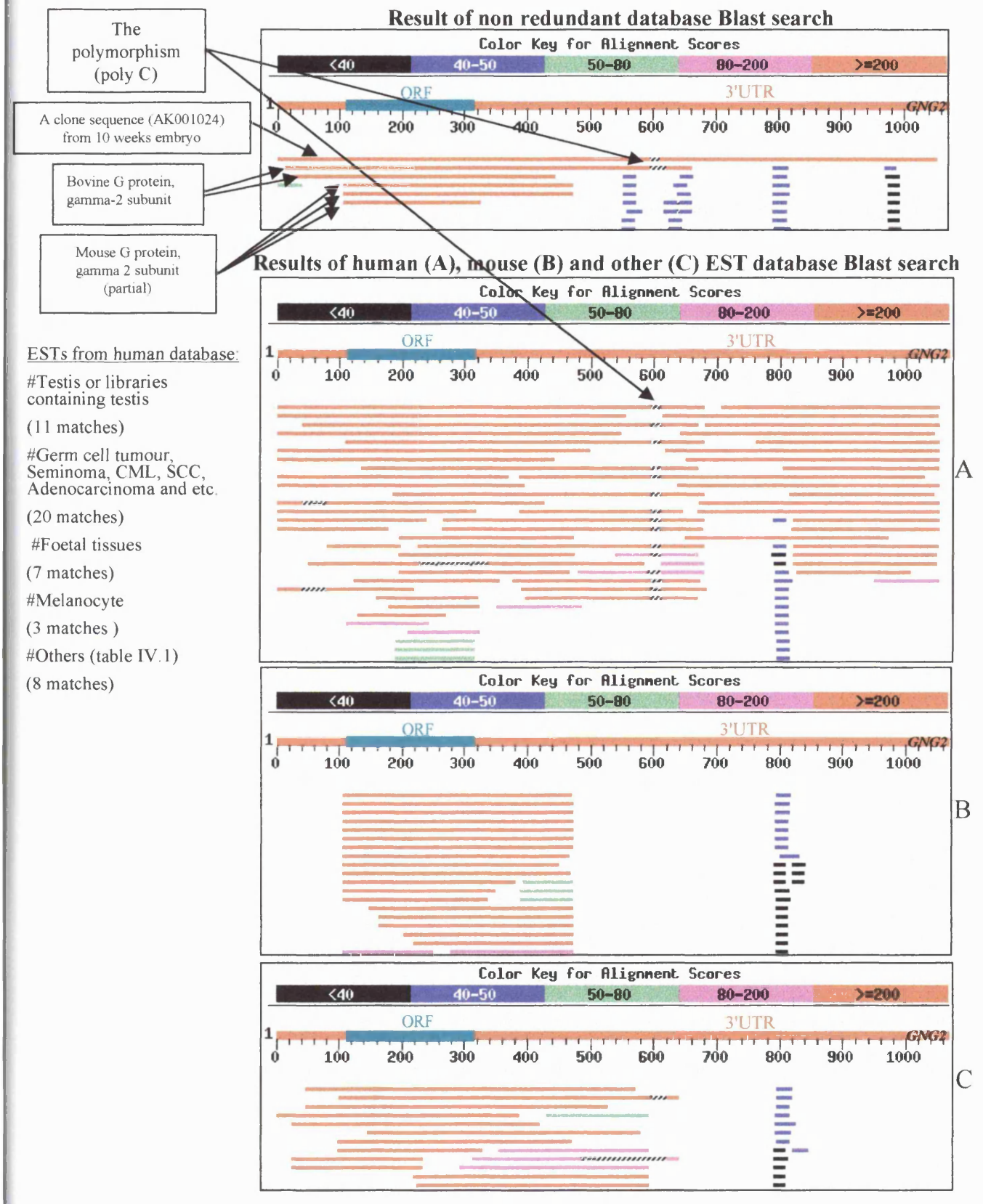


Figure IV.6: Blast search Results for GNG2 sequence.

Accession no.	Length of EST	Source	bp matched and Identity	Extension 5′	3′
AU118419	786	Whole embryo, 10 weeks	591/594 (99%) 63/68 (92%)	105	
BF571711	788	Skin, melanotic melanoma	554/558 (99%)	20	
AW273995	676	libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1)	541/556 (97%) 59/59 (100%)		
BE894025	871	melanotic melanoma	542/552 (98%)	105	
AA868346	579	Soares_testis_NHT	483/485 (99%) 68/68 (100%)		
AA160952	575	Stratagene hNT neuron	493/498 (98%),	32	
AW160428	536	fetal brain	442/443 (99%)	93	
AI818799	446	squamous cell carcinoma, poorly differentiated	436/439 (99%)		0
AI801855	557	Prostate (BPH)	453/458 (98%), 58/58 (100%)		
AA738022	487	pooled germ cell tumors	403/404 (99%)		80
AW955815	696	colon tumor metastasis	412/417 (98%) 208/208 (100%)		
AI963936	457	pooled germ cell tumors	427/438 (97%),		4
AW965369	654	colon tumor metastasis	397/402 (98%)		
AI828852	419	pooled germ cell tumors	411/417 (98%),		3
AI240106	500	libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1)	401/409 (98%) 68/68 (100%)		

Accession no.	Length of EST	Source	bp matched and Identity	Extension 5′	3′
AA776673	441	schizophrenic brain S-11 frontal lobe	390/394 (98%)	47	
AW135377	410	Pooled library	379/381 (99%)		28
AA723261	377	Soares_fetal_heart_NbHH19W	370/371 (99%)		5
BE245271	526	acute lymphoblastic leukemia	346/348 (99%) 40/40 (100%)	178	
N26108	426	Soares melanocyte 2NbHM	380/389 (97%),		28
AI979268	368	well-differentiated endometrial adenocarcinoma	344/346 (99%)		20
AI242192	423	libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1)	328/329 (99%), 68/68 (100%)		
AW957531	551	colon tumor metastasis	355/371 (95%)	113	
N93852	418	Soares_fetal_lung_NbHL19W	325/333 (97%), 66/66 (100%)		
AA586469	291	bulk germ cell seminoma	289/291 (99%)		0
AA333511	324	Embryo, 8 week	313/324 (96%),		
W40325	447	Soares_fetal_lung_NbHL19W	313/322 (97%),	126	
AI015456	459	Soares_testis_NHT	349/371 (94%), 68/68 (100%)		
AW806179	423	uterus	275/279 (98%),		
N36745	278	Soares melanocyte 2NbHM	273/278 (98%)		
AI223156	423	Soares_testis_NHT	247/247 (100%) 177/178 (99%)		

Accession no.	Length of EST	Source	bp matched and Identity	Extension 5′	3′
BE246434	379	acute lymphoblastic leukemia	238/239 (99%)	140	
BF941999	260	Thyroid, medullary carcinoma	230/231 (99%)		28
BE502454	314	pooled germ cell tumors	230/231 (99%)		
BF942131	229	Thyroid, medullary carcinoma	228/229 (99%)		0
AW075750	470	myeloid cells, 18 pooled CML cases	231/233 (99%)		0
BF942165	228	Thyroid, medullary carcinoma	227/228 (99%)		0
BF942147	228	Thyroid, medullary carcinoma	227/228 (99%)		0
AA382030	312	Activated T-cells	263/275 (95%),		
BF092397	247	testis_normal	228/232 (98%),		
AA332532	264	Embryo, 8 week	207/209 (99%) 34/35 (97%),		
AA609126	308	Soares_testis_NHT	204/217 (94%) 62/62 (100%)		
AA353597	291	Activated T-cells	174/177 (98%)	114	
BE774617	314	uterus	202/209 (96%), 71/73 (97%)		
AA732747	276	Soares_testis_NHT	188/198 (94%), 58/58 (100%)		
BE171926	454	head_neck	174/181 (96%),		
AA355501	286	Jurkat T-cells	139/140 (99%) 40/40 (100%)	146	

Accession no.	Length of EST	Source	bp matched and Identity	Extension 5′	3′
BF594583	163	colon tumor	156/163 (95%)		
H97059	145	Soares melanocyte 2NbHM	141/145 (97%)		
AA620960	435	Soares_testis_NHT	118/119 (99%)		

Table II.1: Human EST matches with the GNG2 sequences. The columns labelled 'extension' show if an EST begins upstream or ends downstream of the sequence deduced in this thesis. The columns give the number of additional nucleotides and blanks are internal.

IV.5 Nomenclature and the symbol of the gene, *TG2*

The system used for naming the DDRT-PCR fragments at the start of this study is mentioned in the Appendix, section A.2.1. This DDRT-PCR fragment was originally named TG2, before it was realised that it is a homologue of a known gene encoding the bovine G protein γ2 subunit. After this the gene symbol, ***GNG2*** (guanine nucleotide binding protein (G protein), gamma 2), was approved by the HUGO Nomenclature Committee. TG2, the name of the DDRT-PCR fragment that has been used in this study, is substituted by *GNG2* for future use in this thesis.

IV.6 Expression of *GNG2* in testis and other tissues

Expression of the gene was demonstrated using combinations of different primers and reverse transcription-PCR with different RNAs. RT-PCR amplification of RNA made from some foetal (16-20 weeks) and tumour tissues using GSP2 and R3 primers for *GNG2* indicated expression in a range of tissues (see figure IV.7-B&C). cDNA from a range of normal tissues which were previously used to confirm the expression of DDRT-PCR fragment, TG2, with the addition of white blood cells and lung were also used with these primers as shown in figure IV.7-A. A control RT-PCR amplification using Phosphoglucomutase 1 (PGM) primers was carried out to check all the RNAs that were used (figure IV.7). In addition, expression was seen in RNA extracted from a lymphoid line cell culture (figure IV.7-C). The following table is a summary of the expression pattern of GNG2:

Expression in normal tissues: Positive: Normal testis (1&2), adrenal gland, brain, white blood cell and lung Negative: Azoospermic testis, infertile testis, sperm, liver, muscle and prostate	Figure IV.7-A
Expression in foetal tissues: Positive: Limbs, stomach, intestine, kidney, spleen, lung, sternum and brain Negative: Testis and heart	Figure IV.7-B
Expression in tumour tissues: Positive: Thyroid tumour, parotid tumour and squamous cell carcinoma Negative: Granulation tissue from a surgical scar (very faint band)	Figure IV.7-C
Expression in cell lines: Positive: Lymphoid cell line	Figure IV.7-C

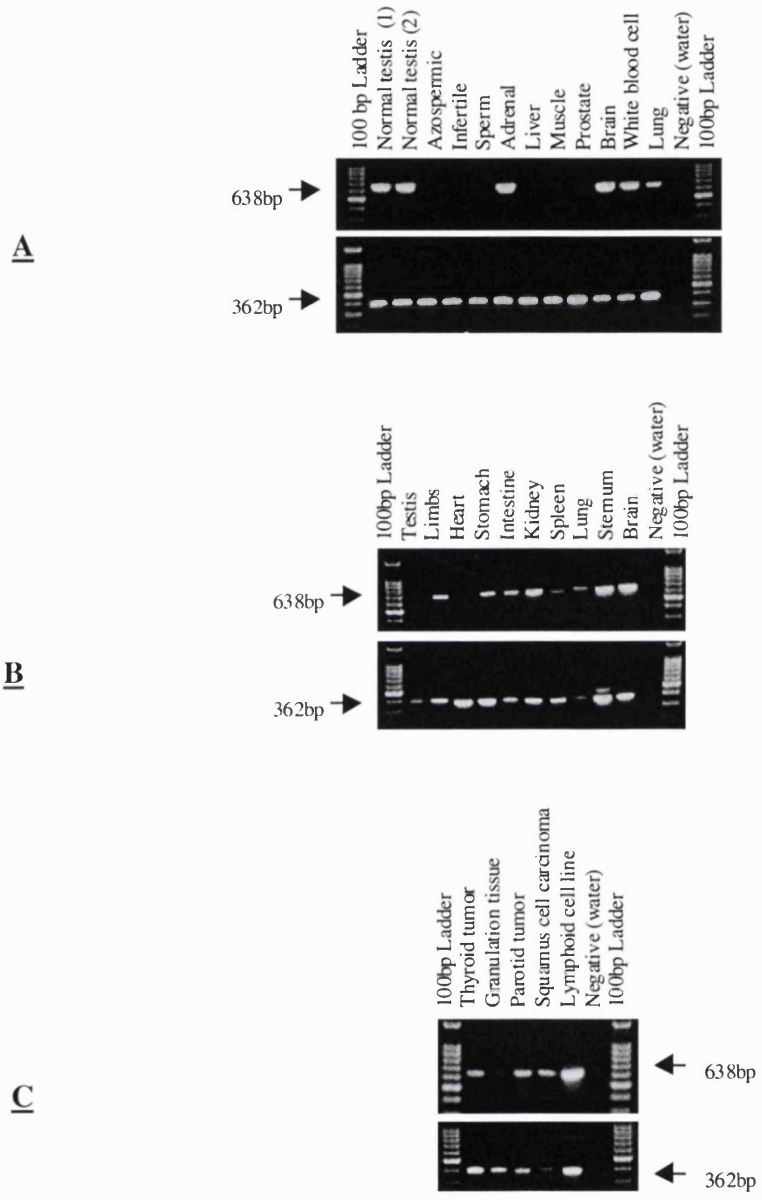


Figure IV.7: RT-PCR amplification of cDNA made from adult human tissues (A), fetal (about 16 weeks) tissues (B) and some tumor tissues (C). Amplification of these cDNAs with primers, GSP2, and R3, specific for GNG2 found the above expression pattern in the various tissues. The expected PCR product size (indicated by an arrow) was 638bp. A control RT-PCR amplification using Phosphoglucomutase 1 (PGM1) primers, indicated that all the cDNAs that were tested, gave a product of the correct size (362bp) as shown.

IV.7 5′end RT-PCRs

The cDNA sequence AK001024 was noticed in the database (see figure IV.6). It is derived from a 10 weeks embryo. It has an additional 105 nucleotide at the 5′end. A primer, R5 (see figure IV.9-A) was designed from the 5′flanking region of *GNG2* gene, and was located in the 5′end of the human cDNA clone AK001024. PCRs were carried out with primers R5&GSP5, R5&GSP6 and R5&GSP4 (see figure IV.9-A for location of the primers) using cDNAs made from different tissues and a genomic DNA control. RT-PCR results shown in figure IV.9 B, C and D show absence of the RT-PCR product from this 5′ region in the testis, but its presence in RT-PCR from WBC, brain and lymphoid cells demonstrating that this gene has alternative promoters.

IV.8 Screening of genomic library

An RT-PCR product between GSP2 and R3 (figure IV.10) was used to screen a PAC genomic library and identified eight positive clones (6I21, 17E9, 57H17, 150F24, 195M1, 229K1, 229M1, 233K14) (figure IV.8-A). The presence of this gene rather than a related gene was confirmed by PCR using GSP1 and R1 primers (located in the 3′ UTR) (figure IV.8-B). The PAC clone 229M1 was partially sequenced using GSP2, GSP3, IF1, IR1, GSP5, R3, IF2, GSP4 and R4 (location of primers is shown in figure IV.9-A). No sequencing results were found using PAC clone 229M1 and primers GSP4 and R4. Therefore, this PAC was shown to contain the 3′UTR, ORF, intron 3 and part of the second intron of the gene. Searching "gss database"(Genome Survey Sequence database which contains BAC end sequences) for the 3′end of intron 2 (figure IV.9-A) using sequence obtained from the PAC clone 229M1 revealed a BAC genomic clone (2649J15) containing this gene. Negative results in figure IV.9-D from PAC 229M1 and BAC 2649J15 suggest that they can not include the 5′end of the *GNG2* gene. At the time of this experiment there was no genomic clone containing this gene in the databases. Searching the recently released sequence database, High-Throughput Genome Sequences (htgs), at NCBI revealed a working draft sequence, an incomplete and unlocated clone (Accession no. AC025872) which contained exon 3, intron 3 and exon 4 of *GNG2*

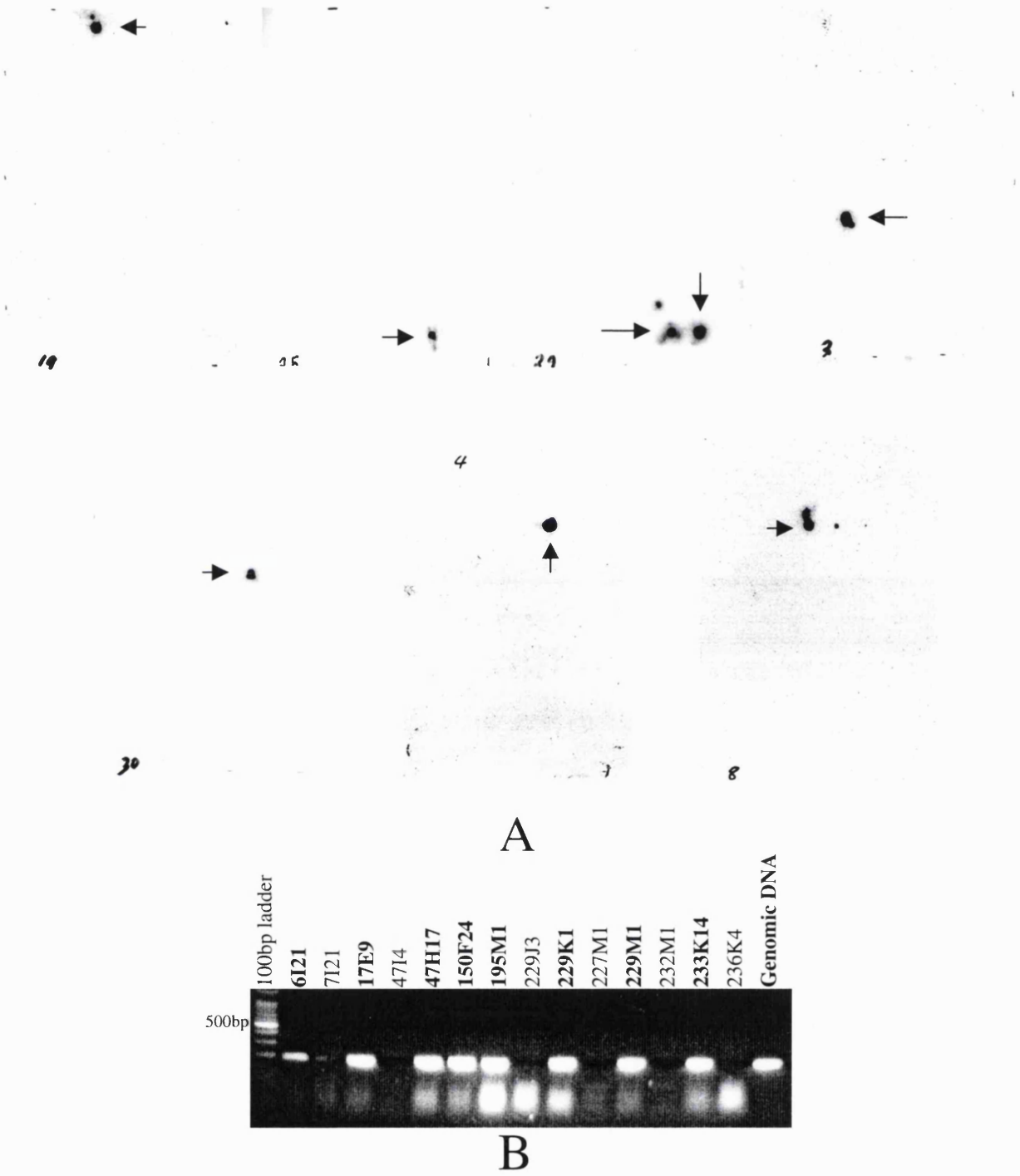


Figure IV.8: A) Eight positive PACs were identified by hybridisation of PCR products to the PAC library filters (41 filters). B) All eight PAC clones were confirmed by PCR. Clones which are shown not in 'bold' were either close to the real clones on the filters and indistinguishable (using X-ray films) or were false positive from the radioactive screen.

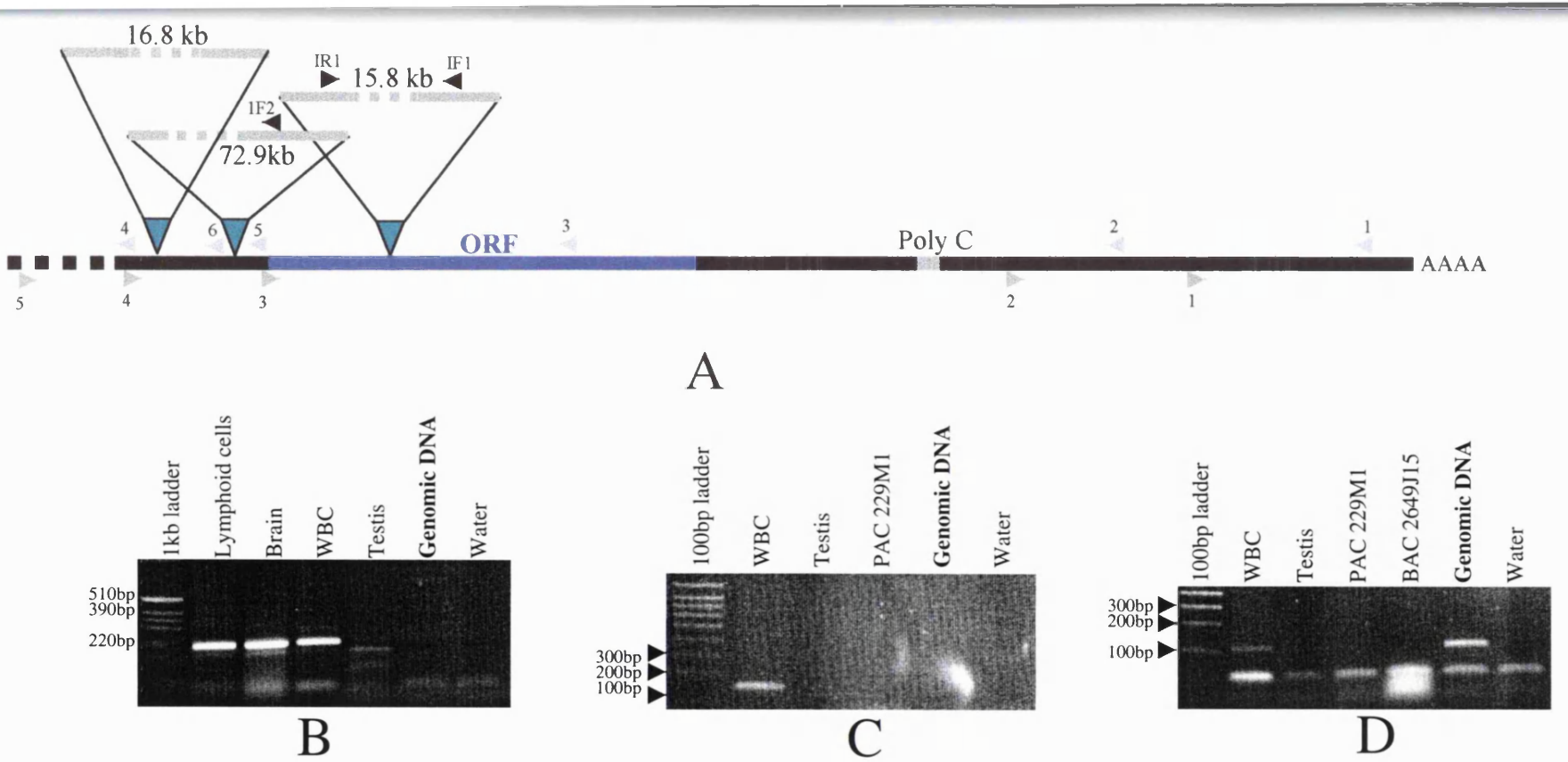


Figure IV.9: A) Schematic of complete GNG2 gene. Locations of GSP ◄ primers and R ► primers are shown on the gene. Three primers which were used for sequencing or amplifications were located in introns. The ORF, Polymorphic region, polyA tail and the location of introns have been indicated on the gene. The dashed line indicate the 5′flanking region.

B, C and D are results of RT-PCR using R5&GSP5, R5&GSP6 and R5&GSP4 primers, respectively. The shorter bands in D are primer dimers and the band of 117 bp in length is the correct size band.

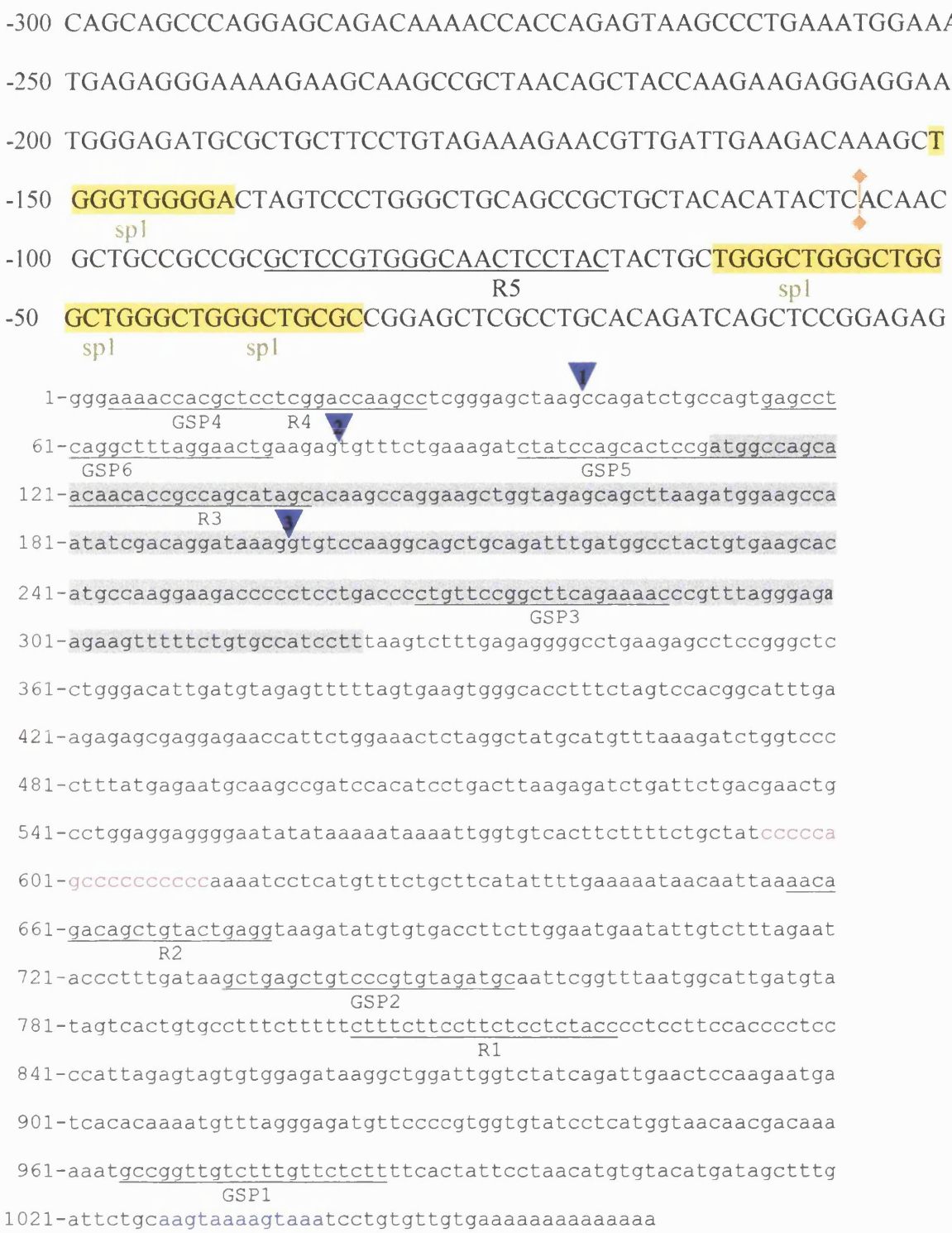


Figure IV.10: Nucleotide sequence and the ORF (grey) of the human gamma 2 subunit of G protein. Primers are mentioned and underlined. 5`flanking region of the gene is in upper case and the putative transcription factor sites are indicated in yellow. A polyadenylation signal sequence -like is blue sequence and the polymorphic site is red. Exon/intron boundaries in the sequence are indicated by blue triangles. shows the 5'end of cDNA clone AK001024.

gene. Another working draft sequence (Accession no. AL358333) was released during the last few months containing the whole gene. Two introns were found in similar positions to the mouse γ2 gene (Downes et al, 1998) (figure IV.9-A), one is 72.9kb in the 5′ UTR and one of 15.8kb is located within the ORF However, the mouse full-length cDNA of the γ2 subunit has not been published or registered in GenBank. Another intron is 16.8kb long and is located just 42bp before the second intron detected from this clone sequence which was not similar to mouse intron positions. The sequences of the predicted splice junctions conform to the consensus.

