REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author Benton, B

Investigation of *TSGA10* gene expression, localization and protein interaction in human and mouse spermatogenesis

by

Babak Behnam

A thesis submitted for the Degree of Doctorate of Philosophy in the University of London

May 2005

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

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

Abstract

Isolation of the testis-specific human gene *TSGA10* was reported in 2001 (Modarressi *et al.*). In this project, I have carried on the work by further characterising the gene and its product by cloning and sequencing the mouse homologue (*Tsga10)* (GenBank Accession no. AF530050[1]). *Tsga10* was localised (by FISH) to mouse chromosome 1 band B, and RNA *in situ* hybridization and RT-PCR experiments showed that *Tsga10* transcripts were developmentally regulated and expressed in pachytene cells of the testis. It was also expressed in some other actively dividing cells such as foetal tissues, primary tumors and astrocytes. Interestingly, the program Pfam (which predicts protein structural motifs) suggests that the Tsga10 protein contains a 'myosin like' coiled-coil domain. A Green fluorescent protein (GFP)-Tsga10 fusion gene was constructed and cell transfection resulted in the formation of short thick protein filaments. Antibodies raised individually against the N-terminus and C-terminus parts of the Tsga10 protein were used to localise Tsga10 protein expression by immunohistochemistry. In collaboration with Hossein Modarressi and Frans Van der Hoorn, using immuno-blotting and staining techniques, I discovered that *Tsga10* encodes a 82 kD precursor protein in spermatids which is incorporated into sperm tails. Upon passage of sperm into the epididymis this is processed to a 27 kD protein which is associated with the fibrous sheath, a major tail structure. Using *Tsga10* as bait in a yeast two-hybrid experiment,atestis cDNA library was screened for proteins that interacted with Tsga10. Three candidate interacting proteins were found; *Odf2*, a previously described protein of the fibrous sheath, the rat homologue of *FLJ32880*, (a hypothetical human protein whose gene is expressed predominantly in spermatocytes, spermatids

and mature sperm cells), and the mitochondrial protein *cytochrome c1*. Using immunocytochemistry I then localised FLJ32880 (newly named as TSGA10 interacting protein, TSGA10IP) in the sperm tail as another fibrous sheath protein. These results support the theory that *Tsga10* has a role to play in the fibrous sheath of the sperm tail, and also suggest a possible functional involvement with a mitochondrial protein. Antibody was raised against the mouse homologue of TSGA10IP for further study. I suggest that *Tsga10* plays multiple roles, one in sperm tail formation and possibly a second role in cell cycle regulation.

[1] All of the nucleotide numbers given in this thesis are based on *Tsga10* sequence (AF530050) which I submitted into the GenBank database.

Acknowledgements

I would like to express my gratitude to my supervisor and mentor, **Dr. Jonathan Wolfe** without whose continuous help and supervision I could not have completed this project. Also to **Professor Sue Povey** for her gracious advice and productive discussions in weekly Lab meetings, and to Dr. Kay Taylor for all her kind attention and useful comments, and to Professor Dallas Swallow for her help. I would like to thank Professor Frans Van der Hoorn (University of Calgary, Canada) for providing a rat testis cDNA library and Dr. Mohammad Hossein Modarressi for his support and providing the GFP plasmid and a rat testis blot. I would also like to thank Dr. Aldamaria Puliti (University of Pisa, Italy) for providing the neural crest cDNA, Professor Ming Du for providing slides of germ cell tumour, Dr. Hazel Smith for her helpful comments in RNA In *situ* hybridisation and Professor John Wood for offering the use of the cryosection facility in his lab to me. Thanks also to Sarah Manistre for her help with screening the yeast colonies screening and for some of her figures; to Lynne Vinall, Drs. Rosemary Ekong, Karrine Rousseau and Claire Willoughby for being there ready whenever I needed their help, and to everyone else in the lab (Marina, Ranji and Ian) for their support and interest. Finally I would like to offer my thanks to my sponsor Iran University of Medical Sciences (IUMS) where I graduated in medicine. I am very grateful to them for this award.

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

TABLE OF CONTENTS

Table of figures

Table of tables

Fifth Philosopher's Song

by Aldous Huxley

(Issue 10 · posted June 13, 1997)

A million million spermatozoa,
All of them alive:
Out of their cataclysm but one poor Noah
Dare hope to survive.

And among that billion minus one
Might have chanced to be
Shakespeare, another Newton, a new Donne
-
But the One was Me.

Aldous Leonard Huxley (1894-1963) authored 47 books, many still considered classics today. Grandson of Thomas Henry Huxley, the great biologist who helped develop the theory of evolution, Huxley began his career as a magazine writer and published several books of poetry before his first novel appeared. He is probably most famous for his dystopian Brave New World *(1931), a nightmare future of biologically controlled humanity. His brother Sir Julian Huxley was a staunch advocate of Darwinian evolution and helped bring the field in line with population genetics.*

Abbreviations

aa	amino acids
Acr	acrosin
AKAP	A-kinase anchor protein
AMH	anti-Mullerian hormone
AZF	azoospermia factor
BLAST	basic local alignment search tool
bp	base pair
cDNA	complementry deoxyribonucleic acid
c-Kit	c-kit proto-oncogene
CLASP	*CLIP*-associated protein
CLIP	cytoplasmic linker protein
Dax1	a gene is required for testis formation and located on X chromosome
DAZ	deleted in AZoospermia
DDRT	differential display reverse transcription reaction
DEPC	diethylpyrocarbonate
DFS	dysplasia of fibrous sheath
dH_2O	distilled water (milli-RQ plus)
DMRT1	double-sex and mab-3 related in testis 1
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate

DTT	dithiothreitol
EDTA	Ethamine diaminotetra-acetic acid disodium salt
FITC	fluorescein isothiocyanate
FSH	follicle stimulating hormone
FS	fibrous sheath
GAPDS	glyceraldehydes 3-phosphate dehydrogenase-S
GDNF	glial cell line-derived neurotrophic factor
HCl	hydrochloric acid
HTGS	high-Throughput Genomic Sequences
ISH	in situ Hybridization
kb	kilobase
LC	light chain
LH	luteinizing hormone
LIF	leukocyte migration inhibitory factor
MAPK	mitogen activated protein kinase
ml	millilitre
mM	millimolar
mRNA	messenger RNA
MGF	mast cell growth factor
MIS	mullerian inhibiting substance
MS	mitochondrial sheath
NaCl	sodium chloride
NCBI	national Centre of Biotechnology Information

NRE	negative regulatory element
ODF	outer dense fibre
ORF	open reading frame
PAC	P1 artificial chromosome
PCR	polymerase chain reaction
PGC	primordial germ cell
PhK	Phosphorylase kinase
PI3-K	phosphatidylinositol 3-kinase
PEG 3350	polyethylene glycol 50%
PKA	protein kinase A
PKB	protein kinase B
PKC	protein kinase C
PRM	protamine
RACE	rapid Amplification of cDNA Ends
Rab6IP2	Rab6-Interacting protein 2
RNA	ribonucleic acid
RNAse	ribonuclease
RPM	revolutions per minute
RT-PCR	reverse transcription polymerase chain reaction
SF1	steroidogenic factor 1
SCF	stem cell factor
SOX3	SRY box-related gene 3
SOX9	SRY box-related gene 9

Spag4	sperm-associated antigen 4 protein
SPAM1	sperm adhesion molecule 1
SRY	sex-determining region Y
TESPA	triethoxysilylpropylamine
UTR	untranslated region
VDAC	voltage-dependent anion-selective channel protein
VEGF	vascular endothelial growth factor
Wnt4	wingless-type 4, a gene name
WT1	wilm's Tumour 1 (Tumour suppressor gene)

Chapter I:

Introduction

1.1 Testis and Seminiferous tubules

The testis lies within a serous sac named *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 pyramid contains one to three tightly packed and highly convoluted *seminiferous tubules,* which contain gametes in various stages of development and supportive Sertoli cells. The tubules in each lobule join to form a short, straight tube called the *straight rectus tubule.* From all lobules, these tubules connect to form a collecting network towards the posterior portion of the testis called *retes testis,* which empties into the highly convoluted *efferent ductule.* The *tunica 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 testicular lobules. The testis can be considered as a gland with two closely related functions; the first, carried out by somatic cells, is its function in exocrine and endocrine secretion producing semen and male sex hormones (androgens) in mature testis. The second role is the production of germ cells and their maturation to fertile spermatozoa. Before puberty, the number of the most immature germ cells (spermatogonia) is relatively small and is present in no later stages of maturation. After puberty meiosis commences leading to the production of spermatids which look nothing like the sperm they will develop into over the next 70-odd days. They move towards the Sertoli cells rapidly and bury their heads within them, like ostrich heads in the sand.

85-90% of the testis is composed of seminiferous tubules. Each seminiferous tubule is surrounded by a well-defined cellular basement membrane (figure 1.1).

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.

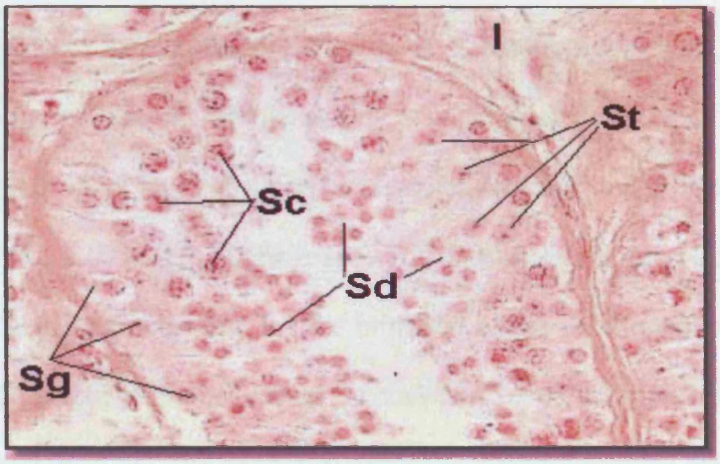


Figure 1.1- Mouse Seminiferous Tubules, Hematoxyline and Eosin (H&E) staining, x40 (**I** = interstitial space, **Sg** = spermatogonia **Sc** = primary spermatocytes, **Sd** = spermatids, **St** = Sertoli cell nuclei)

1.2 Differentiation of the Mammalian testis

The first signs of testicular organisation can be recognised at approximately 42 days after fertilization 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. A Y chromosomal gene, *SRY* (sex determining region on the Y chromosome), acts as a switch signal for testis differentiation which involves several steps controlled by other non-Y-linked genes, such as *WT1, EMX2, LIM1, SF1, SOX9* (figure 1.2). Since other genes, such as *Wnt4* in mouse, and *DAX1*, are necessary for the initiation of the female pathway in sex determination, female development cannot be considered a default process (Sinici *et al.*, 2003). Expression of the *SRY* gene is restricted to gonadal somatic cells at 10.5-12.5 days post coitum (dpc) in the mouse (Sinclair *et al.*, 1990). However, even if normal *SRY* exists, the haploinsufficienty of *SOX9* or *KTS+* splicing form of *WT1* can cause male-to-female sex reversal. Furthermore, duplication of the partial region on the X chromosome including the *DAX1* gene can also cause male-to-female sex reversal. The sex-determining system seems to be sensitive both for gene dosage and level of gene expression. In males, *SF1* has a role during sexual development in regulating expression of Mullerian inhibiting substance (MIS) or anti-Mullerian hormone (AMH) (Behringer *et al.,* 1994; Lahbib *et al.,* 1997). WT1 is a zinc-finger transcription factor, which associates and synergizes with SF1 (Nachtigal *et al.,* 1998). *DAX1* is an X-linked gene, thought only to be activated in the absence of *SRY* expression and is important for female sex development (review, Swain and Lovell-Badge, 1997; Swain *et al.*, 1998). However, although Dax1 (an orphan nuclear

receptor) was originally proposed to act as an 'anti-testis' factor, *Nr0b1* (also called *Dax1* and *Ahch*, which encodes Dax1) has in fact been known to be required for testis differentiation (Meeks *et al.*, 2003). Most of these genes involved in sex determination are transcription factors and are able to switch on or off other genes. DAX1 and WT1 oppose each other to modulate SF1 mediated transactivation of male-specific genes. In addition, there are other genes both X-linked (*SOX3*) and autosomal (*SOX9)* which are involved in sex determination (Jimenez and Burgos, 1998). The detail of these genes interactions is still unknown. In the absence of *SRY* gene expression, the bi-potential gonad differentiates into an ovary over 100 days (weeks 6-20), however, when it is expressed; it develops into a testis (in 70 days). 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 salk 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 precursors of the Sertoli cells are derived from cells of the primitive sex cords. The outer part of the testicular sex cords form 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 be degenerated resulting in development of the Wolffian ducts into the seminal vesicles, epididymis and vas deferens. Differentiation of the male external genitalia at this stage depends on androgenic hormone secreted from the foetal testis (Huhtaniemi, 1994). Hormonal activity of the Leydig cells is finished by about the 18th week and these cells then remain undeveloped and are not reactivated

until puberty when they stimulate spermatogenesis.

All of these genes that play a role in mammalian testis differentiation affect the *DMRT1* gene (mapped in 9p24.3) and its encoded protein as the final target (Boyer *et al.*, 2002). *DMRT1* is demonstrated to be implicated in human testis differentiation, and has a gonad-specific and sexually dimorphic expression profile during embryogenesis in mammals, birds, and a reptile (*Alligator mississippiensis*) (Smith *et al.*, 1999). High *DMRT1* expression is necessary for testicular differentiation, whereas lower expression is compatible with ovarian differentiation. Since *DMRT1* represents an ancient, conserved component of the vertebrate sex-determining pathway, it seems that it is conserved during evolution. It has been shown that haploinsufficiency of this dosage-sensitive gene is responsible for the failure of testicular development and feminization in XY patients with monosomy for 9p (Smith *et al.*, 1999).

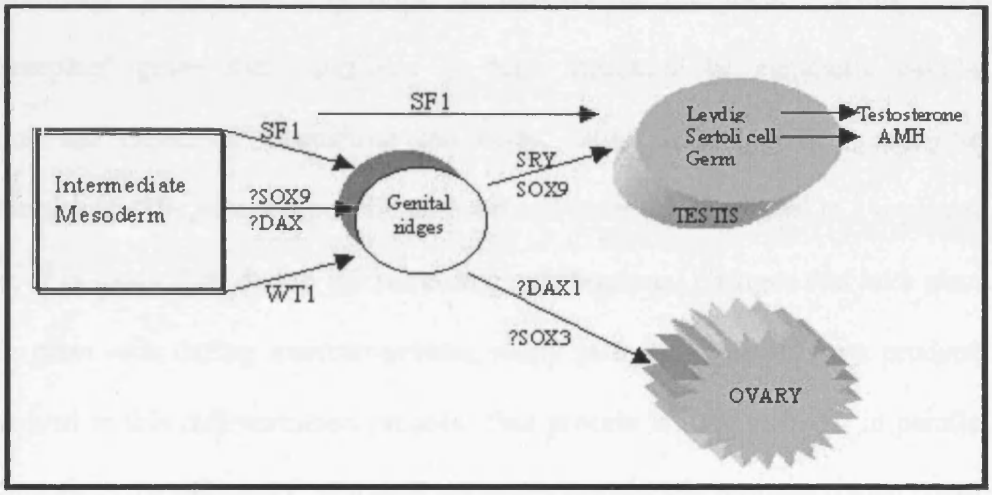


Figure 1.2- Genes and the pathway involved in gonadal development and differentiation. Chromosomal location of candidate genes in the testicular development cascade are: SRY: Yp11.3, DAX1: Xp21-p22, SF1: 9q33, WT1: 11p13, SOX9: 17q24-q25 and AMH: 19p13.3, DMRT1: 9p24.3 (DMRT1 has synergistic effects to SRY and SOX9 secondary to androgenic hormonal stimulation).

1.3 The regulation of germ cell differentiation (spermatogenesis and spermiogenesis), and sperm structure

The pathway of sperm development is an excellent system for studying cellular division and differentiation because it is possible to isolate cells at various stages of differentiation all the way from the stem cells (spermatogonia) to mature sperm cells. The process of spermatogenesis is like a mini-system of development in terms of differentiation that goes on in the life of the cell. It is one of the few systems that allow us to examine human differentiation. Cell differentiation begins with totipotent stem cells, in a cycle of continual self-renewal, that eventually become committed to a very specific pathway, finally leading to mature spermatozoa.

The mitotic division of germ cells is similar to many mitotically active undifferentiated tissues such as those in tumours or the foetus. While some 'housekeeping' genes that contribute to basic structural or metabolic cellular functions are expressed throughout the body, other tissue-specific genes that contribute to specific 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).

The spermatogenic cells are actively replicating cells at various stages of a complex differentiating process (from the division of a spermatogonia to the production of a

completed spermatozoa) called spermatogenesis. The premature germ cells become spermatogonia of the germinal epithelium. There are three sequential major phases in spermatogenesis: growth of the spermatogenic tissue by mitosis, the meiotic divisions and finally morphological changes into a sperm. Spermiogenesis is the process, which morphologically changes a spermatid into a spermatozoon. This transformation includes nuclear condensation, acrosome formation, loss of most of the cytoplasm, elongation, development of the tail and arranging mitochondria into the upper part of the tail to power it. The processes of spermatogenesis and spermiogenesis also require a contribution from the somatic cells (Hecht, 1995). Androgens produced by somatic cells are necessary for the production and successful delivery (via development of normal sexual behaviour and secondary sexual characteristic) of spermatozoa. In an average man, three quarters of sperm cells developed do not reach ejaculation, and only half of those that do are functional. Throughout spermatogenesis, a somatic cell group, the Sertoli cells, nurtures the developing germ cells. The Sertoli cells are responsible for sustaining the germ cells, and providing them with nutrients, as well as secreting hormonal factors and proteins that regulate their development. Sertoli cells also have biochemical interaction and a metabolic influence in relation to germinal cells (review, Griswold, 1998). These cells are also phagocytic (Carr, *et al.*, 1968). Sertoli cells also secret *retes testis* fluid and androgen binding protein under the control of testosterone and/or FSH. They also contain a high concentration of a carbohydrate storage compound, glycogen, which supplies energy to the developing sperm. Tight junctions connect their membranes. These are leak proof junctions that form a barrier between the

circulatory system and the lumen of the seminiferous tubules. This barrier prevents germ cells provoking the immune system and is important because spermatozoa are produced during puberty, long after the period of self-recognition by the immune system. The spermatogenic developmental pathway occurs from the germinal epithelium of the seminiferous tubules. Up to their meiotic stage, the developing germ cells are found on the external side of the Sertoli tight junction barrier (Holstein *et* al. 2003).

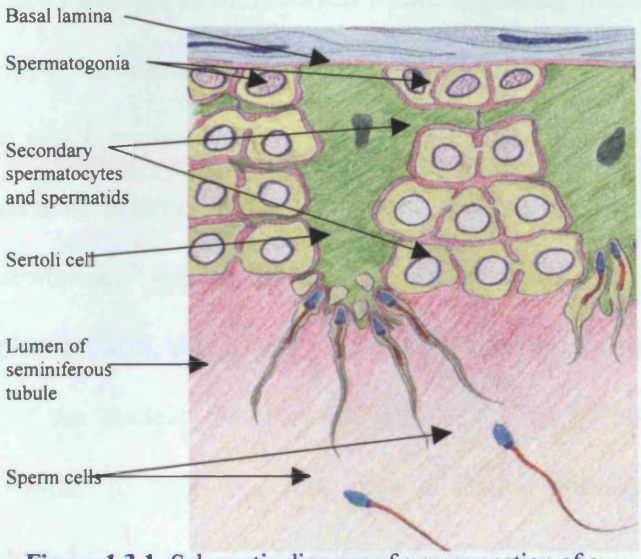


Figure 1.3.1- Schematic diagram of a cross-section of a seminiferous tubule, showing developing sperm cells (by Sarah Manistre)

At puberty, spermatogonia proliferate. Type A_{single}, (A_S) spermatogonia may divide to produce separate daughter cells of their own kind, or $A_{primary}$ (A_{pr}), and then A_{al} spermatogonia. Cytoplasmic cleavage is not complete in the formation of A_{pr} this is the only difference between $A_{secondary}$ (A_s) and A_{pr}. Similarly, when A_{pr} divide, a chain of cells linked by cytoplasmic bridges is formed, A_{al}. A_S, A_{pr} and A_{al} are undifferentiated spermatogonia. Under a light microscope, these cells appear darker

than other spermatogonial cells when stained with Hematoxyline and Eosin (H&E). The type A_S, A_{pr} and A_{al} spermatogonia are moved to specific sites along the seminiferous tubules probably as a result of signalling from other cells of the interstitium. In mice they have been seen to space themselves so that they are well spaced out as they develop in a cyclic manner (Chiarini-Garcia *et.* al. 2001).

Type A_{al} spermatogonia divide to produce type A_1 spermaotognia. These cells may be distinguished from type A_{al}, as they have larger, more spherical nuclei. The RNA binding protein *DAZ* is thought to be involved in the signalling pathway underlying differentiation of type A_{al} into type A_1 (Bianca et. al. 2001). Type A_1 divides to produce type A_2 in which nuclear vacuoles (detectable in A_1 spermatogonia) are not seen. A small amount of heterochromatin (condensed chromatin) is adjacent to the internal nuclear membrane. Type A_2 divides to produce type A_3. In these cells, the nuclei are more spherical still, and a larger amount of heterochromatin can be seen around the inside of the nuclear membrane. Type A_3 divides to produce type A_4. These cells are similar to type A_3, but have a further increased amount of heterochromatin around the inside of the nuclear envelope, and the most spherical nuclei of the A type spermatogonia. As they develop, type A spermatogonia spread laterally along the base of the seminiferous tubule and remaining on the blood-side of the barrier formed by the tight junctions of the Sertoli cell membranes (Chiarini-Garcia *et. al.,* 2001 b) (figure 1.3.2).

Up until this point, the cells are thought to be stem cells, not committed to a particular fate. The dosage of a protein produced by the Sertoli cells, Glial cell line-derived neurotrophic factor (GDNF) determines whether or not the stem cells

differentiate (Meng et al., 2002). In mice, low levels of GDNF promote differentiation causing the mice to be fertile but to have a decreased rate of spermatogonial cell proliferation later in life. At high levels, GDNF causes mice to be infertile and large cell clusters form in the seminiferous tubules. The clustered cells do not express the c-kit marker for differentiating spermatogonia, which suggests that high levels of GDNF prevent differentiation of spermatogonial stem cells and promote self-renewal. This explains why there is a build up of undifferentiated cells even though the proliferation rate is not especially high. The clusters remain until they are broken down during puberty and taken up by Sertoli cells.

At high concentrations, self-renewal or A type accumulation occurs, at low concentrations, the type A_4 spermatogonia divide to form intermediate spermatogonia, a cell type committed to become sperm cells (Meng et. al. 2000). Intermediate spermatogonia are similar in appearance to type A_4, again with a darker nucleus due to the heterochromatin, now covering virtually the whole interior of the nuclear membrane (figure 1.3.2). Intermediate spermatogonia divide to produce type B spermatogonia which have more oval shaped nuclei and which stain more darkly than any of their precursors.

The heterochromatin on the inner membrane of the nuclear envelope is seen under the light microscope to form into clumps attached to the membrane, with less of the surface being covered than in intermediate spermatogonia (Chiarini-Garcia et. al. 2001). The type B spermatogonia in turn divide to form the primary spermatocytes, the cells in which meiosis I (the first meiotic division) occurs. Primary spermatocytes

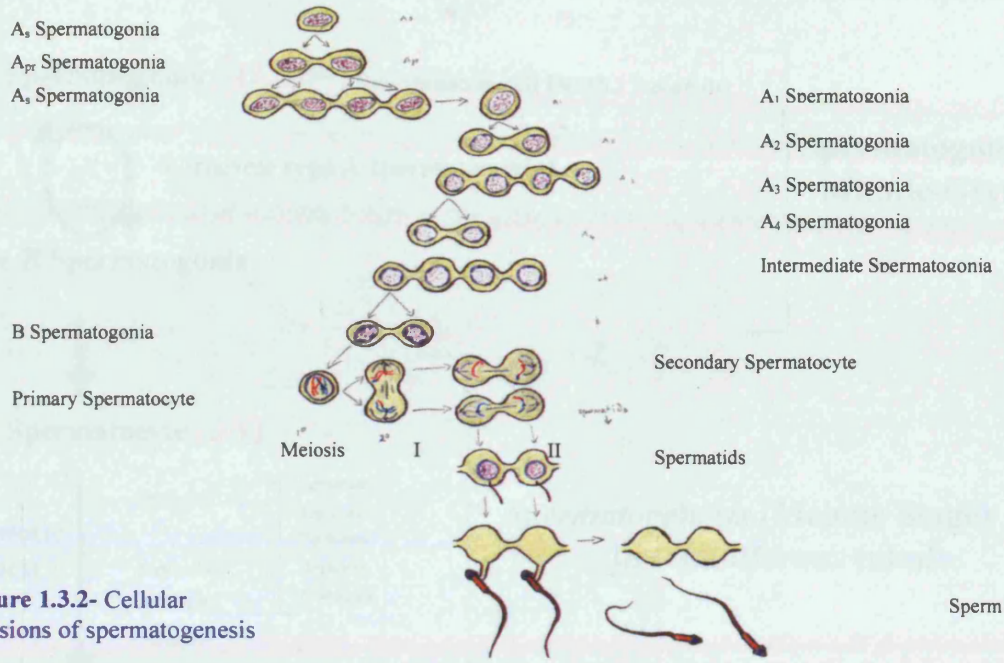


Figure 1.3.2- Cellular divisions of spermatogenesis

are the largest of the germinal epithelial cells. They are also the first cells to move within the Sertoli cell barrier, passing down within recesses in the Sertoli cells (Holstein et. al. 2003) (figure 1.3.2).

The primary spermatocytes go through the first division of meiosis, forming the secondary spermatocytes. It is during this division that chiasmata form, leading to genetic recombination. The secondary spermatocytes then complete meiosis. The products of the second meiotic division are the spermatids, which go through the process of spermiogenesis (figure 1.3.3).

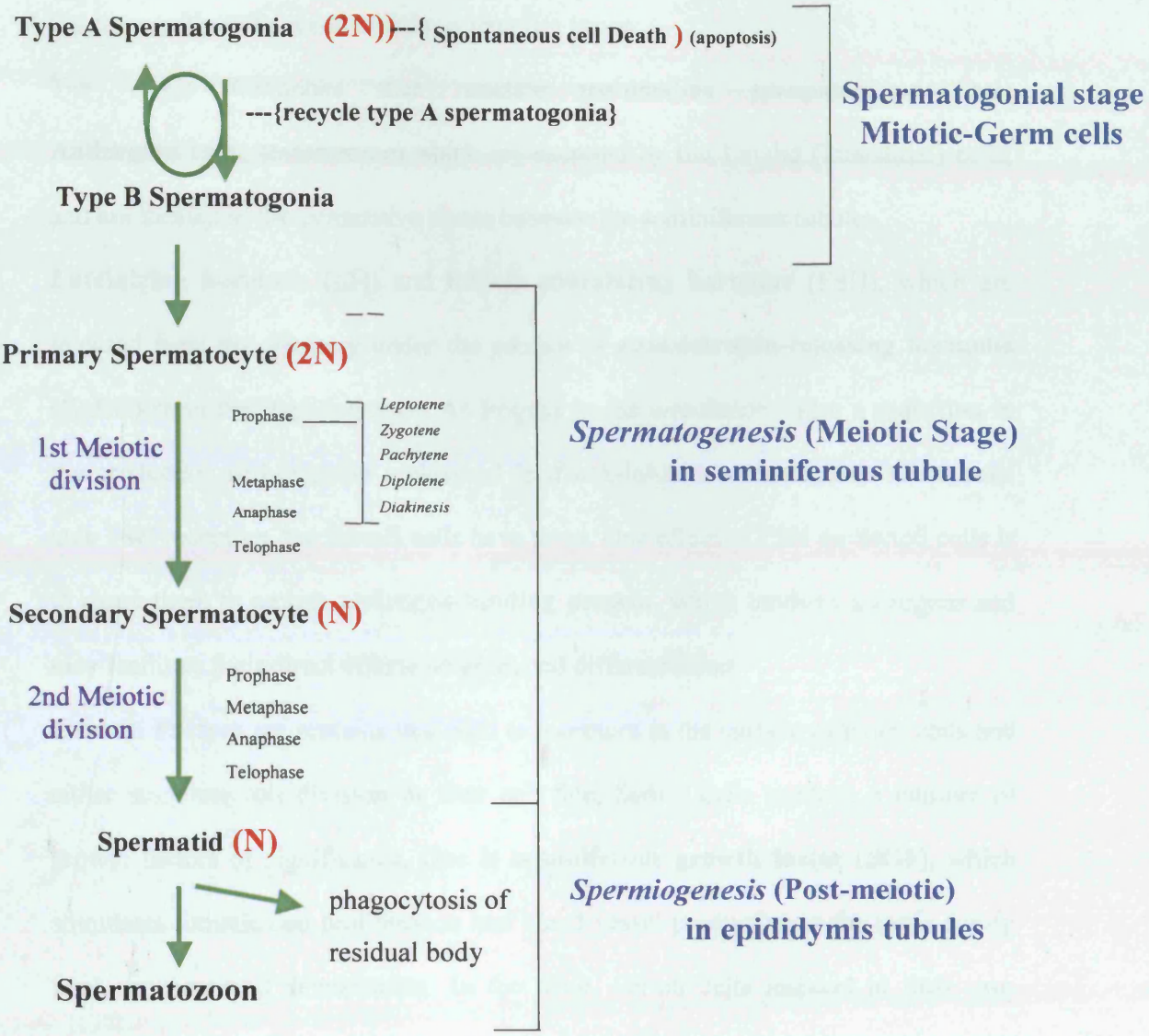


Figure 1.3.3- Schematic showing the stages of mitotic and meiotic division during the differentiation of germ cells. The ploidy of each cell type is indicated. The whole process of sperm maturation from spermatogonia to presence of sperm in an ejaculate will take at least 80 days and depends on hormonal signals combined with a mechanical process (Jequier, 2000).

At the end of meiosis, the spermatids, while having haploid nuclei, do not have the structural characteristics of mature sperm cells. The cells, still linked together by cytoplasmic bridges, appear rounded, with no defined head and tail. Head and tail differentiation occur simultaneously during the second stage of sperm production,

spermiogenesis. Throughout this stage, intricate and highly specific modifications in the composition of several cellular organelles occur.

The major hormones that regulate mammalian spermatogenesis are: **Androgens** (e.g., testosterone),which are secreted by the Leydig (interstitial) cells, and are located in the connective tissue between the seminiferous tubules.

Luteinizing hormone (LH) and **follicle stimulating hormone** (FSH), which are released from the pituitary under the control of **gonadotropin-releasing hormone** (GnRH) from the hypothalamus. Androgens in the circulation cause a reduction in the production of LH under a classical feedback-inhibition mechanism. Germ cells lack FSH receptors, but Sertoli cells have them. 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.

Growth Factors are proteins that bind to receptors in the surface of target cells and either stimulate cell division or alter cell fate. Sertoli cells produce a number of growth factors of significance. One is **seminiferous growth factor** (**SGF**), which stimulates somatic cell proliferation and blood vessel production in the testis during fetal and postnatal development. In the adult, Sertoli cells respond to their own production of SGF by producing **sulfated glycoprotein-2** (SGP-2; also called Clusterin or Apolipoprotein J). This is an **autocrine interaction**. SGP-2 is the major secretory product of adult Sertoli cells. It, in turn, binds to the membranes of spermatozoa. This is a **paracrine interaction**. SGP-2 is thought to be involved in protecting cells from apoptotic cell death (Clark et al, 1997; Clark and Griswold, 1997; Viard et al, 1999).

The first step towards the development of the sperm head is the formation of the acrosomal vesicle. This is a membrane-bound sac containing specialised enzymes involved in digesting a pathway to, and fusion with the egg membrane. It has both an inner and an outer membrane and forms a cap over the anterior of the nucleus in the mature spermatid.

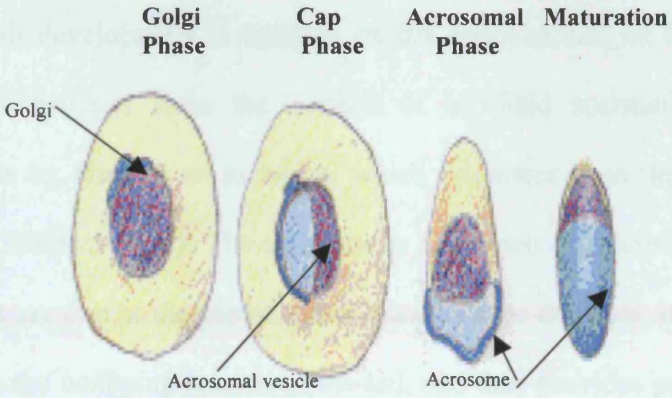


Figure 1.3.4- Formation of sperm acrosome

The acrosomal vesicle or acrosome is derived from the flattened stacks of membrane sacs that make up the golgi apparatus. This has been seen to take place in four stages (Abou-Haila et al. 2000), (Figure 1.3.4). The first stage is the golgi stage, during which small membrane bound granules form the stacks of the golgi. These then fuse together to form a large membrane bound sac, in close contact with the nuclear envelope. Next is the cap phase, in which the newly formed acrosomal vesicle becomes enlarged, and as it does so, flattens out over the nucleus. Third is the acrosomal phase, in which the acrosome adopts a concave structure. The final stage is maturation, when the characteristic cap-like structure, covering a large portion of the surface of the nucleus is established. The nucleus and the acrosomal cap rotate, then they are facing the anterior (head end), end of the future sperm cell. This leaves

the posterior (future tail end), facing into the lumen of the seminiferous tubule (Gilbert et. al. 2003). A transient structure named *'manchette'* is seen in sperm head consisting of microtubules that encircle the condensing nucleus. The manchette consists of the microtubule mantle, a perinuclear ring, and dense plaques at the distal ends of the mantle (Tres and Kierszenbaum, 1996).

Meanwhile, tail development is initiated on the opposite side of the nucleus. The flagellum first projects from the surface of a round spermatid during early spermiogenesis by forming an axoneme, which originates from the distal centriole (Fawcett and Phillips, 1969). The axoneme is composed of microtubules, arranged such that there are two at the centre, surrounded by nine doublets. It is this structure that facilitates the whipping motion of the tail, and thus provides propulsion for the sperm cell. The developing flagellum extends down into the lumen of the seminiferous tubule.

During the next stage, changes occur in the nucleus, so that it becomes smaller and flatter. The chromatin, usually organised in a solenoid structure, maintained by histones, becomes further condensed. The histones are replaced by protamines. The protamines interact with the minor groove of the DNA molecule, effectively cross-linking and neutralising the phosphodiester backbone of the molecule. Residues on the protamine complexes form hydrogen bonds, hydrophobic associations and disulphide bonds with each other (Balhorn 1982). This results in condensation, aiding the formation of a more compact, crystalline chromatin structure. Under these conditions, transcription is prevented from taking place, thus any subsequent protein

synthesis is reliant on mRNA, transcribed prior to this stage and stored as messenger ribonucleic protein, mRNP (Schmidt et. al. 1997).

Also at this stage, the mitochondria are localised towards the posterior of the developing cell's cytoplasm. These mitochondria are distinct from somatic cell mitochondria, as they contain large vacuoles (Hecht 1998). They accumulate around the flagellum near the centriole. Here they become more crescent shaped, and organise themselves end to end in such a way that they coil around the axoneme, forming a sheath (Metz et. al. 1985). This becomes the midpiece of the sperm. From this position they can supply the ATP required for the protein dynein in the axoneme to produce mechanical movement of the microtubules. In other words, the mitochondria in the midpiece provide the energy for the sperm cell to swim (Figure 1.3.5).

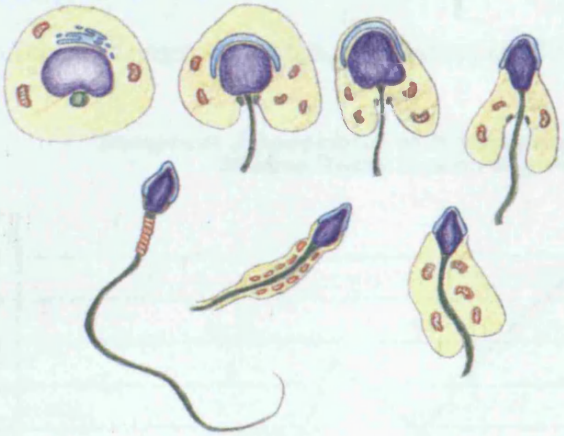


Figure 1.3.5- Schematic diagram representing different stages of sperm maturation during spermiogenesis

Finally, the residual cytoplasm is discarded, releasing the fully formed sperm into the lumen of the seminiferous tubules (Gilbert et. al. 2003). The sperm then pass along the seminiferous tubules to the epididymis mediated by peristaltic contraction of the seminiferous tubules.

In the proximal region of the epididymis, the sperm cells mature further. Modifications of cell surface proteins occur, enabling the sperm cell to bind to the zona pellucida, and subsequently penetrate the oocyte. Also, it is in the epididymis that the sperm cells acquire self-propulsion. The fully functional sperm cells are then stored in the terminal region of the epididymis until they are needed. Spermatogenesis is generally considered to be about 8 stages but the number differs with species and investigator. The number of divisions (and their rate) undergone by spermatogonia determines the rate of sperm production. Briefly, spermatogenesis is controlled (synchronized) by Sertoli cells and Leydig cells (Interestitial cells) via androgen secretion under LH control (Yoshida and Oshima, 1994). Spermatogenic cells require testosterone especially at the primary spermatocyte stage. Some believe FSH has a role in later stages of spermiogenesis (Moura and Erickson, 2001), too. Inhibin (produced by Sertoli cells and which inhibits pituitary release of FSH) may be involved as a negative feedback (Chada et al., 2003).