IV.9 Chromosomal localisation of GNG2

The chromosomal location of the human γ2 gene, *GNG2*, was determined by three methods. Firstly by PCR amplification, using gene specific primers GSP1 and R1, of DNA from a panel of hamster or mouse cell hybrids that each contains a single human chromosome. All hybrids gave rodent products but a specific human product of 181bp was seen only in the chromosome 14 cell line (PCR results not shown but summary of results are shown in chapter 3, table III.4). Secondly, to refine the localisation of the γ2 gene, PAC 229M1 was used for Fluorescence in situ hybridisation to human metaphase chromosomes. A specific signal was observed at chromosome band 14q21 (figure IV.11). Thirdly, the draft browser program at http://genome.ucsc.edu/goldenPath/septTracks.html was used to examine the BAC genomic clone with accession no. AL35833 (containing the entire *GNG2* gene). It places the clone at chromosome position 14q21 and the result also displayed some additional information that is shown in figure IV.12.

IV.10 Polymorphism

In the course of this study, the nucleotide sequence of the gene was determined from at least 4 independent sources. Nucleotides 602 to 611 in the 3′ UTR were found to be polymorphic. One allele of 13 Cs was sequenced from the PAC genomic clone 229M1 (data not shown), alleles of 9 and 10 Cs were observed in sequence derived by RT-PCR (figure IV.13). Comparing EST sequences at this point C_{10}, $C_{11}$and C_{13} alleles were seen. Figure IV.10 shows the polyC in the

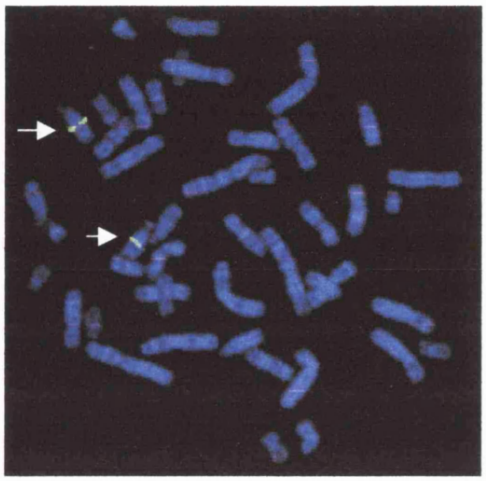


Figure IV.11: Localisation of PAC 229M1 (containing GNG2) to band 14q21 by FISH.

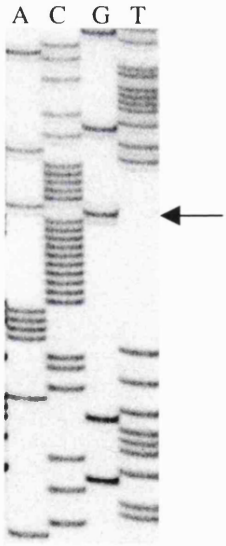


Figure IV.13: Sequencing of the polymorphic region from RT-PCR product shows two alleles in the PCR product, all bands above the arrow are duplicated. The shorter allele seems to be expressed at a lower level than the longer for an unknown reason.

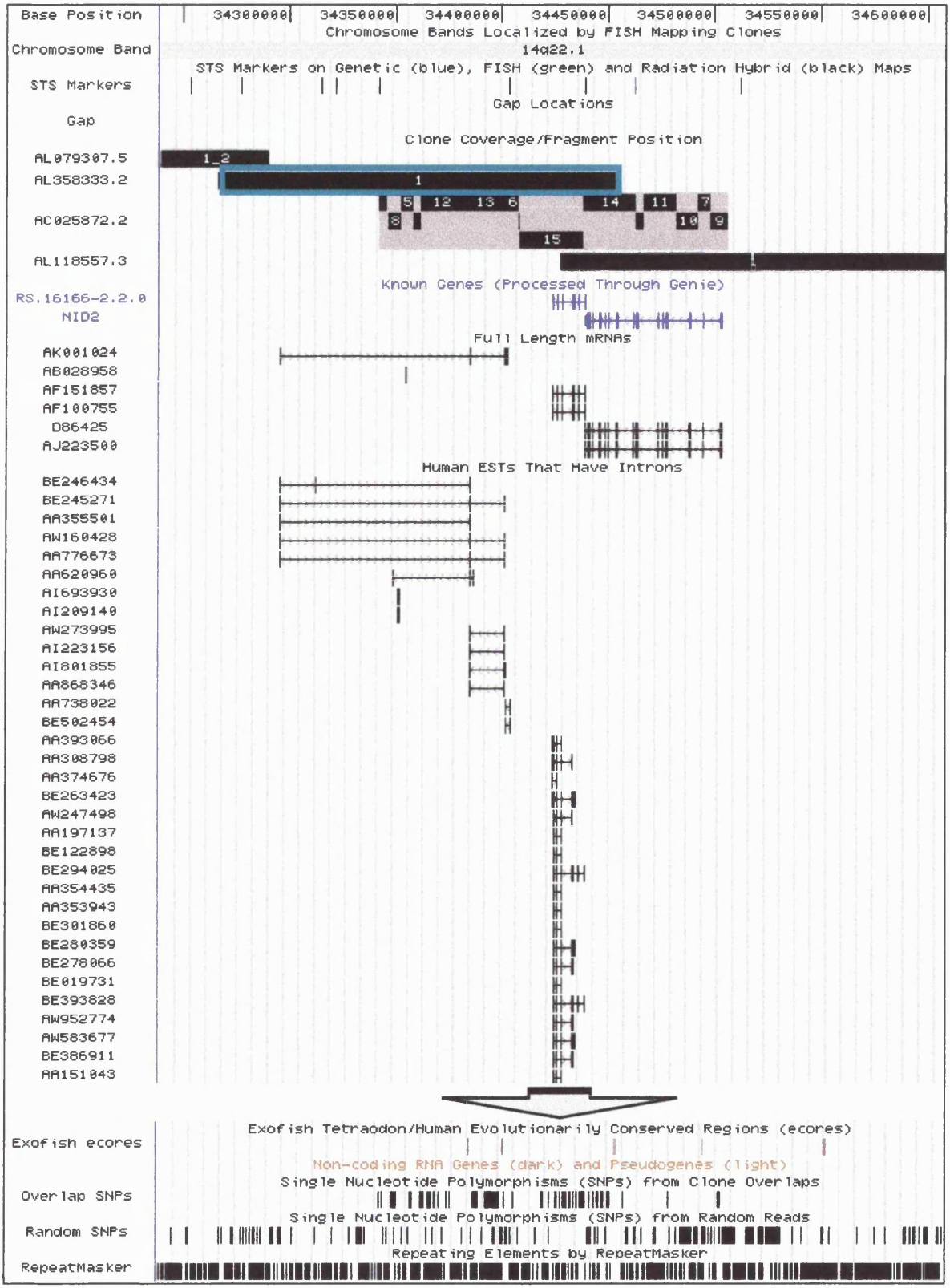


Figure IV.12: The sequenced human genomic clone AL358333 (green) was used to display the above draft. Some ESTs have been deleted from this figure due to limitation of space.

sequence and this region also has been indicated in the result of databases search (figure IV.6). In addition, there were several unrelated sequences at this point in clones which were otherwise >95% identical to our sequence. It is possible that the run of C residues has led to sequencing artefacts in some of the machine-read ESTs.

IV.11 Discussion

IV.11.1 Expression pattern and function

Guanine nucleotide binding (G) proteins interact with membrane bound receptor proteins and enable the activation or inhibition of intracellular enzymes, transcription factors and also ion channels. Most neurotransmitters and a large number of hormone signalling pathways in mammals have effects which are mediated by heterotrimeric G proteins (Gilman 1987) and members of the G protein super family play very important roles in determining the specificity and temporal characteristics of the cellular responses to external signals. Testis is an organ that not only produces a large number of hormone signals but also is the target for many signals (signalling system in the testis discussed in chapter I, section). G proteins are composed of α, β and γ subunits (figure IV.14). The α subunit is a GTPase. The α subunit is activated by receptor catalysed guanine nucleotide exchange of GTP for GDP bound to the α subunit (figure IV.15). This causes a structural change in the switch region of the α subunit-GTP complex which lowers the α subunit's affinity for the $\beta\gamma$ complex and leads to their dissociation. Separation of α and $\beta\gamma$ subunits allows them both to interact independently with, and activate a number of effector molecules. (Clapham and Neer 1997). As shown in figure IV.15 effectors may be enzymes, ion channels, transcription factors and MAP kinase pathway. The latter is related to cell cycle control and differentiation and is explained later (also see chapter I, section 1.3.2). 16 mammalian α subunits have been identified and categorised into four subfamilies (αs, αi, αq and α12) (figure IV.16) based on their amino acid sequences (Simon, *et al*, 1991). It has been shown that the β and γ subunits act together as a heterodimer forming a functional unit which is not

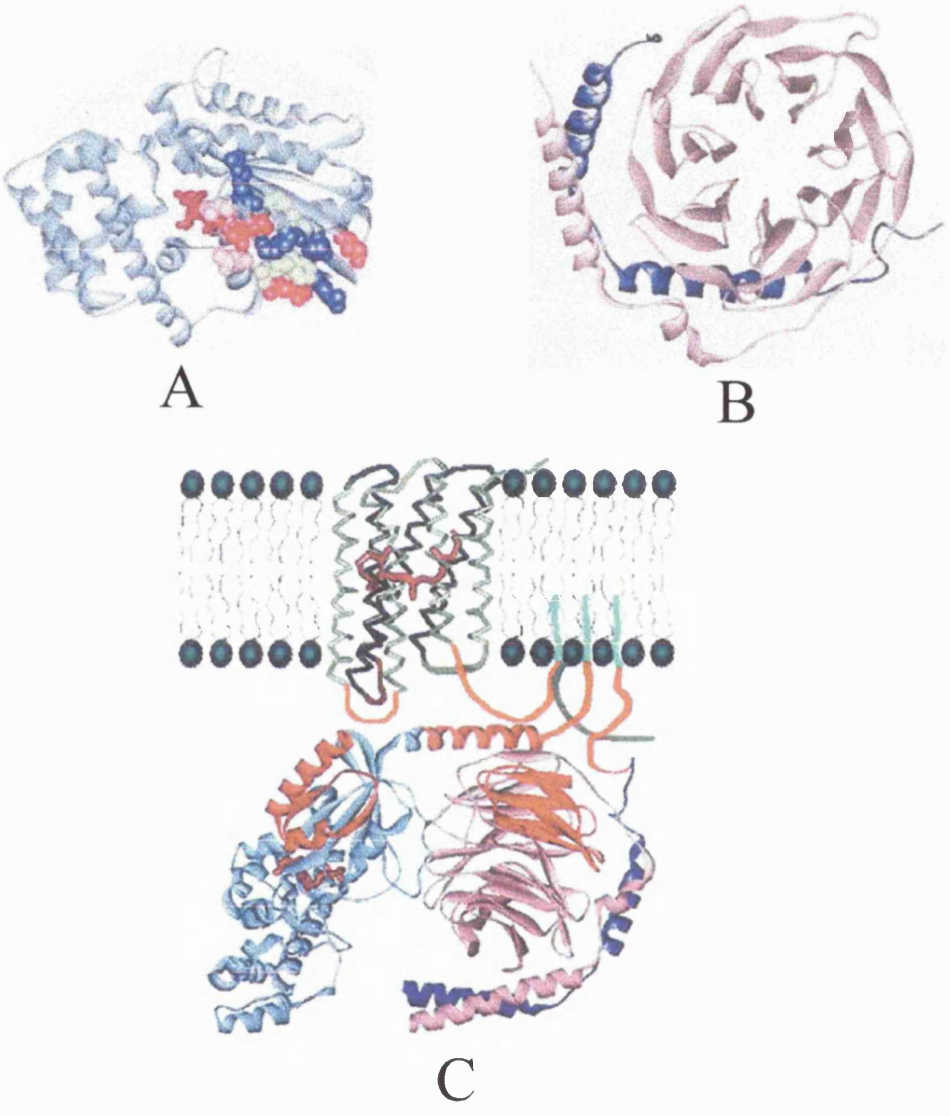


Figure IV.14: Structure of G protein subunits. The Gα subunit is medium blue (A), the Gβ subunit is pink, and the Gγ is blue (B= βγs subunits). Picture C shows the configuration of the α helices of rhodopsin (red and green) that interact with a heterotrimeric G protein (Schertler and Baldwin et al.).

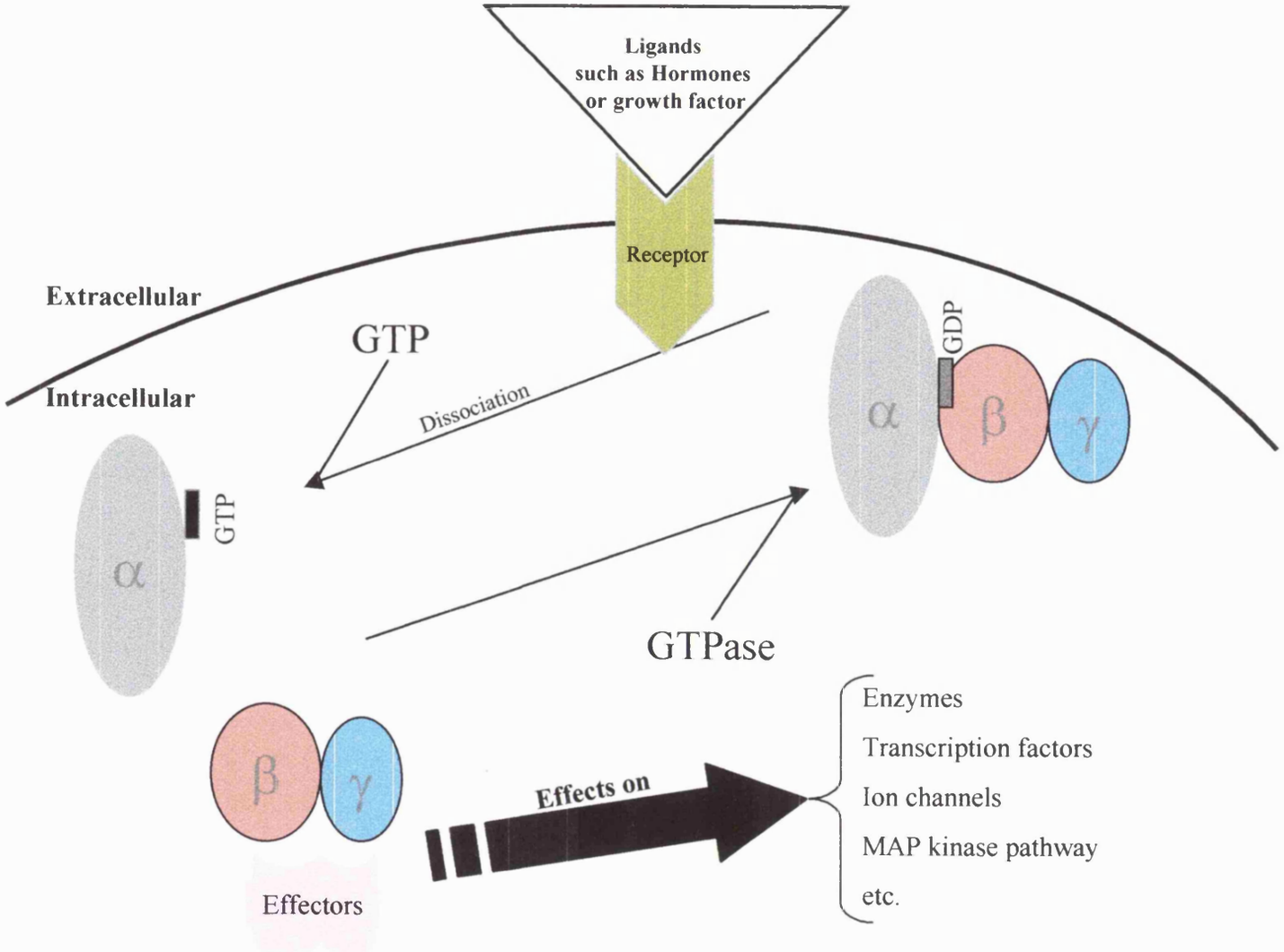


Figure IV.15: The process of signal transdution from extracellular receptors to intracellular effectors by G Proteins.

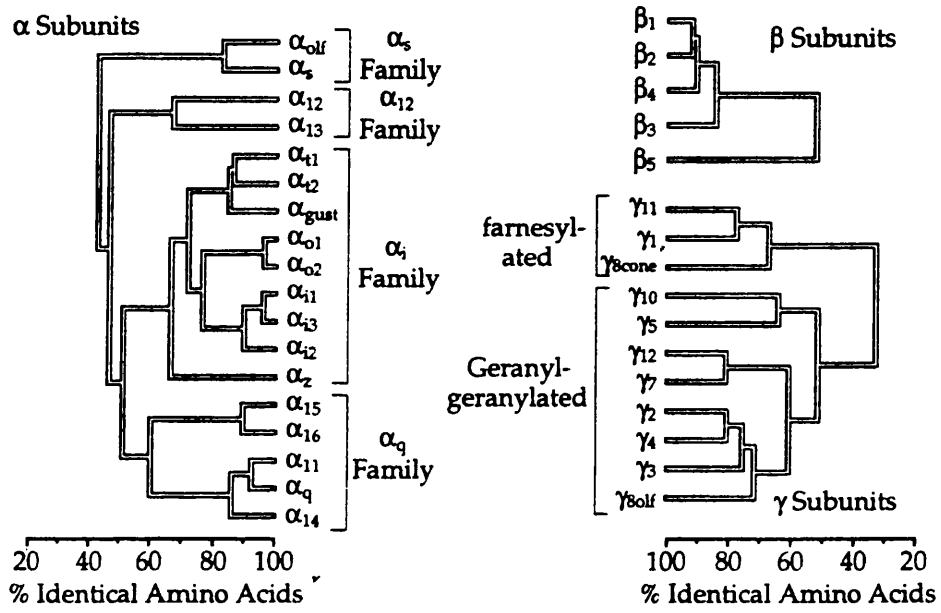


Figure IV.16: Diversity in G α, β and γ subunits proteins. Family groups are generated as a result of identity of protein sequences (figure copied from review by Hildebrandt 1997).

dissociable except by denaturation and which also enhances receptor interactions with α subunits. Direct binding interactions between receptors and the βγ subunit have also been reported (Taylor et al., 1996, Yasuda et al., 1996). In addition, the C terminal region of the γ subunit of G proteins has been shown to be a site of membrane attachment thought to be involved in receptor coupling and specificity (Yasuda et al., 1996, Resh 1996). Different G protein subunits provide a variety of combinations of signal transduction options and regulatory patterns depending on what subtypes of metabolic effectors and G proteins are expressed in any cell or tissue (Clapham and Neer 1997). Although many different α, β and γ subunits may be present in a cell only some specific combinations will form a heterotrimeric G protein complex because of variation in the affinities of different forms of β and γ for each other. Consequently, the mechanisms that control the formation of specific forms of βγ complex in a cell will also regulate properties of that cell in response to external signals.

Five distinct β subunits have been described which show about 90% identity of protein sequence and twelve γ subunits with 30% to 77% identity (Ray et al., 1995, Clapham and Neer, 1997) (figure IV.16). As well as their increased variation in amino acid sequence, the γ subunits also show more variation than the β subunits in their tissue distribution. The greater diversity of γ subunits compared to α or β subunits suggests that the functional specificity of the βγ heterodimer is determined to a large extent by its γ subunit (Wilcox et al., 1995). It has been shown that the γ subunit may interact directly with the receptor (Kisselev and Gautam 1993). If different γ subunit types specifically interacted with various receptors, this might determine a primary level of control over which G protein would be active in a cell (Yan and Gautam 1996). However, the exact role of any γ subunit is not known precisely.

The patterns of protein expression of various γ subunits have been determined by SDS-PAGE and immunoblot analysis. Based on both mRNA and protein expression, it has been reported that γ1, γ3 and γ4 subunits are selectively expressed and enriched in brain (Cali J.J. et al. 1992, Tao, L. et al 1993) and γc or γ8 subunit

is restricted in expression to just one tissue or cell line type (Ryba and Tirindelli 1995, Ong et al., 1995), whereas γ5, γ7, γ10, γ11 and γ12, subunits are distributed in many tissues (Ray, K. et al. 1995, Cali, J.J., 1992, Fisher, K. J. 1992, Asano et al., 1995, Cali et al., 1992, Gautam et al., 1998, Ray et al., 1995). The γ3 mRNA is also expressed at low level in testis (Downes G. 1998). In table IV.2 the patterns of GNG gene expression, have been summarised. Bovine γ2 had previously shown to be expressed in bovine brain and adrenal gland (Robishaw JD et al. 1989, Cali JJ et al. 1992), and this current human study demonstrated that it was also highly expressed in human adult testis, adrenal gland, brain, leukocyte and lung as well as a range of foetal and tumour tissues (figure IV.7). In addition, it has been observed that human ESTs homologous to γ2 have been predominantly isolated from libraries constructed from human testis, seminoma, germ cell tumour and foetal tissues with a smaller proportion of ESTs derived from other tumours such as CML, adenocarcinoma and squamous cell carcinoma (table IV.1). The common factor here is that all sources contain a large number of dividing cells.

It has been reported in many published 5′RACE results and also in my previous 5′RACE experiment (chapter III, figure III.10) that no result is often seen in the first 5′RACE-PCR product. However, it was seen in the result of TG2 5RACE. Therefore, it can be expected that the level of expression of the gene is high.

As I mentioned above and also in chapter I, section 1.3, there are a lot of studies to support a role for G protein coupled receptors via G proteins in normal proliferation and abnormal growth (Dhanasekaran N. and Prasad M.V. 1998, Clapham and Neer 1997). Two classes of receptors can trigger mitogenic pathways in cells: tyrosine kinase receptors and G protein coupled receptors. Both of these receptor-mediated pathways can stimulate the mitogen-activated protein kinase (MAPK) cascade through the activation of Ras (Van Biesen T. et al. 1995). The MAPK cascade is activated by extracellular factors such as growth hormones and proto-oncogene products (Cobb M.H.et al 1995) and is implicated in both regulated and deregulated cell proliferation (induced by growth factors and Ras transformation, etc.) as well as the control of differentiation (Schonwasser D.C. et

Gene symbol	Encoded protein	Source	Tissue distribution	Genome location	Human genomic location	References
GNGT1	γ1	Human	Retinal rod cells		[#1] 7q21-3, 7q31-32	(Hs. 73112)& (Sherer *et al.*, 1996)
GNGT2	γ cone	Human	Retinal cone cells		7q21	(Ong *et al.*, 1997)
GNG2	γ2	Bovine	Predominantly in brain, adrenal	14	[#2] **14q21**	
GNG3	γ3	Mouse	Abundant in brain, low in testis	19	11	(Hs. 179915)
GNG4	γ4	Mouse	Brain and other tissues	13	1	(Hs. 32976)
GNG5	γ5	Human	Ubiquitous		1p22	(Hs. 5322)
GNG7	γ7	Human	Brain and other tissues		19p or [#3] 12q	(Hs. 127828)& (Watson *et al.*, 1991)
GNG8	γ8	Rat	Olfactory neuroepitelial, low in brain	----		
GNG10	γ10	Human	Ubiquitous		[#1] 15 & 9	(Hs. 79126)& (Downes *et al.*, 1999)
GNG11	γ11	Human	Several		7q31-32	
GNG12	γ12	Bovine	Ubiquitous	----	1p31-33	(Hs. 118520)
GNG13	γ13					

Table IV.2: The mammalian G protein γsubunits genes, distribution and location. [#1] There is a discrepancy between the data in the publication and those from database. [#2] Mapped in this thesis. [#3] Loci predicted based upon mouse mapping data.

al 1998). MAP kinases phosphorylate and activate nuclear transcription factors that in turn activate genes controlling the cell cycle. It has been shown by Mattingly and Macara (1996) that β1γ2 can cause a constitutive activation in Ras exchange activity which can in turn lead to activation of the MAP kinase. Experimentally, the β1γ2 dimer markedly activates the MAPK pathway in COS cells (Zhang S. et al. 1996).

Testis is a proliferative tissue with cell division and differentiation happening during different stages of spermatogenesis and sperm production. G protein-coupled receptors and probably G protein containing β1γ2 may potentially affect many aspects of germ cell physiology in the testis, including cell growth, proliferation, and differentiation during stages of spermatogenesis by activation of the Ras signaling pathway and MAP Kinase cascade. Additionally, numerous receptors related to sex and growth hormone function are involved in spermatogenesis and the role of G proteins during this signal transduction has also been demonstrated (see chapter I, sections 1.3.1 and 1.3.2). Given the absence of the gene product from the testis of two infertile males, it will be interesting to investigate this further with a view to considering the role of human γ2 (with other subunits of G protein) in male fertility and testicular tumour formation.

IV.11.2 cDNA sequence analysis

IV.11.2.1 3′end

Using two different methods the 3′end was confirmed. Firstly, the *GNG2* sequence originally is from DDRT-PCR in which 3′ends are captured using oligo(dT) anchor primers. Secondly, the 3′end of the sequence was proved by 3′RACE. However, the cDNA clone (Accession no. AK001024) from the database has a 3′end that, in the genomic clone AL358333, is 967 nucleotides downstream of this *GNG2* gene end (see section A.6 in Appendix).

Two putative polyadenylation signals "AAGTAAAAGTAAA" are present 12bp upstream from the polyadenylation tail for GNG2, but which differ from the well-known consensus polyadenylation signal sequence AAUAAA. Previously, it

was demonstrated that single base substitutions in AAUAAA reduce the efficiency of cleavage after transcription and also interfere with the addition of poly(A) to RNAs at the poly(A) addition site (Sheets *et al.*, 1990). However, some testis specific genes with modified polyadenylation signals show a strong expression despite these modifications (Nayernia, *et al.*, 1999; Burfeind and Hoyer-Fender, 1991; Gastman *et al.*, 1993). No putative polyadenylation signals were found at the 3′end of the clone AK001024.

IV.11.2.2 5′end

Exon one in the clone AK001024 has 105 nucleotides more than this exon in the *GNG2* gene sequence at the 5′end (figure IV.10). The actual 5′end of the GNG2 gene was obtained by 5′RACE is seen in figure IV.2. In addition RT-PCR using R5, located in the 105 nucleotide 5′flanking region, demonstrated that this transcript is present in a lymphoid cell line, brain and WBC but not in the testis. This result suggests the possibility of tissue specific alternative promotor for this gene. In addition, table IV.1 shows that some ESTs extend from the 5′ end but none of these originate from testis.

Interestingly, searching the 5′flanking region for transcription factor binding sites using a program at the web site: http://searchlauncher.bcm.tmc.edu/seq-search/gene-search.html revealed four Sp1 sites and a GC-rich binding zinc finger region (see figure IV.10). It is suggested that promoters carrying GC-rich sequences are activated by Sp1 or Sp1-like factors through their interaction with proteins from the basal transcriptional machinery to upregulate gene expression (review by Cook *et al.*, 1999). The Sp1 transcription factor is a testis-"enriched" protein that is absent from most somatic tissues. The highest levels of Sp1 expression during spermatogenesis correlate with maximum expression of the human testis-specific lactate dehydrogenase c gene (Bonny *et al.*, 1998) (see chapter I section 3.1.2). Three putative Sp1 binding sites are located between -33bp to -63bp upstream of the 5′end of the GNG2 gene sequence and one is at -141 upstream of the 5′end of the GNG2 (-36bp of the 5′flanking region of the clone AK001024)(see figure IV.10).