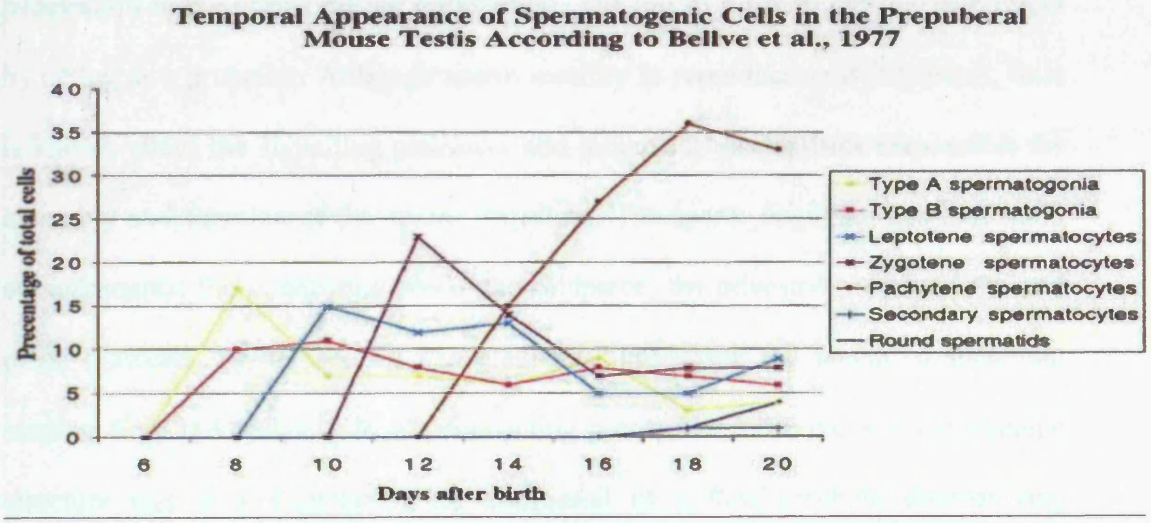


Figure 1.3.6

Mature sperm structure and motility

Each spermatogonium lining the tubules divides rapidly so enables each testicle to produce between 300 and 600 sperm per gram of testis per second. On average, sperm production rate is 1,500 per second per testicle. It takes 74 days to form sperm and 26 day for them to mature as they pass through the epididymis and vas deferens. After maturation a sperm swims at a rate of 3 mm (1/10 in) per hour and lashes its tail 800 times to swim one cm (1/3 in). Then it must travel through 30-40 cm of male and female 'plumbing' to reach the Fallopian tube (more than 100,000 times its length). After ejaculation it reaches the Fallopian tubes within 30-60 minutes (which is helped along by eddy currents). The average survival for a sperm in the female reproductive tract is 3-4 days.

Sperm cells (figure 1.3.7) are made up of a head and a tail. The head contains the nucleus and has an acrosome situated at the tip. A number of digestive enzymes are housed in the acrosome. These break down the surface of the egg allowing penetration of the sperm during fertilisation. The tail of a sperm enables it to travel by acting as a propeller. Although sperm motility in reproduction is important, little is known about the signalling pathways and molecular mechanisms responsible for assembly and function of the sperm flagellum. The sperm flagellum has four main compartments: the connecting piece, the midpiece, the principal piece and the end piece (Fawcett, 1975). The axoneme extends throughout the length of flagellum starting from the centriole in the connecting piece. The axoneme is a cytoskeletal structure and is a flagellar motor composed of a 9-microtubule doublet ring surrounding a central pair of microtubules. Mammalian sperm ultrastructure is highly

conserved. Axonemes are found not only in sperm cells but in all flagellated and ciliated cells from *Chlamydomonas* flagella to the inner-ear hair cells of mammals (Turner, 2003). Sperm motility is driven by the protein dynein, which hydrolyses ATP to provide the energy required for the sliding of adjacent outer doublet microtubules. Sperm mitochondria are present only in the mitochondrial sheath (MS) of the midpiece (figure 1.3.8). This MS is not present in the principal piece, therefore, to supply the energy needs of the axonemal dyneins in the distal flagellum segment, either other regions of the sperm flagellum must be able to produce ATP or the ATP produced by sperm mitochondria would need to travel some distance.

Dynein arms project from each of the outer 9 doublets (Figure 1.3.8). One spoke originates from each of the 9 outer microtubular doublet pairs going toward the central pair in a helical fashion. In the midpiece, 9 outer dense fibres (ODFs) lay outside each of the outer axonemal microtubule doublets and a sheath of mitochondria encloses the ODFs and the axoneme. The ODFs develop during midspermiogenesis. It has been suggested that the ODFs' functions in sperm motility are structural, providing passive elasticity to the motile flagellum (Fawcett, 1975). In the principal piece an MS is not present and the ODFs associated with outer axonemal doublets 3 and 8 are replaced by 2 longitudinal columns of the fibrous sheath (FS). The FS is a cytoskeletal structure with two roles in sperm motility: a mechanical role by providing a rigid support for the flagellum and a second role by creating an "I-beam"-like structure, along which the microtubules can slide, by supporting outer doublets 3, 8 and the central pair of microtubules (figure 1.3.8). In activated axonemal sliding with ATP, the FS slides proximally toward the

connecting piece while doublets 1, 2, 4-7 and 9 (but not 3 and 8) are extruded distally. Fibrous sheath sliding is cAMP-dependent, which suggests protein kinase A (PK-A) involvement (Si and Okuno, 1993, 1995). The morphogenesis of ODFs and FS along the axoneme is well defined: ODFs are assembled in a proximal-to-distal direction in step 8 to 19 spermatids while the FS is assembled distally to proximally along the flagellum length in step 2 to 17 spermatids (Irons and Clermont, 1982; Oko and Clermont, 1989; Clermont et al, 1990).

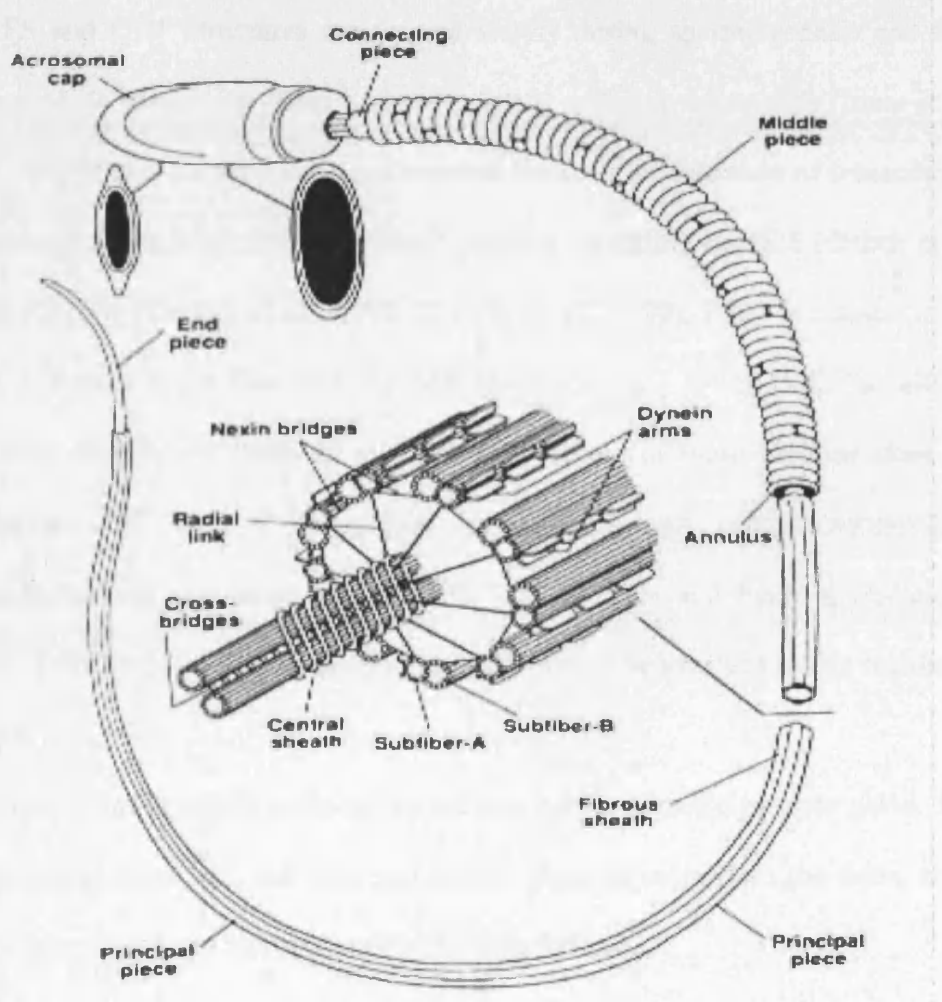


Figure 1.3.7 (http://www.ucalgary.ca/~fvdhoorn/picture.1999.html)

Many ODF and FS proteins are likely to be encoded by testis-specific genes expressed in the flagellum of the spermatid (Oko and Clermont, 1989). It is very likely that the functions of these components rely on the networks, which are formed by interactions between individual ODF and FS proteins (Shao et al., 1999).

Another role for FS has also been proposed in protecting sperm from oxidative stress that interferes with sperm motility (Fulcher *et al*, 1995).

In the distal end of the flagellum, the FS and ODFs are tapered and terminated, until the axoneme is surrounded only by the plasma membrane to form the end piece. Both FS and ODF structures are formed slowly during spermiogenesis and they continue to be made after transcription ceases in midstage spermatids (Irons et al., 1982). Significant delays have been reported between the initiation of transcription and translation for several fibrous sheath proteins, including GAPDS (Bunch et al., 1998), AKAP4 (Carrera et al., 1994, El-Alfy et al., 1999), FS39 (Catalano et al., 2001), and outer dense fibre proteins Odf1 (Morales et al., 1994), Odf2 (Schalles et al., 1998), and Spag4 (Shao et al., 1999). Synthesis of these proteins does not commence until after their mRNA transcription has ceased necessitating posttranscriptional regulation of their mRNAs. A testis and brain RNA-binding protein, Translin (TB-RBP/translin) has been shown to be involved in this regulation (Yang et al., 2003).

The sperm swim in a fluid produced by the seminal vesicles and prostate gland. This fluid provides them with nutrients and enables them to swim from the testis, along the epididymis and vas deferens, and to the ejaculatory duct.

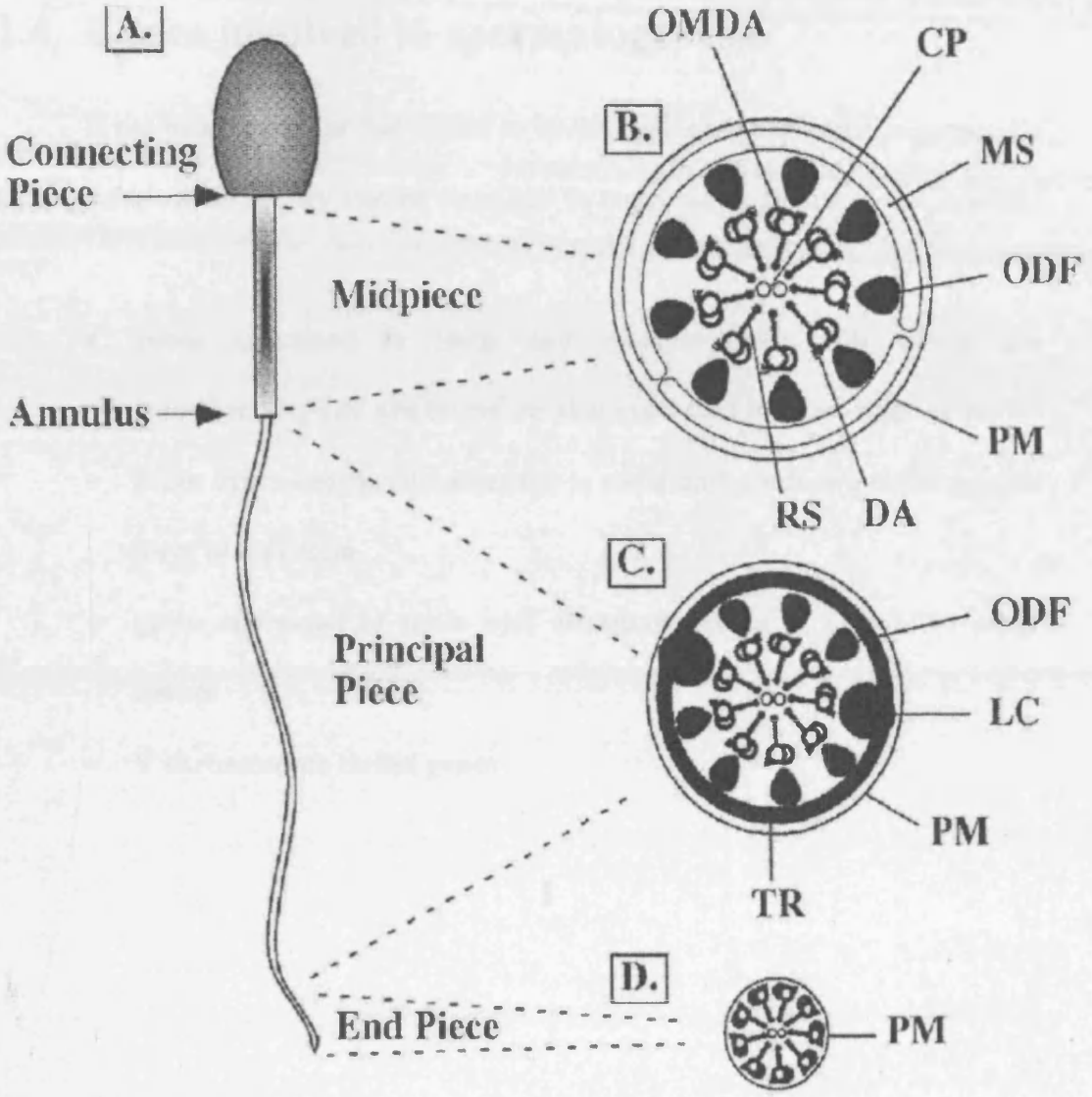


Figure 1.3.8 Schematic feature of a mammalian sperm and the ultrastructure of the flagellum.
(A) Mammalian sperm flagellum consisting of connecting piece, midpiece, principal piece and end piece.
(B) Cross section of a midpiece segment showing the plasma membrane (PM) and mitochondrial sheath (MS) surrounding the 9 outer dense fibres (ODFs). The 9 outer microtubule doublets of the axoneme (OMDA) with dynein arms (DA), radial spokes (RS) and the central pair of microtubule doublets (CP) are within the ODFs.
(C) Cross section of a principal piece segment showing ODFs 3 and 8 are replaced by fibrous sheath (LC). Two longitudinal columns (LCs) are connected by transverse ribs (TR). **(D)** Cross section of an end piece segment. (By Regina Turner, J Androl. 2003 Nov-Dec;24(6):790-803)

1.4 Genes involved in spermatogenesis

It has been estimated that 15,000 to 20,000 genes are expressed in germ cells (Zhang *et al.*, 1997). They can be classified in many ways. Below is one possible way:

- **genes expressed in testis and spermatogenic cells which are housekeeping and are therefore also expressed in most other cells**
- **genes expressed (predominantly) in testis and produce a testis specific form of a protein**
- **genes expressed in testis with alternative form of protein in other tissues**
- **Y chromosome linked genes**

1.4.1 Testis specific genes expressed predominantly in testis

There are about twenty thousand different transcripts in germ cells, relatively few of them have been characterized in detail but some which include:

PRM: Protamine is a main nuclear protein in sperm and is encoded by the *PRM* gene located in 16p13.3. PRM is associated with sperm nuclear shape and initiates condensation of the chromatin and packaging into the sperm head (Lee et al., 1995).

ACR: (Pro)acrosin is the zymogen form of the serine protease acrosin and is expressed throughout early stages of spermatogenesis (in pachytene spermatocytes) (Nayernia et al., 1994), then starts to accumulate into the Golgi system (Ventela et al., 2000).

SPRM1 is a member of the POU-domain gene family that is exclusively expressed in spermatogenesis, predominantly in the haploid spermatid. It is a transcription factor. Both human and mouse genes are testis specific in their expression.

SPAM1: sperm adhesion molecule 1 (SPAM1) gene is located in the same region as the cystic fibrosis gene (7q31.3) and is involved in sperm-egg adhesion by aiding the penetration of the egg's cumulus cell layer (Jones *et al.*, 1995). The gene expression is mainly limited to testis although it has been recently shown in vas deferens and efferent ducts as well (Zhang et al., 2004).

PGK-2: PGK-2 is expressed specifically during spermatogenesis. Its expression is initiated with the onset of male germ cell meiosis, and continues into the later stages of spermatocytes and in round spermatids.

1.4.2 Genes expressed in testis with alternative form of protein in other tissues (somatic cells) or altered expression

The reason for the presence of testis-specific isoforms of proteins in testis may be due to the specific physiology of the testis because of having non-differentiated and dividing cells. Surprisingly, expression has been found in male germ cells genes unexpected to be active during spermatogenesis such as the smooth muscle gamma actin gene (Gu et al., 1996). Also, many proto-oncogenes have altered expression during germ cell division; e.g. the mouse genes c-abl and c-mos (a protein tyrosine kinase and a protein serine/threonine kinase), are both expressed in many tissues. However, they are present in the testis with a different transcript size compared to those seen in somatic tissues. 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).

Testis specific promoters have been described for a number of genes expressed during spermatogenesis. It has been proposed that the choice of promoter is governed by DNA methylation (review, Herman and Baylin 2000).

HSP70: encodes Hsp70-2 protein in humans and Hsc70t protein in mice, which are molecular chaperones and possibly are linked to mechanisms that inhibit apoptosis of spermatocytes (Dix *et al.*, 1996; Eddy, 1999).

LDHC: encodes lactate dehydrogenase-C (LDHc) isozyme which has been found in mature testes of many species including men (Zinkham *et al.*, 1965).

H1t- The mammalian testis-specific histone H1t gene is transcribed in primary spermatocytes during spermatogenesis (Doenecke *et al.*, 1994). It is likely that H1t is essential for dramatic changes in chromatin structure and the changing patterns of gene transcription seen during spermatogenesis.

G3PD- Glyceraldehyde 3 phosphate dehydrogenase also has a post meiotic isozyme (GAPD-S) which is expressed only in male germ cells and is associated with the fibrous sheath of the flagellum (Welch *et al.*, 1995).

ACE: Angiotensin Converting Enzyme gene encodes a testis-specific isozyme found only in differentiating spermatids and mature sperm (Sibony *et al.*, 1994). It has a haploid expression and post-meiotic translation.

1.4.3 Y chromosome linked genes

There are several families of Y linked genes. In 1996 it was shown that a deletion in the distal part of the long arm of the Y chromosome (Yq11) can result in azoospermia. This area of the Y chromosome is referred to as the azoospermia factor (AZF) region. At least two gene families encode RBP (RNA Binding Protein) in the AZF region. The first of them is RBMY (RNA Binding Motif) (Chai, 1997&1998) consisting of approximately 30 genes and psuedogenes spread over both arms of the Y chromosome. RBM is a nuclear protein that is expressed exclusively in human male germ cells of foetal, prepubertal and adult testis (Cooke and Elliot, 1997). AZFb deletion results in reduced levels of the RBMY protein and spermatogenic impairment at the meiotic and post meiotic transition resulting in type II Sertoli Cell Only Syndrome and limited spermatogenesis (review Krausz and McElreavy, 1999). The RBMY family has an X-linked homologue in both human and mouse. A deletion in the AZFc region located within this region is the gene family '*Deleted in azoospermia*' (DAZ) associates with observed severe hypospermatogenesis more often than with Sertoli cell-only syndrome (Roijen et al., 1995; Vereb et al., 1997). Deletion of this gene family on its own is not sufficient to cause complete loss of the spermatogenic line (Ferlin *et al.*, 1999).

1.5 The regulation of Testis-specific gene expression through common promoter elements

Studies on the mechanisms that regulate testis-specific gene expression have been focused mainly at the transcriptional level because it is so frequently responsible for dynamic regulation of the gene product. Many studies have drawn analogy to the mechanisms that have been shown to regulate transcription in testis. Some genes such as *CREMT* have a similar temporal expression pattern during germ cell differentiation tend to contain the same transcriptional regulatory motifs. The binding site for the cAMP response modulator element (CREMT) is present in the promoters of many testis-specific transcripts (review, Goldberg, 1996; Sassone-Corsi, 1997). This transcription factor plays an important role in the nuclear response to cAMP signal transduction pathway in neuroendocrine cells. In mouse and human, lack of the CREM gene has a severe effect on spermatogenesis (post-meiotic arrest) and germ cell apoptosis (Nantel and Sassone-Corsi P. *et al.*, 1996) and may lead to oligozoosperma (Peri *et al.*, 1998).

Signalling mechanisms often require a signal source (hormones), a signal reception system (e.g., G protein-coupled receptors) and an intracellular signal response system. Paracrine signals originate from nearby cells and there are varying paracrine interactions between the germ cells and Sertoli cells due to their intimate location. It has been reported 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 (Grandijean, 1997). In the classic pathway of activation, steroid hormones pass through the cell membrane

and bind to intracellular receptors. These receptors are hormone-regulated transcription factors that activate migrating components into the nucleus and regulate gene expression. Also steroids are able to act in another way as a 'non genomic action' (review, Wehling, 1997). In addition, when estradiol makes a complex with an intracellular receptor it activates the tyrosine kinase and MAP-kinase (Mitogen-activated protein) pathway in human mammary cancer MCF-7 cells and acts in a similar fashion to a peptide mitogen (Migliaccio et al., 1996). The first part of spermatogenesis is mitosis and the best-known pathway of stimulation of mitosis (RAS → RAF → MAPK cascade) is strongly stimulated by mitogens, growth hormones and proto-oncogene products (Cobb et al., 1995). MAP kinases phosphorylate and activate nuclear transcription factors that in turn activate genes controlling the cell cycle. Testis specific promoters have been described for a number of genes expressed during spermatogenesis (see above).

The transformation in the sperm 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. Some genes have been identified that are not transcribed until after the meiotic reduction (review, Hecht, 1990). There remain many unanswered questions: Does any translation occur in sperm right up until fertilization and what is the exact process of translation in the very small volume of cytoplasm, which remains in sperm during spermatogenesis? Where are the origins of transcripts for protein production? Are these transcripts made post-meiotically or are they stored mRNAs transcribed at the pre-meiotic stage? The expression of proteins in the acrosome and sperm Golgi apparatus

together with the importance of the ER-chaperone protein calreticulin are likely signs of translation in sperm (review, Okabe et al., 1998). Repression and activation of translation require proteins or antisense RNAs that bind to regulatory elements in the RNA and regulate the translatability of messages. mRNA binding proteins (RBPs) bind to specific sequence sites of a transcript and prevent translation thus creating stored 'paternal' mRNAs. After transcription of protamine 1 gene 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. Protamine 1 contains a "Y box"-binding site for RBP attachment. 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). Synthesis of a translational activator such as a protein kinase or phosphatase that modifies the mRNP might be necessary. Evidence that mRNPs and polysomes can associate with cytoskeletal elements supports the presence of a mechanism for organised protein synthesis (Braun, 2000). Y box proteins are sequence-specific RNA binding proteins. Mys1 (a mouse Y box protein) may be associated with untranslated RNAs (Tafuri et al., 1993). The mouse p48 and p52 box proteins are highly expressed 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). Shortening of the poly A tail of mRNAs from about 100-200 nucleotides to about 30 occurs in spermatogenesis and it may not be required for translational initiation. In testis, migration, proliferation and differentiation of the most immature germ cells (spermatogonia) can be regulated by the c-kit ligand (Bhasin et al., 1994; Griswold 1993). 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, 1998). The CSF-1 growth factor stimulates the growth and survival of myelomonocytic lineage cells such as macrophages. Steel factor (SLF) is a hematopoieitic and tissue growth factor produced by Sertoli cells that binds to the receptor encoded by *c-Kit* proto-oncogen (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 results in activation of mitosis. The above-mentioned growth factors belong to the PDGF family which are the major protein growth factor in human serum.

Gonadotropin receptors belong to a subgroup of the super G protein-coupled receptors. A large N-terminal extracellular domain (responsible for binding of the hormone) containing leucine-rich repeats characterizes these receptors. An important secretion of Sertoli cells are transforming growth factors B (TGF-B), a superfamily of growth factors, whose receptors are also expressed in mammalian testis. They regulate a variety of developmental processes including spermatogenesis (review, Massague, 1998). Inhibin and bone morphogenesis protein (BMP) are two members of this family that play roles in spermatogenesis.

1.6 *TSGA10*: mRNA, cDNA, amino acid sequences and predicted protein preliminary analysis

TSGA10 was originally identified by differential display RT-PCR (DDRT-PCR) in the Galton laboratory in 2001 (Modarressi *et al.*, 2001). The DDRT-PCR experiment was carried out on a range of adult human tissues, and *TSGA10* was one of a limited number of cDNA fragments that appeared to be specifically expressed in testis. *TSGA10* is composed of 19 exons and the whole gene extends over more than 120kb (figure 1.6.1). It is shown that the gene is expressed in adult normal testis but not in other adult tissues, including those, which are functionally related to testis. Interestingly, it is not expressed in the testes of some infertile patients. Northern analysis confirmed the specific expression of *TSGA10* in testis and showed that the mRNA length is about 3.2 kb (Modarressi et al., 2001).

Searching NCBI GenBank database with the human *TSGA10* sequence revealed just one homologous rat sequence for the gene in April 2001. In addition, *TSGA10* sequence were found in cDNA libraries made from some neoplastic cells, this suggested that the protein may play a role in cell cycle. The *TSGA10* gene has a long 5' untranslated region (UTR) whereas many 5'UTR sequences are only 20-100-nucleotides long among eukaryotic mRNAs (Kozak, 1987). The efficiency of translation from the first AUG initiation codon can be increased by the presence of a long 5'UTR via loading an added 43S-pre-initiation complex termed pre-loading (Kozak, 1991). Conversely, it is impaired by shortened 5'UTR of reporter mRNAs to less than 12 nucleotides (Sedman *et al.*, 1990). However, many cellular mRNAs with long 5'UTRs are poorly translated as a result of upstream AUGs, uORFs and/or

```
   1-caggtatctaacctatctttgtatttttggaatttaatatagttcctgacataatcattg   -R12
  61-ttcaatgggtctaaTTTTTTTTTcctttaaaggcatttggataccttgtgatagaaaag
 121-gccctttttcttcatccattgatggtttgatagtgggctgggaaggaaagctgtgttcctc   -GSP12
 181-cacattaggcagcaaatacttgattgatttgatatttagcttagtcaaattgaagatctc   -GSP11
 241-aaacagacaaatcatggcttggaagaatgtgttaggaaactcttggatagtaaggaggtg   -R11
 301-gtaagcagtcaagtagatgatttaaccagccacaatgagcatctttgtaaagaattgatt
 361-aaaattgaccaactagcagagcaactcgaaaaagagaaaaattttgtggtggattccgcc
 421-aacaaggaacttgaagaagccaagattgatctcatttgccagcaaaataatatttatagta
 481-ttggaagatacaatagaaaaggcttaaatctacttgattcctttgtcaagactttggaagc
 541-agacaaagatcactataagagtgaagctcaacattgagaaagATGatgcgaagtaggtct   -GSP10 -MET
 601-aaaagtccaagacgcccatcaccaactgcccggggtgcaaactgtgatgtagaacttttg   -R10
 661-aagacaacaacaagagatcgtgaagaacttaaatgcatgctggaaaaatatgagcgccat
 721-ttggcagaaattcagggtaatgtcaaggttcttaaatctgagagagacaagatcttcctt
 781-ctttatgaacaggcacaggaagaaattacccgacttcgacgagaaatgatgaaaagctgt
 841-aagagtcctaaatcaacaacggcacatgctattctccggcgagtggagactgaaagagat   -GSP9
 901-gtagcctttactgatttacgaagaatgaccacagaacgagatagtctaagggagaggcta
 961-aagattgctcaagagacagcatttaatgagaaggctcacctggaacaaaggatagaggag
1021-ctggagtgtacagttcataatcttgatgatgaacgtatggagcaaatgtcaaatatgact
1081-ttgatgaaggaaaccataagcactgtggaaaaagaaatgaaatcactagcaagaaaggca
1141-atggataccgaaagtgaacttggcagacaaaaagcagagaataattctttgagacttttg   -R8
1201-tatgaaaacacagaaaaagatctttctgatactcagcgacaccttgctaagaaaaaatat
1261-gagctacagcttactcaggagaaaattatgtgcttggatgaaaaaattgataactttaca
1321-aggcaaaatattgcacagcgagaagaaatcagcattcttggtggaaccctcaatgatctg   -GSP6
1381-gctaaagaaaaggaatgcctgcaagcatgtttggataaaaaatctgagaatattgcatcc
1441-cttggagagagtttggcaatgaaagaaaagaccatttcaggcatgaagaatatcattgct   -R6
1501-gagatggaacaggcatcaagacagtgtactgaggccctaattgtgtgtgaacaagacgtt
1561-tccagaatgcgtcggcaattggatgagacaaatgatgagctggcccagatcgccaggggaa
1621-agagatatcttggctcatgacaatgacaatctccaggaacagtttgctaaagctaaacaa
1681-gaaaaccaggcactgtccaaaaaattgaatgacactcataatgaacttaatgacataaaa
1741-cagaaggttcaagatactaatttggaggttaacaagctgaagaatatattaaagtctgaa
1801-gaatctgagaaccggcaaatgatggaacaacttcgaaaagccaatgaagatgctgaaaac   -R5
1861-tgggaaaataaaagcccgtcaatcagaggcagataacaatacccctcaaactggaacttatc
1921-actgctgaggcagagggtaacagattaaaagaaaaagtagattccctcaacagagaggtt
1981-gagcaacacttaaatgcagaaaggtcttacaagtcccagatttctacottacataaatct
2041-gttgtaaaaatggaagaggagcttcagaaggttcagtttgaaaaagtgtccgctcttgca
2101-gatttgtcttctactagggaactctgtattaaacttgactcaagcaaagaacttcttaat   -GSP5
2161-cgacagctggttgctaaagatcaagaaatagaaatgagggagaatgagttagattctgct
2221-cattctgaaattgaactcctgaggagtcagatggcaaatgagagaatctccatgcagaat
2281-ctagaagctttgctggtggccaatcgagacaaagaatatcagtctcagatagcacttcaa   -R4
2341-gaaaaagaatctgaaattcagcttcttaaagaacacctttgtttggcagaaaataaaatg
2401-gccatccagagtagagatgtggcccagttcagaaatgttgtcacacaattggaagctgat
2461-ttagacattaccaaaagacaactaggaacgagcgctttgaaagggagagggccgtacaa   -GSP4
2521-gaacttcgccgccaaaattattcaagtaatgcttatcatatgagttctacaatgaagcca
2581-aatacaaaatgtcattcaccagaacgtgctcaccatcgatctcctgaccgaggcctagat
2641-cgatcattagaagagaatctttgctacagagatttcTGAcacctgaaatgattcttcaca   -STOP
2701-tccctgagaaaggtcaaagttacaaactgaTTTTTTTTTTTTTTgctacatgagtgcatt   -GSP3
2761-tatcttttaaatgcttggcaatgttaaatgtatttattaactttgtgtctctgaatctct   -R1
2821-gttctaatgtgccatgttgcagtgactgagatgacttatgaaaacaaaatgtatatgctct   -GSP2
2881-tctatccatgcaatgatagtgagtgtaaaatctgcttacttcactattgaacactattct   -GSP1
2941-gttcactatctggagtaaataaagaagcttattaaaacagggaaaaaaaaaaaa
```

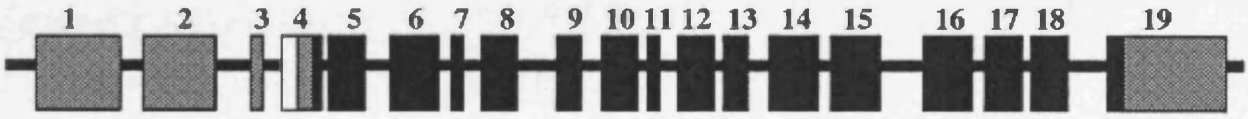


Figure 1.6.1- The sequence of the *TSGA10* gene including complete coding sequence (the initiating methionine at position 583 and the stop codon at position 2677 are indicated on the right). Some of primers used are underlined and mentioned at the right, as is the polyadenylation signal. Polymorphic 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′). The *TSGA10* gene contains 19 exons (solid black boxes are coding regions and solid grey 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, 634, 793, 964, 1042, 1192, 1309, 1465, 1521, 1690, 1801, 1987, 2197, 2400, 2505 and 2655, respectively. The length of introns are 791, >8744, 1337 or 1293, 397, >2971, 90, 1240, >12234, >2532, >5493, >1163, 2707, 3763, >18793, >1192, >1625, >427 and 19976 base pairs, respectively.

secondary structure. Long 5'UTRs are present in mRNAs encoding proto-oncogenes, transcription factors, growth factors and their receptors (Kozak 1987, 1991) suggesting a role in tightly controlled translation.

In 2001, human TSGA10 protein sequence composed of 698 amino acids (see appendix) showed to have a domain homologous to myosin tail.

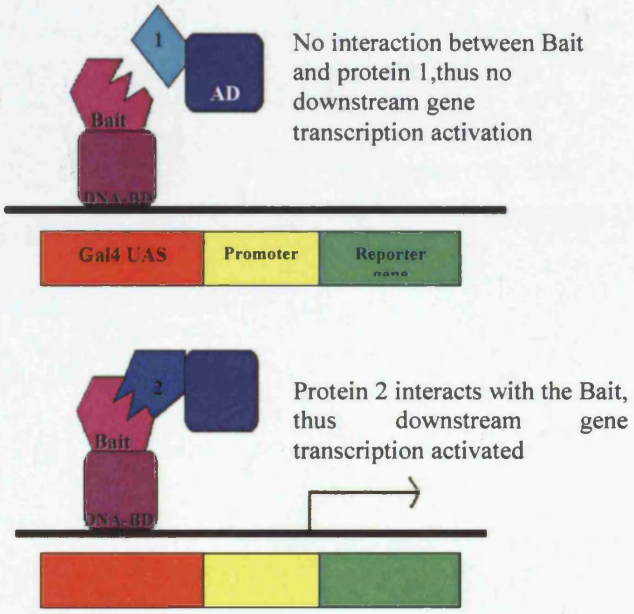
1.7 The investigation of protein interaction

The interaction of one protein with another protein forms the molecular basis for many cellular processes. Thus an important part of a protein's functional profile has to be which, if any, other proteins it interacts with. An important part of this study was to investigate which other proteins react with TSGA10. The yeast two-hybrid system provides an *in-vivo* system for the detection of protein interaction. It enables detection of weaker interactions that other methods, such as co-immunoprecipitation, may not be sensitive enough to pick up. It has the further advantage of identifying not just the interacting proteins, but also the genes that encode them (Chien et. al. 1991). Since this approach has been used in the work described in this thesis, it is explained here.

This is usually carried out in yeast as they are able to express most mammalian genes and performing the test *in vivo* increases the accuracy of the results. The system utilises properties of a yeast *Saccharomyces cerevisiae* protein called GAL4. GAL4 is a transcriptional activator, which switches on transcription of genes whose protein products are involved in galactose metabolism. The GAL4 protein itself consists of two functional domains: The N-terminal domain that binds to upstream activating sequences (UASs) of the genes that it regulates, GAL4 DNA-BD; and the C-terminal domain, that activates transcription of the downstream genes, GAL4 AD. The two domains are functionally distinct from one and other, and both are required to activate transcription of the galactose metabolising enzyme genes. This is exploited in the yeast two-hybrid system devised by Stanley Fields and Ok-kyu Song in 1989.

The yeast two-hybrid system works by separating the two functional domains of the GAL4 protein. The separate functional domains are each then used to create a hybrid protein. One hybrid protein is comprised of GAL4 DNA-BD and protein A, and the other is comprised of GAL4 AD and protein B. Interaction between proteins A and B can then be tested. If A and B form an interaction complex, the UAS-binding domain and the activating domain of GAL4 are brought into close proximity. The UAS binding domain binds to the UAS, and the proximate activating domain activates transcription of downstream genes. Thus the interaction can be screened for by selecting for expression of genes whose transcription is activated by GAL4, e.g. galactose selection (Fields and Song 1989). The system can be used to screen cDNA libraries, enabling isolation of clones encoding proteins that interact with a protein of interest.

Figure 1.7- Yeast two-hybrid interaction activates transcription of downstream selectable marker genes



The hybrid proteins are encoded on two yeast plasmids. On one plasmid, GAL4 DNA-BD is encoded, upstream and in-frame of the cDNA encoding the protein of interest. This is the bait plasmid construct. On the other plasmid, GAL4 AD is encoded, upstream and in-frame with a cloning site

into which library cDNA is cloned. This is the library plasmid. The two plasmids are introduced into the same yeast, *Saccharomyces cerevisiae,* cell. The plasmids have yeast promoters upstream of the hybrid protein gene construct, and thus once in the yeast cell, both hybrid proteins are produced.

The principle drawback of the yeast two-hybrid system is that it can throw up a large number of false positives. It is therefore important to consider yeast strains, and vectors carefully to try to minimise the number of false positives.

In the case of *TSGA10*, this information could reveal how the protein affects sperm motility and how mutations and abnormalities may cause infertility. Interactions between proteins are studied using the two-hybrid system to discover whether any other proteins found in the testis interact with the product of the *Tsga10* gene.

1.8 Aims

The main aim of this project was to investigate the role of the testis-specific *TSGA10* gene and the function of its encoded protein in order to draw some conclusions about how the gene may be involved in male fertility. Briefly the following strategy was followed in this work. Each step was designed to obtain information that would relate to the function of the gene.

1- Clone the mouse homologue of the *TSGA10* gene (*Tsga10*) to work on mouse as an animal model because the gene is highly conserved in mammals. Examine the DNA sequence in terms of gene structure, transcription factor binding sites and signalling to other genes.

2- Confirm the expression pattern in testis and other organs using RT-PCR and *In Situ Hybridisation.*

3- Raise an antibody against a portion of the *Tsga10* protein and examine the tissue-specific expression pattern of the protein by western blotting and immunohistochemistry.

4- Use the antibody to study the protein localization within the cell (e.g. sperm, astrocytes) by immunofluorescence and immunocytochemistry.

5- Transfect a *Tsga10* coding sequence linked to a reporter gene (e.g. GFP) into a wide spectrum of cells to study its subcellular localization.

6- Examine the interaction of Tsga10 with other proteins using the yeast two-hybrid system.

Chapter II:

Materials and methods

2.1 RNA extraction

All experiments including RNA extraction were done in a RNase free environment. DEPC treated double distilled water was used for making all solutions. The slides for RNA in *situ* hybridization were dried in a fume hood and stored in a dust free container at room temperature. All plastics used were siliconised and all glassware was baked at 180°C overnight

2.1.1 Solid tissues (testis, embryos)

The protocol used was described by Chomczynski and Sacchi (1987). RNA was extracted from mature mouse testis, developmental stages of rat testis (after birth) and whole mouse embryos (E7, E10.5, E12.5, E13.5, E14.5, E15.5, E16.5, E17.5, E18.5 and E11.5, E15.5 head embryo) with Trizol (Gibco BRL).