Details of methylation and the role of CpG islands during spermatogenesis are discussed in chapter I section 3.2. A published CpG island sequence, 235 nucleotides, (Accession no. Z63227) (Cross et al., 1994) was aligned with the 5'flanking region and exon one of the *GNG2* gene:

```
          -104
GNG2 sequence CAAC  CTGCCGCCGCGCTCCGTGGGCAACTCCTACTACTGCTGGGCTGGGC
CpG sequence  CAAC  CTGCCGCCGCGCTCCGTGGGCAACTCCTACTACTGCTGGGCTGGGC
           59

          -54
GNG2 sequence CTGGGCTGGGCTGG  C  GCGCCGGAGCTCGCCTGCACAGATCAGCTCCGG
CpG sequence  CTGGGCTGGGCTGG  C  GCGCCGGAGCTCGCCTGCACAGATCAGCTCCGG
          109

           -4      0                                              exon/intron
GNG2 sequence AGAG][GGGAAAACCACGCTCCTCGGACCAAGCCTCGGGAGCTAAG/   A  AT
CpG sequence  AGAG][GGGAAAACCACGCTCCTCGGACCAAGCCTCGGGAGCTAAG/   A  AT
          159
```

CpG islands are enriched in the dinucleotide CpG and are frequently associated with housekeeping genes (Oi and sit 2000; Siegfried *et al.*, 1999) as well as some tissue-specific genes in the mammalian genome (kundu and Rao, 1999; Cho and Hedrick, 1997) which supports the idea that *GNG2* might be expressed in many tissues.

IV.11.3 Gene structure and localisation

This study showed that the human *GNG2* gene is similar to other reported *GNG* gene structures, both human and from other species. All reported genes have an intron about 80 - 100bp into the ORF and many have an intron 1 – 30 bp upstream of the start codon (Liu and Aronson, 1998, and Downes et al., 1998) but here there is one more intron (figure IV.9-A and figure IV.10).

I have shown that the human *GNG2* gene maps to 14q21. This is surprising because it is not the position which would have been predicted from the mapping of the murine *gng2* gene to mouse chromosome 14 in a region which corresponds to

human chromosome band 10q24 (Downes, et al. 1999; minireview by Downes and Gautam, 1999).

The genomic location of some of the known isoforms of γ subunits described so far have been determined (Gautam et al., 1998) (table IV.2). The human chromosomal location for the γ1, γc and γ5 subunits have been published respectively as chromosomes 17q21.3, 17q21 and 1p22 (Scherer et al., 1996, Ong et al., 1995, Ahmad 1995). Also the map localization of γ4 (GNG4 to ch.1), γ10 (GNG10 to ch.9&15) and γ11 (GNG11 to ch.7q31-32) is reported in the UniGene Databases at the National Center for Biological Information, http://www.ncbi.nlm.nih.gov/UniGene/index.html (Schuler et al., 1996). The γc, γ2, γ3, γ4 and γ12 subunits have also been mapped within the mouse genome to chromosomes 11, 14, 19, 13 and 6 respectively (Downes et al., 1998, Kalyanaraman et al., 1998, Downes et al., 1999). Gng2-ps1 is a mouse γ2 pseudogene (EMBL / GeneBank Accession No. AF098487) which has been mapped to the proximal region of mouse chromosome 4 (Downes et al., 1999).

In summary, I have described the cDNA sequence and complete genomic structure of the human γ2 subunit gene (which consists of four exons and three introns) and I have mapped it to chromosome band 14q21.

Chapter V

Identification, characterisation and mapping of a human testis-specific gene or pseudogene, *C4ORF2*

One of the results from DDRT-PCR was a fragment generated by using oligo dT-VG and OP3 primers (see table III.2 and figure V.1). The process of DDRT-PCR and directly sequencing from the re-amplified DDRT-PCR excised band, T2G3, was explained in chapter III in detail. In this chapter, the procedures of obtaining full-length cDNA, characterisation and localisation of the gene, after step 9 of the figure III.1 are discussed.

V.1 Expression pattern

T2G3 was sequenced and compared to sequences in GenBank, EST, repeat and other databases by using the BLAST program at www.ncbi.nlm.nih.gov/cgi-bin/BLAST. It was not present in any database screened. As mentioned previously, one disadvantage of DDRT-PCR has been false positive results so, to confirm the expression pattern of T2G3, RT-PCR was carried out to re-examine the expression of the DDRT cDNAs. Specific primers, GSP1 and R1, were designed from the DDRT fragment sequence and used to perform RT-PCR using RNA from two normal testes, two different infertile patients' testes, sperm, adrenal gland, liver, muscle, prostate and brain. Positive results were seen with normal testis and also, more weakly, with sperm (figure V.2).

V.2 To obtain full length cDNA

In an attempt to obtain full length cDNA of the DDRT 3'UTR fragment T2G3, a 5'RACE experiment (Methods section II.2.6) was carried out by converting human adult testis RNA to cDNA using a gene-specific primer, R1. PCR results from the second amplification (second RACE-PCR) using different dilutions (1/20, 1/200 and 1/500) of the first RT-PCR reaction showed two bands in the gel (figure V.3, lanes 4, 5 and 6). Both bands were excised and sequenced with the results shown in figures V.3 and V.4 but there was no similarity between the two. The

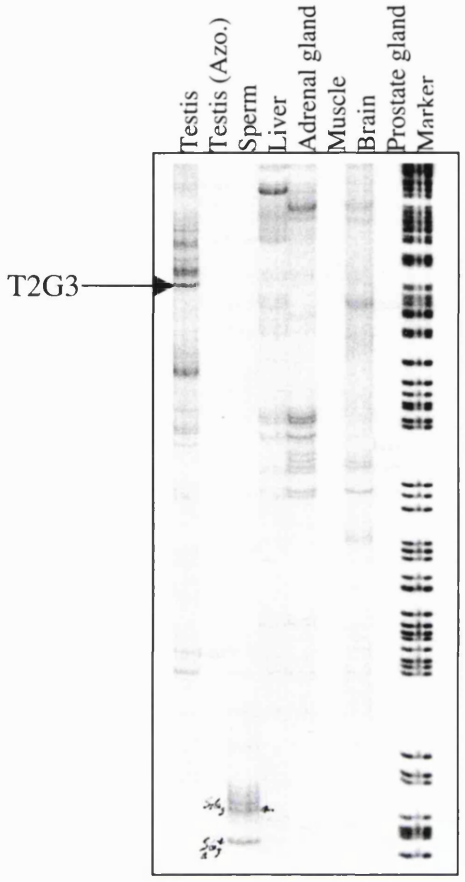


Figure V.1: DDRT-PCR result and T2G3 band is shown with an arrow.

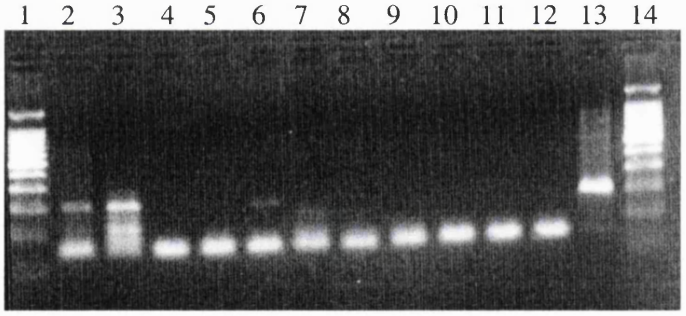


Figure V.2: RT-PCR using GSP1 and R1. Lanes numbered from the left 1-100 bp ladder, 2- and 3- Normal testis 4- Infertile patient's testis 5- Azoospermic patient's testis 6- Sperm 7- Adrenal 8- Liver 9- Muscle 10- Prostate 11- Brain 12- Negative (water) control 13- Positive (genomic DNA) control 14- 100 bp ladder
The expected band size was 204 bp which can be seen in the two normal testis samples and, more weakly, in sperm. The short band present in all samples is result of primer dimer.

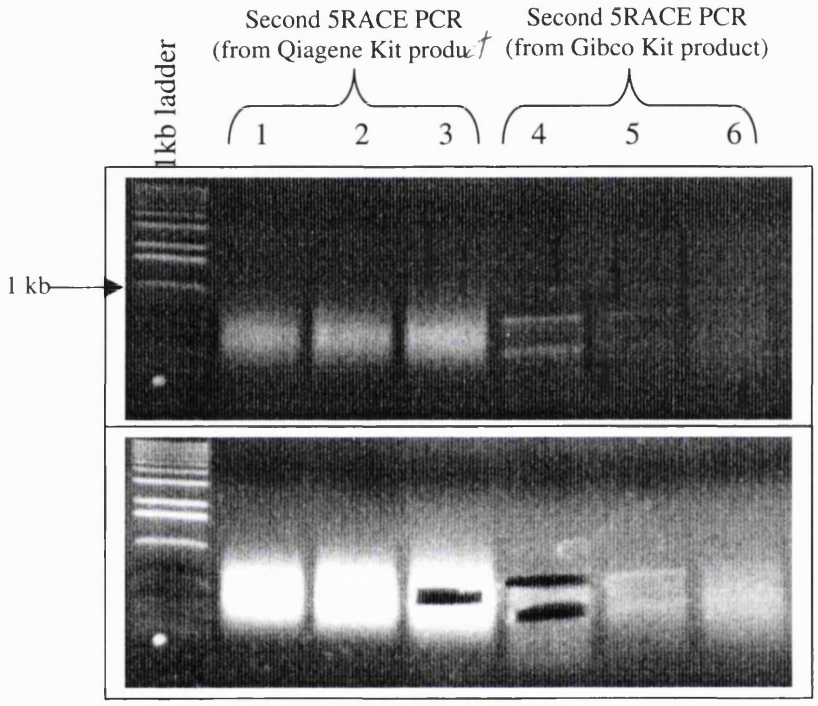


Figure V.3: Second 5′RACE PCR products.

A

T2G3 Sequences:

5′end

AGCTGAGATTGCACATAGTTGTATTACATCCCAAAATTTACAATCTTTTT **R2**
TTTAGATCATATTACTGTGAATATAGGAATCTCAGAGTGAGGGAGCAGGA
TAGAAGAGGATATATTACTTAGCTAAACACTTAGGAAAAGAAAACCAGTA
CAAAGCGAAGGAAGATATTAAAGAACCAC**ATGCTTTCTTATGGATCTTTC**
TTTCCTGAGTTCTCTTGCCTCTCTTCCTGCTGTATTCTAACCTGTACCCG
AGGCCTGAGCAAACTTTCTCAACCAAGGTCCTGGTCATCTCACTACCTGG
CTCTAAAATTTCCCCACTACTCACGAGGAAATTGAGCATTTCTTAGAGTA
GCATTCAGTATCCCCTCACTCTGGTGTTACCACATACCCATCCCAGCATT
AACCCCCTTAGTCTTAGACCCAGATCCCTCCTCTTGGCCTGACTCGTCCT **GSP3**
CCAACATACCAGAGCATGTGACATTTGACTATATTCTCTTTTTTGCTAAG
GCTTTTATGTGTAACATCTTTCTACCCTCACCACCACCCTTTGAACCTAT **R1**
ACCCCTTTTCTGCAAGCTTCAAAAGTTTAGATTTTAATCAAAAAGCCCTC
CTGGTTCTTCTCTGTAAAAATAAAAAAGGAACCAGGATCCCATAAATGTG **GSP2**
GATGTGTCAATAGCTAGGATATACAGACAATCTTTGAGAATCTTTGACTT
CAAGAGGCTTATAGGTTTGAATGCCAAAAAAAAAAAA **GSP1**

B

5′end

ATCACTGGGCTCAGCCTAGGCAGTGGCCTCTGTGCTTGACATTTTTCTTC
CTCATTCTGCTCCCAGAGTCTCCATCCTGTGTCTCCTCATTCTTTCAAGC
CTCAGTTCTTTCCTTTGGGTTCTAACATTTAGCATTGGCCTAGGTGTGCG
TGGTAGAGGCTTGGGGCCTGTGGGTTGAACCCAGGTGAAGGCACAAGGGG
AAGACAGAGTGAAGGAGAAGGAAAACCAAAAGTGGCAGAGAAACACAAGA
GGATGTAGGGAAAGGGGAGAGATAGAGGGCCATTCaaaagaagaaagaac
aaaggtgagtctccaagacaggatcccagagtgtgggg
GTCCTAGGGT-T=ACACC-<<**GSP2primer**

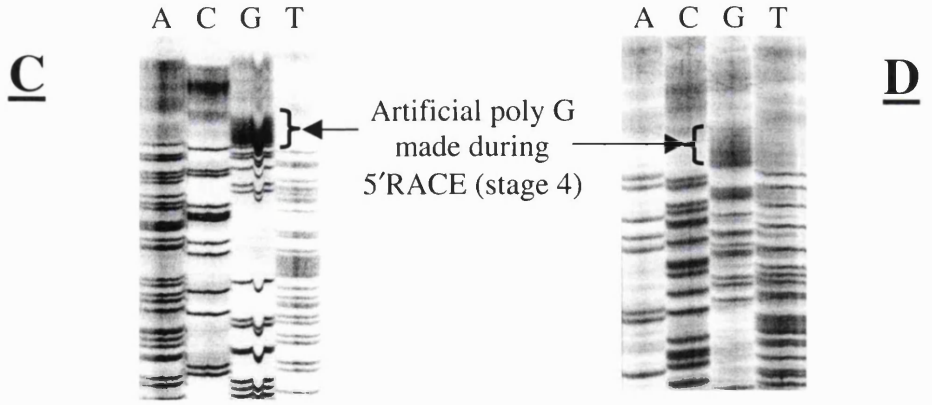


Figure V.4: A) T2G3 full-length mRNA. Primer locations are underlined and named on the right. Two sequences in boxes are shown the location of the DDRT-PCR primers, oligo(dt)-VG and OP3 primers. ORF is in bold letters.

B) Sequence of the short band (see figure V.3) from 5′RACE. The box shows similarity of GSP2 primer to sequence retrieved from database (hsgt) which is downstream of this piece.

C) 5′end of the T2G3 sequence (mRNA).

D) 5′end of the sequence from the short band (mRNA).

larger band consisted of 735 nucleotides but only contains an open reading frame of 153 nucleotides (figure V.4-A). The sequence of the short band, shown in figure V.4-B, showed no ORF.

V.3 The small band from 5′RACE

Searching htgs database using the small band sequence, 285bp, revealed two working draft clones, AC069259 and AC020649, containing the sequence. The clones were located on chromosome 3. By comparing more sequence of both ends which were retrieved from those clones with primers used during the second PCR of the 5′RACE, no similarity was found to AUAP primer but one similarity with about 10 nucleotides of the 3′end of GSP2 was seen. Figure V.4-D (lane G) shows the artificial poly C tail (in cDNA and poly G in the mRNA) which was made during 5RACE and it seems the small band is PCR product of GSP2 (mismatch) and AUAP primers from another cDNA which was probably generated by the GSP1 during cDNA synthesis.

V.4 5′RACE modifications

1- During the third step of 5′RACE (see figure III.10 and IV.2), when the single strand cDNA was cleaned and all dNTP and primers were removed from the reaction, another parallel experiment, was carried out to see whether the cleaning stage can be critical during 5′RACE. Qiagen PCR cleaning kit was used instead of Gibco BRL cleaning kit to remove the dNTPs and primers from cDNA reaction. The second PCR result using Qiagen PCR cleaning kit was smeary in the agarose gel (see figure V.3, lanes 1, 2 & 3) whereas the second 5′RACE PCR from cDNA which was cleaned using the Gibco BRL kit showed those two bands (see figure V.3). Bands and a part of smear were excised from the gel and sequenced. The result from smear was not readable.

2- A recent study has adapted the well-established 5′RACE technique to facilitate the identification of mRNA 5′ ends even for transcripts with a low level of expression (Flouriot *et al.*, 1999). Biotinylated single stranded DNA template is used as template to prepare a highly labelled long primer by extension from a short

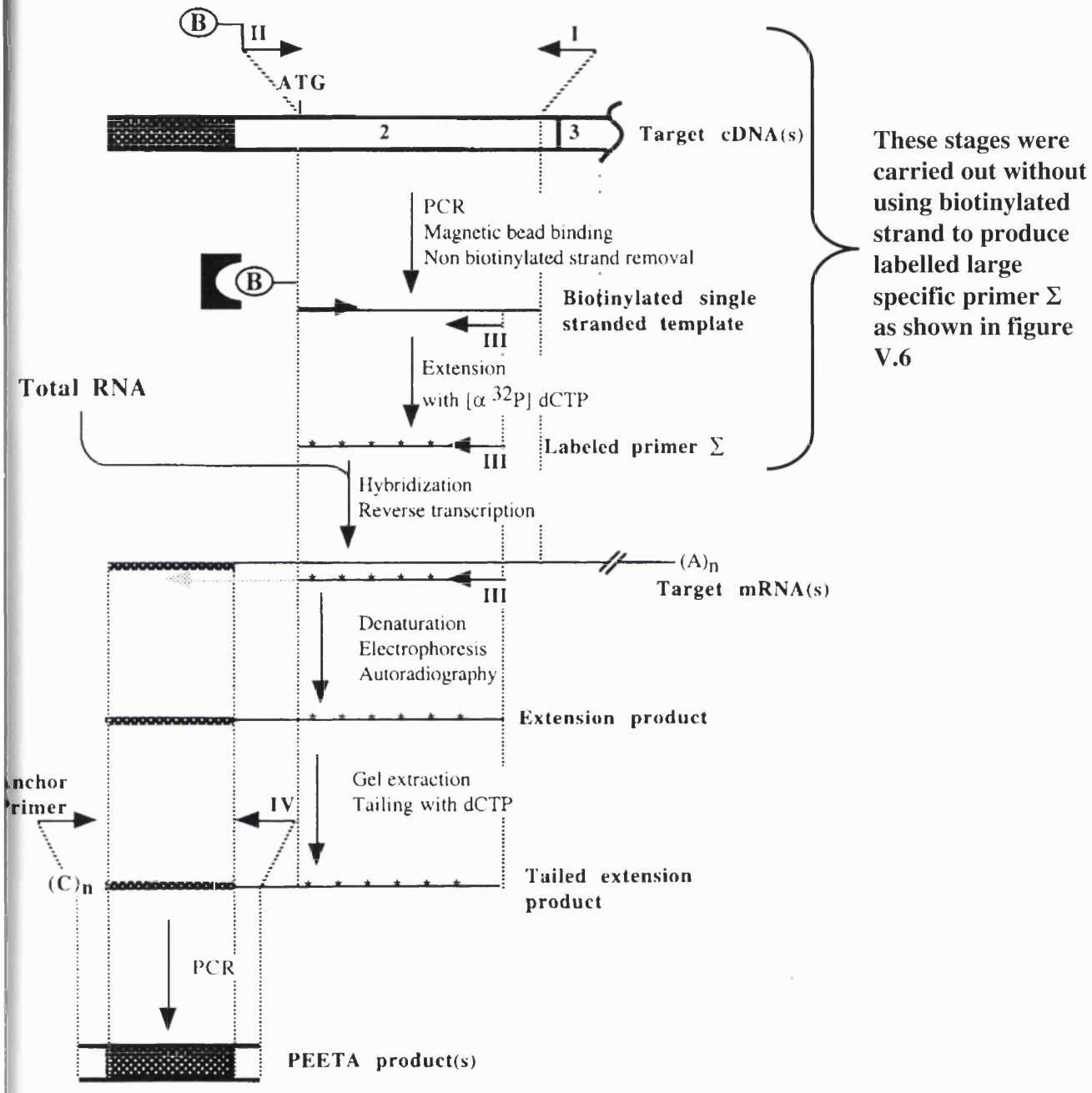


Figure V.5: Schematic representation of the modified 5´RACE method, called PEETA, suggested by Flouriot et a. (1999). Primer II is biotinylated. Primer IV is "nested" to primer III which is "nested" to primer I. PrimerΣ is a large gene-specific single stranded (antisense) oligonucleotide which can be used for specific hybridisation to the mRNA. Open rectangles represent known sequences of the target cDNAs whereas shaded rectangles correspond to unknown sequences.

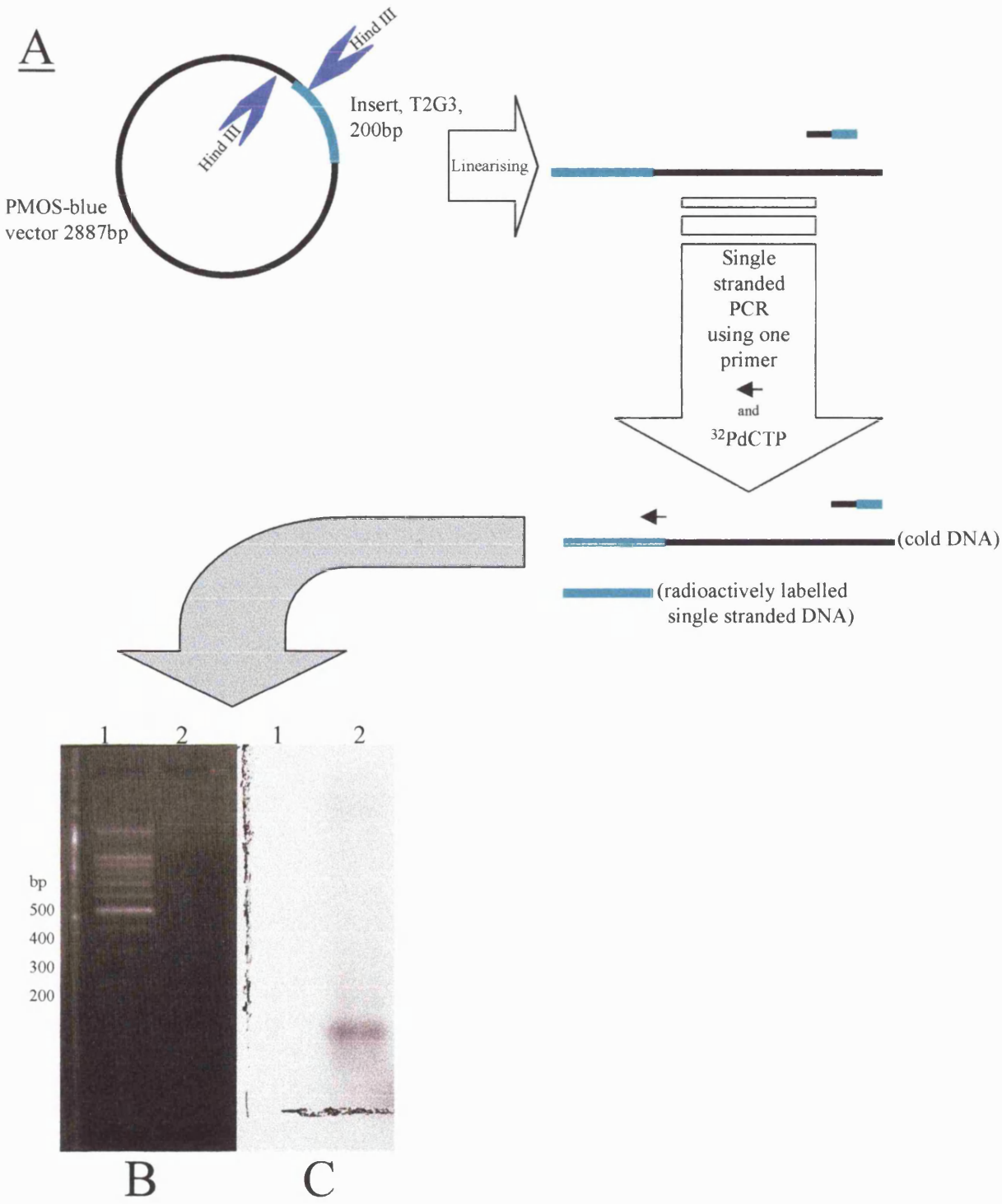


Figure V.6: A) The procedure of making anti-sense single stranded DNA (radioactive). B) An ethidium bromide stained gel. C) Autoradiograph photographed to the same scale. The 100bp ladder is seen in lane 1 but no band was seen in lane 2 on the 1.8% agarose gel. After 4 hours exposure onto an Xray film a band was visible in lane 2 at the right size. The band was excised for hybridisation with the testis RNA and further experiments.

specific sequence by T7 DNA polymerase in the presence of [α-^{32}P] dCTP (Flouriot *et al.*, 1997). Twenty nanograms of the long specific primer (Σ) are labelled with [α-^{32}P]dCTP in the presence of T7DNA polymerase. This labelled primer is allowed to hybridise overnight with 30μg of total RNA. Reverse transcription is then carried out resulting in ^{32}P labelled single stranded cDNAs. Extension products are isolated and a poly C tail added to allow for subsequent amplification using a combination of the gene-specific reverse primer with an oligo (dG) anchor primer (Flouriot *et al.*, 1999). A schematic of this procedure is shown in figure V.5. This experiment was carried out for T2G3 DDRT-PCR fragment but without using biotinylated single stranded DNA template (see figure V.6). Instead, the resulting RT-PCR product was cloned in a vector. A clone was digested and linearised using a restriction enzyme with a site in the 5′ end of the cDNA. Anti-sense DNA was generated by 40 cycle single-stranded PCR using only one primer and [α-^{32}P] dCTP. The product was electrophoresed on a 1.8% agarose gel. The product was not visible but could be identified using Xray film after 4 hours exposure (figure IV.6). The hot band was cut out, extracted and cleaned from the gel as explained in the Methods (section II.2.1.6). The quantity and quality of the ^{32}P labelled single stranded DNA were checked by spectrometer and by Cherenkov radiation counting using a scintillation counter.

Total c.p.m in the PCR reaction:

1.6×10^8 c.p.m

Total c.p.m in product:

1.3×10^8 c.p.m

Percentage of incorporation of [α-^{32}P] dCTP is:

about %81 = $[(1.3 \times 10^8) / 1.6 \times 10^8)]$

~2 pmol of gene specific primer (20 to 25nt) is usually used in 5′RACE. The generated product extracted from the gel was 150nt (about six to seven times longer than normal primer) with 29 cytosine. 1/100 of dCTP (4μl of 7.5μM) in the linear PCR reaction was [α-^{32}P] dCTP.

Calculation of the product concentration from used dCTP

$(4 \times 10^6) \times (7.5 \times 10^6) \times \%81 / 29 = \sim 1$pmol

Therefore, the product (approximately 1pmol) was sufficient to hybridise overnight with the RNA to carry out 5'RACE as mentioned above (figure V.5).

This experiment was not pursued for lack of time but would probably have been very useful in overcoming the non-specific amplification which complicated the earlier 5'RACE experiment.

The advantages of this system are that it generates sequences up to/or very close to the 5` extremities of the mRNA (Flouriot *et al.*, 1999) and also using a large oligonucleotide instead of a specific primer to produce cDNA increases the specificity. However, this system requires about 30μg of total RNA, in comparison with other RACE systems, which in general require only 1μg of starting template. The requirement for radioactive labelling makes this a less than desirable procedure for routine use.

V.5 Hybridisation of the transcript to Northern blot

A RT-PCR product using GSP1 and R2 primers was radiolabelled and hybridised to Northern blot. No positive hybridisation results were seen. The blot had also previously been used for four other experiments, β-actin and *TSGA10* gene (see figure III.14). It was stripped according to the manufacturer's instructions before hybridisation. This hybridisation, negative, is shown in figure V.8.

It is possible that no positive result was achieved due to the low expression of the DDRT cDNA; this will be discussed in section V.10.1.

V.6 Genomic clones

A human genomic library, 41 PAC filters derived from a gridded PAC library of 125k clones was screened as described in Methods. Three positive PACs were identified by hybridisation of the GSP1-R2 PCR product to one quarter of the PAC library (figure V.8). All three PAC clones were confirmed by PCR.

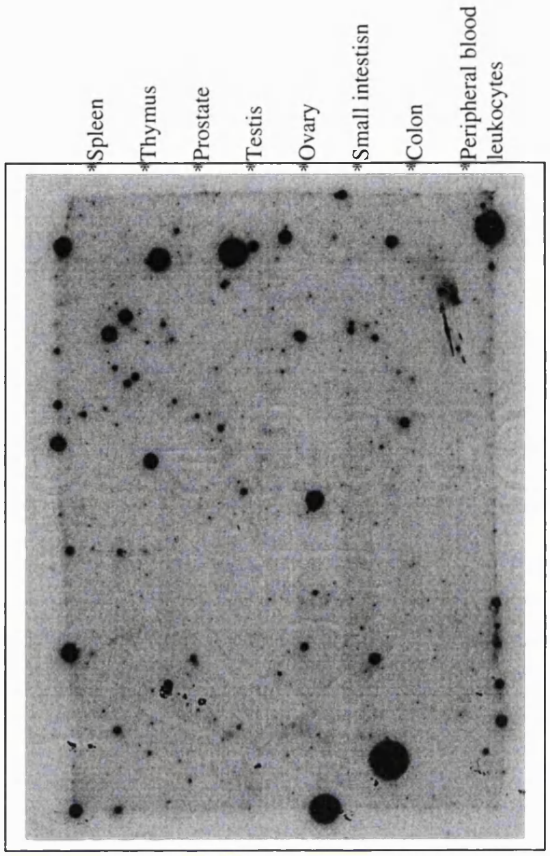


Figure V.7: Northern analysis shows negative results (two days exposure).

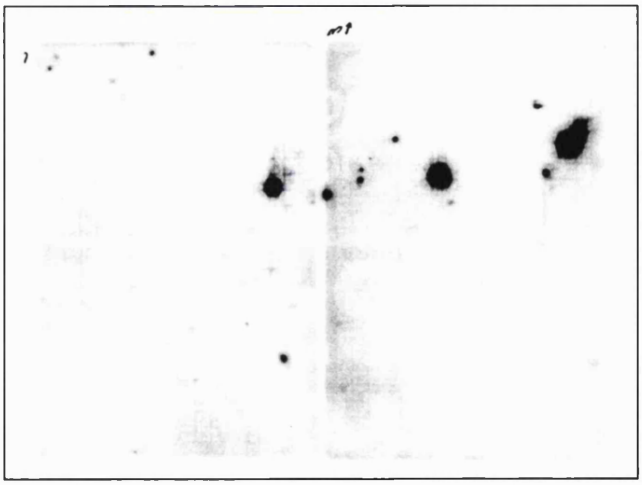


Figure V.8: Results of screening of a gridded PAC human genomic library (a quarter of the library was screened). Three strongest signals were positive clones.

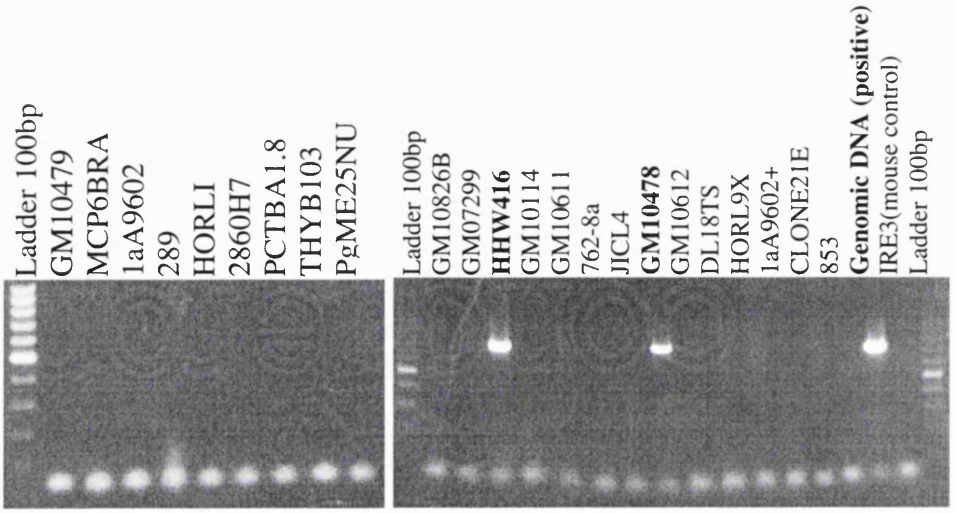


Figure V.9: The chromosomal location of* T2G3 *was determined by PCR amplification using gene specific primers, GSP1 and R2 on a panel of hamster or mouse cell hybrids that each contain a single human chromosome.
The PCR product (7XXbp) is present in the cell hybrid containing human chromosome 4 (HHW416) and chromosome 20 (GM10478) but it has been demonstrated that this cell hybrid, GM10478, also contained part of chromosome 4 (see table III.4).

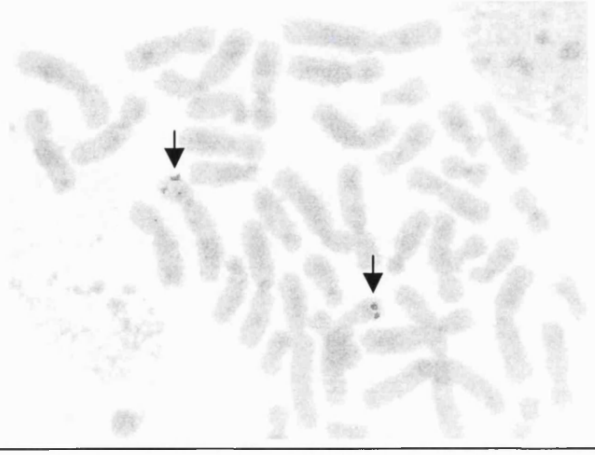


Figure V.10: T2G3 mapped to chromosome 4p16 by FISH.

V.7 Localisation and mapping

Firstly, the gene was mapped by PCR on a panel of somatic cell hybrids to chromosome 4 and no evidence was observed to suggest the existence of a second locus (see table III.4 and figure V.9).

Secondly, PAC dj64b8 was mapped using FISH to 4p16 (figure V.10). The 4p16 region has been extensively mapped (in the search for Huntingdon's disease amongst others) but no genes relating to infertility or testis function have been assigned to it.

V.8 Presence of the gene in other species

The primer pair R2 and GSP1 gave a similar sized PCR product from human, chimpanzee, gorilla and orang-utan genomic DNAs suggesting that the gene have been conserved during evolution (figure V.11).

V.9 Nomenclature and the symbol of the gene, *T2G3*

This DDRT-PCR fragment was originally named T2G3, but the gene has been given the approved symbol *C4ORF2* by the HUGO Nomenclature Committee. Therefore, this name is used instead of the name of the DDRT-PCR fragment, T2G3, henceforth in this thesis.

V.10 Discussion

V.10.1 Expression of the gene

It has been demonstrated by DDRT-PCR and RT-PCR on eight adult human tissues that *C4ORF2* may be specifically expressed in testis (figures V.1, V.2 and V.12). There was no result from the Northern blotting. There was results from the blot using a housekeeping gene, β-actin, before this experiment (data not shown). This negative result might be a technical problem, the result of re-using a worn out blot. It has been previously observed that it can be difficult to achieve a satisfactory hybridisation result of Northern blotting with DDRT-PCR products (Ognian *et al.* 1996). However I was using a much longer probe than just the DDRT-PCR fragment. More likely, *C4ORF2* could be a messenger RNA which is expressed at

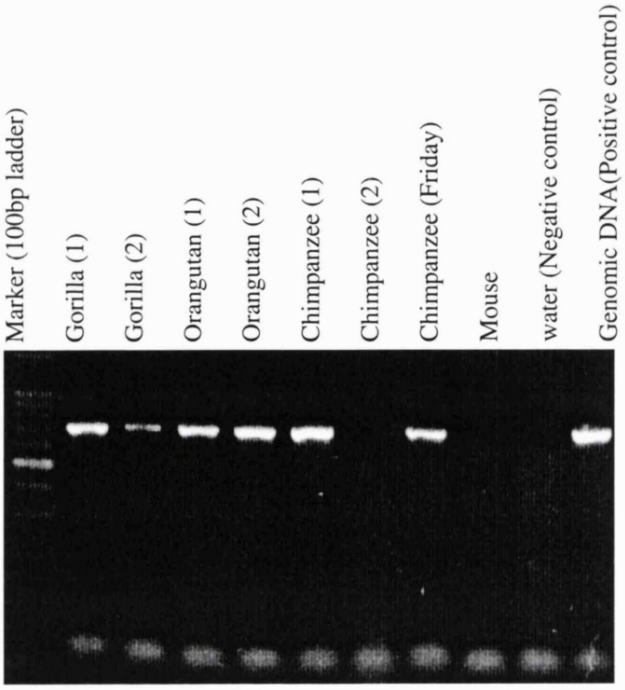


Figure V.11: PCR product using genomic DNA from different species. Primers in this PCR were GSP1 and R2.

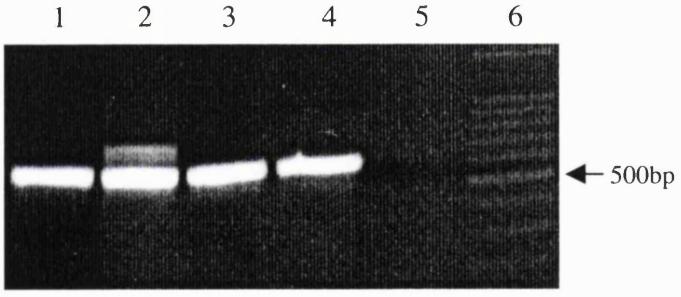


Figure V.12: RT-PCR using GSP2 and R2. Templates in lanes were:
1- the specific cDNA from testis which was used for 5′RACE, 2- the first PCR of 5′RACE, 3&4- genomic DNA (male and female), 5- water (negative control) and the last lane is a marker, 100bp ladder.

a very low level, below the detection limit of a Northern. DDRT-PCR allows the direct comparison of genes expressed even at levels as low as one in 10^4 to 10^5 molecules of total RNA (Ozaki et al., 1996).

No significant match was found by searching nr and est databases using the gene sequence (see figure V.13). A part of the 3′UTR sequence, between nucleotides 573 and 592 (aagtttagattttaatcaaa) was observed to be present in many cDNAs including Glutamine Synthetase. It was not thought to be part of a longer repetitive sequence, therefore it is potentially a controller sequence site.

V.10.2 Is this a pseudogene?

The genome contains a large number of pseudogenes which were created by duplication events or by the reintegration of reverse transcribed mRNA. Many of these pseudogenes possess promoter regions identical or very similar to those of the active ancestral genes.

PCR with primers R2 and GSP1 near to the 5′ and 3′ ends respectively of the cDNA showed that the amplified fragment size from genomic DNA was the same as from the cDNA suggesting that the gene is intronless. The possibility exists that the PCR product from genomic DNA is derived from a processed pseudogene. However, this is unlikely because both RT-PCR results and 5′RACE, two different experiments, suggest the existence of a transcript.

Transcription of pseudogenes in the testis has been directly demonstrated. There are some functional pseudogenes or intronless genes expressed in testis such as *PGK*-1 (MaCarrey and Thomas 1987), a testis-specific poly(A) polymerase, *Papt* (Lee and Chung, 2000), a *SOX5* pseudogene located on 8q21.1, whereas the *SOX5* gene itself, which contains a minimum of five introns, maps to 12p12.1 (Wunderle *et al.*, 1996).

An X-linked processed pseudogene of *PGK-1* (psi *PGK-1*) in humans has been identified and shown to contain premature termination codons in all reading frames. It was therefore unexpected to find that the intronless autosomal *PGK* sequence reported here is not a pseudogene, but is rather a functional gene that has

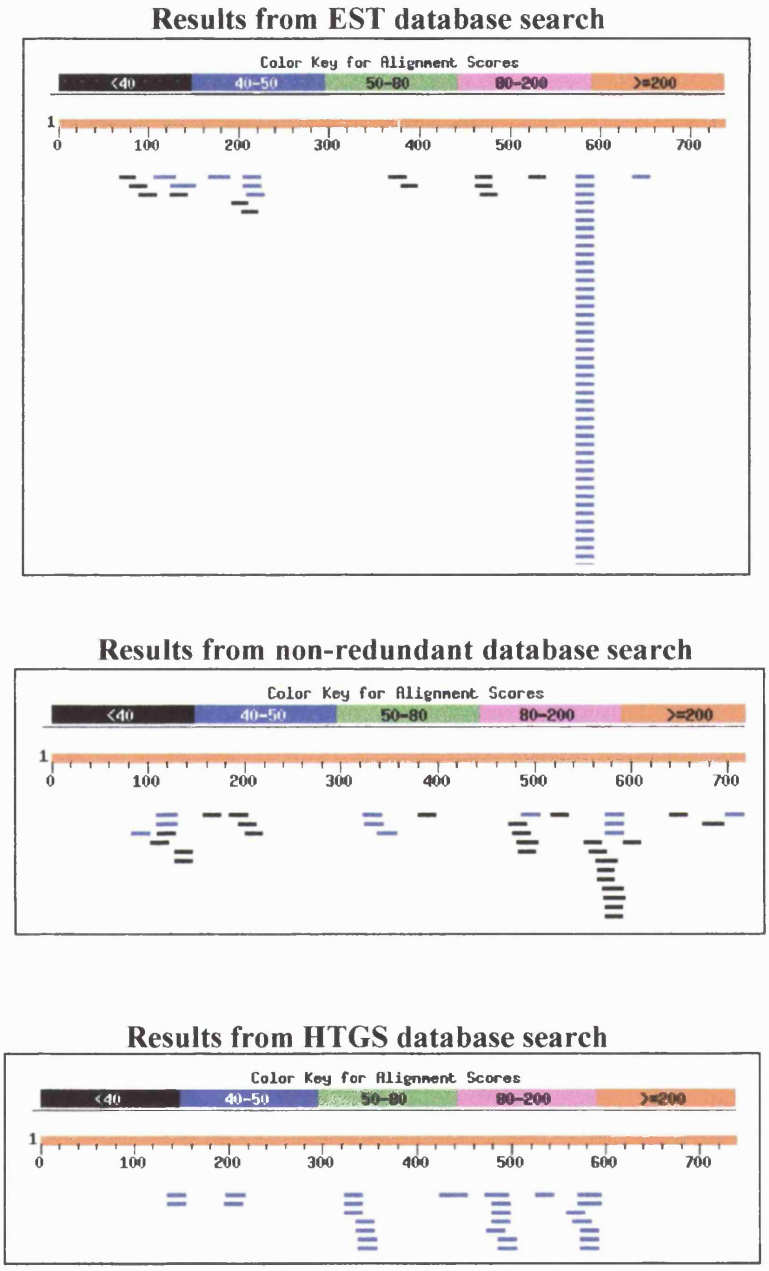


Figure V.13: Results of searching C4orf2 sequence from different databases.

retained a complete open reading frame, and is actively expressed in mammalian spermatogenesis (McCarrey and Thomas, 1987).

Another example is the creatine transporter gene, *SLC6A8* (Ch.X28) and the autosomal (Ch.16) pseudogene ψ*SLC6A8*. The pseudogene is not expressed in any tissue except testis (Iyer *et al.*, 1996). This tissue specificity led to the hypothesis that it might serve to compensate for the inactivation of the X-chromosomal copy during certain stages of spermatogenesis (Iyer *et al.*, 1996). Whereas the promoter regions of all the genes are essentially unmethylated, the pseudogene ψ*SLC6A8* was found to be highly methylated in somatic tissue, but not in testis (Grunau et al., 2000).

One more example of an intronless testis specific gene is the haspin protein which is sort to play a role in cell cycle arrest in haploid spermatids. The genomic structure of human haspin was shown to be intronless and the whole transcription unit was found to be located in an intron of the *integrin alphaE2* gene (Tanaka *et al.*, 2001).

In summary, *C4ORF2* is a novel intronless gene or pseudogene obtained from a testis using DDRT-PCR. The full-length cDNA was obtained using the 5′RACE technique. The expression was confirmed by RT-PCR. The number of clones from genomic library screening is consistent with the number expected from a single copy locus. The gene mapped to 4p16 by FISH.

REFFERENCES

Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C., (1993). Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat Genet.* **4**, 373-80.

Ahmad W., Li S., Chen H., Tuck-Muller C.M., Pittler S.J., Aronson N.N., (1995). Lysosomal chitobiase (CTB) and the G-protein gamma 5 subunit (GNG5) genes co-localize to human chromosome 1p22. *J. Cytogenet Cell Genet.* **71**, 44-46.

Afzelius, B.A., (1985). The immotile-cilia syndrome: a microtubule-associated defect. *CRC Crit Rev Biochem.* **19**, 63-87. Review.

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.

Aitken R.J., Vernet P. (1998). Maturation of redox regulatory mechanisms in the epididymis. *J Reprod Fertil Suppl.* **53**:109-18.

Akhondi M.A., Chapple C., Moore H.D. (1997). Prolonged survival of human spermatozoa when co-incubated with epididymal cell cultures. *Hum Reprod.* **12**, 514-22.

Albig, W., Kioschis, P., Poustka, A., Meergans, K., Doenecke, D. (1997). Human histone gene organization: nonregular arrangement within a large cluster. *Genomics.* **40**, 314-322.

Albig, W., Drabent, B., Kunz, J., Kalff-Suske, M., Grzeschik, K.-H., Doenecke., D. (1993). All known human H1 histone genes except the H1(0) gene are clustered on chromosome 6. *Genomics.* **16**, 649-654.

Alcivar A.A., Trasler J.M., Hake L.E., Salehi-Ashtiani K., Goldberg E., Hecht N.B. (1991). DNA methylation and expression of the genes coding for lactate dehydrogenases A and C during rodent spermatogenesis. *Biol Reprod.* **44**:527–535.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J. Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Staudt, L.M., (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* **3**, 503-11.

Allan D.J., Harmon B.V., (1992). Spermatogonial apoptosis has three morphologically recognizable phases and shows no circadian rhythm during normal spermatogenesis in the rat. *Roberts SA Cell Prolif.* **25**, 241-50.

Allanson, J.E., (1987). Noonan syndrome. *J Med Genet.* **24**, 9-13. Review.

Amer, M., Soliman, E., el-Sadek, M., Mendoza, C., Tesarik, J., (1997). Is complete spermiogenesis failure a good indication for spermatid conception? *Lancet.* **12**, 116.

Andersen B., Pearse R.V., Schlegel PN., Cichon Z., Schonemann M.D., Bardin C.W. and Rosenfeld M.G.. (1993). Sperm 1: A POU-Domain Gene Transiently Expressed Immediately Before Meiosis I in the Male Germ Cell *Proc. Natl. Acad. Sci. USA.* **90**, 11084-11088.

Arbizu T., Santamaria J., Gomez J.M., Quilez A., Serra J.P. (1983). A family with adult spinal and bulbar muscular atrophy, X-linked inheritance and associated testicular failure *Neurol Sci.* **59,** 371-82.

Ariel, M., Cedar, H., McCarrey, J., (1994). Developmental changes in methylation of spermatogenesis-specific genes include reprogramming in the epididymis. *Nat Genet.* **7**, 59-63.

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.

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.

Asano T., Morishita R., Ohashi K., Nagahama M., Miyake T., Kato K., (1995). Localization of various forms of the gamma subunit of G protein in neural and nonneural tissues. *J Neurochem.* **64**,1267-1273.

Averback, P., Wight, D.G., (1979). Seminiferous tubule hypercurvature: A newly recognised common syndrome of human male infertility. Lancet. **27**,181-3.

Baker H.W. (1998). Reproductive effects of nontesticular illness. *Endocrinol Metab Clin North Am.* **27**, 831-50.

Baker S.M., Bronner C.E., Zhang L., Plug A.W., Robatzek M., Warren G., Elliott E.A., Yu J., Ashley T., Arnheim N., et al. (1995). Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. *Cell.* **28**, 309-19.

Baker S.M., Plug A.W., Prolla T.A., Bronner C.E., Harris A.C., Yao X., Christie D.M., Monell C., Arnheim N., Bradley A., Ashley T., Liskay R.M. (1996). Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nat Genet.* **13**, 336-42.

Baldi, E., Luconi, M., Bonaccorsi, L., Forti, G., (1998). Nongenomic effects of progesterone on spermatozoa: mechanisms of signal transduction and clinical implications. *Front Biosci* **1**, 3:D, 1051-9

Baldi, E., Krausz, C., Luconi, M., Bonaccorsi, L., Maggi, M., Forti, G., (1995). Actions of progesterone on human sperm: a model of non-genomic effects of steroids. *J Steroid Biochem Mol Biol.* **53**,199-203

Baldocchi R.A., Flaherty L. (1997). Isolation of genomic fragments from polymorphic regions by representational difference analysis. *Methods.* **13**, 337-46

Barbosa A.S., Ferraz-Costa T.E., Semer M., Liberman B., Moreira-Filio C.A. (1995) XY-gonadal dysgenesis and gonadoblastoma: a study in two sisters with a cryptic deletion of the Y chromosome involving the SRY gene. *Hum.Genet.* **95**, 63-66.

Bauer, D., Muller, H., Reich, J., Riedel, H., Ahrenkel, V., Warthoe, P., and Strauss, M. (1993). Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nuc. Acids Res.* **18**, 4272-4280.

Berns A. (2000). Gene expression in diagnosis. *Nature.* **403**, 491-2.

Berruti G., (1998). Signaling events during male germ cell differentiation: bases and perspectives. *Front Biosci.* **1**, 3:D, 1097-108

Berchtold, M.W., (1989). A simple method for direct cloning and sequencing cDNA by the use of a single specific oligonucleotide and oligo(dT) in a polymerase chain reaction (PCR) *Nucl. Acids. Res.* **17**, 453

Bernstein, J., Sella, O., Le, S-Y, Elroy-Stein, O., (1997). PDGF2/c-sis mRNA leader contains a differentiation-linked internal ribosomal entry site (D-IRES). *J. Biol. Chem.* **272**, 9356–62.

Bernstein, J., Shefler, I., Elroy-Stein, O., (1995). The translational repression mediated by the platelet-derived growth factor 2/c-sis mRNA leader is relieved during megakaryocytic differentiation. *J. Biol. Chem.* **270**, 10559–65

Bertioli, D.J., Schlichter, U.H., Adams, M.J., Burrows, P.R., Steinbiss, H.H, Antoniw, J.F., (1995). An analysis of differential display shows a strong bias towards high copy number mRNAs. *Nucleic Acids Res.* **11**, 4520-3.

Besmer, P., Manova, K., Duttlinger, R., Huang, E.J., Packer, A., Gyssler, C., Bachvarova, R.F., (1993). The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Dev.* **Suppl.**,125-37. Review.

Bestor, T.H., (1998). Cytosine methylation and the unequal developmental potentials of the oocyte and sperm genomes. *Am J Hum Genet.* **62**,1269-73. Review.

Bestor, T.H., (1992). Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J.* **11**, 2611-7.

Bhat, M.A., Philip, A.V., Glover, D.M., and Bellen, H.J. (1996). Chromatid segregation at anaphase requires the *barren* product, a novel chromosome associated protein that interacts with topoisomerase II. *Cell.* **87**, 1103-1114.

Bhasin S., De Kretser D.M., Baker H.W.G. (1994). Pathophysiology and natural history of male infertility. *J Clin. Endocrinol Metab.* **79**, 1525-1529.

Blendy J.A., Kaestner K.H., Weinbauer G. F., Nieschlag E. & Schutz G.. (1996). Severe impairment of spermatogenesis in mice lacking the CREM gene. *Nature.* **380**, 162-5.

Boitani, C., Geremia, R., Rossi, R., Monesi, V., (1980). Electrophoretic pattern of polypeptide synthesis in spermatocytes and spermatids of the mouse. *Cell Differ.* **9**, 41-9.

Brazma A., Robinson A., Cameron G., Ashburner M. (2000). One-stop shop for microarray data.*Nature* **403**, 699-700

Braun, R.E., (2000). Temporal control of protein synthesis during spermatogenesis. *Int J Androl.* **23**, Suppl, 92-4.

Braun, R.E. (1998). Post-transcriptional control of gene expression during spermatogenesis. *Semin Cell Dev Biol.* **9**, 483-9.

Braun B.S., Frieden R., Lessnick S.L., May W.A., Denny C.T. (1995). Identification of target genes for the Ewing's sarcoma EWS/FLI fusion protein by representational difference analysis. *Mol. Cell Biol.* **15**, 4623-30

Brawerman, G. (1981). The Role of the poly(A) sequence in mammalian messenger RNA. *CRC Crit. Rev. Biochem.* **10**, 1-38.

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

Brown, G.M., Furlong, R.A., Sargent, C.A., Erickson, R.P., Longepied, G., Mitchell, M,, Jones, M.H., Hargreave, T.B., Cooke, H.J., Affara, N.A., (1998). Characterisation of the coding sequence and fine mapping of the human DFFRY geneand comparative expression analysis and mapping to the Sxrb interval of the mouse Ychromosome of the Dffry gene. *Hum Mol Genet.* **7**, 97-107.

Burfeind P., Hoyer-Fender S. (1991). Sequence and developmental expression of a mRNA encoding a putative protein of rat sperm outer dense fibers. *Dev Biol.* **148**, 195-204.

Boguski M.S., Lowe T.M., Tolstoshev C.M. (1993). dbEST--database for "expressed sequence tags". *Nat Genet.* **4**, 332-3

Bonny C., Cooker L.A., Goldberg E. (1998). Deoxyribonucleic acid-protein interactions and expression of the human testis-specific lactate dehydrogenase promoter: transcription factor Sp1 plays a major role. *Biol Reprod.* **58**, 754-9

Bundey S. (1982). Clinical evidence for heterogeneity in myotonic dystrophy. *J Med Genet.* **19**(5):341-8

Bunch D.O., Welch J.E., Magyar P.L., Eddy E.M., O'Brien D.A. (1998). Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis. *Biol Reprod.* **58**, 834-41.

Burgoyne P.S. (1998). The role of Y-encoded genes in mammalian spermatogenesis. *Semin Cell Dev Biol.* **4**, 423-32.

Burgoyne, P.S., (1987). The role of the mammalian Y chromosome in spermatogenesis. *Development.* **101**, Suppl: 133-41.

Callard G.V., Jorgensen J.C., Redding J.M., (1995). Biochemical analysis of programmed cell death during premeiotic stages of spermatogenesis in vivo and in vitro. *Dev Genet.***16**, 140-7.

Callard, D., Lescure, B., Mazzolini, L., (1994). A method for the elimination of false positives generated by the mRNA differential display technique. *Biotechniques.* **16**, 1096-7, 1100-3.

Cameron, J., (1998). Identification by Differential display RT-PCR and cDNA selection, of genes expressed in human testis. *The University of London, Thesis.*

Capel., B. (1996). The role of Sry in cellular events underlying mammalian sex determination. *Current Topics in Developmental Biology.* **32**, 1-37.

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

Cabello, O.A., Baldini, A., Bhat, M., Bellen, H., and Belmont, J.W. (1997). Localization of BRRN1, the human homologue of Drosophila barr, to 2q11.2. *Genomics.* **46**, 311-3.

Cali J.J., Balcueva E.A., Rybalkin I., Robishaw J.D., (1992). Selective tissue distribution of G protein gamma subunits, including a new form of the gamma subunits identified by cDNA cloning. *J Biol Chem.* **267**, 24023-24027.

Carlson, M., (1994). Human Embriology and Developmental Biology. Mosby-year Book.

Carlsen E., Giwercman A., Keiding N., Skakkebaek N.E. (1992). Evidence for decreasing quality of semen during past 50 years. *BMJ* **12**, 609-13.

Carr I., Clegg E.J., Meek G.A. (1968). Sertoli cells as phagocytes: an electron microscopic study. *J Anat.* **102**, 501-9.

Cate R.L., Mattaliano R.J., Hession C., Tizard R., Farber N.M., Cheung A., Ninfa E.G., Frey A.Z., Gash D.J., Chow E.P., (1986). Isolation of the bovine and human genes for Mullerian inhibiting substance and expression of the human gene in animal cells. *Cell* **45**, 685-98.

Cedar, H., (1988). DNA methylation and gene activity. *Cell* **8**, 3-4.

Chabot, D.A. Stephenson, V.M. Chapman, P. Besmer & A. Berstein. (1988). The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature.* **335**, 88-9.

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.

Chieffi, G, Pierantoni, R, Fasano, S. (1991). Immunoreactive GnRH in hypothalamic and extrahypothalamic areas. *Int Rev Cytol.* **127**, 1-55.