2.1.2 Cell cultures

This protocol is a modification of the solid tissue preparation above. The amount of guanidine was increased to keep the molarity (1.5-2 M) correct for a liquid system. RNA extraction was performed on mouse neuro2A (a neural crest specific cell line) and NB4 cells (an acute promyelogenic leukemia cell line) from a starting sample of 250 µl of culture medium. The biological samples were first lysed and homogenised in the presence of a highly denaturing guanidine isothiocyanate (GITC)-containing solution that immediately inactivates RNase to ensure isolation of intact RNA. In

order to do this, after thawing out both the cells and the solution containing GITC, 400 μL of this solution was added to the sample and mixed on a roller for 15 minutes. Then 65 μL (1/10 of the volume of the sample and GITC-containing solution) of 2 M sodium acetate was added to the solution and thoroughly vortex mixed. Sodium acetate helps precipitate RNA and DNA. A 650 μl aliquot (equivalent to the volume of the sample plus GITC-containing solution) of water-saturated phenol was added and the whole was thoroughly vortex mixed. The resulting solution was passed back and forth several times through a 19-gauge hypodermic needle using a 1-ml syringe to shear the DNA. Then 130 μl (2 × the volume of sodium acetate) of chloroform was added, vortex mixed and put on ice for 15 minutes. The solution was then centrifuged at 13000-rpm (1000g) for 20 minutes. The upper aqueous layer was taken off into a fresh 1.5-mL tube and mixed with an equal volume of isopropanol (propan-2-ol). This precipitated the RNA. The solution was incubated at –70°C for at least two hours (overnight was satisfactory). The frozen sample was centrifuged at 13000-rpm (1000g) for 15 minutes and a pellet was formed. The supernatant was pipetted off. Finally, any lasting impurities were removed by washing the RNA pellet in 100 μL 70% (v/v in water) ethanol. After a further five minutes centrifugation at 13000-rpm (1000g), the ethanol was removed and the RNA pellet dried for 10 minutes in a vacuum desiccator. The tube was spun down briefly to ensure the pellet was at the bottom of the tube. The RNA pellet was dissolved in 10-20 μl DEPC-treated water, depending on the pellet's size. The RNA was stored at -70°C and used as the substrate for the ribozyme and DNase in the cell free system cleavages, or for making cDNA. Total RNA extraction was also performed in the same way on K562 cells to provide a

control substrate in the cleavage reactions. For the cells of neural crest origin, the RNA has been extracted from primary cell cultures of neural crest cells obtained from mouse embryos (8.5 dpc) and cultured for 6 days.

2.2 Reverse Transcription (RT) reaction

The nucleic acid pellet obtained using the method above was depleted of any contaminating DNA by treatment with DNase I (Gibco BRL). The concentration of the remaining RNA was determined by spectrophotometry (absorbance at OD 260) and 1 μg of this total RNA was reverse transcribed using MMLV Reverse transcriptase (Invitrogen, Burlington) and oligohexamer primers with RT buffer at 55°C for 50 min. The amount of single stranded cDNA was determined by checking on agarose gel.

2.3 Cloning protocols

2.3.1 Vectors

In all cloning experiments the vector was digested by the appropriate restriction enzyme, in the presence of BSA if necessary. The linear vector was electrophoresed through a 1% agarose gel buffered with TAE, and was extracted from a gel slice using the QIAquick gel extraction kit (QIAgen). The purified DNA was then dephosphorylated using calf intestine alkaline phosphatase (Amersham Pharmacia Biotech), and the concentration of DNA was measured by

spectrophotometry. Vectors used were pTAG, EGFP-c2, pGBKT7, pGADT7 and pGAD424 (Clontech) (See Appendix for vector information).

2.3.2 Inserts

DNA fragments for cloning were produced in several ways; they were either amplified by PCR from BAC R-138D5 containing both ends of the human *TSGA10* gene or were produced by restriction digestion of the same PAC, or of the *Tsga10* mouse homologue cDNA (clone P2), or by PCR amplification of the mouse testis reverse transcription product. All the DNA fragments to be cloned were extracted from 1% agarose gels (usually buffered with TAE, Appendix 4.2) using the QIAquick gel extraction kit (Qiagen) and glass wool purification method. After extraction of inserts from the gel, the DNA concentration was determined by spectrophotometry (CIBA-CORNING 2800 Spectrascan). Table 2.3.2 shows the details of the constructs, which were made and used in this project.

Table 2.3.2- Tsga10 constructs and relevant vectors

Vector Name	*Total Length*	*Insert Length*	*start*	*End*	*Promoter*
pBluescript II/Tsga10 (P2)	5.7 Kb	2.732 Kb	-533	2732	Lac
EGFP-c2/Tsga10	6.8 Kb	2.117 Kb	1	2117	CMV
pGBKT7/Tsga10	10.4 Kb	2.117 Kb	1	2117	GaL4
pGBKT7/Tsga10-N	9.35 Kb	1.042 Kb	1	1041	GaL4
pGBKT7/Tsga10-C	9.42 Kb	1.123 Kb	994	2117	GaL4

2.3.3 Ligation of digested vectors and the DNA fragments to be inserted

Typically 100-250 ng of digested vector DNA and a 3 to 5-fold molar excess of the fragment were used with 2-4 Weiss units of T4 DNA ligase (New England BioLabs or Roche) in a total volume of 10μl at 16°C or 4°C (respectively according to the recommendation of enzyme manufacturer) for 16 hours. Alternatively, the ligation reactions were carried out using the *Rapid Ligation Kit* (Promega) for a 12μl reaction and incubated at 24°C for 1.5hours:

2x Rapid ligation buffer	5μl
Digested vector DNA	3μl (1μg)
Digested insert DNA	3μl (600ng)
T4 DNA ligase	3μ

Samples of the ligations were then electrophoresed on a gel described previously.

For blunt-ended ligations, the digested fragment was first phosphorylated using T4 polynucleotide kinase enzyme (New England Biolabs) by incubation at 37°C for 30 minutes. To fill in overhangs, the fragment was incubated at 12-16°C for 20 minutes with 20 units of T4 DNA polymerase (New England Biolabs) and a low (1 mM) concentration of dNTPs. The reaction was stopped by incubation at 75°C for 10 minutes.

2.3.4 Bacterial transformation and PCR of colonies

5-10 μl of the ligation reactions were mixed with 50 μl of competent E.coli DH5α, JM109 or XL-Blue cells (Gibco BRL), incubated on ice for 30 minutes, heat shocked

at 42°C for 60-90 seconds and then returned to ice. After a couple of minutes Lauria-Bertani (LB) (see appendix) medium (950 µl) was added to each transformation and the tubes were placed in a shaking incubator at 37°C for 1 hour, at 250 RPM. Aliquots of 100 and 500 µl of the culture were spread onto LB plates containing the appropriate antibiotic marker (ampicillin for pTAG and kanamycin for EGFP-c2 and pGBKT7) (and, where appropriate, IPTG and Xgal), and incubated at 37°C overnight. Once colonies of *E. coli* had grown on the agar plates, 6 colonies were taken from each plate. PCR reactions were then set up with each colony to test for the presence of the inserted DNA. The reactions were set up using the same method as before but the PCR program was modified so that the first denaturing cycle ran for 5min instead of 3min. This was to ensure the bacterial cell walls would be broken to release the DNA inside. Following the PCR reaction, the samples were electrophoresed on an agarose gel and photographed.

2.3.5 Small-scale plasmid DNA preparation

4.5ml of LB containing appropriate antibiotics was inoculated with a picked colony and the suspension was grown at 37°C overnight. DNA was prepared from a 4 ml of culture using the QIAprep Spin Miniprep Kit (Qiagen). The plasmids were analysed by restriction analysis or by PCR, and positive samples were sequenced to confirm the positive clones. In some experiments, before growing the colonies, PCR was used to screen the colonies on the plate and find a positive clone that was then grown overnight. In some cases plasmids were prepared and purified manually as:

2ml of overnight bacterial culture grown from the colonies was transferred to fresh tubes. Then bacteria were pelleted by centrifugation for 1min at 10,000 rpm, supernatants were discarded and 200µl of Solution 1 (resuspension buffer: 50mM glucose, 25mM tris pH8, 10mM EDTA pH8) was added to the DNA pellet and mixed well by vortexing. Then 200µl of Solution 2 (0.2M NaOH, 1% SDS) and 200µl of Solution 3 (precipitation buffer: 60ml 5M potassium acetate, 11.5µl 37% acetic acid, 28.5µl water) was added and mixed gently after each addition. The mixture was centrifuged for 3min at 10,000 rpm and the supernatant was transferred to fresh tubes. The pellet was discarded. 500 µl of isopropanol was added to the supernatant, mixed and the tube was centrifuged for 5min at 10,000 rpm and the supernatant was discarded. Then the DNA pellet was resuspended in 500µl of 70% ethanol and the tube was centrifuged for 1min at 10,000 rpm. The remaining supernatant was discarded, and this centrifugation step was repeated. Finally the DNA resuspended in 50µl of distilled water. DNA samples were then electrophoresed on an agarose gel.

2.4 Bacterial cultures

The Escherichia coli (E.coli) strains DH5α, XL-Blue and JM109 were used in the experiments. Cultures were grown in LB medium. Agar plates were poured from LB containing 1.5% (w/v) agar and antibiotic (25-35 µg/ml). Where blue/white screening was used, Xgal (40 µg/ml) and IPTG (0.1 mM) were also added to the agar by spreading the solutions on top of the agar after it had set and been dried in a fume cupboard.

2.5 Large-scale preparation of plasmid and PAC DNA

DNA was prepared from 500 ml overnight cultures of PAC D2/ containing the mouse *Tsga10* gene using a Qiagen-tip 500 and following the manufacturer's protocol for preparation of large-construct DNA without removal of genomic DNA. The same Qiagen-tip 500 kit was used for plasmid preparation from similar overnight cultures of vectors to be used in subcloning experiments, sequencing and/or yeast transformation.

2.6 Digestion with restriction enzymes

All restriction digest experiments were carried out at 37°C for a minimum of 2 hours and sometimes as long as 10 hours. Restriction digests were set up using the following method for a 25µl reaction:

10x buffer	2.5µl
Restriction enzyme 1	30U
Restriction enzyme 2	30U
Purified PCR product	6µl (1-10 ug)
ddH_2O	14µl

The digested DNA was run on a 1% TAE agarose gel and the required DNA was purified by using the QIAquick gel extraction kit (Qiagen). Sometimes DNA was purified on a PCR purification column (Qiagen).

2.7 Ethanol Precipitation

i) 20µl PCR product for each sample was mixed with 2.5 µl 5M NaCl and 100µl of 100% ethanol in 1.5 ml tubes.

ii) The tubes of solution were then placed in the –70° C freezer for 30 min.

iii) The tubes were then centrifuged at 14000 rpm for 10 minutes, and the supernatant discarded

iv) The pellets were resuspended in 150µl of 70 % ethanol, centrifuged again at 14000 for 10 minutes, and once again the supernatant was discarded.

v) The remaining purified DNA pellets were then resuspended in approx. 20-30 µl of distilled water.

2.8 Polymerase chain reaction (PCR)

When the purpose of PCR was to generate template for sequencing part of a gene to be used to search the htgs database, or to screen bacterial colonies, Taq DNA polymerase was used. When PCR was used to amplify different fragments to be inserted into a construct, *Pwo* DNA polymerase which is a high fidelity DNA polymerase due to it's 3' to 5' exonuclease (proofreading) activity, was used (Roche). In both cases PCR was performed using a 25µl reaction and a *hot start* protocol. Whenever Taq was used as the polymerase the protocol was as below:

Reaction mixture:	Taq buffer (10 x)	2.5 µl
	dNTPs (2mM)	2.5 µl
	Taq DNA polymerase (5 U/µl)	0.1 µl
	DMSO	1.25 µl
	Downstream primer (10 pmol/µl)	1.25 µl
	Upstream primer (10 pmol/µl)	1.25 µl
	Template DNA (200ng/µl or a bacterial colony)	1.0 µl
	ddH_2O	15.15 µl
	Total	25 µl

Hot start with:	94°C	3 min
Then 30 cycles of:	94°C	30 seconds
	50-69°C	30 seconds (Ta of primers)
	72°C	1- 3 min (dependent on fragment length)
Then one cycle of:	68°C	7 min

When *Pwo* DNA polymerase was used, 2µl of each primer (10 pmol/µl) was used to give a 500 nM concentration of each primer. Otherwise, the protocol was as for Taq DNA polymerase.

2.9 Oligonucleotides

Oligonucleotides for the PCR reactions were designed using the programme Primer3 at http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi. All oligonucleotides were ordered from Invitrogen (Table 2.9). All of the nucleotide locations given in the table and in this thesis are based on the sequence of mouse *Tsga10* which I submitted to the GenBank database (AF530050). The sequence is numbered from the first of the two-methionine residues at the 5' end.

2.10 Electrophoresis protocols

Sufficient electrophoresis-grade agarose (Gibco BRL), to produce a 1% gel was heated in an appropriate volume of TAE or TBE. Once cast and set, the gel was immersed in the same buffer in a gel tank and run at approximately 4-4.5 V/cm voltage gradient. Four sets of markers were used on agarose gels depending on the relative sizes of the experimental DNA to be determined. These were λ DNA/Hind III fragments and 1kb ladder (Gibco BRL) (range 23130-564 bp), 500 bp ladder (Gibco BRL) (range 8000-500 bp), and 100 bp ladder (Gibco BRL) (range 2075-100 bp). In each case 0.5 to 1.5 μg was used depending on the size of the well in the gel.

Table 2.9: Oligonucleotide sequences

Oligonucleotide	Sequence (5' → 3') # Nucleotide AF530050)
1. MHF4'	CTG GAT TCC TTT GTC AAG ACT (459-479)
2. MR4	ATG GGC GTC TCG GAC TCT TAG A (570-549)
3. MF5'	GCT GCA AAC TGT GCT GTA GA (585-604)
4. MR5'	AAG GAG GAA GGT CTT GTC T (734-716)
5. MF6	GCT CGA CTT CGA CGA GAA AT (759-778)
6. MF19	TGC ACA GGA CCT GGA GTG TA (2606-2625)
7. MR19	TGA AAA CCC TTT ATT TTT GGT G (2717-2696)
8. MR6'	GCA TGT GCT GTG GTT GAC TT (819-800)
9. RpGEX	AGA GCC TCG AGA TTC TGC ATG *(XhoI)* (2242-2222)
10. FpGEX	ATT TAG AAT TCA TGA TGA GAA ATA GAT CTA AG *(EcoRI)* (535-566)
11. MF4-TC10clone	ATT CAA GCT TAA TGA TGA GAA ATA GAT CTA AG *(Hind III)* (534-554)
12. MR19-TC10clone	TCG AAT TCT CAA ATC TCA CTG TGA ACA TG *(EcoRI)* (2650-2629)
13. MIF4	GCC AGC TCT GGC TTC ATA GT
14. MG4	GTT GCC TTT TGG TAA TGT C (2433-2415)
15. MR12	GGT TAA CAC CAA AGA ACA TCA G (1-22)
16. MR5	GCA TAC ACT TTA GCT CTT CAC G (650-629)
17. MR6	GCT TTT CAT CAT TTC TCG TCG (788-768)
18. MF5	GTC AAG GTT CTC ACA TCT GAG A (693-714)
19. MF4	ATC TAA GAG TCC GAG ACG CCC ATC (548-571)
20. M32880F	CGG CTA CAC AGG CAG TTA CA

Table 2.9: Cont'd

Oligonucleotide	Sequences 5' → 3'	# Nucleotide (AF530050)
21. M32880R	AGT GAG TCT GGC GTT GGT CT	
22. GAL4AD	TAC CAC TAC AAT GGA TG	
23. pGBKT7-5'	TCA TCG GAA GAG AGT AGT	
24. pGBKT7-3'	GTC ACT TTA AAA TTT GTA T	
25. pGAD424-5'	CTA TTC GAT GAT GAA GAT ACC CCA CCA AAC CC	
26. pGAD424-3'	AGT GAA CTT GCG GGG TTT TTC AGT ATC TAC GA	
27. T3Odf2	ATC GAA ATT AAC CCT CAC TAA AGG CAG GCA ATG AGG TGG ATT CT *(T3 Promoter)*	
28. T7Odf2	ATC GAT AAT ACG ACT CAC TAT AGG ACC ATG TCT GCC TCA TCC TC *(T7 Promoter)*	
29. T7Mtc10	ATC GAT AAT ACG ACT CAC TAT AGG ACT CAG CGA CAT CTT GCT *(T7 Promoter)*	(1182-1199)
30. T3Mtc10	ATC GAA ATT AAC CCT CAC TAA AGG AGA CTT TTT ATC CAG ACA *(T3 Promoter)*	(1358-1341)
31. YTHM	CAC TCT CCC ATG GGA AGC TTA ATG ATG AGA AAT *(NcoI)*	(527-546)
32. MFE17	GAG ACG TGG CAC AGT TCA GA	(2365-2384)
33. MRE18	GCT CTG GTG AAT GGC ATT TT	(2550-2531)
34. R6H	TCA TTG CTG AGA TGG AAC AG	(1444-1463)
35. F-HIF1a-rat	CTG TGG GGT TTC GTT TCT GT	
36. R-HIF1a-rat	TGT TTT CTT TTG AGC CGG TAA	
37. FLJ32880F	CGG CTA CAC AGG CAG TTA CA	
38. FLJ32880R	AGT GAG TCT GGC GTT GGT CT	

2.11 RNA In Situ Hybridization

2.11.1 Pretreatment of slides

Slides were incubated for 30 seconds in coplin jars containing 2% APES (3-aminopropyltriethoxy-silane) in methanol and 9% methanolone in dH2O.

2.11.2 Preparation of tissue slides

To prepare mouse tissue, a male from an outbred transgenic strain (CBAxC51Bl/6) was selected for dissection. The mouse was the result of back crossing (strain carrying the smalleye mutation) for several generations. Organs were dissected and fixed in 4% paraformaldehyde for 6-18 hours and then in 20% sucrose in PBS overnight.

2.11.3 Preparing Cryostat Sections

A 7-ml bijou was used to mould aluminium foil into an embedding chamber wrapped (3-4 times) round the bottom, trimmed and then filled with OCT (BDH) at room temperature. A fixed mouse testis was placed onto a plastic petri dish to allow most of the sucrose solution to drain off and then it was dropped onto the surface of the OCT and allowed to sink. The chamber was put onto a bed of dry ice to be frozen. The content of the block was then placed onto a drop or two of OCT on the chuck of the microtome and was frozen in place using the bed of dry ice. Cutting was started at 30 micrometres per section until the tissue was reached then was continued at 10-12 micrometres per section. TESPA (Triethoxysilylpropylamine)-coated slides were pressed onto the leading edge of a section which was lifted onto the slide. When all sections were cut, the slides were dried for 1-2 hours in a fume hood. Slides were then placed in a light-tight plastic box which was sealed with tape and stored at -70°C.

2.11.4 Preparation of labeled RNA probe

A program via www.genebee.msu.su/services/rna2-reached.html was used to predict the secondary structure of the mouse *Tsga10* mRNA (figure 1.16). A 195-nucleotide DNA sequence was identified which was unlikely to form any stable secondary structure and oligonucleotides were designed to amplify this region from cDNA (figure 2.11.4). A T7 and T3 promoter sequence was added to the 5' end of the forward (T7Mtc10 5'**ATC GAT AAT ACG ACT CAC TAT AGG** ACT CAG CGA CAT CTT GCT3') and reverse (T3Mtc10 5'**ATC GAA ATT AAC CCT CAC TAA AGG** AGA CTT TTT ATC CAG ACA3') primers respectively. 1 µg DNA of PCR product was used after purification (with Qiagen PCR purification kit) as the template for RNA synthesis using either T7 or T3 RNA polymerase. The RNA probe (riboprobe) was prepared by incubating at 37°C for 3 hours with Digoxigenin conjugated UTP (Amersham) using the T7 labeling kit (Ambion) according to the manufacturer's instruction. The Digoxigenin labeled riboprobe was purified by ethanol precipitation according to the manufacture instruction. The riboprobe in DEPC treated water, was collected in a 0.5 ml tube and treated with Dnase to remove the template clone DNA. 2µl was then taken for running in an agarose gel while the rest (16µl) was stored at -20°C until required.

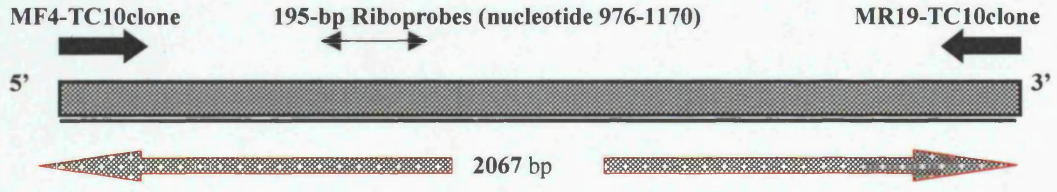


Figure 2.11.4- Murine *Tsga10* cDNA (ORF) and location of different primers used to amplify cDNA fragments to be used as probe in ISH, northern blotting or for screening of bacterial colonies and plasmids.

2.11.5 (In Situ) Hybridization

The compositions of solutions are described in the Appendix. Tissue sections, sections were thawed while the probe was diluted (~200-1000 fold dilution) in hybridisation buffer and denatured at 70°C for 5 minutes. 100 µl of hybridization mix (recipe in appendix) was added to each slide, which was covered with a coverslip. The slides were hybridised overnight at 65°C in a sealed box (with Kimwipes on the base moistened with 1× salts/50% formamide, two wooden sticks were placed inside the box to support the slides).

The slides were then transferred to a coplin jar and placed in washing solution (recipe in appendix) preheated to 65°C for one hour, and the coverslips were allowed to fall off. The slides were washed two further times in washing solution preheated to 65°C for 30 minutes each. Then they were washed in MABT (100 mM Maleic acid, 150 mM NaCl, 0.1% Tween-20) for 2 × 30 min at room temperature and blocked in M-Block (MABT + 2% Blocking Reagent + 20% heat-inactivated serum) for at least one hour at room temperature (40µl of M-Block per section) without a coverslip. After dilution of anti-DIG antibody in M-Block (1:1500) 75 µl was placed on the section and a coverslip was placed over it. The slide was then incubated in the humidified chamber overnight at 4°C.

2.11.6 Signal Detection

The complex of DIG-labelled riboprobe and anti-DIG antibody was detected by using the alkaline phosphatase method. Positive signals were developed by using nitro-blue tetrazolium chloride (NBT) / 5-bromo-4-chloro-3'-indolyphosphate p-toluidine salt (BCIP) as alkaline reaction detector. Next day, coverslips were removed and the sections were washed 4-5 times at room temperature in MABT, for 20 min total and then washed in staining buffer (recipe in the appendix) for 2 × 10 min.
The slides then transferred into a coplin jar wrapped in foil full (80 mls) of staining buffer plus NBT/BCIP. (40ml staining buffer, 40ml 20% PVA, 360µl NBT, 280µl BCIP). The staining reaction was stopped with 2 washes in PBT (PBS + 0.1% Tween 20), dehydrated in 100% ethanol (x2), then histocleared (x2) in histoclear solution and mounted in DPX.

2.12 Sequencing

Templates for DNA sequencing were given to the DNA sequencing department, Wolfson Institute, University College London (UCL), who carried out all sequencing reactions using an ABI PRISM® 3100 Genetic Analyzer sequencing machine.

2.13 Quantification of DNA

To determine the concentration of DNA, 5 or 10 µl of sample was added with dH_2O to a final volume of 1 ml and the absorbance at 260 nm was measured. The purity of the sample was determined by measuring the absorbance at 260 nm/280 nm.

2.14 Calcium Phosphate-mediated DNA Transfection of cells

2.14.1 DNA preparation

The desired amount of DNA was ethanol precipitated, washed in 70% ethanol, dried and resuspended in sterile water or TE at a concentration of 1-3.5 mg/ml.

2.14.2 Cells

The following cell lines were used into study: HeLa, NIH 3T3 mouse fibroblast and hamster fibroblast (64063a12) hybrids containing chromosome 9 (a gift from Carol Jones) were used.

2.14.3 Preparing the calcium phosphate precipitate

The calcium phosphate co-precipitation method (Promega) was used for the transfection of adherent cells. Cells were seeded on 10 cm dishes 1 day prior to the transfection. For stable transfection 1×10^6 cells were plated on each 10 cm dish. For transient transfection anywhere from 1 to 5×10^5 cells are plated. The confluence of the cells was 50-90% on the day of transfection. 0.5-3 hours prior to transfection the medium was renewed and the transfection was carried out in the following order:

- In one tube (tube A) 5-10μg of plasmid was mixed with sterile distilled water to a total volume of 240 μl, and 31μl of 2M calcium phosphate solution was added.
- 240μl (i.e., the same volume as tube A) of 2 X HBS (1.5 mM $NaHPO_4$, 280

mM NaCl, 50 mM HEPES with a pH of 7.05-7.12) was prepared in another tube (tube B).

- Solution A was added dropwise into tube B while tube B was slowly vortexed. The precipitated particles appeared while the solutions were mixed. The quick addition of solution A can result in poor precipitation.
- The transfection solution was left at room temperature for twenty minutes, gently mixed, and then added dropwise into the cell culture plate.
- The plate was moved gently back and forth to distribute the transfection solution.
- The plate was incubated at 37°C with 5% CO_2 for 2-12 hours.
- Calcium phosphate containing medium was removed from the plate and the cells were washed with fresh medium.
- 48-72 hours after transfection the selection drug was added to the cultures and the cells were incubated for selection by the drug.

2.14.4 Versin/trypsin treatment of the cells

This step enhances nuclear uptake of the introduced DNA (Chu and Sharp, 1981). The medium was aspirated and each dish was washed 2 times with 5 ml of PBS to remove most of the precipitates. Typically, cells were maintained in nonselective medium for 1-2 days post-transfection, and then trypsinized by removing the final wash and adding enough trypsin/versin solution to cover the cell monolayer. Dishes were swirled whilst being washed, and were then checked under the microscope to see whether the cells had become detached from the dish.

2.14.5 Transient assay and stable transfection

Cells were harvested, stained or analyzed 48 to 60 hours after the precipitate was placed on the cells. For stable transfection selective medium (G418) was placed on the cells one day after the transfection and cells were fed with fresh selective medium every 3 days.

2.15 Northern Blot

2.15.1 Radiolabelling DNA probes by random priming

DNA to be used as probes was labelled by random priming (Feinberg and Vogelstein; 1984) using the rediprime DNA labelling system (Amersham Pharmacia Biotech). This system contains a dried, stable labelling mix of dATP, dGTP, dTTP, Klenow enzyme and random primers (9mers). Approximately 25ng of DNA was 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-32P]dCTP (10mCi/mL, 3000 $Ci.mmol^{-1}$). After incubation at 37°C for 20-60 minutes, unincorporated [α-32P]dCTP and primers were removed by passing the labelling reaction through a Sephadex G-50 column.

2.15.2 Northern Blotting

Blots were prepared as described in Maniatis containing 10µg RNA per lane. Probes were prepared as described above.

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 2×SSC and 0.05% SDS for 30' at room temperature and once in 0.1×SSC and 0.1% SDS for 40' at 55°C. The blot was exposed to X-ray film (Kodak-Biomax-MR) for 36 hours.

2.16 Raising Antibodies against Tsga10 and FLJ32880 proteins

2.16.1 Peptide design strategy

The amino acid sequences of mouse Tsga10 protein and the mouse homologue of human hypothetical FLJ32880 protein were analyzed. After considering important factors for selecting peptides (described in section 3.4.1), the antibodies were raised against them (see following section).

In addition, I hoped to raise antibodies which would recognize both the rodent and human homologues of this protein. This meant that the immunising peptides had to be conserved, which was not necessarily the best strategy for raising an antibody in another mammal (rabbit). In the Tsga10 protein sequence, the antigen prediction file generated by the program "Peptide Structure (GCG)" shows that it is not the easiest protein to work with due to the lack of proline residues apart from the N-terminus region. However, the N terminus looked as though it is completely exposed and was worth considering. (Provided there is no reason to think it is chopped off post translationally.) Regions containing Proline and Arginine, and some hydrophobic amino acids are likely to be antigenic.

2.16.2 Immunisation program

Eurogentec (Belgium) generated peptide polyclonal antibodies to the selected peptide sequences using multiple antigenic peptides (MAP method) in the following immunization procedures. The peptides were synthesized and conjugated to a carrier protein through the amino- or carboxy terminus depending on the position of the peptide within the native protein and the antigenicity profile. Two rabbits were used for each immunization. A proprietary immunization protocol (table 2.16.2), which allowed the production of 70-80 ml serum per rabbit within 80 days followed. One-month extension was requested based on the ELISA result for Tsga10 antisera. Both animals came from the same high antibody responding breed and were subjected to strict veterinarian controls. Finally both animals were bled.

Immunization schedule							
Day	0	14	28	38	56	66	80
Injection	1st	2nd	3rd		4th		
Bleed	pre-immune			2 ml		2+20 ml	50-60 ml

Table 2.16.2 - Immunization Protocol: two rabbits - 80 days

2.16.3 Antibody titre by enzyme-linked immunosorbent assay (ELISA)

The company, which made the antisera, also carried out ELISA results to assay the titre of the antibody using the following procedures:

1. Making conjugate with horseradish peroxidase: first, 2 mg Horseradish peroxidase (HRP) was dissolved in 1 ml water (solution A) and 21.4 mg NaIO in 1 ml water solution (B). 100 µl of solution B was added to solution A (color changed to dark

green), and left for 10 min at room temperature. The reaction was then put into the dialysis tube (such as Molecular cut off 20,000). At this stage the tube was put into 5 mM NaAcetate buffer, pH 4.0 in a 2 to 3 *l* flask and dialysed overnight (color changed to gold). To raise the pH of the HRP solution to pH9.0, 0.2 M NaCarbonate buffer, pH 9.5 was added and the pH was monitored using a universal pH paper.

2. Antibody purification by protein G: the antibody solution was mixed (8 mg of IgG in 1 ml), which had been pre-dialyzed to 0.01 M NaCarbonate buffer, pH 9.0 overnight, and the mixture was incubated for 2 hr at room temperature. At this stage, 100 µl of freshly prepared, 0.1 M NaHBr4 in water was added to the solution, and this was incubated at 4°C for 2 hr. The mixture was then put into a dialysis tube and dialyzed against PBS overnight.

3. 96-well plate: The solid phase reaction was was carried out in a 96-well plate (Corning' Nunc').

4. Buffers and other reagents consist of:

a) Plate buffer: 0.1 M Sodium carbonate buffer, pH 9.5,

b) Reaction buffer: 0.01 M Sodium phosphate buffer, pH 7.2, 0.15 M NaCl (PBS), 0.5% BSA,

c) Washing buffer: 0.05% Tween-20, 0.01 M Sodium,

d) phosphate buffer, pH 7.2 or 0.05% Tween-20, 0.15 M NaCl,

e) Developing buffer: 0.05 M Sodium acetate buffer, pH 5.5,

f) TMB stock solution: Tetramethylbenzidine 1 mg/ml in DMSO.

2.17 Immunohistochemistry (IHC) and Immunocytochemistry

2.17.1 Preparing slides

8-12 microns (μ) thick were cut from paraffin embedded tissues and placed on the positive charged SuperFrost glass slides (BDH). Then they were baked for half an hour at 55°C to fix them to the slides to prevent the tissue from detaching during immunohistochemistry. The slides were then stored in a cold room at 4°C until immunohistochemistry was performed.

2.17.2 Tissue fixation and processing

Sections of mouse, rat, boar and pig testes, mouse brain, various developmental stages of mouse embryos and human germ cell tumor were selected for immunohistochemical analysis. All human cells and tumour sections were obtained from patients who had given informed consent for their use in research.

Sections were prepared from mouse embryos from 10.5-dpc (days post coital) (E10) to newborn (19.5-dpc). The mothers of the embryos were killed and then the embryos were dissected out and kept in 1% Phosphate Buffered Saline (PBS) for less than 1 hour. Embryos were processed for sectioning and paraffin embedding by being placed in 4% Paraformaldehyde, made with 1% PBS, and were left overnight in the cold room. The larger embryos of 17.5 and 18.5dpc were injected as well as being immersed so that all the tissue was exposed to the preservative.

Alternatively for the fixation, Carnoy's or Bouine's solutions were used. The recipe for Carnoy's fixative is:

Ethyl alcohol (100%)	70ml
Chloroform	30ml
Acetic acid	10ml

Again the larger embryos were injected with Cornoy's and left overnight. The following morning they were washed and stored in 70% ethanol.

Individual embryos were embedded in paraffin using the standard NIH method (http://cgap-mf.nih.gov).

2.17.3 IHC procedure and DAB staining

The first stage of the immunohistochemistry protocol involved deparaffination and hydration of the slides, so that the tissue on the slides would be effectively treated in the later stages of the protocol. The slides were bathed in xylene for twenty minutes, changing the xylene after the first ten minutes. The xylene dissolves the paraffin. The slides were then bathed in ethanol at 100%, 90% and 70% concentrations respectively, each for ten minutes. Antigen retrieval is then performed to combat post translational modifications of TSGA10, in particular glycosylation so that the antibodies can recognise the antigen. Antigen retrieval involves 'baking' the slides in 0.01 M sodium citrate buffer (pH=6). The slides were 'baked' in the microwave for five minutes on each of high (900 W), medium (600 W) and low (300 W) settings consecutively. It was necessary to ensure that the citrate buffer did not dry out during the 'baking' so extra was added as needed so that the tissues were always covered. When the slides came out of the microwave they were immediately bathed in ice cold 1% Phosphate

Buffer Solution. The slides were then treated with 6% Hydrogen Peroxide in PBS for twenty to thirty minutes to remove endogenous peroxidase activity. Hydrogen Peroxide is light sensitive so the slides were kept in the dark while the reaction was occurring. The reaction was stopped by rinsing with PBS three times each for ten minutes, and then the slides were put in PBS on a moving tray for 3 minutes to ensure they were rinsed properly, this stops the Hydrogen Peroxide reaction. At this point in the protocol the slides were incubated in 3% Normal Goat Serum (NGS) overnight. NGS blocks all sites to which the primary and secondary antibodies might bind non-specifically later in the procedure. After blocking with NGS the slides were separated into three groups and specific peptides were added to block certain parts of the protein. In one group the N terminus is blocked, in the other the C terminus and the third group the whole protein is blocked, this is the negative control. The peptides were added to the NGS solution for one hour and the primary antibody was added at a concentration of 0.2% at 37°C for one and a half hours. The slides were then rinsed three times, each for ten minutes, in PBS with 0.1% 'Tween 20'. The tissue slides were then incubated in the secondary antibody, Goat Anti- Rabbit IgG conjugated to peroxidase (CHEMICON International, CA, USA). The secondary antibody recognises the primary antibody and the peroxidase is recognised by the staining kit. The secondary antibody was added at a dilution factor of 1:660 (1.5μg/ml) for thirty minutes. The slides were again rinsed in PBS three times each for ten minutes. Once the antigen (TSGA10) had been recognised specifically by the primary antibodies and the primary antibodies had been recognised by the secondary antibodies, it was necessary to stain the tissues in order to detect the labelled protein under the microscope. To do this I

used the Di-amino Benzidine (DAB) kit and so care had to be taken when using this. Using distilled water the DAB solution was prepared based on the kit (2 drops of buffer stock solution to 5ml of distilled water, then mixing, adding 4 drops of DAB stock solution, and then adding 2 drops of hydrogen peroxide solution and mixing well. The slides were lined up on a tray and the DAB was added to the slides for between 2-10 minutes until the tissue on the slides darkened. The slides were then rinsed by shaking off excess DAB and then rinsing in tap water to remove the remaining DAB. They were then incubated in 0.05M sodium bicarbonate for ten minutes. An enhancer was then used to amplify the DAB signal. The DAB enhancing solution was pipetted over the slides as they were lined up in a tray and then excess was shaken off after ten to twenty seconds and the slides were re-immersed in tap water to stop the reaction. The DAB kit and DAB enhancing solution were supplied by Vector Laboratories. The slides were again placed on a tray and Hematoxylin and Eosin (H&E) counterstain was added. This was left on the slides for thirty seconds then the excess was shaken off and the slides were submersed in tap water. The slides were then dehydrated by submersing them in ethanol at concentrations of 70%, 90% and 100% respectively each for thirty seconds. The slides were left to dry, and then DPX was used to fix the cover slips. The slides were then stored at room temperature until being analysed under the microscope. Using Di-amino Benzidine and Hematoxylin & Eosin (H&E) to stain and counterstain the slides, the immune signal will show as brown (stained by DAB) and the background will be blue (stained by H&E). Phalloidin conjugated to TRITC (Sigma) was also used (according to the manufacturer's instruction) to stain actin filaments.

2.17.4 Indirect Immunofluorescence Analysis

For immunostaining of mouse, bovine and boar sperm, immunocytochemistry was performed. Mature bovine and boar spermatozoa were collected from their epididymis and washed in PBS. Spermatozoa were pelleted and resuspended in demembranization buffer (5 mM 1,4-dithiothreitol [DTT], 50 mM Tris-HCl [pH 9.0], and 2% v/v Triton) followed by fixation in 4% formaldehyde. The cells were spread on cover slips treated with poly-L-lysine (Sigma) and incubated with primary antibody diluted 1:25 (v/v) in blocking solution for 1 h at 37°C and washed in PBS. Sperm were incubated for 1 h at 37°C with secondary antibody (FITC-conjugated goat anti-rabbit IgG; Jackson Immunoresearch Laboratory, Inc., West Grove, PA) diluted 1:100 (v/v) in blocking solution and washed in PBS before staining with DAPI (as described above). Cover slips were mounted on glass slides with Permount SP15-100 (Fisher Scientific, Edmonton, AB, Canada). Slides were viewed with a Zeiss confocal fluorescent microscope.

2.17.5 Taking photos

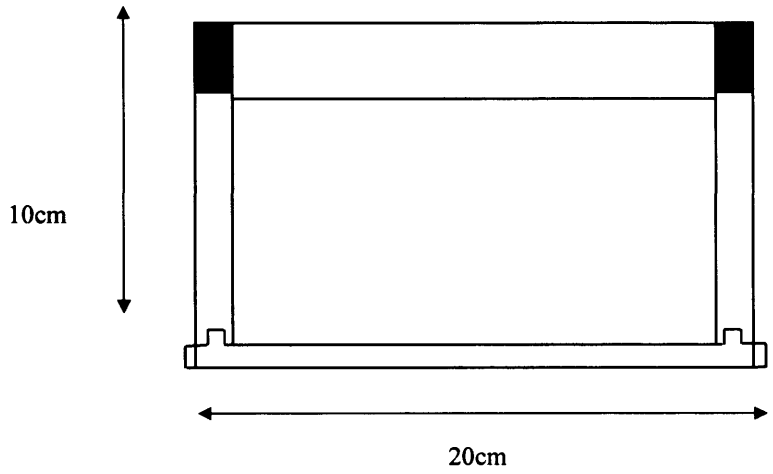
A Zeiss 100M light microscope was used to detect the signals. The photographs were taken with a Kodak MDS camera attached to the microscope, and coupled to a monitor was used to take pictures of the slides under the microscope. Normal Panasonic and Gold KODAK films were used and developing tools in Boots/Jessop photo shops were used for the photos developing.