Cho, K. R. & Hedrick, L. (1997). Genetic alterations in human tumors. *Curr. Top. Microbiol. Immunol.* **221**, 149-176.

Choi Y.C., Aizawa A., Hecht NB. (1997). Genomic analysis of the mouse protamine 1, protamine 2, and transition protein 2 gene cluster reveals hypermethylation in expressing cells. *Mamm Genome.* **8**, 317-23.

Choi Y.C., Gu W., Hecht N.B., Feinberg A.P., Chae C.B. (1996) Molecular cloning of mouse somatic and testis-specific H2B histone genes containing a methylated CpG island. *DNA Cell Biol.* **15**, 495-504

Chomczynski, P. & Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* **162**, 156-9.

Choavaratana R., Suppinyopong S., Chaimahaphruksa P. (1999). ROSI from TESE the first case in Thailand: a case report.*J Med Assoc Thai.* **82**, 938-41.

Cheng MH, Maines JZ, Wasserman SA. (1998). Biphasic subcellular localization of the DAZL-related protein boule in Drosophila spermatogenesis. *Dev Biol.* **15**, 567-76.

Choi Y.C., Chae C.B. (1991). DNA hypomethylation and germ cell-specific expression of testis-specific H2B histone gene. *J Biol Chem.* **266**, 20504–20511CB.

Chu C.C. and Paul W.E. (1997). Fig1, an interleukin 4-induced mouse B cell gene isolated by cDNA representational difference analysis *Proc. Natl. Acad. Sci.* **94**, 2507-2512.

Clapham D.E., Neer E.J., (1997) G protein beta gamma subunits. *Annu. Rev. Pharmacol. Toxico.* **37**, 167-203.

Clermont, Y., Rambourg, A., (1978). Evolution of the endoplasmic reticulum during rat spermiogenesis. *Am J Anat.* **151**, 191-211.

Cobb M.H., Goldsmith E.J., (1995). How MAP kinases are regulated. *J Biol Chem.* **270**,14843-14846.

Conkright M.D., Wani M.A., Anderson K.P., Lingrel J.B. (1999). A gene encoding an intestinal-enriched member of the Kruppel-like factor family expressed in intestinal epithelial cells. *Nucleic Acids Res.* **1**, 1263-70.

Cortes D. Scand J. (1998). Cryptorchidism--aspects of pathogenesis, histology and treatment. *Urol Nephrol Suppl.* **196**,1-54.

Cook T., Gebelein B., Urrutia R. (1999). Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. *Ann N Y Acad Sci.* **30**, 94-102. Review.

Cooke H.J., Hargreave T., Elliott D.J. (1998). Understanding the genes involved in spermatogenesis: a progress report. *Fertil Steril.* **69**, 989-95.

Cooke, H.J & Elliott, D.J. (1997). RNA-binding proteins and human male infertility. *Trends Genet.* **13**, 87-9.

Crew, F.A.E. (1921). A suggestion as to the cause of the aspermatic condition in the imperfectly descended testis. *J. Anat.* **56**, 98–106.

Cross, S.H., Charlton, J.A., Nan, X., Bird, A.P., (1994). Purification of CpG islands using a methylated DNA binding column. *Nat Genet.* **6**, 236-44.

Dacheux J.L., Druart X., Fouchecourt S., Syntin P., Gatti J.L., Okamura N., Dacheux F. (1998) Role of epididymal secretory proteins in sperm maturation with particular reference to the boar. *J. Reprod Fertil Suppl.* **53**:99-107.

Dadze S., Wieland C., Jakubiczka S., Funke K., Schroder E., Royer-Pokora B., Willers R., Wieacker P.F. (2000). The size of the CAG repeat in exon 1 of the androgen receptor gene shows no significant relationship to impaired spermatogenesis in an infertile Caucasoid sample of German origin. *Mol Hum Reprod.* **6**, 207-14.

De. Rooij D.G., Vergouwen R.P. (1991) The estimation of damage to testicular cell lineages. *Prog Clin Biol Res.* **372**,467-80.

De. Felici M., Dolci S., Pesce M. (1992). Cellular and molecular aspects of mouse primordial germ cell migration and proliferation in culture. *Int J. Dev Biol.* **36**, 205-13.

de Die-Smulders C.E., Howeler C.J., Mirandolle J.F., Brunner H.G., Hovers V., Bruggenwirth H., Smeets H.J., Geraedts J.P. (1994). Anticipation resulting in elimination of the myotonic dystrophy gene: a follow up study of one extended family. *J Med Genet.* **31**, 595-601.

de Kretser, D.M., Phillips, D.J., (1998). Mechanisms of protein feedback on gonadotropin secretion. *J Reprod Immunol.* **39**, 1-12

De Smet C., Lurquin C., Lethe B., Martelange V., Boon T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol. Cell Biol.* **19**:7327–7335.

De Smet C., De Backer O., Faraoni I., Lurquin C., Brasseur F., Boon T. (1996). The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc Natl. Acad. Sci.* **93**:7149–7153.

Debouck,C., (1995). *Curr. Opin. Biotechnol.* **6**, 597-599.

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

Delbridge M.L., Lingenfelter P.A., Disteche C.M., Graves J.A. (1999). The candidate spermatogenesis gene RBMY has a homologue on the human X chromosome. *Nat Genet* **22**, 223-4.

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.

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.

DeRisi, J.L., Iyer, V.R., Brown, P.O., (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science.* **24**, 680-6.

DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996). Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**, 457-460

Devoto E., Madariaga M., Lioi X. (2000). Causes of male infertility. The contribution of the endocrine factor. *Rev Med Chil.* **128**, 184-92.

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., C. Mori., 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.* **93**, 3264-8.

Dix D.J., Allen J.W., Collins B.W., Poorman-Allen P., Mori C., Blizard D.R., Brown P.R., Goulding E.H., Strong B.D. & Eddy E.M. (1997). HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development.* **124**, 4595-603.

Dhanasekaran N., Prasad M.V., (1998). G protein subunits and cell proliferation. *Biol Signals Recept.* **7**, 109-117.

Doenecke D., Drabent B., Bode C., Bramlage B., Franke K., Gavenis K., Kosciessa U., Witt O. (1997). Histone gene expression and chromatin structure during spermatogenesis. *Adv Exp Med Biol.* **424**, 37-48.

Doenecke D., Albig W., Bouterfa H., Drabent B.J. (1994). Organization and expression of H1 histone and H1 replacement histone genes. *Cell Biochem.* **54,** 423-31.

Dolan J.W., Kirkman C., Fields S. (1989). The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc Natl Acad Sci.* **86**, 5703-7

Downes, G.B., Gautam, N., (1999). The G protein subunit gene families. *Genomics.* **15**, 544-52. Review.

Downes G.B., Gilbert D.J., Copeland N.G., Gautam N., Jenkins N.A., (1999). Chromosomal mapping of five mouse G protein gamma subunits. *Genomics.* **57**,173-176.

Downes G.B., Copeland N.G., Jenkins N.A., Gautam N. (1998). Structure and mapping of the G protein gamma3 subunit gene and a divergently transcribed novel gene, gng3lg. *Genomics.* **53**, 220-230.

Dowsing, A., Yong, E.L., Clark, M. (1999). Linkage between male infertility and trinucleotid repeat expansion in the androgen receptor gene. *Lancet.* **354**, 640–643.

Dym Martin. (1994) Spermatogonial stem cells of the testis. *Proc Natl Acad Sci U S A.* 22;91(24):11287-9

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.

Edelmann W., Cohen P.E., Kane M., Lau K., Morrow B., Bennett S., Umar A., Kunkel T., Cattoretti G., Chaganti R., Pollard J.W., Kolodner R.D., Kucherlapati R. (1996). Meiotic pachytene arrest in MLH1-deficient mice. *Cell.* **28**,1125-34.

Eddy, E.M., (1999). Role of heat shock protein HSP70-2 in spermatogenesis. *Rev Reprod.* **4**, 23-30. Review.

Eddy, E.M., (1998). Regulation of gene expression during spermatogenesis. *Semin Cell Dev Biol.* **9**, 451-7. Review.

Eichler, E.E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N.A., Moyzis, R.K., Baldini, A., Gibbs, R.A. and Nelson, D.L. (1996). Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.* **5**, 899–912.

Elliot, D.J. & Cooke, H.J. (1997). The molecular genetics of male infertility. *BioEssays* **19**, 801-809.

Elsawi,M.M., Pryor,J.P, Klufio,G., Barnes,C., and Patton,M. (1994). Genital tract function in men with Noonan syndrom. *Journal of Medical Henetics.* **31**, 468-470.

Erickson, R.P. (1990). Post-meiotic gene expression. *Trends Genet* **6**, 264-9

Faiman C., Hoffman D.L., Ryan R.J., Albert A. (1968). The "fertile eunuch" syndrome: demonstration of isolated luteinizing hormone deficiency by radioimmunoassay technique. *Mayo Clin Proc.* **43**, 661-7.

Faisst, S., and Meyer, (1992)Compilation of vertebrate-encoded transcription factors *Nucl. Acids. Res.* **20**, 3-26.

Feig, L.A., Bellve, A.R., Horbach Erickson, N. & Klagsbrun, M. (1980). Sertoli cells contain a mitogenic polypeptide. *Proc. Natl. Acad. Sci.* **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.

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.

Ferlin A., Moro E., Garolla A., Foresta C. (1999). Human male infertility and Y chromosome deletions: role of the AZF-candidate genes DAZ, RBM and DFFRY. *Hum Reprod.* **14**, 1710-6.

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 K.J., Aronson N.N. Jr., (1992). Characterization of the cDNA and genomic sequence of a G protein gamma subunit (gamma 5). *Mol Cell Biol.* **12**, 1585-1591.

Fisher, R.C., Scott, E.W., (1998). Role of PU.1 in hematopoiesis. *Stem Cells.***16**, 25-37. Review.

Fishel, S., Green, S., Bishop, M. et al. (1995). Pregnancy after intracytoplasmic injection of spermatid. *Lancet.* **345**, 1641–1642.

Flouriot, G., Brand, H., Gannon, F., (1999). Identification of differentially expressed 5'-end mRNA variants by an improved RACE technique (PEETA). *Nucleic Acids Res.* **1,** e8.

Flouriot,G., Pope, C., Kenealy, M.R., Gannon, F., (1997). Improved efficiency for primer extension by using a long, highly-labeled primer generated from immobilized single-stranded DNA templates. *Nucleic Acids Res.* **15**, 1658-9.

Foresta C., Ferlin A., Garolla A., Moro E., Pistorello M., Barbaux S., Rossato M. (1998.) High frequency of well-defined Y-chromosome deletions in idiopathic Sertoli cell-only syndrome. *Hum Reprod.* **13**(2):302-7

Foulkes N.S., Mellstrom B., Benusiglio E. & Sassone-Corsi P., (1992). Developmental switch of CREM function during spermatogenesis: from antagonist to activator. *Nature.* **355**, 80-4.

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.

Frohman, M.A., (1993). Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol.* **218**, 340-56.

Frohman, M.A., Martin, G.R., (1989). Cut, paste, and save: new approaches to altering specific genes in mice. *Cell.* **27**, 145-7. Review.

Frohme M., Scharm B., Delius H., Knecht R., Hoheisel J.D. (2000). Use of representational difference analysis and cDNA arrays for transcriptional profiling of tumor tissue. *Ann N Y Acad Sci.* **910**, 85-104

Fu, Y.-H., Pizzuti, A., Fenwick, R. G., Jr., King, J., Rajnarayan, S., Dunne, P. W., Dubel, J., Nasser, G. A., Ashizawa, T., de Jong, P., Wieringa, B., Korneluk, R., Perryman, M. B., Epstein, H. F., Caskey, C. T. (1992). An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science.* **255**, 1256-1258,

Gartler, S. M., Riley, D. E., Lebo, R. V., Cheung, M.-C., Eddy, R. L., Shows, T. B. (1986). Mapping of human autosomal phosphoglycerate kinase sequence to chromosome 19. Somat. Cell Molec. *Genet.* **12**, 395-401.

Gautam N., Downes G.B., Yan K., Kisselev O., (1998). The G-protein betagamma complex. *Cell Signal.* **10**, 447-455.

Gautam N., Baetscher M., Aebersold R., Simon M.I., (1989). A G protein gamma subunit shares homology with ras proteins. *Science.* **244,** 971-4

Gebara M.M., McCarrey J.R., (1992). Protein-DNA interactions associated with the onset of testis-specific expression of the mammalian Pgk-2 gene. *Mol Cell Biol.* **12**, 1422-31

Ghazzawi, I. M., Alhasani, S., Taher, M., Souso, S. (1999) Reproductive capacity of round spermatids compared with mature spermatozoa in a population of azoospermic men. *Hum Reprod.* **14**, 736-740.

Ghiselli, G., Siracusa, L., and Iozzo, R. V. (1999). Complete cDNA cloning, genomic organization, chromosomal assignment, functional characterization of the promoter, and expression of the murine Bamacan gene. *J. Biol. Chem.* **274**, 17384-17393.

Gillett, G.T., McConville, C.M., Byrd P.J., Stankovic, T., Taylor, A.M., Hunt, D.M., West, L.F., Fox, M.F., Povey, S., and Benham, J.F. (1993). Irradiation Hybrids for Human Chromosome 11: Characterization and use for Generating Region-Specific Markers in 11q14-q23. *Genomics*. **15**, 332-341.

Gilman A.G., (1987) G proteins: transducers of receptor-generated signals. *Annu. Rev. Biochem.* **56**,615-649

Giuili G., Shen W.H., Ingraham H.A. (1997). The nuclear receptor SF-1 mediates sexually dimorphic expression of Mullerian Inhibiting Substance, in vivo. *Development* **124**, 1799-807.

Giwercman, Y.L., Xu, C., Arver, S. *et al.* (1998). No association between the androgen receptor gene CAG repeat and impaired sperm production in Swedish men. *Clin. Genet.* **54**, 435–436.

Gnessi L., Fabbri A., Spera G. (1997). Gonadal Peptides as mediators of development and functional control ofthe testis: an integrated system with hormones and local environment. *Endocr Rev.* **18**, 541-609.

Goldberg E. (1996). Transcriptional regulatory strategies in male germ cells. *J Androl.* **17**, 628-32.

Gotz M.J., Johnstone E.C., Ratcliffe S.G. (1999). Criminality and antisocial behaviour in unselected men with sex chromosomeabnormalities. *Psychol Med.* **29**, 953-62.

Graf, D., Fisher, A.G., Merkenschlager, M., (1997). Rational primer design greatly improves differential display-PCR (DD-PCR). *Nucleic Acids Res.* **1**, 2239-40.

Grandjean, V., Sage, J., Ranc, F., Cuzin, F., and Rassoulzadegan, M. (1997). Stage-specific signals in germ line differentiation: control of Sertoli cell phagocytic activity by spermatogenic cells. *Dev. Biol.* **184**, 165-174.

Grays Anatomy. 38th Edition. Churchill Livingstone.

Gray, N.K., Wickens, M., (1998). Control of translation initiation in animals. *Annu Rev Cell Dev Biol.***14**, 399-458. Review.

Gray J.W., Collins C. (2000). Genome changes and gene expression in human solid tumors. *Carcinogenesis.* **21**, 443-52.

Greenfield, A., (2000). Applications of DNA microarrays to the transcriptional analysis of mammalian genomes. *Mamm Genome.* **11**, 609-13. Review.

Greenfield, A., (1998). Genes, cells and organs: recent developments in the molecular genetics of mammalian sex determination. *Mamm Genome.* **9**, 683-7. Review.

Greenfield, A., Carrel, L., Pennisi, D., Philippe C., Quaderi, N., Siggers, P., Steiner, K., Tam, P.P., Monaco, A.P., Willard, H.F., Koopman, P. (1998). The UTX gene escapes X inactivation in mice and humans. *Hum Mol Genet.* **7**, 737-42.

Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M.B., de Fatima Bonaldo, M., Smith, L., Tymowska-Lalanne, Z., Wells, C., Greenfield, A., (2000). Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mousegonad prior to overt differentiation suggests a role in mammalian sexual development. *Hum Mol Genet.* **12**, 1553-60.

Griswold, M.D., (1993). Actions of FSH on mammalian Sertoli cells. In: Russell, L.D. and Griswold, M.D., (eds) *The Sertoli Cell.* Cache River Press, Clearwater, FL 493-508

Grimes S.R. Jr., van Wert J., Wolfe S.A. (1997). Regulation of transcription of the testis-specific histone H1t gene by multiple promoter elements. *Mol Biol Rep.* **24**,175-84.

Grunau C., Hindermann W., Rosenthal A. (2000) Large-scale methylation analysis of human genomic DNA reveals tissue-specific differences between the methylation profiles of genes and pseudogenes. *Hum Mol Genet.* **9**, 2651-63.

Gu W., Kwon Y.K., Hecht NBIn. (1996). postmeiotic male germ cells poly (A) shortening accompanies translation of mRNA encoding gamma enteric actin but not cytoplasmic beta and gamma actin mRNAs *Mol. Reprod. Dev.* **44**, 141-5.

Guerin, J.F., Czyba, J.C., Perrin, P., Rollet, J., (1981). Congenital or acquired obstructions of the human epididymis: study of the motility of the spermatozoa above the obstruction. *Bull Assoc Anat (Nancy).* **65**,297-306.

Gurskaya,N.G., Diachenko, L, Chenchik,A., Siebert,P.D., Khaspekov,G.L., Lukyanov,K.A., Vagner,L.L., Ermolaeva,O.D., Lukyanov,S.A. and Sverdlov,E.D. (1996). Equalizing cDNA subtraction based on selective suppression of polymerasechain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem.* **240**, 90-97.

Hagaman J.R., Moyer J.S., Bachman E.S., Sibony M., Magyar P.L., Welch J.E., Smithies O., Krege J.H., O'Brien D.A. (1998). Angiotensin-converting enzyme and male fertility. *Proc Natl Acad Sci U S A.* **3**, 2552-7.

Hann, S.R., Sloan-Brown, K., Spotts, G.D., (1992). Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. *Genes Dev.* **6**, 1229–40

Hann, S.R., King, M.W., Bentley, D.L., Anderson, C.W., Eisenman, R.N., (1988). A non-AUG translational initiation in c-myc exon 1 generates a N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphoma. *Cell* **52**, 185–95

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.

Harley V.R., Jackson D.I., Hextall P.J., Hawkins J.R., Berkovitz G.D., Sockanathan S., Lovell-Badge R., Goodfellow P.N. (1992). DNA binding activity of recombinant SRY from normal males and XY females. *Science*. **24**, 453-6.

Hathout E.H., Thompson K., Baum M., Dumars K.W. Am J. (1998). Association of terminal chromosome 1 deletion with sertoli cell-only syndrome. *Med Genet* **4**, 396-8

Hecht, N.B. (1998). Molecular mechanisms of male germ cell differentiation. *BioEssays* **20**, 555-561.

Hecht, N.B. (1995). The making of a Spermatozoon: A molecular prospective. Dev. *Genetics*. **16**, 95-103.

Hecht N.B. (1990). Regulation of 'haploid expressed genes' in male germ cells. *J Reprod Fertil.* **88**, 679-93.

Herman J.G., Baylin S.B. (2000). Promoter-region hypermethylation and gene silencing in human cancer. *Curr Top Microbiol Immunol.* **249**, 35-54.

Hesketh, J., (1994). Translation and the cytoskeleton: a mechanism for targeted protein synthesis. *Mol Biol Rep.* **19**, 233-43. Review.

Hess R.A., (1998). Effects of environmental toxicants on the efferent ducts, epididymis and fertility *J Reprod Fertil Suppl.* **53**, 247-59.

Hirano, T., Mitchison, T.J., (1994). A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro. Cell. **79**, 449-458

Hirano, T., Kobayashi, R., Hirano., M., (1997). Condensins, chromosome condensation protein complexes containing XCAP-C, XCAP-E and a Xenopus homolog of the Drosophila Barren protein. *Cell.* **89**, 511-521

Hiort O., Holterhus P.M., Nitsche E.M. (1998). Physiology and pathophysiology of androgen action. *Baillieres Clin Endocrinol Metab.* **12**, 115-32.

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.

Hornigold N., van Slegtenhorst M., Nahmias J., Ekong R., Rousseaux S., Hermans C., Halley D., Povey S., Wolfe J., (1997). A 1.7-megabase sequence-ready cosmid contig covering the TSC1 candidate region in 9q34. *Genomics.* **41**, 385-389.

Hoog, C., (1995). Expression of a large number of novel testis-specific genes during spermatogenesis coincides with the functional reorganization of the male germ cell. *Int J Dev Biol.* **39**, 719-26.

Hoog C. (1991). Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. *Nucleic Acids Res.* 25;**19**(22):6123-7.

Hoshi, K., Yanagida, K., Katayose, H., Yazawa, H., (1994). Pronuclear formation and cleavage of mammalian eggs after microsurgical injection of freeze-dried sperm nuclei. *Zygote.* **2**, 237-42.

Houston, D.W., Zhang, J., Maines, J.Z., Wasserman, S.A., King, M.L., (1998). A Xenopus DAZ-like gene encodes an RNA component of germ plasm and is a functional homologue of Drosophila boule. *Development.* **125**, 171-80.

Howard T.E., Shai S.Y., Langford K.G., Martin B.M., Bernstein K.E. (1990). Transcription of testicular angiotensin-converting enzyme (ACE) is initiated within the 12th intron of the somatic ACE gene. *Mol Cell Biol.* **10**, 4294-302.

Hsueh, R.C., Scheuermann, R.H., (2000). Tyrosine kinase activation in the decision between growth, differentiation, and death responses initiated from the B cell antigen receptor. *Adv Immunol.***75**, 283-316. Review.

Hubank M., Schatz D.G. (1994). Identifying differences in mRNA expression by representational difference analysis of cDNA. Nucleic *Acids Res.* **22**, 5640-8.

Huhtaniemi, I., Eur J., (1994). Fetal testis--a very special endocrine organ. *Endocrinol.* **130**, 25-31

Hutson J.M., Baker M., Terada M., Zhou B., Paxton G. (1994). Hormonal control of testicular descent and the cause of cryptorchidism. *Reprod Fertil Dev.* **6**, 151-6.

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.

Ivanova,N. and Belyavsky,A. (1995). Identification of differentially expressed genes by restriction endonuclease-based gene expression fingerprinting.*Nucleic Acids Res.* **23**, 2954-2958

Iyer, G.S., Krahe, R., Goodwin, L.A., Doggett, N.A., Siciliano, M.J., Funanage, V.L. and Proujansky, R. (1996). Identification of a testis-expressed creatine transporter gene at 16p11.2 and confirmation of the X-linked locus to Xq28. *Genomics*, **34**, 143–146.

Jakubiczka, S. Schnieders, F. & Schmidtke, J. (1993). A bovine homologue of the human TSPY gene. *Genomics* **17**, 732-735.

Jamieson, C. R., van der Burgt, I., Brady, A. F., van Reen, M., Elsawi, M. M., Hol, F., Jeffery, S., Patton, M. A., Mariman, E., (1994). Mapping a gene for Noonan syndrome to the long arm of chromosome 12. *Nature Genet.* **8**, 357-360.

Jansen-Durr P., Mondesert G., Kedinger C. (1989). Replication-dependent activation of the adenovirus major late promoter is mediated by the increased binding of a transcription factor to sequences in the first intron. *J Virol.* **63**, 5124-32

Jansen, G., Mahadevan, M., Amemiya, C., Wormskamp, N., Segers, B., Hendriks, W., O'Hoy, K., Baird, S., Sabourin, L., Lennon, G., Jap, P. L., Iles, D., Coerwinkel, M., Hofker, M., Carrano, A. V., de Jong, P. J., Korneluk, R. G., Wieringa, B., (1992). Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. *Nat Genet.* **1**, 261-6.

Jansen G., Bachner D., Coerwinkel M., Wormskamp N., Hameister H., Wieringa B. (1995). Structural organization and developmental expression pattern of the mouse WD-repeat gene DMR-N9 immediately upstream of the myotonic dystrophy locus. *Hum Mol Genet.* **4**, 843-52n.

Jenster, G., van der Korput, H.G.M., Trapman, J. (1995). Identification of two transcription activation sites in the N-terminal domain of the human androgen receptor. *J. Biol. Chem.* **270**, 7341–7346.

Janulis, L., Bahr, J.M., Hess, R.A., Janssen, S., Osawa, Y., Bunick, D., (1998). Rat testicular germ cells and epididymal sperm contain active P450 aromatase. *J Androl.* **19**, 65-71.

Janulis, L., Bahr, J.M., Hess, R.A., Bunick, D., (1996). P450 aromatase messenger ribonucleic acid expression in male rat germ cells: detection by reverse transcription-polymerase chain reaction amplification. *J Androl.* **17**, 651-8.

Jequier, A.M., (2000) Varicocele and its Rol in Male Infertility. *Male Infertility*. Blackwell Science. 136-153.

Jimenez R., Burgos M. (1998) Mammalian sex determination: joining pieces of the genetic puzzle. *Bioessays*. 20(9):696-9

Jin, H., Cheng, X., Diatchenko, L., Siebert, P.D., Huang, C.C., (1997). Differential screening of a subtracted cDNA library: a method to search for genespreferentially expressed in multiple tissues. *Biotechniques.* **23**, 1084-6

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.

Johnson B.L., Campbell C. (1996). Identification of a human RAD52 pseudogene located on chromosome 2. *Gene*. **9**, 229-32.

Johnson, P.F. and McKnight,S.L. (1989). Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**, 799-839. Review.

Jones R., (1998). Plasma membrane structure and remodelling during sperm maturation in the epididymis. *J Reprod Fertil Suppl.* **53**, 73-84. Review.

Jones, R., (1999). Some misconception of the human epididymis. In: *Male fertility and infertility*. Glover, T. D. and Barratt C.L.R. p 85-105.

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.

Josso N., Racine C., di Clemente N., Rey R., Xavier F., (1998). The role of anti-Mullerian hormone in gonadal development. *Mol Cell Endocrinol.* **25**, 3-7

Jue, K., Bestor, T.H., Trasler, J.M., (1995). Regulated synthesis and localization of DNA methyltransferase during spermatogenesis. *Biol Reprod.* **53**, 561-9.

Jurisicova, A., Lopes, S., Meriano, J., Oppedisano, L., Casper, R.F., Varmuza, S., (1999). DNA damage in round spermatids of mice with a targeted disruption of the Pplcgamma gene and in testicular biopsies of patients with non-obstructive azoospermia. *Mol Hum Reprod.* **5**, 323-30.

Kafri T., Ariel M., Brandeis M., Shemer R., Urven L., McCarrey J., Cedar H., Razin A. (1992). Developmental pattern of gene-specific DNA methylation in the mouse embryo and germ line. *Genes Dev.* **6**, 705-14.

Kalyanaraman S., Copeland N.G., Gilbert D.G., Jenkins N.A., Gautam N., (1998) Structure and chromosomal localization of mouse G protein subunit gamma 4 gene. *Genomics.* **49**,147-151.

Kato, K., (1996). RNA fingerprinting by molecular indexing. *Nucleic Acids Res.* **15**, 394-5.

Keegan K.S., Holtzman D.A., Plug A.W., Christenson E.R., Brainerd E.E., Flaggs G., Bentley N.J., Taylor E.M., Meyn M.S., Moss S.B., Carr A.M., Ashley T., Hoekstra M.F. (1996). The Atr and Atm protein kinases associate with different sites along meiotically pairing chromosomes. *Genes Dev.* **1**, 2423-37.

Kelsell D.P., Rooke L., Warne D., Bouzyk M., Cullin L., Cox S., West L., Povey S., Spurr N.K., (1995). Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. *Ann Hum Genet.* **59**, 233-241.

Kent-First M., Muallem 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.

Kerr J.B., (1991). Ultrastructure of the seminiferous epithelium and intertubular tissue of the human testis. *J Electron Microsc Tech.* **19**, 215-40

Kisselev O., Gautam N., (1993). Specific interaction with rhodopsin is dependent on the gamma subunit type in a G protein. *J. Biol. Chem.* **268**, 24519-24522.

Konietzko U. and Kuhl D. (1998). A subtractive hybridisation method for the enrichment of moderately induced sequences. *Nucleic Acids Res.* **26**, 1359-61.