2.18 Western Blot (SDS-PAGE method)

2.18.1. Setting up the gel.

1 large (19 x 20cm) and 1 small glass plates were used. The best side of each plate was chosen (ie. with no cracks or chips), cleaned with detergent, rinsed well with ddH_2O, dried and then wiped over with ethanol.

Both sides of the spacer gaskets (use the thicker ones for protein gels) were smeared with Vasoline and placed onto the large plate as shown below with the black rubber facing upwards.



The small plate was placed on top and secured at the bottom with bulldog clips.

The plates were placed into a tray, water was added to check for leaks, if OK this was emptied and the apparatus was dried with a paper towel.

2.18.2 Pouring the Gradient

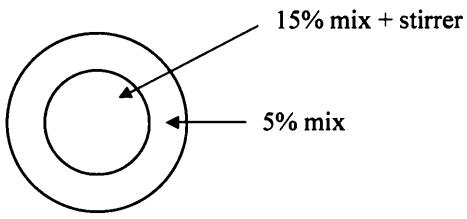
The gradient maker was cleaned and checked for blockages and a magnetic stirrer was placed into the middle compartment.

Acrylamide solutions were prepared as below.

	5%	**15%**
50% glycerol	0	3.6ml
dH_20	10.55ml	1.2ml
Acrylamide/Bis, 37.5:1, 30%T	2.79ml	8.36ml
1.5M Tris-HCl pH8.8	3.34ml	3.34ml
TEMED	15µl	15µl
10% SDS	200µl	200µl
10% w/v (APS)	50µl	25µl

The ammonium persulphate was added last (just before use), and mixed briefly.

Making sure the lever on the gradient maker was in the closed position, the 5% mix was poured into the outer ring of the gradient maker and the 15% mix into the inner ring.



The glass plates were placed inside a plastic tray, to contain any spills. The solution was pumped into the apparatus and covered with layers of water. Gels could be stored at 4°C.

2.18.3 Pouring the stacking gel

The apparatus was allowed to come to room temperature. The water was removed. The solution was prepared as below:

	3%
dH_2O	15.5ml
0.5M Tris-HCl pH6.8	2.5ml
Acrylamide/Bis, 37.5:1, 30%T	2.0ml
TEMED	20µl
10% SDS	200µl
10% w/v (ammonium persulphate)	200µl

The ammonium persulphate was added last as before, mixed gently then poured on top of the set gradient gel using a plastic pasteur pipette. When the gel reached the top of the small glass plate the spacer comb was added at an angle to avoid trapping air bubbles. The gel was left to set for about an hour.

2.18.4 Sample Preparation.

30µl of sample was mixed with 30µl of dissolving buffer. For markers I used 7µl of Rainbow marker (Amersham) mixed with 7µl of sample buffer.

Samples were boiled for 2 minutes and markers for 30 seconds before loading on the gel. 100-500µl culture (according to the condition of growth) was taken and centrifuged for 5min, its supernatant discarded and 50-100 µl 2 x treatment buffer added. In this step, the cells were ruptured by freezing and thawing three times using

liquid nitrogen and then sonication with high frequency for 1 minute at ice cold temperature. Just before SDS-PAGE and western blotting it was boiled for 5 min and centrifuged briefly.

2.18.5 Tank set up

Freshly boiled protein samples were electrophoresed through an SDS gel in a BioRad apparatus. The gels were electrophoresed at 180V 45mA until the blue dye reached the bottom of the gel (~ 8 hrs) and then continued to run for a further hour.

2.18.6 Western Blotting

Proteins were transferred from the SDS gel to the blots in transfer buffer at an initial current of 0.1-0.18A, 45-90V for 17 hours onto an Amersham polyvinylidene fluoride (PVDF) filter.

2.18.7 Antibody staining

The sandwich was removed from the Western blotting apparatus and the gel and the PVDF were placed together on the glass plate. The position of the gel was marked on the PVDF including any orientation marks.

The gel was discarded and the PVDF was placed into a glass tray containing blocking buffer (5% non-fat milk) and incubated at 4°C with shaking for a minimum of 6 hours.

The membrane was transferred to a plastic bag and antibody was added at an appropriate dilution (for Tsga10 antisera: 1:1000 in blocking solution). The bag was sealed and incubated overnight (or weekend) at 4°C with shaking.

The antibody solution was removed from the bag and the membrane was washed in a glass tray in PBS for 5× 5min at 4°C with shaking.

Then GOAT anti Rabbit IgG conjugated to peroxidase (diluted 1:4000 in PBS) was added to the membrane in the bag and incubated for 2 hours at 4°C with shaking. The membrane was removed from the bag and washed again in a glass tray in PBS for 5 × 5 minutes at 4°C with shaking.

For chemiluminescence signal detection, the ECL Plus Amersham kit was used by applying 2ml/10cm^2 detection solution and was left for a maximum of 5-6 minutes. Longer exposure led to high background.

2.19 Yeast two hybrid assay

2.19.1 Yeast strains, MATCHMAKER yeast two hybrid assay

The *Saccharomyces cerevisiae* yeast strain, AH109, was used for transformation or co-transformation with plasmid containing bait plasmids (*Tsga10*, either of the two halves of *Tsga10)* and *Saccharomyces cerevisiae* yeast strain, Y187, was used for the mating protocol.

Table 2.19.1- Bait strain - AH109 - characteristics

Genotype	Phenotype
MATa	Mating type a
trp1-901	Deletion mutation, cells cannot synthesise tryptophan, and are thus unable to grow on minimal media lacking it.
leu2-3,112	Double mutation, cells cannot synthesise leucine, and are thus unable to grow on minimal media lacking it.
ura3-52	Deletion mutation, preventing cells from synthesising uracil. The lacZ construct, containing the functional gene, is integrated into the yeast genome at this location, enabling selection of recombinants through growth on -ura media.
his3-200	Deletion mutation, cells cannot synthesise histidine, unless in the presence of GAL4. (see below)
gal4Δ	Deletion mutation
gal80Δ	Deletion mutation
LYS2::$GAL1_{UAS}$-$GAL1_{TATA}$-HIS3	Reporter gene construct. HIS3 reporter gene is activated when GAL4 binds to the corresponding upstream activating sequence, $GAL1_{UAS}$.
$GAL2_{UAS}$-$GAL2_{TATA}$-ADE2	Reporter gene construct. ADE2 reporter gene is activated when GAL4 binds to the corresponding upstream activating sequence, $GAL2_{UAS}$.
URA3::$MEL1_{UAS}$-$MEL1_{TATA}$-lacZ	Reporter gene construct. lacZ reporter gene is activated when GAL4 binds to the corresponding upstream activating sequence, $MEL1_{UAS}$.
MEL1	Gene encoding α-galactosidase, naturally activated by GAL4

Table 2.19.2- Library strain - Y187 characteristics

Genotype	Phenotype
MATα	Mating type α
ura3-52	(see Table 2.19.1)
his3-200	(see Table 2.19.1)
ade2-101	Deletion mutation, cells cannot synthesise adenine, and are thus unable to grow on minimal media
trp1-901	(see Table 2.19.1)
leu2-3,112	(see Table 2.19.1)
gal4Δ	(see Table 2.19.1)
gal80Δ	(see Table 2.19.1)
met-	Met- phenotype in this strain is unstable and was not used in these experiments.
URA3::$GAL1_{UAS}$-$GAL1_{TATA}$-lacZ	Reporter gene construct. lacZ reporter gene is activated when GAL4 binds to the corresponding upstream activating sequence, $GAL1_{UAS}$.
MEL1	(see Table 2.19.1)

2.19.2 Bait and prey domains

Bait plasmid - pGBKT7-Tsga10: Mouse Tsga10 cDNA was cloned into the multiple cloning site of pGBKT7, between the *NcoI* and *EcoRI* restriction sites.

This plasmid encodes the GAL4 DNA binding domain (GAL4 DNA-BD) upstream and in frame with the multiple cloning site. Thus when Tsga10 was inserted into the multiple cloning site, a fusion protein of Tsga10-GAL4 DNA-BD, i.e. the bait protein is encoded by the plasmid. pGBKT7 contains both *S. cerevisiae* and *E. coli* origins of replication, and thus can be transformed into both. It also contains selectable markers for both the kanamycin resistance gene for selection in *E. coli* and the metabolic marker gene *TRP1* for selection in *S. cerevisiae*.

Also the 5′ and 3′ halves of *Tsga10* were amplified and separately cloned into plasmid vectors. Figure 2.19.2 shows the schematic of the primers that were designed for the PCR reaction to make Tsga10-N and Tsga10-C constructs as bait plasmids to see their possible interactions:

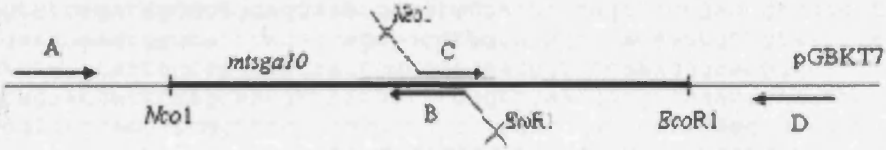


Figure 2.19.2A- A Primer design to make Tsga10-N,C halves bait construct (not to scale)

The two halves of the mouse *Tsga10* gene were amplified by PCR using primers A and B, and C and D (figure 2.19.2, A+B). Once suitable primers had been picked, using Primer3, restriction sites were added to the beginnings of primers B and C. An *EcoR1* site was added to primer B and an *Nco1* site was added to primer C. Two random nucleotides were also added to the beginning of the *EcoR1* restriction site and

Tctactgatttttcctcgagaagaccttgacatgattttgaaaatggattctttacaggatataaaagc
attgttaacaggattatttgtacaagataatgtgaataaagatgccgtcacagatagattggcttcagt
ggagactgatatgcctctaacattgagacagcatagaataagtgcgacatcatcatcggaagagagtag
taacaaaggtcaaagacagttgactgtatcgccggaatttgtaatacgactcactatagggcgagccgc
catcatggaggagcagaagctgatctcagaggaggacctgcatatggcactct
cactctccatgg(gaagctt)atgatgagaaat
atgatgagaaatagatctaaga
gtccgagacgcccatcgccaacttcccgggctgcaaactgtgctgtagagcttttgaagtcaactgcaa
gagaccgtgaagagctaaagtgtatgctggaaaaatatgaacgtcatttggcagaaatccagggcaacg
tcaaggttctcacatctgagagagacaagaccttcctcctttatgagcaggcacaggaagaaattgctc
gacttcgacgagaaatgatgaaaagctgtaagtctcctaagtcaaccacagcacatgctattcttcgtc
gggtagagacggagagagatgtagccttcactgacctacgaagaatgaccacagagcgagacagtctga
gagaaaggctcaagattgctcaagagacagcgttcaacgagaaggctcacttggaacagcggatagagg
agctggagtgcacagttcacaaccttgatgatgagcgcatggaacagatggcaaacatgactttgatga
aggaaaccataaccactgtggaaaaagaaatgaaatcattagcaagaaaggcaatggacaccgagagtg
agcttggcagacaaaaagcagagaataattctttgagacttttatatgaaaacacagaaaaagatcttt
ctgatactcagcgacatcttgctaagaaaaaatatgagctacagcttactcaggagaaaattatgtgct
tggatgaaaaaattgataatttcacgaggcaaaatattgcacagcgagaagaaatcagcattcttggtg
caaccctcaatgacctggctaaagaaaaggaatgcctgcaagcgtgtctggataaaaagtctgagaaca
ttgcatcccttggcgagagcttggcaatgaaagaaaaaaccatttcaggcatgaagaatatcattgctg
agatggaacaggcatcaagacagtctactgaagccctcattatgtgcgaacaagatatttccaggatgc
gcgatgcgccggcagttggacgagacaaatgatgagctgggtcagattgccagggagagagatat
gattccatgggcagttggacgagacaa
at gtcagattgccagggagagaga(attcca)

tatcttggctcatcagaataccaatcttcaagaacagtttgccaaagtcaaacaagaaaaccaggcact
gtccaaaaaactgaatgatactcatactgaactcagtgacataaagcagaaggtccaggacacgaatct
ggaggttaacaagttgaagaacatattaaagtttgaggaatctgagaaccggcaaataatggaacaact
ccgaaaagccaatgaagatgctgaaaactgggaaaataaggcccgccagacagaggcagaaaacaacac
cctcaaactggaacttatcactgctgaagcagagggcaacagattaaaggaaaaagttgatgcccttaa
cagagaggttgagcaacacctaaacgcagagcgctcttacaaatcccagattgcaactttacacaagtc
tcttgtgaagatggaggaggagcttcagaaggttcagtttgagaaggtgtctgctctcgcagatttgtc
ttccacaagggaactctgcataaaactcgactcaagcaaagaacttcttaatcgacagctggttgccaa
agatcaggaaatagaaatgatggagaatgagctggactcagcgcgctctgaaattgaactgctccggag
tcagatgacaaacgagaggatctccatgcagaatctcgaggctctgctggtggccaaccgggacaaaga
gtaccagtctcagatagcactgcaggaaaaggagtctgagatccagttgctgaaggagcacctctgcct
ggctgagaacaaaatggccatccagagcagagacgtggcacagttcagaaatgttgtaacgcagttaga
agcagatttggacattaccaaaaggcaactcgggaagaacgttttgaaagggaaagggctgtccaactt
cgccgccagaattattcaagcaatgcttataatttcggtccaatggagccaaatacaaaatgccattca
ccagagcgtgttcaccatcggtttcctgaccgaggcttcgaccgatcattggaagagtaagtgtgcaca
ggacctggagtgtagc*catgttcacagtgagatttgag*gaagcgaggaaactgatactaaacctggtta
ataagacatgatgacaccgaattcccggggatccgtcgacctgcagcggccgcataactagcataaccc
cttggggcctctaaacgggtcttgaggggttttttgcgcgcttgcagccaagctaattccgggcgaatt
tcttatgatttatgatttttattattaaataagttataaaaaaaataagtgtatacaaattttaaagtg
actcttaggttttaaaacgaaaattcttattcttgagtaactctttcctgtaggtcaggttgctttctc
aggtatagcatgaggtcgctcttattgaccacacctctaccggcatgcaagcttggcgtaatcatggtc
atagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaa
gtgtaaagcctggggtgcctaatgagtgaggtaactcacattaatt

Figure 2.19.2B- Partial sequence of pGBKT7 plasmid with *Tsga10* insert (1st and 2nd portions). Key: Plasmid sequence, *Tsga10* sequence, Primers A and B, Primers C and D Restriction sites for *Nco*1 and *Eco*R1

four random nucleotides were added before the *Nco*1 site. These extra nucleotides enabled the enzymes to bind to the sequences and cleave them. This cleavage was necessary to create complementary ends and enable the fragments to be ligated into fresh plasmid. The positions of the *NcoI* site in primer C was adjusted such that the junction would be in the correct reading frame for translation of the second half of the sequence to produce a functional protein.

The sequences of the primers are shown below:

A (YTHM)	CACTCTCCCATGGGAAGCTTAATGATGAGAAAT *(NcoI)*
B	TGGAATTCTCTCTCCCTGGCAATCTGAC *(EcoRI)*
C	GATTCCATGGGCAGTTGGACGAGACAAAT *(NcoI)*
D (MR19TC10clone)	TCGAATTCTCAAATCTCACTGTGAACATG *(EcoR I)*

Prey plasmid - pGAD424: This plasmid encodes the GAL4 activating domain (GAL-AD) upstream and in frame with the multiple cloning site. When rat testis cDNA library was inserted into the multiple cloning site, one third of the clones coded for a fusion protein of testis protein-GAL4-AD. pGAD242 contains both *S. cerevisiae* and *E. coli* origins of replication, and thus can be transformed into both. It also contains selectable markers for both the *bla* ampicillin resistance gene for selection in *E. coli* and the metabolic marker gene *LEU2* for selection in *S. cerevisiae*.

2.19.3 The rat cDNA Library

Since the protein was known to be testis specific, proteins that interact with it must be expressed in the testes. Thus a testis cDNA library was screened using the yeast two-hybrid system. So Tsga10 protein interaction was studied by rat testis library screening pretransformed in pGAD424 vector (Clontech). Approximately 200,000 rat testis cDNA independent clones were cloned into the multiple cloning sites of pGAD242, between the *Eco*RI and *Sal*I restriction sites. The library was a gift of Professor Frans Van der Hoorn, University of Calgary, Alberta, Canada.

2.19.4 Yeast transformation

The protocol outlined below, based on the Clontech MATCHMAKER Two-Hybrid System, was used for the yeast co-transformations. The protocol was scaled up four times so that enough yeast cells could be plated onto the 32 plates required for the experiment. For yeast transformation, 2-3 yeast colonies (2-3mm in diameter) were added to 4ml YPD medium, vortexed to mix and incubated at 30°C overnight with shaking at 250rpm. Then the culture was transferred into 30ml YPD medium, incubated at 30°C for 3hours with shaking at 250rpm and centrifuged for 5min at 2000 × g. After discarding the supernatant, the DNA pellet was resuspended in 20ml of water and vortexed to mix. After centrifuging for 5min at 2000 × g and discarding the supernatant again, the DNA was pelleted in 15ml of 1xTE/LiAc, centrifuged for 5min at 2000 × g, and the supernatant discarded. Then the DNA pellet was resuspended in 1ml of 1×TE/LiAc and the following were added to each tube: 100ng DNA-binding domain plasmid, 100ng activation domain plasmid (or library plasmid), 5μg Herring

testes carrier DNA, 0.1ml yeast competent cells. Then after vortexing, 0.6ml of PEG/LiAc solution (0.5ml 10xTE, 0.5ml 10xLiAc, 4ml PEG4000) was added, vortexed to mix, and incubated at 30°C for 45min with shaking at 200rpm. One tenth of the volume of the transformation of dimethylsulphoxide (DMSO) was added to each tube and gently inverted to mix before heat shock for 15min in a 42°C water bath. Then it was chilled on ice for 1-2min, centrifuged for 15sec at 14000rpm and the supernatant was discarded. Finally, the cell pellet was resuspended in 0.5ml TE or water just before plating out onto SD medium (yeast nitrogen base without amino acids, agar, water, 40% glucose, and with the appropriate dropout supplement) containing amino acids lacking either 'Leu, Trp' or 'Leu, Trp, His, Ade'.

250µl of each transformation was plated out and incubated at 30°C for two days.

To estimate transformation efficiency, yeast colonies were counted in the control plate by using the formula:

$$\text{Transformation efficiency} = \frac{\text{no. colonies x total suspension volume(µl)}}{\text{Volume plated(µl) x dilution factor x amount DNA(µg)}}$$

2.19.5 PCR of colonies

Following the transformation, samples were taken from four yeast colonies on each SD/-Trp plate. The samples were tested in a PCR reaction to see if the plasmids contained the inserted DNA and would therefore be suitable for the yeast mating. The first denaturing stage of the PCR reaction was increased from 3min to 5min to allow the DNA to be released from the yeast cells. The PCR products were electrophoresed on an agarose gel.

2.19.6 Yeast Mating

Yeast mating is a convenient method of introducing two different plasmids into the same host cells, and, in some applications, can be used as a convenient alternative to yeast co-transformations (Bendixen et al., 1994; Harper et al., 1993; Finley & Brent, 1994).