Kozak, M., (1991). An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* **115**, 887–903

Kozak, M.. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **26**, 8125-48. Review.

Krausz C. and Forti G. (2000). Clinical Aspect of Male Infertility. In: *The genetic basis of male infertility* McElreavey K.

Krausz, C. and McElreavey, K., (1999). Y chromosome and male infertility. *Frontiers in Bioscience.* **4**, e1-8, 15

Krebber, H., Bastians, H., Hoheisel, J., Lichter, P., Ponstingl, H., and Joos, S. (1997). Localization of the gene encoding the Ran-binding protein RanBP2 to human chromosome 2q11-q13 by fluorescence in situ hybridization. *Genomics.* **43**, 247-8.

Kundu T.K., Rao M.R. (1999). CpG islands in chromatin organization and gene expression. *J Biochem* **125**, 217-22

Kwon, Y.K., Hecht, N.B., (1993). Binding of a phosphoprotein to the 3' untranslated region of the mouse protamine 2 mRNA temporally represses its translation. *Mol Cell Biol.* **13**, 6547-57.

Kwon, Y.K., Hecht, N.B., (1991). Cytoplasmic protein binding to highly conserved sequences in the 3' untranslated region of mouse protamine 2 mRNA, a translationally regulated transcript of male germ cells. *Proc Natl Acad Sci U S A.* **1,** 3584-8.

Lahn, B.T. & Page, D.C. (1997). Functional coherence of the human Y chromosome. *Science* **278**, 675-680.

Langford, K. G., Zhou, Y., Russell, L. D., Wilcox, J. N. & Bernstein, K. E. (1993). Regulated expression of testis angiotensin-converting enzyme during spermatogenesis in mice. *Biol. Reprod.* **48**, 1210-1218.

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.

Layman L.C., (1999). Genetics of human hypogonadotropic hypogonadism. *Am J Med Genet.* **29**, 240-8.

Le Coniat M., Soulard M., Della Valle V., Larsen C.J., Berger R. (1992). Localization of the human gene encoding heterogeneous nuclear RNA ribonucleoprotein G (hnRNP-G) to chromosome 6p12. *Hum Genet.* **88**, 593-5.

Lee Y.J., Lee Y., Chung J.H. (2000) An intronless gene encoding a poly(A) polymerase is specifically expressed in testis. *FEBS Lett.* **487**, 287-92

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.

Legius, E., Schollen, E., Matthijs, G., Fryns, J.-P., (1998). Fine mapping of Noonan/cardio-facio cutaneous syndrome in a large family. *Europ. J. Hum. Genet.* **6**, 32-37

Lemkin P.F., Thornwall G.C., Walton K.D., Hennighausen L. (2000). The Microarray Explorer tool for data mining of cDNA microarrays: application for the mammary gland.*Nucleic Acids Res.* **28**, 4452-4459

Li, W.B., Gruber, C.E., Lin, J.J., Lim, R., D'Alessio, J.M., Jessee, J.A., (1994). The isolation of differentially expressed genes in fibroblast growth factor stimulated BC3H1 cells by subtractive hybridization. *BioTechniques*, **16**, 722-729

Li, E., Bestor, T.H., Jaenisch, R., (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* **12**, 915-26.

Liang, P., Pardee, A.B., (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* **257**, 967-71.

Liang P., Pardee A.B., (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science.* **257**, 967-971.

Liang P., Averboukh L., Pardee A.B., (1993). Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Res.* **21**, 3269-3275.

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.

Lin Y.C., Chang T.C., Tseng Y.J., Lin Y.L., Huang F.J., Kung F.T., Chang S.Y., (2000). Seminal plasma zinc levels and sperm motion characteristics in infertile samples. Chang Keng I *Hsueh Tsa Chih.* **23**, 260-6.

Lindquist S. (1988). The heat-shock proteins. *Annu. Rev. Genet.* **22**, 631-77.

Linnekin D. (1999). Early signaling pathways activated by c-Kit in hematopoietic cells Int *J Biochem Cell Biol.* **31**(10):1053-74.

Lisitsyn N., Lisitsyn L., Wigler M. (1993). Cloning the differences between two complex genomes. *Science.* **259**, 946-51.

Liu, C., Calogero,A., Ragona,G., Adamson,E. and Mercola,D. (1996). *Crit. Rev. Oncogen.* **7**, 101-125.

Liu B., Aronson N.N. Jr., (1998). Structure of human G protein G gamma 5 gene GNG5. *Biochem Biophys Res Commun.* **251**, 88-94.

Lopes, S., Sun, J.G., Jurisicova, A. et al. (1998). DNA fragmentation in spermatozoa: Implication for failed fertilization in intracytoplasmic sperm injection (ICSI). *Fertil. Steril.* **69**, 528–532.

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 U S A.* **88**, 9628-32

Lubahn, D.B., Brown, T.R., Simental, J.A. *et al.* (1989). Sequence of the intron/exon junctions of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity. *Proc. Natl Acad. Sci.*, **86**, 9534–9538

Lund, L. and Nielsen, K.T. (1996). Varicocele testis and testicular temperature. *Br. J. Urol.* **78**, 113–115.

Luo X., Ikeda Y., and Parker K. L. (1994). Cell-specific Nuclear Receptor Is Essential For Adrenal and Gonadal Development and Sexual Differentiation *Cell, Vol.* **77**, 481–490.

Schlessinger, J., (1994). SH2/SH3 signaling proteins. *Curr. Opin. Genet. & Dev.* **4**, 25-30.

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. (1993). A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell.* **75**,1287-95

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.

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.

Massague J., (1998). TGF-beta signal transduction. *Annu Rev Biochem.* **67**, 753-91.

Mattingly R.R., Macara I.G., (1996). Phosphorylation-dependent activation of the Ras-GRF/CDC25Mm exchange factor by muscarinic receptors and G-protein beta gamma subunits. *Nature.* **382**, 268-272.

Matsuda R., Takahashi T., Nakamura S., Sekido Y., Nishida K., Seto M., Seito T., Sugiura T., Ariyoshi Y., Takahashi T., (1993). *et al* Expression of the c-kit protein in human solid tumors and in corresponding fetal and adult normal tissues. *Am J Pathol.* **142**, 339-46

Matz M.V., Lukyanov S.A. (1998). Different strategies of differential display: areas of application. *Nucleic Acids Res.* **26**, 5537-43.

Matz, M., Usman, N., Shagin, D., Bogdanova, E., Lukyanov, S., (1997). Ordered differential display: a simple method for systematic comparison of geneexpression profiles. *Nucleic Acids Res.* **15**, 2541-2.

Melchior, F., Guan, T., Yokoyama, N., Nishimoto, T., Gerace, L., (1995). GTP hydrolysis by Ran occurs at the nuclear pore complex in an early step of protein import. *J. Cell Biol.* **131**, 571-581.

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.

McElreavey, K., Krausz, C. and Bishop, C.E., (2000). The human Y chromosome and male infertility. In: *The genetic basis of male infertility*, McElreavey (Ed.) Springger-Verlag.

McCarrey, J.R., Kumari, M., Aivaliotis, M.J., Wang, Z., Zhang, P., Marshall, F., Vandeberg, J.L., (1996). Analysis of the cDNA and encoded protein of the human testis-specific PGK-2 gene. *Dev Genet.***19**, 321-32.

McCarrey JR, Thomas K. (1987). Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature.* **2-8**, 501-5.

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.

Mitchell, P.J. and Tjian,R. (1989). Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science.* **245**, 371-378. Review.

Michael D., (1998). The central role of Sertoli cells in spermatogenesis. *Griswold.* 411-416.

Miele, G., MacRae, L., McBride, D., Manson, J., Clinton, M., (1998). Elimination of false positives generated through PCR re-amplification of differential display cDNA. *Biotechniques.* **25**, 138-44.

Migliaccio, A., Di Domenico, M., Castoria, G., de Falco, A., Bontempo, P., Nola, E. and Auricchio, F., (1996). Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. *EMBO J.* **15**, 1292-300

Melchior, F., Guan, T., Yokoyama, N., Nishimoto, T., Gerace, L., (1995). GTP hydrolysis by Ran occurs at the nuclear pore complex in an early step of protein import. *J Cell Biol.* **131**, 571-81.

Monesi, V., (1965). Synthetic activities during spermatogenesis in the mouse RNA and protein. *Exp Cell Res.* **39**, 197-224.

Money,T., Reader,S., Qu,L.J., Dunford,R.P. and Moore,G. (1996). AFLP-based mRNA fingerprinting. *Nucleic Acids Res.* **24**, 2616-2617

Moore, F.L., Reijo-Pera, R.A., (2000). Male sperm motility dictated by mother's mtDNA. *Am J Hum Genet.* **67**, 543-8.

Moore, H.D., (1998). Contribution of epididymal factors to sperm maturation and storage. *Andrologia.* **30**, 233-9

Moore H.D., Akhondi M.A., (1996). In vitro maturation of mammalian spermatozoa. *Rev Reprod.* **1**, 54-60

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.

Moreno R.D., Hoshi M., Barros C. (1999) Functional interactions between sulphated polysaccharides and proacrosin: implications in sperm binding and digestion of zona pellucida. *Zygote.* **7**, 105-11.

Mori C., Nakamura N., Welch J.E., Gotoh H., Goulding E.H., Fujioka M., Eddy E.M., (1998). Mouse spermatogenic cell-specific type 1 hexokinase (mHk1-s) transcripts are expressed by alternative splicing from the mHk1 gene and the HK1-S protein is localized mainly in the sperm tail. *Mol Reprod Dev.* **49**, 374-85.

Nantel F., Monaco L., Foulkes N.S., Masquilier D., LeMeur M., Henriksen K., Dierich A., Parvinen M. & Sassone-Corsi P.. (1996). Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. *Nature.* **380**, 159-62.

Nachtigal M.W., Hirokawa Y., Enyeart-VanHouten D.L., Flanagan J.N., Hammer G.D., Ingraham H.A. (1998). Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell.* **1**, 445-54.

Nayernia K., Nieter S., Kremling H., Oberwinkler H., Engel W. (1994). Functional and molecular characterization of the transcriptional regulatory region of the proacrosin gene. *J Biol Chem.* **23**, 32181-6.

Nayernia K., Adham I., Kremling H., Reim K., Schlicker M., Schluter G., Engel W. (1996). Stage and developmental specific gene expression during mammalian spermatogenesis. *Int J Dev Biol.* **40**, 379-83. Review.

Nayernia, K., von Mering, M.H., Kraszucka, K., Burfeind, P., Wehrend, A., Kohler, M., Schmid, M., and Engel, W., (1999). A novel testicular haploid expressed gene (*THEG*) involved in mouse spermatid-sertoli cell interaction. *Biol. Reprod.* **60**, 1488-95.

Niederberger C.S., Shubhada S., Kim S.J., Lamb D.J. (1993) Paracrine factors and the regulation of spermatogenesis. *World J Urol.* 11:120-128

Nieschlag, E., (1997). Classification of andrological disorders. In: Nieschlag, E. and Behre, H.M. (eds), Andrology. *Male Reproductive Health and Dysfunction. Springer*, Berlin, Germany, pp. 81–83.

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.

Nishiyama H., Danno S., Kaneko Y., Itoh K., Yokoi H., Fukumoto M., Okuno H., Millan J.L., Matsuda T., Yoshida O., Fujita J.Am J. (1998). Decreased expression of cold-inducible RNA-binding protein (CIRP) in male germ cells at elevated temperature. *Pathol.* **152**, 289-96.

Nistal, M., Paniagua, R., Picazo, M.L., (1983). Testicular and epididymal involvement in Fabry's disease. *J Pathol.* **141**, 113-24.

Nordgren A., Arver S., Kvist U., Carter N., Blennow E. (1997). 5Trisomy 5q12-->q13.3 in a patient with add(13q): characterization of an interchromosomal insertion by forward and reverse chromosome painting. *Am J Med Genet.* **19**, 351

Nothwang, H.G., Rensing, C., Kubler, M., Denich, D., Brandl, B., Stubanus, M., Haaf, T., Kurnit, D., and Hildebrandt, F. (1998). Identification of a novel Ran binding protein 2 related gene (*RANBP2L1*) and detection of a gene cluster on human chromosome 2q11-q12. *Genomics.* **47**, 383-92.

Nudell, D., Castillo, M., Turek, P.J., Pera, R.R., (2000). Increased frequency of mutations in DNA from infertile men with meiotic arrest. *Hum Reprod.* **15**, 1289-94.

Okabe, M., Ikawa, M., and Ashkenas, J. (1998). Male infertility and the genetics of spermatogenesis. *Am. J. Hum. Genet.* **62**, 1274-81.

Oko, R., Korley, R., Murray, M.T., Hecht, N.B., Hermo, L., (1996). Germ cell-specific DNA and RNA binding proteins p48/52 are expressed at specific stages of male germ cell development and are present in the chromatoid body. *Mol Reprod Dev.* **44**, 1-13.

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.

Ong O.C., Yamane H.K., Phan K.B., Fong H.K., Bok D., Lee R.H., Fung B.K., (1995). Molecular cloning and characterization of the G protein gamma subunit of cone photoreceptors. *J Biol Chem.* **270**, 8495-8500

Osman, R.A., Andria, M.L., Jones, A.D., Meizel, S., (1989). Steroid induced exocytosis: the human sperm acrosome reaction. *Biochem Biophys Res Commun* **28**, 828-33.

Osoegawa K., Tateno M., Woon P.Y., Frengen E., Mammoser A.G., Catanese J.J., Hayashizaki Y., de Jong P.J. (2000). Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10**, 116-28.

Ozaki, K., Kwroki ,T., Hayashi, S., and Nakamuro, Y. (1996) Isolation of three testis-specific genes (*TSA303*, *TSA806* and *TSA903*) by Differential mRNA Display Methode. *Genomics*. **36**, 316-319.

Paduch D.A., Niedzielski J., (1996). Semen analysis in young men with varicocele: preliminary study. *J. Urol.* **156**, 788-90.

Parola, A.L., Kobilka, B.K., (1994). The peptide product of a 5' leader cistron in a ß2 adrenergic receptor mRNA inhibits receptor synthesis. *J. Biol. Chem.* **269**, 4497–505.

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.

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 U S A.* **88**, 9623-7.

Parker K.L., Schimmer B.P., (1997). Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocr Rev.* **18**, 361-77.

Pearse R.V., Drolet D.W., Kalla K.A., Hooshmand F., Bermingham J.R. Jr., Rosenfeld M.G. (1997). Reduced fertility in mice deficient for the POU protein sperm-1. *Proc Natl Acad Sci.* **8**, 7555-60.

Peri A., Krausz C., Cioppi F., Granchi S., Forti G., Francavilla S. & Serio M., (1998). CREM gene expression in germ cells of normo- and oligozoospermic men. *J. Clin. Endocrinol. Metab.* **83**, 3722-6.

Pennarun G., Escudier E., Chapelin C., Bridoux A.M., Cacheux V., Roger G., Clément A., Goossens M., Amselem S., and Duriez B A.M. (1999). Loss-of-Function Mutations in a Human Gene Related to Chlamydomonas reinhardtii Dynein IC78 Result in Primary Ciliary Dyskinesia *J. Hum. Genet.* **65**:1508-1519.

Pontiggia A., Rimini R., Harley V.R., Goodfellow P.N., Lovell-Badge R., Bianchi M.E., (1994). Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* **13**, 6115-24

Pritchard-Jones K., Fleming S., Davidson D., Bickmore W., Porteous D., Gosden C.J., Buckler A., Pelletier J., Housman D., (1990). The candidate Wilms' tumour gene is involved in genitourinary development. *Nature*. **12**, 194-7

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.

Proudfoot, N.J., and Brownlee, G.G. (1976). 3' Non-coding region sequence in eukaryotic messenger RNA. *Nature*. **263**, 211–214.

Purvis K., Christiansen E. (1992). Male infertility: current concepts. *Ann Med.* **24**, 258-272.

Qi L., Sit K.H. (2000). Housekeeping genes commanded to commit suicide in CpG-cleavage commitment upstream of Bcl-2 inhibition in caspase-dependent and -independent pathways. *Mol Cell Biol Res Commun* **3**, 319-27

Rao, C.D., Pech, M., Robbins, K.C., Aaronson, S.A., (1988). The 5' untranslated sequence of the c-sis/platelet-derived growth factor 2 transcript is a potent translational inhibitor. *Mol. Cell. Biol.* **8**, 284–92

Ray K., Kunsch C., Bonner L.M., Robishaw J.D., (1995). Isolation of cDNA clones encoding eight different human G protein gamma subunits, including three novel form designated the gamma 4, gamma 10, and gamma 11 subunits. *J Biol Chem.* **270**, 21765-21771.

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.

Resh M.D. (1996). Regulation of cellular signalling by fatty acid acylation and prenylation of signal transduction proteins. *Cell Signal.* **8**, 403-412.

Richter, J.D., (1991). Translational control during early development. *Bioessays.* **13**, 179-83. Review.

Robishaw J.D., Kalman V.K., Moomaw C.R., Slaughter C.A., (1989). Existence of two gamma subunits of the G proteins in brain. *J Biol Chem.* **264**, 15758-15761.

Rodriguez T.A., Burgoyne P.S. (2000) Evidence that sex chromosome asynapsis, rather than excess Y gene dosage, is responsible for the meiotic impairment of XYY mice.*Cytogenet Cell Genet.* 89(1-2):38-43

Rossi P., Dolci S., Albanesi C., Grimaldi P., Ricca R., Geremia R. (1993) Follicle-stimulating hormone induction of steel factor (SLF) mRNA in mouse Sertoli cells and stimulation of DNA synthesis in spermatogonia by soluble SLF. *Dev Biol.* 155(1):68-74

Rubin C.M., VandeVoort C.A., Teplitz R.L., Schmid C.W. (1994) Alu repeated DNAs are differentially methylated in primate germ cells. *Nucleic Acids Res* **25**, 5121-7

Ruggiu M., Speed R., Taggart M., McKay S.J., Kilanowski F., Saunders P., Dorin J., Cooke H.J. (1997) The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. *Nature.* **4**, 73-7

Ryba N.J., Tirindelli R., (1995). A novel GTP-binding protein gamma-subunit, G gamma 8, is expressed during neurogenesis in the olfactory and vomeronasal neuroepithelia. *J Biol Chem.* **270**, 6757-6767.

Saez J.M., (1994). Leydig cells: endocrine, paracrine, and autocrine regulation. *Endocr Rev.* **15**, 574-626.

Sakai, I., Sharief, F. S., Li, S. S.-L., (1987). Molecular cloning and nucleotide sequence of the cDNA for sperm-specific lactate dehydrogenase-C from mouse. *Biochem. J.* **242**, 619-622,

Sambrook, J., Fritch, E.F., Maniatis, T., *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press, 1989.

Sanger, F., Nicklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**, 5463-5467.

Sar M., Hall S.H., Wilson E.M., French F.S., (1993). Androgen regulation of Sertoli cells. In: Russell LD, Griswold MD (eds) *The Sertoli Cell.* Cache River Press, Clearwater. Fl 509-516.

Sargent, T.D. (1987). Isolation of differentially expressed genes. *Methods Enzymol.* **152**, 423-432.

Sasagawa I., Nakada T., Ishigooka M., Hashimoto T., Izumiya K., Adachi Y. (1992). An azoospermic male with reciprocal translocation t(3;4)(p12;q21). *Urol Int.* **48**, 425-7

Sassone-Corsi, P., (2000). CREM: a master-switch regulating the balance between differentiation and apoptosis in male germ cells. *Mol Reprod Dev.* **56**, (Suppl) 228-9. Review.

Sassone-Corsi P. (1995). Transcription factors responsive to cAMP. *Annu. Rev. Cell Dev. Biol.* **11**, 355-77.

Sassone-Corsi P. (1997). Transcriptional checkpoints determining the fate of male germ cells. *Cell.* **24**, 163-6.

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.

Sazer, S., (1996). The search for the primary function of the Ran GTPase continues. *Trends cell Biol.* **6**, 81-85.

Schafer M., Nayernia K., Engel W., Schafer U. (1995). Translational control in spermatogenesis. *Dev Biol.* **172**, 344-52.

Schena M., Shalon D., Davis R.W., Brown P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray.*Science.* **270**, 467-70.

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.

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.

Sheets, M.D., Ogg, S.C., and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nuc. Acids Res.* **18**, 5799-805.

Scherer S.W., Feinstein D.S., Oliveira L., Tsui L.C., Pittler S.J., (1996). Gene structure and chromosome localization to 7q21.3 of the human rod photoreceptor transducin gamma-subunit gene (GNGT1). *Genomics.* **35**, 241-243.

Schonwasser D.C., Marais R.M., Marshall C.J., Parker P.J., (1998). Activation of the mitogen-activated protein kinase/extracellular signal-regulated kinase pathway by conventional, novel, and atypical protein kinase C isotypes. *Mol Cell Biol.* **18**, 790-798.

Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tome P., Aggarwal A., Bajorek E., Bentolila S., Birren B.B., Butler A., Castle A.B., Chiannilkulchai N., Chu A., Clee C.,

Cowles S., Day P.J., Dibling T., Drouot N., Dunham I. Duprat S., East C., Hudson T.J., (1996) A gene map of the human genome. *Science.* **274**, 540-546.

Shaper N.L., Wright W.W., Shaper J.H. (1990) Murine beta 1,4-galactosyltransferase: both the amounts and structure of the mRNA are regulated during spermatogenesis. *Proc Natl Acad Sci.* **87**, 791-5

Sharpe RM 1994 Regulation of Spermatogenesis. In: Knobil E, Neill JD (eds) *The Physiology of Reproduction.* 2nd edition. Raven Press Ltd, New York, 1363-1434

Shen, W.H., Moore, C.C., Ikeda, Y., Parker, K.L., Ingraham, H.A., (1994). Nuclear receptor steroidogenic factor 1 regulates the mullerian inhibiting substance gene: a link to the sex determination cascade. *Cell.* **3**, 651-61.

Shi, H., Zhang, Z., Wang, X., Liu, S., Teng, C.T., (1999). Isolation and characterization of a gene encoding human Kruppel-like factor 5 (IKLF): binding to the CAAT/GT box of the mouse lactoferrin gene promoter. *Nucleic Acids Res.* **15**, 4807-15.

Sibony, M., Segretain, D. & Gasc, J.-M. (1994). Angiotensin-converting enzyme in murine testis:step-specific expression of the germinal isoform during spermiogenesis *Biol. Reprod.* **50**, 1015-1026.

Sibony, M., Gasc, J.-M., Soubrier, F., Alhenc-Gelas, F. & Corvol, P. (1993). Gene expression and tissue localization of the two isoforms of angiotensin I converting enzyme. *Hypertension* **21**, 827-835.

Siegfried Z., Eden S., Mendelsohn M., Feng X., Tsuberi B.Z., Cedar H. (1999). DNA methylation represses transcription in vivo. *Nat Genet* **22**, 203-6

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., and 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.

Shaban S.F., (1999). Male infertility overview, (Assessment, Diagnosis, and Treatment) at Atlanta Reproductive Health Center web site.

Shalom S., Don J. (1999). Tlk, a novel evolutionarily conserved murine serine threonine kinase, encodes multiple testis transcripts. *Mol Reprod Dev.* **52**, 392-405.

Sheets M.D., Ogg S.C., Wickens M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* **11**, 5799-805.

Starborg M., Brundell E., Hoog C. (1992). Analysis of the expression of a large number of novel genes isolated from mouse prepubertal testis. *Mol Reprod Dev.* **33**, 243-51.

Simon, M.I., Strathmann, M.P., Gautam, N., (1991). Diversity of G proteins in signal transduction. *Science.* **10**, 802-8. Review.

Singal, R., vanWert, J., Bashambu, M., Wolfe, S. A., Wilkerson, D. C., Grimes, S. R., (2000). Testis-Specific Histone H1t Gene Is Hypermethylated in Nongerminal Cells in the Mouse. *Biol Reprod.* **63**:1237-1244.

Singal, R., Ginder, G. D., (1999). DNA Methylation. *Blood.* **93**: 4059-4070.

Sitaram N., Nagaraj R., (1995). Seminal plasmin.Bioessays. **17**, 415-22

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.

Schmidt R.J., Holley J.L., (1998). Fertility and contraception in end-stage renal disease. *Adv Ren Replace Ther.* **5**, 38-44.

Schmiesing, J.A., Ball, A.R.J., Gregson, H.C., Alderton, J.M., Zhou, S., Yokomori, K. (1998). Identification of two distinct human SMC protein complexes involved in mitotic chromosome dynamics. Proc. Natl. Acad. Sci. USA. **95**, 12906-12911

Shoemaker D.D., Schadt E.E., Armour C.D., He Y.D., Garrett-Engele P., McDonagh P.D., Loerch P.M., Leonardson A., Lum P.Y., Cavet G., Wu L.F., Altschuler S.J., Edwards S., King J., Tsang J.S., Schimmack G., Schelter J.M., Koch J., Ziman M., Marton M.J., Li B., Cundiff P., Ward T., Castle J., Krolewski M., Meyer M.R., Mao M., Burchard J., Kidd M.J., Dai H., Phillips J.W., Linsley P.S., Stoughton R., Scherer S., Boguski M.S. (2001). Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922-7

Slee R., Grimes B., Speed R.M., Taggart M., Maguire S.M., Ross A., McGill N.I., Saunders P.T., Cooke H.J. (1999). A human DAZ transgene confers partial rescue of the mouse Dazl null phenotype. *Proc Natl Acad Sci* **6**, 8040-5.

Soto, A.M., Michaelson, C.L., Sonnenschein C., (1998). Screening of chemicals for estrogen and androgen agonist and antagonist activity. "In culture" assays. In: In vitro germ cell developmental toxicology, from science to social and industrial demand. *del Mazo J., Plenum Press Corporation*, NY

Soulard M., Della Valle V., Siomi M.C., Pinol-Roma S., Codogno P., Bauvy C., Bellini M., Lacroix J.C., Monod G., Dreyfuss G., et al (1993). hnRNP G: sequence and characterization of a glycosylated RNA-binding protein. *Nucleic Acids Res* **11**, 4210-7.

Sompayrac, L., Jane, S., Burn, T.C., Tenen, D.G., and Danna K.J. (1995). Overcoming limitations of the mRNA differential display technique. *Nuc. Acids Res.* **22**, 4738-4739.

Sorrentino V., Giorgi M., Geremia R., Besmer P., Rossi P., (1991). Expression of the c-kit proto-oncogene in the murine male germ cells. *Oncogene.* **6**, 149-51.

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.

Steen, R.L., Cubizolles, F., Le Guellec, K., Collas, P., (2000). A kinase-anchoring protein (AKAP)95 recruits human chromosome-associated protein (hCAP)-D2/Eg7 for chromosome condensation in mitotic extract. *J Cell Biol.* **1**;149(3):531-6

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.

Sun Y.., Hegamyer G. and Colburn N.H. (1994). Molecular cloning of five messenger RNAs differentially expressed in preneoplastic or neoplastic JB6 mouse epidermal cells: one is homologous to human tissue inhibitor of metalloproteinases-3 *Cancer Research*, **54**, 5 1139-1144.

Suzuki,H., Yaoi,T., Kawai,J., Hara,A., Kuwajima,G. and Watanabe,S. (1996) Restriction landmark cDNA scanning (RLCS): a novel cDNA display system using two-dimensional gel electrophoresis. *Nucleic Acids Res.* **24**, 289-294.

Swain A., Lovell-Badge R. (1997) A molecular approach to sex determination in mammals. *Acta Paediatr Suppl.* **423**, 46-9.

Swain A., Narvaez V., Burgoyne P., Camerino G., Lovell-Badge R., (1998). Dax1 antagonizes Sry action in mammalian sex determination. *Nature* **391**, 761-7.

Tafuri, S.R., Familari, M., Wolffe, A.P., (1993). A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes. *J Biol Chem.* **5**, 12213-20.

Tanaka H., Iguchi N., Nakamura Y., Kohroki J., de Carvalho C.E., Nishimune Y. (2001). Cloning and characterization of human haspin gene encoding haploid germ cell-specific nuclear protein kinase. *Mol Hum Reprod.* **7**, 211-218.

Tao L., Pandey S., Simon M.I., Fong H.K., (1993). Structure of the bovine transducin gamma subunit gene and analysis of promoter function in transgenic mice. *Exp Eye Res.* **56**, 497-507.

Taylor J.M., Jacob-Mosier G.G., Lawton R.G., VanDort M., Neubig R.R., (1996). Receptor and membrane interaction sites on Gbeta. A receptor-derived peptide binds to the carboxyl terminus. *J. Biol. Chem.* **271**, 3336-3339

Temple-Smith P.D., Zheng S.S., Kadioglu T., Southwick G.J. (1998). Development and use of surgical procedures to bypass selected regions of the mammalian epididymis: effects on sperm maturation. *J Reprod Fertil Suppl.* **53**, 183-95.

Tesarik, J., Greco, E., Cohen-Bacrie, P. and Mendoza, C., (1998). Germ cell apoptosis in men with complete and incomplete spermiogenesis failure. *Mol. Hum. Reprod.* **4**, 757–762.

Tesarik, J., Mendoza, C., (1996). Spermatid injection into human oocytes. I. Laboratory techniques and special features of zygote development. *Hum Reprod.* **11**, 772-9.

Teyssier M., Pousset G., (1994) XYYY mosaicism case report and review of the literature. *Genet Couns.* **5**, 357-61.

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.

Tournaye. H., (1999). ICSI: the revolution and the portents. In: Glover T.D. and Barratt C.L.R., *Male fertility and infertility.* Cambridge University Press. 149-161.

Toussaint C., Bousquet-Lemercier B., Garlatti M., Hanoune J., Barouki R., (1994). Testis-specific transcription start site in the aspartate aminotransferase housekeeping gene promoter. *J Biol Chem.* **6**, 13318-24.

Trasler J.M. (1998). Origin and roles of genomic methylation patterns in male germ cells. *Semin Cell Dev Biol.* **9**, 467-74.

Trasler, J.M., Alcivar, A.A., Hake, L.E., Bestor, T., Hecht, N.B., (1992). DNA methyltransferase is developmentally expressed in replicating and non-replicating male germ cells. *Nucleic Acids Res.* **25**, 2541-5.

Trasler, J.M., Hake, L.E., Johnson, P.A., Alcivar, A.A., Millette, C.F., Hecht, N.B., (1990). DNA methylation and demethylation events during meiotic prophase in the mouse testis. *Mol Cell Biol.* **10**, 1828-34.

Tremblay K.D., Duran K.L., Bartolomei M.S. (1997). A 5' 2-kilobase-pair region of the imprinted mouse H19 gene exhibits exclusive paternal methylation throughout development. *Mol Cell Biol.* **17**, 4322-9.

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., Nishida T., Fujimoto H. (1999). The Hsp70 homolog gene, Hsc70t, is expressed under translational control during mouse spermiogenesis. *Mol Reprod Dev.* **52**, 383-91

Tut T.G., Ghadessy F.J., Trifiro M.A., Pinsky L., Yong E.L. (1997). Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility.*J Clin Endocrinol Metab.* **82**, 3777-82.

Tycko, B., Trasler, J., Bestor, T., (1997). Genomic imprinting: gametic mechanisms and somatic consequences. *J Androl.* **18**, 480-6. Review.

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 Wijnen, A.J., Stein, J.L., and Stein, G.S. (1987). A nuclear protein with affinity for the 5` flanking region of a cell cycle dependent human H4 histone gene in vitro. *Nuc.* Acids Res. **15**, 1679-1698.

Van Biesen T., Hawes B.E., Luttrell D.K., Krueger K.M., Touhara K., Porfiri E., Sakaue M., Luttrell L.M., Lefkowitz R.J., (1995). Receptor-tyrosine-kinase- and G beta gamma-mediated MAP kinase activation by a common signalling pathway. *Nature.* **376**, 781-784.

Vanderzwalmen, P., Lejeune, B., Nijs, M., (1995). Fertilization of an oocyte microinseminated with a spermatid in an in-vitro fertilization programme. *Hum. Reprod.* **10**, 502–503.

Vanderzwalmen P, Zech H, Birkenfeld A, Yemini M, Bertin G, Lejeune B, Nijs M, Segal L., Stecher A., Vandamme B., van Roosendaal E., Schoysman R. (1997). Intracytoplasmic injection of spermatids retrieved from testicular tissue: influence of testicular pathology, type of selected spermatids and oocyte activation. *Hum Reprod* **12**, 1203-13

Vanderzwalmen, P., Zech H., Birkenfeld, A., (1997). Intracytoplasmic injection of spermatids retrieved from testicular tissue: influence of testicular pathology, type of selected spermatids and oocyte activation. *Hum. Reprod.*, **12**, 1203–1213.

Vaux D.L., Strasser A., (1996). The molecular biology of apoptosis. *Proc Natl Acad Sci U S A.* **93**, 2239-44. Review.

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.

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.

Vogel T., Dittrich O., Mehraein Y., Dechend F., Schnieders F., Schmidtke J. (1998). Murine and human TSPYL genes: novel members of the TSPY-SET-NAP1L1 family. *Cytogenet Cell Genet.* **81**, 265-70.

Vogel, T., Borgmann ,S., Dechend, F., Hecht, W. & Schmidtke, J. (1997). Conserved Y-chromosomal location of TSPY in Bovidae. *Chromosome Res.* **5**, 182-5.

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. 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.

Von Stein,O.D., Thies,W.-G. and Hofmann,M. (1997). A high throughput screening for rarely transcribed differentially expressed genes. *Nucleic Acids Res.* **25**, 2598-2602

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **11**, 4407-14.

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.

Wallrapp C., Muller-Pillasch F., Micha A., Wenger C., Geng M., Solinas-Toldo S., Lichter P., Frohme M., Hoheisel J.D., Adler G., Gress T.M. (1999). Novel technology for detection of genomic and transcriptional alterations in pancreatic cancer. *Ann Oncol.* 10 Suppl. **4**:64-8

Wan, J.S., Sharp, S.J., Poirier, G.M., Wagaman, P.C., Chambers, J., Pyati, J., Hom, Y.L., Galindo, J.E., Huvar, A., Peterson, P.A., Jackson, M.R., Erlander, M.G., (1996). Cloning differentially expressed mRNAs. *Nat Biotechnol.* **14**, 1685-91.

Wang, C., McDonald, V., Leung, A. et al. (1997) Effect of increased scrotal temperature on sperm production in normal men. Fertil. *Steril.* **68**, 334–339.

Wehling M., (1997). Specific, nongenomic actions of steroid hormones. *Annu. Rev. Physiol.* **59**, 365-93.

Weighardt, F., Biamonti, G. & Riva, S. (1996). The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. **18**, 747-56.

Welch, J.E., Brown, P.R., O'Brien, D.A., Eddy, E.M., (1995). Genomic organization of a mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells. *Dev Genet.***16**, 179-89.

Welsh J., Chada K., Dalal S.S., Cheng R., Ralph D., McClelland M., (1992). Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res* **20**, 4965-70.

Wieacker P., Jakubiczka S. (1997). Genetic causes of male infertility. *Andrologia.* **29**, 63-9. Review.

Wilcox M.D., Schey K.L., Busman M., Hildebrandt J.D., (1995). Determination of the complete covalent structure of the gamma 2 subunit of bovine brain G proteins by mass spectrometry. *Biochem Biophys Res Commun.* **212**, 367-374

Willison, K. & Ashworth, A. (1987). Mammalian spermatogenic gene expression. *TIG.* **3**, 351-355.

Witte O.N. (1990) Steel locus defines new multipotent growth factor. *Cell.* **63**: 5–6.

Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., Lockhart, D.J., (1997). Genome-wide expression monitoring in Saccharomyces cerevisiae. *Nat Biotechnol.* **15**, 1359-67.

Wolffe, A.P., (1994). The role of transcription factors, chromatin structure and DNA replication in 5 S RNAgene regulation. *J Cell Sci.* **107**, 2055-63. Review.

Wollin, M., Marshall, F.F., Fink, M.P., Malhotra, R., Diamond, D.A., (1987). Aberrant epididymal tissue: a significant clinical entity. *J Urol.* **138**, 1247-50.

World Health Organization (1992) *WHO Laboratory Manual for the Examination of Human Semen and Sperm–Cervical Mucus Interaction.* 3rd edn. Cambridge University Press, *Cambridge. UK.*

Wormington, M., (1993). Poly(A) and translation: development control. *Curr Opin Cell Biol.* **5**, 950-4. Review.

Wunderle V.M., Critcher R., Ashworth A., (1996) Goodfellow PN Cloning and characterization of SOX5, a new member of the human SOX gene family. *Genomics.* **36**, 354-8

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 W., Cooper G.M. (1995). Identification of a candidate c-mos repressor that restricts transcription of germ cell-specific genes. *Mol Cell Biol.* **15**, 5369-75.

Xu Y., Ashley T., Brainerd E.E., Bronson R.T., Meyn M.S., Baltimore D. (1996). Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **1**, 2411-22

Yamamoto Y., Sofikitis N., Miyagawa I. (1999) Ooplasmic injections of rabbit round spermatid nuclei or intact round spermatids from fresh, cryopreserved and cryostored samples.*Hum Reprod.* **14**, 1506-15

Yan K., Gautam N., (1996). A domain on the G protein beta subunit interacts with both adenylyl cyclase 2 and the muscarinic atrial potassium channel. *J. Biol. Chem.* **271**, 17597-17600

Yasuda H., Lindorfer M.A., Woodfork K.A., Fletcher J.E., Garrison J.C., (1996). Role of the prenyl group on the G protein gamma subunit in coupling trimeric G proteins to A1 adenosine receptors. *J Biol Chem.* **271**, 18588-18595.

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.

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**, 2013-7.

Yoder, J.A., Walsh, C.P., Bestor, T.H., (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* **13**,335-40. Review.

Young, D., (1970). Surgical treatment of male infertility. *J. Reprod. Fertil.* **23**, 541-542.

Yuan L., Liu J.G., Hoog C. (1995). Rapid cDNA sequencing in combination with RNA expression studies in mice identifies a large number of male germ cell-specific sequence tags. *Biol Reprod.* **52**, 131-8.

Zhang, L., Zhou, W., Velculescu, V.E., Kern, S.E., Hruban, R.H., Hamilton, S.R., Vogelstein, B., Kinzler, K.W., (1997). Gene expression profiles in normal and cancer cells. *Science.* **23**, 1268-72.

Zhang S., Coso O.A., Lee C., Gutkind J.S., Simonds W.F., (1996). Selective activation of effector pathways by brain-specific G protein beta5. *J Biol Chem.* **271**, 33575-33579.

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.

Zinkham, W. H., Blanco, A., Clowry, L. J., (1964). An unusual isozyme of lactic dehydrogenase in mature testes: localization, ontogeny, and kinetic properties. *Ann. N.Y. Acad. Sci.* **121**, 571-588.

Appendix

A.1 Primers used for Differential Display RT-PCR

Differential Display RT-PCR primers from Operon Technologies Inc. (Purchased from VH Bio):

Anchor kit 2 (SK020)

Primer name	Primer sequence 5′ to 3′
$dT_{12}VA$ or VA	TTTTTTTTTTTTVA
$dT_{12}VC$ or VC	TTTTTTTTTTTTVC
$dT_{12}VG$ or VG	TTTTTTTTTTTTVG
$dT_{12}VT$ or VT	TTTTTTTTTTTTVT

10-mer kit (SK030)

Primer name	Primer sequence 5′ to 3′
OP-1	TACAACGAGG
OP-2	TGGATTGGTC
OP-3	CTTTCTACCC
OP-4	TTTTGGCTCC
OP-5	GGAACCAATC
OP-6	AAACTCCGTC
OP-7	TCGATACAGG
OP-8	TGGTAAAGGG
OP-9	TCGGTCATAG
OP-10	GGTACTAAGG
OP-11	TACCTAAGCG
OP-12	CTGCTTGATG
OP-13	GTTTTCGCAG
OP-14	GATCAAGTCC
OP-15	GATCCAGTAC
OP-16	GATCACGTAC
OP-17	GATCTGACAC
OP-18	GATCTCAGAC
OP-19	GATCATAGCC
OP-20	GATCAATCGC
OP-21	GATCTAACCG
OP-22	GATCGCATTG
OP-23	GATCTGACTG
OP-24	GATCATGGTC
OP-25	GATCATAGCG
OP-26	GATCTAAGGC

A.2 Sequence information and nomenclature for DDRT cDNAs isolated from human adult testis

A.2.1 Nomenclature:
The first letter refers to the testis or sperm and the first number refers to the order the bands were excised from the acrylamide gel (the first letter without number means it was from the first band). The second letter refers to which anchor primer was used and the second number refers to Operon 10-mer primer used. Bands excised from DDRT gels displaying human testis or sperm against other tissues were named in this way. eg. T2G4 or S2G9

A.2.2 DDRT-PCR sequences

Primers were underlined
DDRT-PCR primers are italic

T2A13

*TTTTTTTTTTTCA*CACATAACCACTGCCATATTTAAAGCACTGTACACCGGCC
CTCGGAGCCAGAATTCTTTCCCTCGAGGTTGCCCCAGGAATCCCTTTGACCTC
TGCCATTGGACGTCACAGAAGGCGCGACTAATCGTGGCTGGCTCCCGGAGCT
TCACTCTTCCCTCTCTTCCTGC

23TG2

*TGGATTGGTC*TATCAGATTGAACTCCAAGAATGATCACACAAAATGTTTAGGG
AGATGTTCCCCGTGGTGTATCCTCATGGTAACAACGACAAAAAATGCCGGTT
GTCTTTGTTCTCTTTTCACTATTCCTAACATGTGTACATGATAGCTTTGATTCT
GCAAGTAAAAGTAAATCCTGTGTTG*CGAAAAAAAAAAAA*

SG2 (colonies 3&5)

*TGGATTGGTC*TACATGTCTATCTTTATGTCAACAGCACACTGTTTTAATTACTG
CAGCTTTGTAGTAAGTTTTGCAATCAGAAAATGAGAGTCCTCCATTTTGTTCT
TCTCTTTCAAGATTGTTTTGGCTACTTGGGGTCCTTGAAATTTTATATAAATTT
CAGAATAGGCTTTTCTATTTCTG*CCAAAAAAAAAAA*

SG2

TGATGTCTGTATTTTGTAATTGTGAAAATACCAGATACTGTCATGAAAACAGT
GAACCTAAGTAATACAAAAGTAATCTATACGAGGCTATTTTGTTTTCTTGCCA
TTTTCAATCATTCTTGTCAGTGGTACTTTTTTTCATGGCCAAGTTGGTAACCTT
CTTTATAAGGAACACCCTAGCAACAACTAATATGTGCTATGACCG

4VA2

TTTTTTTTTTTTCAGGACAATACAGCTTTTTATTAAAACTGCAGCACTTCTGAC
TTACAGTCATAAAATGCTTCAAGGAACAGAATGCTCTCCTGATCTGCTGGGG
AGATGAATCAATGTGGAGAGGCTGACAGA

S1G9
CGTAACGTTTTACGACAAGTCTTTTTCACTTTAGGATTGACCCCTATAAGTGT
TTGGGTACCACCGGGTACGACACAACGAACGACCCCTCCCCGGAAAAGAGTG
ACGTGAAAGACAGTAAGAGTGGG

T1G2
GTGTGACAAAATTAATGACGTCGAAACATCATTCAAAACGTTAGTCTTTTACT
CTCAGGAGGTAAACACAAGAAGAGAAAGTTCTAACAAAACCGTGCCCCGGA
A

T1G4
ACATCATGAATTCATACTAATATTTCCCATTTAAATGTAAGATTACAGTGTTT
GTGCAGCTTCTTTGATTTTTTAATTTGTATCTCTCTCCTCTTATGCTGAAAATC
TTGGTTCCTAATGACATCTGCGTGATGACTTACCTGCTTTCTCTCACTGCACAT
GATCTGTGCGACTGATTGGAACGTGGGGAGCCAAA

T2G4
GGCTCCCAGATGAACAAAGAAGTTTACGGAATCTCCACATTTCTAATTTCAAT
TTTTCTTTACATAACTGGGTTTTGTTTTGTATTTATTTATAGCTAAGTTATTCCC
ATTTGTGGCAGCAGAGTGGGACACTGATTTTATGGGCTGACCAATCCCGGGA
TGTACCCAA GAGAAATGAA AACATATGCCCAAAAAAAAAAAA

T3G5
AAGATCCTTTAAGACTATTATGAATGACTCGATACACACAAACTATAAAATC
TAGAGGAAATGAATTTCTGGAAACACACATCTCCCAAGATTGAATCAAGAAG
AAATTGAAACTCTGAATAGGCCAATATTGAGTACTGAAATTGAATCCATAAT
ATAAATA

TG2 (colony26)
*TGGATTGGGC*AACAAATAAATCGAGTGTAAAAGAAGTCTGATTCCTCATTGCT
GACAGGAAGACAGCCATGGATTGGGAAAAGATTAGAGTGAATCCTGTGATGT
TGGATTGGTCCAAAGGCATTCAGATTATAATAACGTACCGCAATGATACATC
GAGCAGTTACTAAATGTCGGACTTCCCTG*CCAAAAAAAAAAAA*

A.3 A sequence was obtained during 5RACE of TC10 using GSP7

TC10-GSP7
1-TTCTGGGTCTTCTTCCTACTTCTTTGGCTTCTTGTTACTC
41-AGTCCTGGATCCTCTTCTCATTCTCTATATTTCCCTAGGT
81-AATTTCACGGTTTTAAAATATCATTTATATGCTGATGTGT

121-CCCAAATATGTATCTTTCTCTTCTGAGACACAAACTTAAA
161-TATTCAGCTGCTTACCTTAGATCTTTATTGGGACTTAAAA
201-TACCCAAATGTGAGTTCTCCTTCAATATTATCTTTTATATT
241-TCTGTTTTAATATTTCCCGGTTTGATAAATAGTATCTCC
281-ATCTACCTAGTTGGTTATACCAGACATTCTTGAACTATAT
321-AATAGCAAGTGTAACAAT

A.4 Mouse homologue sequence of *TSGA10*

Primers were underlined.

5`Flanking
155 bases
GACAGCTTGGTAACTCTGCTTTACTGATGGAAAGTTAAGAACAATAGAAA
ATATGAACGAAATCCATTGTTACTAGTTTTCAAAACTTACTGATGTGAGA
AGTCACAGCTCTGTTTTAAATAGAGCTCATTCTTATATCTCATTCTTTTT
GTTAG

5'end
>*tsga10* 2731 bases
GGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTTTCTGAAG MTC10-G12
CTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAA MTC10-R12
GATAGATTCATCTCTACACTGAAATTTACAGATTGAAGATCTCAAACAGA
CAAATCATGACTTGGAAGAATATGTTAGGAAACTCTTGGATAGTAAAGAG
GCGGTAAGCACTCAAGTAGATGACTTAGCCAACCACAATGAGCACCTTTG
TAAAGAGTTGATTAAACTTGACCAACATGCAGAGAAATTACAAAAAGAAA
AAAATTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGCAAAGATTG
AACTCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAA
AGGCTTAAGTCCATAATTTTAGAGACTGAAAAGGCACAAAATACATCTCC
TTCTAGACTGGATTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTATT
ATAAGACTGAAGCTCAGAATTTAAGAAAAATGATGAGAAATAGATCTAAG
AGTCCGAGACGCCCATCGCCAACTTCCCGGGCTGCAAACTGTGCTGTAGA
GCTTTTGAAGTCAACTGCAAGAGACCGTGAAGAGCTAAAGTGTATGCTGG
AAAAATATGAACGTCATTTGGCAGAAATCCAGGGCAACGTCAAGGTTCTC
ACATCTGAGAGAGACAAGACCATTCCTCCTTTATGAGCAGGCACAGGAAG
AAATTGCTCGACTTCGACGAGAAATGATGAAAAGCTGTAAGTCTCCTAAG
TCAACCACAGCACATGCTATTCTTCGTCGGTAGAGACGGAGAGAGATGTA
GCCTTCACTGACCACGAAGAATGACCACAGAGCGAGACAGTCTGAGAGAA
AGGCTCAAGATTGCTCAAGAGACAGCGTTCAACGAGAAGGCTCACTTGGA
ACAGCGGATAGAGGAGCTGGAGTGCACAGTTCACAACCTTGCTGATGAGC
GCATGGAACAGATGGCAAACATGACTTTGATGAAGGAAACCATAACCACT
GTGGAAAAAGAAATGAAATCATTAGCAAGAAAGGCAATGGACACCGAGAG
TGAGCTTGGCAGACAAAAAGCAGAGAATAATTCTTTGAGACTTTTATATG
AAAACACAGAAAAAGATCTTTCTGATACTCAGCGACATCTTGCTAAGAAA
AAATATGAGCTACAGCTTACTCAGGAGAAAATTATGTGCTTGGATGAAAA
AATTGATAATTTCACGAGGCAAAATATTGCACAGCGAGAAGAAATCAGCA
TTCTTGGTGCAACCCTCAATGACCTGGCTAAAGAAAAGGAATGCCTGCAA
GCGTGTCTGGATAAAAAGTCTGAGAACATTGCATCCCTTGGCGAGAGCTT
GGCAATGAAAGAAAAAACCATTTCAGGCATGAAGAATATCATTGCTGAGA
TGGAACAGGCATCAAGACAGTCTACTGAAGCCCTCATTATGTGCGAACAA
GATATTTCCAGGATGCGCCGGCAGTTGGACGAGACAAATGATGAGCTGGG
TCAGATTGCCAGGGAGAGAGATATCTTGGCTCATGAGAATGACAATCTTC

AAGAACAGTTTGCCAAAGTCAAACAAGAAAACCAGGCACTGTCCAAAAAA
CTGAATGATACTCATAACGAACTCAGTGACATAAAGCAGAAGGTCCAGGA
CACGAATCTGGAGGTTAACAAGTTGAAGAACATATTAAAGTCTGAGGAAT
CTGAGAACCGGCAAATAATGGAACAACTCCGAAAAGCCAATGAAGATGCT
GAAAACTGGGAAAATAAGGCCCGCCAGACAGAGGCAGAAAACAACACCCT
CAAACTGGAACTTATCACTGCTGAAGCAGAGGGCAACAGATTAAAGGAAA
AAGTTGATGCCCTTAACAGAGAGGTTGAGCAACACCTAAACGCAGAGCGC
TCTTACAAATCCCAGATTGCAACTTTACACAAGTCTCTTGTGAAGATGGA
GGAGGAGCTTCAGAAGGTTCAGTTTGAGAAGGTGTtctgCTCTCGCAGAT
TTGTCTTCCACAAGGGAACTCTGCATAAAACTCGACTCAAGCAAAGAACT
TCTTAATCGACAGCTGGTTGCCAAAGATCAGGAAATAGAAATGATGGAGA
ATGAGCTGGACTCGAGCGCGCTCTGAAATTGAACTGCTCCGGAGTCAGAT
GACAAACGAGAGGATCTCCATGCAGAATCTCGAGGCTCTGCTGGTGGCCA
ACCGGGACAAAGAGTACCAGTCTCAGATAGCACTGCAGGAAAAGGAGTCT
GAGATCCAGTTGCTGAAGGAGCACCTCTGCCTGGCTGAGAACAAAATGGC
CATCCAGAGCAGAGACGTGGCACAGTTCAGAAATGTTGTAACGCAGTTAG
AAGCAGATTTGGACATTACCAAAAGGCAACTCGGGACAGAACGTTTTGAA MTC10-G4
AGGGAAAGGGCTGTCCAAGAACTTCGCCGCCAGAATTACTCAAGCAATGC
CTATAATTTGGGTCCAATGAAGCCAAATACAAAATGCCACTCACCAGAGC
GTGCTCACCATCGATCTCCTGACCGAGGCCTCGACCGATCATTGGAAGAG
TAAGTGTGCACAGGACCTGGAGTGTAGCCATGTTCACAGTGAGATTTGAG
GAAGCGAGGAAACTGATACTAAACCTGGTTAATAAGACATGATGACACCA
AAAATAAAGGGTTTTCATCTATAAAAAAAAAA

3' end

A.5 comparisson of following sequences:

MAP Multiple Sequence Alignment Results

```
          1            15 16           30 31           45 46           60 61           75 76           90
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   3 Mouse  ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   4 Rat    ---------------- ---------------- ---------------- ---------------- ----------AGG CGGAGGTGGAAGGAG     18
   5 Human  CAGGTATCTAACCTA TCTTTGTATTTTTGG AATTTAATATAGTTC CTGACATAATCATTG TTCAATGGGTCTAAT TTTTTTTTCCTTTAA     90

          91          105 106          120 121          135 136          150 151          165 166          180
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    ---------------- ---------------- ---CTTTTCTTCATC CATTGATGGTTTGAT AGTGGGCTGGGAAAG AAGGCTGTGGTCCTC     57
   3 Mouse  ----------GGTT AACACCAAAGAACAT CAGTTTGGTCTTGGG GCTCCAGTTTCTGAA GCTGGAAGAAACCAG AATACTTTCCAACTA     79
   4 Rat    AGCCCCGCTCAGGTT ACCACCAAAGAACAT CAGTTTGGTCTCGGG GCTCCAGTTTCTGAA GCTGGAAGAAACCAG AATACTTTCCAACTA    108
   5 Human  AGGCATTTGGATACC TTTGTGATAGAAAAG GCCCTTTTCTTCATC CATTGATGGTTTGAT AGTGGGCTGGGAAGG AAAGCTGTGTTCCTC    180

          181          195 196          210 211          225 226          240 241          255 256          270
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    CACATTAGTCAGCAA ATACTTGATTGATTT GGTATTTAGCTTAGT ----CAAATTGAAGA TCTCAAGCAGACAAA TCATGTCTTGGAAGA    143
   3 Mouse  GAACAAGAAGTGAGA ACCCAAGATAGATTC ATCTCTACACTGAAA TTTACAGATTGAAGA TCTCAAACAGACAAA TCATGACTTGGAAGA    169
   4 Rat    GAACAAGAAGGGAGA ACCCAAGATAGAGTC ATCTCTACACTGAAA TT-ACAGATTGAAGA TCTCAAACAGACAAA TCATGACCTGGAAGA    197
   5 Human  CACATTAGGCAGCAA ATACTTGATTGATTT GATATTTAGCTTAGT ----CAAATTGAAGA TCTCAAACAGACAAA TCATGGCTTGGAAGA    266

          271          285 286          300 301          315 316          330 331          345 346          360
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    ATATGTTAGGAAACT CTTGGATAGTAAGGA AGTGGTAAGCAGTCA AGTAAATGATTTAAC CAGCTACAATGAGCA TCTTTGTCAAGAATT    233
   3 Mouse  ATATGTTAGGAAACT CTTGGATAGTAAAGA GGCGGTAAGCACTCA AGTAGATGACTTAGC CAACCACAATGAGCA CCTTTGTAAAGAGTT    259
   4 Rat    ATGTGTTAGGAAACT CTTGGATAGTAAAGA GGTGGTAAGCACTCA AGTAGATGATCTAAC CAACCACAATGAGCA TCTTTGTAAAGAGTT    287
   5 Human  ATGTGTTAGGAAACT CTTGGATAGTAAGGA GGTGGTAAGCAGTCA AGTAGATGATTTAAC CAGCCACAATGAGCA TCTTTGTAAAGAATT    356

          361          375 376          390 391          405 406          420 421          435 436          450
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    GATTAAAATTGACCA ACTAGCAGAGCAACT AGAAAAAGAGAAAAA ATTTGTGGTGGATAC TGCTGACAAGGAACT TGAAGAAGCAAAGAT    323
   3 Mouse  GATTAAACTTGACCA ACATGCAGAGAAATT ACAAAAAGAAAAAAA TTT-GTGGTGGACAC TGCCGACAAGGAACT TGAAGAAGCAAAGAT    348
   4 Rat    GCTTAAACTTGACCA ACTAGCAGAGCAATT GCAAAAAGAAAAGAA TTTTGTGGTGGACAC TGCCGACAAGGAACT TGAAGAAGCAAAGAT    377
   5 Human  GATTAAAATTGACCA ACTAGCAGAGCAACT CGAAAAAGAGAAAAA TTTTGTGGTGGATTC CGCCAACAAGGAACT TGAAGAAGCCAAGAT    446

          451          465 466          480 481          495 496          510 511          525 526          540
   1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------      0
   2 Pig    TGAAGTCATTTGCCA GCAAAATAATATAAT AGTATTGGAAGATAC AATAAAAAGGCTTAG ATCTATAATTTTAGA GACTGAAAAAGCACA    413
   3 Mouse  TGAACTCATTTGCCA GCAAAATAATATAAC AGTATTAGAAGATAC AATCCAAAGGCTTAA GTCCATAATTTTAGA GACTGAAAAGGCACA    438
   4 Rat    TGAACTCATTTGCCA GCAAAATAACATAAC AGTATTAGAAGATAC AATTAAAAAGCTTAA GTCAATAATTTTAGA GACTGAAAAAGTACA    467
   5 Human  TGATCTCATTTGCCA GCAAAATAATATTAT AGTATTGGAAGATAC AATAAAAAGGCTTAA ATCTATAATTTTAGA TACTGAAAAAGCACA    536
```

```
          541         555 556         570 571         585 586         600 601         615 616         630
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------     0
2 Pig     AAATAAATCTCCTTC CAGACTTGATTCCTT TGTCAAGACTTTAGA GGCAGACAGAGATTA CTAC----------- ---------------   477
3 Mouse   AAATACATCTCCTTC TAGACTGGATTCCTT TGTCAAGACTTTAGA AGCAGACAGAGATTA TTATAAGACTGAAGC TCAGAATTTAAGAAA   528
4 Rat     AAATACATCTCCTTC TAGACTTGATTCCTT TGTCAAGACTTTAGA AGCAGACAGAGATTA TTATAAGGCTGAAGC TCAGAATTTGAGAAA   557
5 Human   AAATAAATCTCCTTC TAGACTTGATTCCTT TGTCAAGACTTTGGA AGCAGACAAAGATCA CTATAAGAGTGAAGC TCAACATTTGAGAAA   626

          631         645 646         660 661         675 676         690 691         705 706         720
1 Bovine  --------------- --------------- --------------- ---TCGGGGTACAAA CTGTGATGTAGAGCT TTTGAAGACAAC---    39
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AATGATGAGAAATAG ATCTAAGAGTCCGAG ACGCCCATCGCCAAC TTCCCGGGCTGCAAA CTGTGCTGTAGAGCT TTTGAAGTCAACTGC   618
4 Rat     AATGATGAGAAATAG ATCTAAGAGTCCAAG ACGCCCATCACCAAC TTCTCGGGCCGCCAA CTGTGATGTAGACCT TCTGAAGTCAACTGC   647
5 Human   GATGATGCGAAGTAG GTCTAAAAGTCCAAG ACGCCCATCACCAAC TGCCCGGGGTGCAAA CTGTGATGTAGAACT TTTGAAGACAACAAC   716

          721         735 736         750 751         765 766         780 781         795 796         810
1 Bovine  AAGAGACCGTGAAGA ACTCAAATGCATGCT GGAAAAATATGAGCG TCATTTGGCAGAAAT TCAGGGTAATGTCAA GGTTCTCACGTCTGA   129
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AAGAGACCGTGAAGA GCTAAAGTGTATGCT GGAAAAATATGAACG TCATTTGGCAGAAAT CCAGGGCAACGTCAA GGTTCTCACATCTGA   708
4 Rat     AAGAGACCGTGAAGA GCTCAAGTGTATGCT GGAAAAATATGAACG CCATTTGGCAGAAAT CCAGGGTAATGTCAA GGTTCTCACATCTGT   737
5 Human   AAGAGATCGTGAAGA ACTTAAATGCATGCT GGAAAAATATGAGCG CCATTTGGCAGAAAT TCAGGGTAATGTCAA GGTTCTTAAATCTGA   806

          811         825 826         840 841         855 856         870 871         885 886         900
1 Bovine  GAGAGACAAGACC-T TTCTTCTGTATGAGC AGGCACAGGAAGAAA TTGCCCGACTTCGAC GAGAAATGATGAAAA GCTGCAAGTCTCCTA   218
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   GAGAGACAAGACCAT TCCTCCTTTATGAGC AGGCACAGGAAGAAA TTGCTCGACTTCGAC GAGAAATGATGAAAA GCTGTAAGTCTCCTA   798
4 Rat     GAGAGACAAGACC-T TCCTCCTTTATGAGC AGGCACAGGAAGAAA TTGCTCGACTTCGAC GAGAAATGATGAAAA GCTGCCAGTCTCCTA   826
5 Human   GAGAGACAAGATC-T TCCTTCTTTATGAAC AGGCACAGGAAGAAA TTACCCGACTTCGAC GAGAAATGATGAAAA GCTGTAAGAGTCCTA   895

          901         915 916         930 931         945 946         960 961         975 976         990
1 Bovine  AATCAACAACAGCAC ATGCTATCCTCCGGC GGGTGGAGACTGAGA GAGATGTGGCCTTCA CTGACTTGCGAAGAA TGACCACAGAACGAG   308
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AGTCAACCACAGCAC ATGCTATTCTTCGTC G-GTAGAGACGGAGA GAGATGTAGCCTTCA CTGACC-ACGAAGAA TGACCACAGAGCGAG   886
4 Rat     AATCAACTACAGCAC ATGCTATTCTTCGCC GCGTCGAGACGGAAA GAGATGTCGCCTTCA CTGATTTACGAAGAA TGACCACCGCGCGAG   916
5 Human   AATCAACAACGGCAC ATGCTATTCTCCGGC GAGTGGAGACTGAAA GAGATGTAGCCTTTA CTGATTTACGAAGAA TGACCACAGAACGAG   985

          991        1005 1006        1020 1021        1035 1036        1050 1051        1065 1066        1080
1 Bovine  ATAGTCTGAGGGAGA GGCTAAAGATTGCCC AAGAGACAGCGTTTA ATGAGAAGGCTCACT TGGAACAAAGGATAG AGGAGCTGGAGTGTA   398
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   ACAGTCTGAGAGAAA GGCTCAAGATTGCTC AAGAGACAGCGTTCA ACGAGAAGGCTCACT TGGAACAGCGGATAG AGGAGCTGGAGTGCA   976
4 Rat     ATAGTCTGAGGGAAA GGCTCAAGATTGCTC AAG------CATTCA ACGAGAAGGCTCACT TGGAACAGCGGATAG AGGAGCTGGAGTGCA  1000
5 Human   ATAGTCTAAGGGAGA GGCTAAAGATTGCTC AAGAGACAGCATTTA ATGAGAAGGCTCACC TGGAACAAAGGATAG AGGAGCTGGAGTGTA  1075

          1081       1095 1096       1110 1111       1125 1126       1140 1141       1155 1156       1170
1 Bovine  CAGTTCATAACGTA- --------------- AAGAAAAGTCACTTT TGCACTTTGGGCTAT AAACTATCAATGGCT CCCCACTTGCCCCTT   472
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   CAGTTCACAACCTTG CTGATGAGCGCATGG AACAGATGGCAAACA TGACTTTGATGAAGG AAACCATAACCACTG TGGAAAAAGAAATGA  1066
4 Rat     CAGTTCACAACCTGG ATGATGAGCGCATGG AACAGATGTCAAACA TGACTTTGATGAAGG AAACCATAACCATTG TGGAAAAAGAAATGA  1090
5 Human   CAGTTCATAATCTTG ATGATGAACGTATGG AGCAAATGTCAAATA TGACTTTGATGAAGG AAACCATAAGCACTG TGGAAAAAGAAATGA  1165

          1171       1185 1186       1200 1201       1215 1216       1230 1231       1245 1246       1260
1 Bovine  TTTCTGTGTTCAGAA AAGT--TGGTTTTCT CAGATAGCATTTGCA GTCCTAAGATGTTTA AGGTGGCGTGTATAC TGT------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AATCATTAGCAAGAA AGGCAATGGACACCG AGAGTGAGCTTGGCA GACAAAAAGCAGAGA ATAATTCTTTGAGAC TTTTATATGAAAACA  1156
4 Rat     AATCCTTAGCAAGAA AGGCAATGGACACCG AGAGTGAGCTTGGCA GACAGAAAGCAGAGA ATAATTCTTTGAGAC TTTTATATGAAAATA  1180
5 Human   AATCACTAGCAAGAA AGGCAATGGATACCG AAAGTGAACTTGGCA GACAAAAAGCAGAGA ATAATTCTTTGAGAC TTTTGTATGAAAACA  1255

          1261       1275 1276       1290 1291       1305 1306       1320 1321       1335 1336       1350
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   CAGAAAAAGATCTTT CTGATACTCAGCGAC ATCTTGCTAAGAAAA AATATGAGCTACAGC TTACTCAGGAGAAAA TTATGTGCTTGGATG  1246
4 Rat     CAGAAAAAGATCTTT CTGATACTCAGCGAC ATCTTGCTAAGAAAA AATATGAACTACAGC TTACTCAGGAGAAAA TTATGTGCTTGGATG  1270
5 Human   CAGAAAAAGATCTTT CTGATACTCAGCGAC ACCTTGCTAAGAAAA AATATGAGCTACAGC TTACTCAGGAGAAAA TTATGTGCTTGGATG  1345

          1351       1365 1366       1380 1381       1395 1396       1410 1411       1425 1426       1440
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AAAAAATTGATAATT TCACGAGGCAAAATA TTGCACAGCGAGAAG AAATCAGCATTCTTG GTGCAACCCTCAATG ACCTGGCTAAAGAAA  1336
4 Rat     AAAAAATTGATAATT TTACGAGGCAAAATA TTGCACAGCGAGAAG AAATCAGCATCCTTG GTGCAACCCTCAATG ACCTGGCTAAAGAGA  1360
5 Human   AAAAAATTGATAACT TTACAAGGCAAAATA TTGCACAGCGAGAAG AAATCAGCATTCTTG GTGGAACCCTCAATG ATCTGGCTAAAGAAA  1435

          1441       1455 1456       1470 1471       1485 1486       1500 1501       1515 1516       1530
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   AGGAATGCCTGCAAG CGTGTCTGGATAAAA AGTCTGAGAACATTG CATCCCTTGGCGAGA GCTTGGCAATGAAAG AAAAAACCATTTCAG  1426
4 Rat     AGGAGTGCCTGCAAA CATGTTTGGATAAAA AGTCTGAGAACATTG CATCCCTTGGAGAGA GTTTGGCAATGAAAG AAAAGACCATTTCAG  1450
5 Human   AGGAATGCCTGCAAG CATGTTTGGATAAAA AATCTGAGAATATTG CATCCCTTGGAGAGA GTTTGGCAATGAAAG AAAAGACCATTTCAG  1525

          1531       1545 1546       1560 1561       1575 1576       1590 1591       1605 1606       1620
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   GCATGAAGAATATCA TTGCTGAGATGGAAC AGGCATCAAGACAGT CTACTGAAGCCCTCA TTATGTGCGAACAAG ATATTTCCAGGATGC  1516
4 Rat     GCATGAAGAATATCA TTGCTGAGATGGAAC AGGCATCAAGACAGT CTACGGAAGCCCTAA TTATGTGCGAACAAG ATATTTCCAGAATGC  1540
5 Human   GCATGAAGAATATCA TTGCTGAGATGGAAC AGGCATCAAGACAGT GTACTGAGGCCCTAA TTGTGTGTGAACAAG ACGTTTCCAGAATGC  1615

          1621       1635 1636       1650 1651       1665 1666       1680 1681       1695 1696       1710
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
3 Mouse   GCCGGCAGTTGGACG AGACAAATGATGAGC TGGGTCAGATTGCCA GGGAGAGAGATATCT TGGCTCATGAGAATG ACAATCTTCAAGAAC  1606
4 Rat     GCCGGCAGCTGGATG AAACGAATGACGAGC TGGGGCAGATCGCCA GGGAGAGAGATATCT TGGCTCATGAGAATG ACAATCTTCAAGAAC  1630
5 Human   GTCGGCAATTGGATG AGACAAATGATGAGC TGGCCCAGATCGCCA GGGAAAGAGATATCT TGGCTCATGACAATG ACAATCTCCAGGAAC  1705

          1711       1725 1726       1740 1741       1755 1756       1770 1771       1785 1786       1800
1 Bovine  --------------- --------------- --------------- --------------- --------------- ---------------   548
2 Pig     --------------- --------------- --------------- --------------- --------------- ---------------   477
```

```
3 Mouse  AGTTTGCCAAAGTCA AACAAGAAAACCAGG CACTGTCCAAAAAAC TGAATGATACTCATA ACGAACTCAGTGACA TAAAGCAGAAGGTCC  1696
4 Rat    AGTTTGCCAAAGTCA AACAAGAAAACCAGG CGCTGTCCAAAAAAC TGAATGACACTCATA ATGAGCTTAGCGACA TAAAGCAGAAGGTCC  1720
5 Human  AGTTTGCTAAAGCTA AACAAGAAAACCAGG CACTGTCCAAAAAAT TGAATGACACTCATA ATGAACTTAATGACA TAAAACAGAAGGTTC  1795

         1801       1815 1816       1830 1831       1845 1846       1860 1861       1875 1876       1890
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  AGGAC----------- ---------------- ---------------- ----------ACGA ATCTGGAGGTTAACA AGTTGAAGAACATAT  1735
4 Rat    AGGACACGAACCTGG AGGTTAACAAGCTGA AGAACATATTAAAGT CTGAGGAATCTACGA ACCTGGAGGTTAACA AGCTGAAGAACATAT  1810
5 Human  AAGA----------- ---------------- ---------------- ----------TACTA ATTTGGAGGTTAACA AGCTGAAGAATATAT  1834

         1891       1905 1906       1920 1921       1935 1936       1950 1951       1965 1966       1980
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  TAAAGTCTGAGGAAT CTGAGAACCGGCAAA TAATGGAACAACTCC GAAAAGCCAATGAAG ATGCTGAAAACTGGG AAAATAAGGCCCGCC  1825
4 Rat    TAAAGTCTGAGGAAT CTGAGAACCGGCAAA TAATGGAACAACTCC GAAAAGCCAACGAAG ATGCTGAAAACTGGG AAAATAAGGCCCGCC  1900
5 Human  TAAAGTCTGAAGAAT CTGAGAACCGGCAAA TGATGGAACAACTTC GAAAAGCCAATGAAG ATGCTGAAAACTGGG AAAATAAAGCCCGTC  1924



         1981       1995 1996       2010 2011       2025 2026       2040 2041       2055 2056       2070
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  AGACAGAGGCAGAAA ACAACACCCTCAAAC TGGAACTTATCACTG CTGAAGCAGAGGGCA ACAGATTAAAGGAAA AAGTTGATGCCCTTA  1915
4 Rat    AACTAGAGGCAGAAA ACAACACACTCAAAC TGGAGCTTATCACCG CGGAAGCCGAGGGCA ACAGACTGAAGGAAA AAGTCGATGCCCTTA  1990
5 Human  AATCAGAGGCAGATA ACAATACCCTCAAAC TGGAACTTATCACTG CTGAGGCAGAGGGTA ACAGATTAAAAGAAA AAGTAGATTCCCTCA  2014

         2071       2085 2086       2100 2101       2115 2116       2130 2131       2145 2146       2160
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  ACAGAGAGGTTGAGC AACACCTAAACGCAG AGCGCTCTTACAAAT CCCAGATTGCAACTT TACACAAGTCTCTTG TGAAGATGGAGGAGG  2005
4 Rat    ACAGAGAAGTTGAGC AACACTTAAACGCAG AGCGCTCTTACAAAT CCCAGATTGCAACTT TACACAAGTCTCTTG TGAAGATGGAGGAGG  2080
5 Human  ACAGAGAGGTTGAGC AACACTTAAATGCAG AAAGGTCTTACAAGT CCCAGATTTCTACCT TACATAAATCTGTTG TAAAAATGGAAGAGG  2104

         2161       2175 2176       2190 2191       2205 2206       2220 2221       2235 2236       2250
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  AGCTTCAGAAGGTTC AGTTTGAGAAGGTGT TCTGCTCTCGCAGAT TTGTCTTCCACAAGG GAACTCTGCATAAAA CTCGACTCAAGCAAA  2095
4 Rat    AGCTTCAGAAGGTTC AGTTTGAAAAGGTGT -CTGCGCTCGCAGAT TTGTCTTCCACAAGG GAACTCTGCATTAAA CTCGACTCAAGCAAA  2169
5 Human  AGCTTCAGAAGGTTC AGTTTGAAAAAGTGT -CCGCTCTTGCAGAT TTGTCTTCTACTAGG GAACTCTGTATTAAA CTTGACTCAAGCAAA  2193

         2251       2265 2266       2280 2281       2295 2296       2310 2311       2325 2326       2340
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  GAACTTCTTAATCGA CAGCTGGTTGCCAAA GATCAGGAAATAGAA ATGATGGAGAATGAG CTGGACTCGAGCGCG CTCTGAAATTGAACT  2185
4 Rat    GAACTTCTTAACCGA CAGCTGGTTGCCAAA GATCAGGAAATAGAG ATGATGGAGAACGAG CTGGATTC-AGCACG CTCTGAAATAGAACT  2258
5 Human  GAACTTCTTAATCGA CAGCTGGTTGCTAAA GATCAAGAAATAGAA ATGAGGGAGAATGAG TTAGATTC-TGCTCA TTCTGAAATTGAACT  2282

         2341       2355 2356       2370 2371       2385 2386       2400 2401       2415 2416       2430
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  GCTCCGGAGTCAGAT GACAAACGAGAGGAT CTCCATGCAGAATCT CGAGGCTCTGCTGGT GGCCAACCGGGACAA AGAGTACCAGTCTCA  2275
4 Rat    GCTCCGGAGTCAGAT GACAAACGAGAGGAT CTCCATGCAGAATCT TGAGGCCCTGCTGGT GGCCAACCGAGACAA AGAGTACCAGTCTCA  2348
5 Human  CCTGAGGAGTCAGAT GGCAAATGAGAGAAT CTCCATGCAGAATCT AGAAGCTTTGCTGGT GGCCAATCGAGACAA AGAATATCAGTCTCA  2372

         2431       2445 2446       2460 2461       2475 2476       2490 2491       2505 2506       2520
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  GATAGCACTGCAGGA AAAGGAGTCTGAGAT CCAGTTGCTGAAGGA GCACCTCTGCCTGGC TGAGAACAAAATGGC CATCCAGAGCAGAGA  2365
4 Rat    GATAGCTCTGCAGGA GAAGGAGTCTGAGAT CCAGCTGCTGAAGGA GCACCTCTGCCTGGC CGAGAACAAAATGGC CATCCAGAGTAGAGA  2438
5 Human  GATAGCACTTCAAGA AAAAGAATCTGAAAT TCAGCTTCTTAAAGA ACACCTTTGTTTGGC AGAAAATAAAATGGC CATCCAGAGTAGAGA  2462

         2521       2535 2536       2550 2551       2565 2566       2580 2581       2595 2596       2610
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  CGTGGCACAGTTCAG AAATGTTGTAACGCA GTTAGAAGCAGATTT GGACATTACCAAAAG GCAACTCGGGACAGA ACGTTTTGAAAGGGA  2455
4 Rat    TGTGGCACAGTTCAG AAACGTTGTAACACA GTTAGAAGCAGATTT GGACATTACAAAAAG ACAACTAGGGACAGA ACGATTTGAAAGGGA  2528
5 Human  TGTGGCCCAGTTCAG AAATGTTGTCACACA ATTGGAAGCTGATTT AGACATTACCAAAAG ACAACTAGGAACAGA GCGCTTTGAAAGGGA  2552

         2611       2625 2626       2640 2641       2655 2656       2670 2671       2685 2686       2700
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  AAGGGCTGTCCAAGA ACTTCGCCGCCAGAA TTACTCAAGCAATGC CTATAATTTGGGTCC ---AATGAAGCCAAA TACAAAATGCCACTC  2542
4 Rat    AAGGGCTGTCCAAGA ACTTCGCCGCCAGAA TTACTCAAGCAATGC TTATCATTTGGGTTC ---AATGAAGCCAAA TACAAAGTGTCACTC  2615
5 Human  GAGGGCCGTACAAGA ACTTCGCCGCCAAAA TTATTCAAGTAATGC TTATCATATGAGTTC TACAATGAAGCCAAA TACAAAATGTCATTC  2642

         2701       2715 2716       2730 2731       2745 2746       2760 2761       2775 2776       2790
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  ACCAGAGCGTGCTCA CCATCGATCTCCTGA CCGAGGCCTCGACCG ATCATTGGAAGAGTA AGTGTGCACAGGACC TG---GAGTGTAGCC  2629
4 Rat    ACCAGAGCGAGCTCA CCATCGATCTCCTGA CCGAGACCTCGACCG ATCATTGGAAGAGAA TCTTTGCTATAGAGA TTTCTGACATGTGAA  2705
5 Human  ACCAGAACGTGCTCA CCATCGATCTCCTGA CCGAGGCCTAGATCG ATCATTAGAAGAGAA TCTTTGCTACAGAGA TTTCTGACACCTGAA  2732

         2791       2805 2806       2820 2821       2835 2836       2850 2851       2865 2866       2880
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
3 Mouse  ATGTTCACAGTGAGA TTTGAGGAAGCGAGG AAACTGATACTAAAC CTGGTTAATAAGACA TGATGACACCAAAAA TAAAGGGTTTTCATC  2719
4 Rat    AAGATTCTCCACAGC CTTGGGAA-GGTCAA AGATACAAATGGATT TTTT--------GCT ATGTGATTGCATTTA TCTTTTGAATGCTTG  2786
5 Human  ATGATTCTTCACATC CCTGAGAAAGGTCAA AGTTACAAACTGATT TTTTTTTTTTTTGCT ACATGAGTGCATTTA TCTTTTAAATGCTTG  2822

         2881       2895 2896       2910 2911       2925 2926       2940 2941       2955 2956       2970
1 Bovine ---------------- ---------------- ---------------- ---------------- ---------------- ----------   548
2 Pig    ---------------- ---------------- ---------------- ---------------- ---------------- ----------   477
```

```
3 Mouse  TATAAAAAAAAAA-- ---------- ---------- ---------- ---------- ----------  2732
4 Rat    ACAGGGTTAAATGTA TTTATTAACTTTGTG CCTCTGAGTCT--GT TCTCCAGTGCCCTAC CACAGAAATCTGGGA TGATTGATGTATGCA  2874
5 Human  GCAATGTTAAATGTA TTTATTAACTTTGTG TCTCTGAATCTCTGT TCTAATGTGCCATGT TGCAGTGATCTGAGA TGACTTATAAAAACA  2912


         2971       2985 2986       3000 3001       3015 3016       3030 3031       3045 3046       3060
1 Bovine ---------- ---------- ---------- ---------- ---------- ----------  548
2 Pig    ---------- ---------- ---------- ---------- ---------- ----------  477
3 Mouse  ---------- ---------- ---------- ---------- ---------- ----------  2732
4 Rat    AACGAAATGCATATG GTTCTTTCTATCCAC ACAGTGATAGTGAGT GTAAAATCTGCTTAC TTCACTACTGAGCAC TGTTTTGTTAACTGT  2964
5 Human  A---AAATGTATATG GCTCTTTCTATCCAT GCAATGATAGTGAGT GTAAAATCTGCTTAC TTCACTATTGAACAC TATTCTGTTCACTAT  2999

         3061       3075 3076       3090 3091       3105 3106       3120 3121       3135 3136       3150
1 Bovine ---------- ---------- ---------- ---------- ------   548
2 Pig    ---------- ---------- ---------- ---------- ------   477
3 Mouse  ---------- ---------- ---------- ---------- ------  2732
4 Rat    CCGGAGTAAATAAAG AAGCTTAAAAAAGAA AAAAAAAAAAAAAAA AAAAAAAAAAAAAAA AAAAAA  3030
5 Human  CTGGAGTAAATAAAG AAGCTTATTAAAACA GGGAAAAAAAAAAAA ---------- ------  3044
```

A.6 3′end of GNG2 sequence

Exon three

GTTGTTTTTGTCTCCCTTTCCAGGTGTCCAAGGCAGCTGCAGATTTGATG
GCCTACTGTGAAGCACATGCCAAGGAAGACCCCCTCCTGACCCCTGTTCC
GGCTTCAGAAAACCCGTTTAGGGAGAAGAAGTTTTTCTGTGCCATCCTTT
AAGTCTTTGAGAGGGGCCTGAAGAGCCTCCGGGCTCCTGGGACATTGATG
TAGAGTTTTTAGTGAAGTGGGCACCTTTCTAGTCCACGGCATTTGAAGAG
AGCGAGGAGAACCATTCTGGAAACTCTAGGCTATGCATGTTTAAAGATCT
GGTCCCCTTTATGAGAATGCAAGCCGATCCACATCCTGACTTAAGAGATC
TGATTCTGACGAACTGCCTGGAGGAGGGGAATATATAAAAATAAAATTGG
TGTCACTTCTTTTCTGCTATCCCCCAGCCCCCCCCCCAAAATCCTCATGT
TTCTGCTTCATATTTTGAAAAATAACAATTAAAACAGACAGCTGTACTGA
GGTAAGATATGTGTGACCTTCTTGGAATGAATATTGTCTTTAGAATACCC
TTTGATAAGCTGAGCTGTCCCGTGTAGATGCAATTCGGTTTAATGGCATT
GATGTATAGTCACTGTGCCTTTCTTTTTCTTTCTTCCTTCTCCTCTACCC
CTCCTTCCACCCCTCCCCATTAGAGTAGTGTGGAGATAAGGCTGGACTGG
TCTATCAGATTGAACTCCAAGAATGATCACACAAAATGTTTAGGGAGATG
TTCCCCGTGGTGTATCCTCATGGTAACAACGACAAAAAATGCCGGTTGTC
TGTGTTCTCTTTTCACTATTCCTAACATGTGTACATGATAGCTTTGATTC ← The 3′end
TGCAAGTAAAAGTAAATCCTGTGTTGTGACTGGTGCTTTCATATATTTGT ← of GNG2
GACAATTTTTGAGTAATATTGCATGAAAATGTCCCTATGTTACATCCATT
CAGAAGTTTTGTTGTTTTACTCTAAAGCTGGGAAAGGAAATGAGAGGGAA
AAGACCCCGGAGAGGGAAGAAAATCTCAGTATTTTGAAAATTGAGATTAC
TTCAGAGCCTTAGCCACACCTAAAATACCTCCTAGTTAATAGTGGTATAA
ATGCCTCTTCAATACGTTTTCCAGAATCCAAAGCATTTTGGTTTATCCAG
GACCCAGGGCAGCACAGCTGTCACCAAGCAGGAGAGTTAAGGATTCACCA
TGAGCTGGGAAATGCTTTTGCCATGAGTATGAGCAAATTCCCTCTTTCCC
TGAATCATGGACATTCTAGATTAAAAGAACATTTTTTTGTGCTCTTAACA
AGAAAACCATGGCCCTCCTTTGTTCAAGTATCAGAAGAAATAAACCCACA
GCTCCAGAGAAGGTGACCATTCTCAGAACTCCAGCTATTCACTCTCCAGG
GAGAAGGACCTCAAATCGCCACTCTTTGGGCGGCAGTGCGGTCCCCACGG
CCGGCTCTACGAGGAAGAGTTCTGGCTCTTTTGTCCACTGAGATGGTCTT
GGTTTTTCACTTAACAAATTTTTTAATGGAATCTTTGTTTTTGTTCTCCA
TCTTGTTTGTTAGAGTCTCTCGGCCTTTATTTACAAATTCCTTGCAACTA
GAGCGCTCCTTCCCCAAGATATGGTAGTGAGAGTAATTTTTCATTGTAGC
TGTAGTCTCCATCAGTAACAGCAGGCCCTGGAAGACTTGATCACCTTTTT
CTGTGTCATTTTCAGTCAAAGAGGGCCTCTCTTACATCTTGTTTGCTTTC
AAATCCCCAAATATCATCTCCATCTCCCAATTAATTTTATTGTCTCTTTC
CTCTCTATTCGCTTTCTCCTGTTTTTTTTTTAAAAAAAAAGCATG **(the 3′end of clone AK001024)** AAAAAGGAAGAGAAAAAGTTTGATA