The yeast mating was carried out to create cells that contained both bait and library plasmids. Both bait and library strains are grown up in an overnight culture in SD medium. The two strains are then combined, YPDA/Kan is added, and the suspension is incubated overnight at 30°C, with gentle swirling. As the two strains are of different mating types (see above), they mate, producing diploid cells in which both plasmids are found. The cells are then plated out on double drop out media, so that only cells containing both plasmids can grow.

The following small-scale protocol worked well for creating diploids by yeast mating.

a. One colony of each type was picked to use in the mating. Only large (2–3-mm), fresh (<2-months old) colonies from the working stock plates were used.

b. Both colonies were placed in one 1.5-ml microcentrifuge tube containing 0.5 ml of YPD medium and tubes were vortexed to resuspend the cells completely.

c. Tubes were incubated at 30°C overnight (20–24 hr) with shaking at 200 rpm.

d. 100 µl aliquots of the mating culture were spread on the appropriate SD minimal media. Double dropout (-Leu/-Trp) was used to select for both plasmids (mating efficiency) and triple dropout (-Leu/-Trp/-Ade) to select for diploids in which a positive two-hybrid interaction is occurring.

2.19.7 Plasmid Isolation from Yeast

The yeast mini-prep was carried out manually according to the technique on page 35 of the Clontech protocol in yeast protocols handbook. The yeast cells must first be stripped of their cell walls. This was done with the digestive enzyme lyticase. The resulting spheroplasts (no cell walls) were then burst open using SDS. The DNA released was purified either by ethanol precipitation (see below) or by the chromaspin columns provided by Clontech. The DNA isolated was then used as a template for PCR.

DNA was isolated from yeast colonies in the first incidence to establish that the plasmids were present in the cells growing on the selective media. In the second incidence, it was necessary to isolate DNA in order to screen for clones that encoded interacting proteins.

For plasmid isolation from yeasts, it was done both manually and using kit (BD Clontech). Necessary materials are:

• Appropriate SD liquid or agar medium to keep selection on the plasmids (Appendix 4.2).

• Sterile, 1.5-ml microcentrifuge tubes.

• 20% SDS

• Lyticase Solution (5 units/µl in TE buffer; store at 4 C for up to 2 months or at –20 C for up to 6 months. If colloidal material precipitated, the solution mixed by inversion before using).

• Using kit: CHROMA SPIN-1000 DEPC-H2O Columns (BD Clontech, #K1334-1) and 2-ml centrifuge tubes for use with the columns

• In cases of not using the kit, phenol:chloroform extraction and ethanol precipitation was performed using: phenol:chloroform:isoamyl alcohol (25:24:1), 10 M ammonium acetate, 95–100% ethanol.

To prepare yeast cultures for lysis from a liquid culture, a large (2–4-mm), fresh (2–4-day-old) yeast colony was inoculated into 0.5 ml of the appropriate SD liquid medium and vortexed vigorously to completely break up the colony and resuspend the cells. Then it was incubated at 30°C overnight with shaking at 230–250 rpm. The following day, the cells were pelleted by centrifugation at 14,000 rpm for 5 min the supernatant was discarded and the pellets were resuspended in the residual liquid (total volume ~50 µl).

Then 10 µl of lyticase solution was added to each tube. The tube contents were mixed by vortexing or repeatedly pipetting up and down and the tubes were incubated at 37°C for 30–60 min with rotary shaking at 200-250 rpm. Optionally a drop of the cell suspension was checked under a phase contrast microscope (400x) for the progress of cell lysis by adding a drop of 20% SDS to the side of the coverslip. As they come into contact with the SDS, most cells should lose their refractile appearance and appear as "ghost-like" spheroplasts. If there were still many intact cells present, the samples were incubated for another 30 min. Then 10 µl of 20% SDS was added to each tube and vortexed vigorously for 1 min to mix. The samples were put through one freeze/thaw cycle (at –20°C) and vortexed again to ensure complete lysis of the cells. The entire contents of the tube were poured onto a prespun CHROMA SPIN-1000 Column and the plasmid DNA was purified according to the CHROMA SPIN User Manual. Finally purified plasmid DNA was eluted from the column. Alternatively, if

cleaning the prep manually (not using CHROMA SPIN Columns), the volume of the sample was brought up to 200 µl in TE buffer (pH 7.0) and 200 µl of the phenol:chloroform:isoamyl alcohol (25:24:1) mixture was added. The tubes were vortexed at highest speed for 5 min, centrifuged at 14,000 rpm for 10 min, and the aqueous (upper) phase was transferred to a fresh tube and 8 µl of 10M ammonium acetate / 500 µl of 95–100% Ethanol were added and placed at –70°C or in a dry-ice/ethanol bath for 1 hr. Finally the tubes were centrifuged at 14,000 rpm for 10 min, supernatants were discarded, the pellets were dried and resuspended in 20 µl of H_2O. Since the amount of plasmid DNA recovered was small relative to the contaminating genomic DNA, it could not be measured by A260 nm or seen on an agarose gel. However, it could be used to transform competent bacterial cells.

2.19.8 Control plasmids

The protein products of *p53* and the *T-antigen* gene are known to interact strongly.

Bait control plasmid - pGBKT7-53 (provided by Clontech): it encodes a fusion protein of Murine-p53 and the GAL4 DNA-BD and is a shuttle vector suitable for use in both *S. cerevisiae* and *E. coli*, as it contains both *S. cerevisiae* and *E. coli* origins of replication, and the kanamycin resistance and *TRP1* metabolic markers.

Library control plasmid - pGADT7-T: The library control plasmid was also provided by Clontech. It encodes a fusion protein of the SV40 large T antigen and the GAL4 AD. Again it is a shuttle vector, suitable for use in both *S. cerevisiae* and *E. coli*, as it contains both *S. cerevisiae* and *E. coli* origins of replication, and the ampicillin resistance and *TRP1* metabolic markers. The library and bait control plasmids provide a positive control for protein interaction as p53 and SV40 large T antigen interact, bringing the GAL4 activation domain (AD) and DNA-binding domain (BD) into close enough proximity to activate the reporter genes downstream of the 'upstream activation sequence' (UAS) GAL4 binds to. Thus cells containing both the library control and the bait control plasmid should be able to grow under -Trp/-Leu/-His/-Ade selection if the experiment is carried out appropriately.

2.19.9 Culture media

All culture media were made up and autoclaved at 121°C for 15 minutes. Based on the type of experiment and the plasmids transformed in the yeast strains, the appropriate selective media was used. After yeast transformation or mating, the product was plated onto either SD/-Leu, SD/–Trp, SD/-Leu/-Trp or SD/-Leu/-trp/-Ade/-His medium.

2.20 Fluorescence In Situ Hybridization (FISH)

2.20.1 Probe synthesis: Labeling and purification of the PAC clone

10µl of standard stock solution of purified PAC 618H14 clone containing the mouse *Tsga10* gene was taken and denatured by boiling for 5 minutes, then snap cooled in ice-water mix. 50 µl of labelling mixture was prepared using Random Prime Kit (Gibco BRL, Gaithersburg, MD) with Avidin-conjugated fluorescein isothiocyanate based on the kit instruction and incubated at 37°C for 3 hours before stopping the reaction by heating to 65°C for 10 minutes. 2µl salmon sperm DNA (10µg/µl stock) was added and the DNA were precipitated using ethanol. After centrifugation, the supernatant was removed, the pellet dried briefly and then resuspended in 40 µl ultrapure formamide (Boehringer Mannheim or Intergene (Cat. No. S4117).

2.20.2 Hybridization and detection

The probe was applied to male mouse metaphase spreads as described (by Gillett et al., 1993) and the slides were covered with parafilm before incubation in a humid chamber at 37°C for 12-16 hours. Then the slides were washed by the proper washing solution (0.1% SCC) at 45°C and room temperature. After washing, the slides were immersed and blocked in blocking buffer (5% non fat milk powder) for 10 minutes. For detection, 50µl of BI buffer containing 4µg FITC-Avidin (Vector Labs, DCS grade) was used for 60 minutes. FITC signals and the DAPI (4,6-diamidine-2-phenylindole dihydrochloride) banding pattern were merged for figure preparation. Images of metaphase preparations were captured with a cooled CCD camera using the Cyto Vision Ultra image collection and enhancement system (Applied Imaging Int.).

Results

Chapter III:

Identification of mouse homologue of the *TSGA10 (Tsga10)* and studying its expression pattern

3.1 Cloning the mouse homologue of the TSGA10 gene (Tsga10) and studies of its transcripts

One of the main aims of this project was to find out more about the *TSGA10* gene, which was known to be expressed in human testis and rat spermatocytes, and, therefore, thought to be involved in spermatogenesis or spermiogenesis. As well as the interesting biology of these processes, genes, which affect these two processes, may later be investigated with the hope that they may be used to design treatments for male infertility and male contraceptives. Recently, a group in Switzerland (Primig et al, http://germonline.unibas.ch/gene_page.php?species_id=7&orf_id=268998#) used U34 A and B microarrays (GeneChips) to determine the expression levels of 11955 transcripts in highly enriched populations of rat spermatogonia, spermatocytes, round spermatids and total testis samples. By comparing data from germ cells to Sertoli cells 1268 mRNAs were identified as being strongly differentially expressed. The loci organised into four expression clusters whose transcriptional peak induction patterns correspond to somatic, mitotic, meiotic and post-meiotic cell types. A subset of meiotically expressed genes was not detected in three somatic control tissues (Sertoli, skeletal muscle and brain). This group includes *Tsga10*, which is a member of the meiotic cluster.

To understand more fully the expression profile of this gene we have isolated and characterised the mouse homologue. A partial cDNA sequence from *Rattus norvegicus* spermatocyte (Accession no. AF092091) was aligned with the human *TSGA10* sequence (Accession no. AF254756) and an existing pair of human primers (GSP5 and R6H) was identified which showed just one internal nucleotide difference

to the rat sequence. These were used in an RT-PCR using mouse testis cDNA and a primer annealing temperature 10°C lower than usual. The resulting RT-PCR products were electrophoresed on an agarose gel, and then, after re-amplification as shown in figure 3.1, a product of the same size as the human control was observed. This band was cut from the gel and sequenced from both ends. Comparison with *TSGA10* sequence revealed 89% similarity between the mouse and human sequences.

Approximately $2x10^5$ p.f.u. of a mouse testis cDNA library, made from purified spermatocytes (Mahadevaiah et al., 1998) were plated out on two 22x22 cm plates and duplicate filter lifts were taken. Replica filters were screened with a radioactively labelled PCR product derived from the mouse RT-PCR product described above using one round of synthesis with Klenow DNA polymerase and the PCR primers. The plaques giving the strongest signals were isolated by replating and rescreening and one (clone P2) was sequenced from the ends using vector (pBluescript SK) primers. The complete *Tsga10* coding region was sequenced using templates either of mouse RT-PCR products or the mouse spermatocyte cDNA clone P2. The mouse sequence which had 89% homology to the human *TSGA10* sequence over most of its length, was submitted into GenBank (accession no. AF530050).

A mouse genomic library consisting of seven gridded filters (128×10^3 PAC clones, 1.5×10^5 bp insert size (Osoegawa et al. 2000) provided by the UK HGMP Sanger Centre) was screened (by Hossein Modarressi) with a human cDNA probe. The probe corresponds to nucleotides 1493 to 2125 of the human gene. Eight positive signals were obtained and the corresponding bacterial colonies were checked by PCR

using primers designed from the mouse testis cDNA clone sequence and located in the 5′ end of the gene.

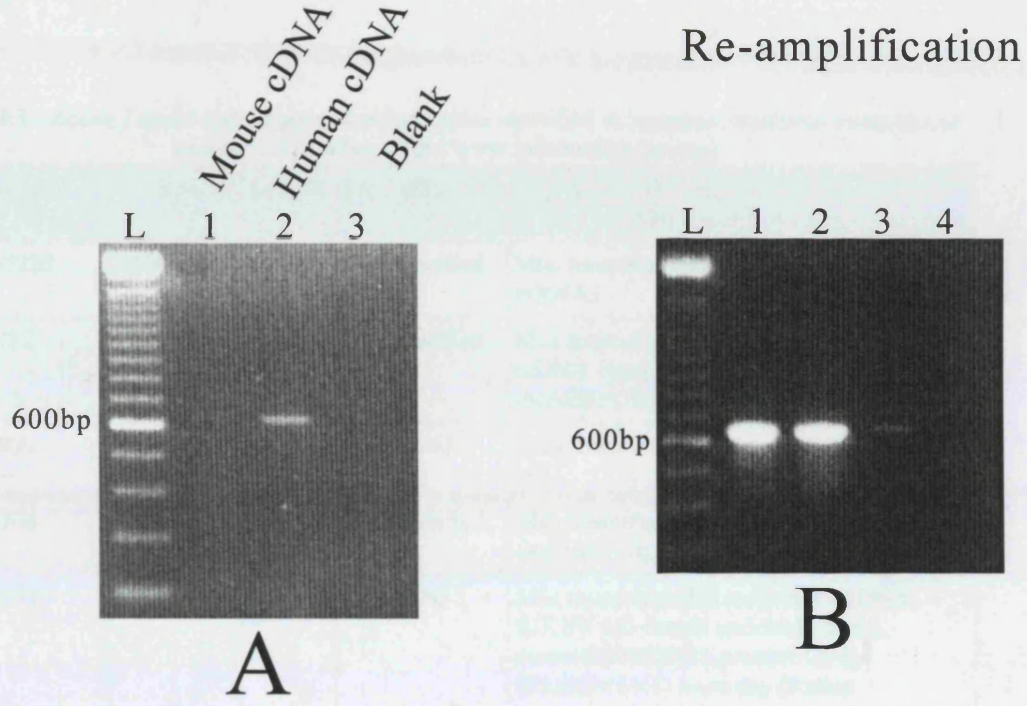


Figure 3.1- RT-PCR using mouse oligo dT primed cDNA and human primers GSP5 and R6 at an annealing temperature of 44 °C. In gel A a band from human cDNA is seen (in lane 2). A very faint band was visible in lane 1 the same size as the human PCR product. Lane 3 is the PCR negative control (water). Gel B shows re-amplification of PCR A products with 26 PCR cycles at an annealing temperature of 54 °C. The template for lane 1 was obtained by pricking the band in gel A lane 1 with a needle. Template in lane 2 was 0.2 µl of the mouse RT-PCR product. The template in lane 3 was from cDNA as in gel A lane 2. The bands in gel B lanes 1 and 2 were excised for sequencing. Lane 4 is a negative (water) control. Lane L is 100bp ladder in both panels.

The PCR results confirmed that colonies 635E6, 552E13, 652L5, 413O2 and 618H14 contained the 5′ end of the mouse gene. The library used in this experiment was equivalent to approximately six fold genome coverage and the hybridisation results (8 colonies) is consistent with the gene being present in a single copy (like the human gene, Modarressi et al. 2001).

In November 2004, apart from the sequence, which was submitted by me, there are 4 other sequences for mouse homologues of *TSGA10* (*Tsga10*). They are listed in table 3.1. The sequences are shown aligned in appendix 8.1.

Table 3.1- Mouse *Tsga10* transcripts and polypeptides identified in sequence databases using mouse genome informatics (http://www.informatics.jax.org)

Sequence	Type	Length	Strain/Species	Description From Sequence Provider
NM_207228	RNA	3499	Not Specified	Mus musculus testis specific 10 (*Tsga10*), mRNA.
BC066782	RNA	3499	Not Specified	Mus musculus testis specific 10, mRNA (cDNA clone MGC:74389 IMAGE:6705127), complete cds.
AF530050	**RNA**	**2732**	**C57BL/6J**	**Mus musculus *Tsga10* mRNA, complete cds.**
AY618264	RNA	2522	Not Specified	Mus musculus *Tsga10* isoform mRNA, complete cds, alternatively spliced.
AK030254	RNA	2701	C57BL/6J	Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933432N21 product:CP431 (FRAGMENT) homolog [Rattus norvegicus], full insert sequence.

Sequences NM_207228, BC066782, AY618264 and AK030254 were searched for matches [to complete cDNA sequence AF530050)

3.1.1 Genomic localisation of Tsga10

The human TSGA10 gene was mapped by fluorescence in situ hybridisation to human chromosome 2q11.2, and mouse *Tsga10* was mapped to the syntenic region on mouse chromosome 1 (figure 3.1.1). Two mouse PAC clones (RP34-178E16 and RP23-21P10) from the HGP Sanger centre cover the whole of the mouse *Tsga10* gene. Subsequent genome analysis using genome browser and BLAST mapviewer (BLAST http://www.ncbi.nlm.nih.gov/mapview) revealed that the *Rattus norvegicus* homologue is located in chromosome 9q21, which is again in the human synteny region.

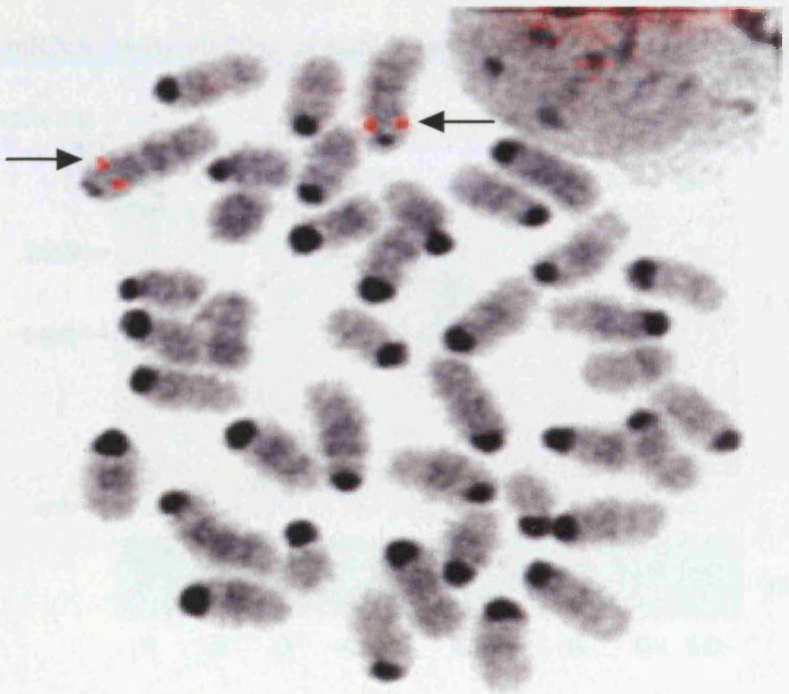


Figure 3.1.1- Localization of mouse *TSGA10* on mouse chromosome 1 (red dots) to the band B by in situ hybridisation with a probe amplified by PCR from intron exon 4 to mouse metaphase chromosomes. The positions of hybridization signals (red dots) are marked by arrows. Original magnification x100 (figure from Modarressi et al., 2004).

3.2 Studying the expression pattern of the *Tsga10* gene using in *situ* Hybridization and Northern blotting

As described previously, male germ cell development proceeds as the cells move in a peripheral-to-central fashion in the seminiferous tubules of the testis. RNA extracted from different mouse tissues was first analyzed by Northern blotting using a mouse *Tsga10*-ORF DNA probe.

The result, shown in figure 3.2A, indicates that *Tsga*10 expression is restricted to testis, in which a single major transcript was observed. Rehybridization of filters with an actin probe confirmed the presence of intact RNA in all instances. The observed size of *Tsga10* mRNA is 3.2 kb, similar to the human Tsga10 transcript. A faint band above the major mRNA was visible, possibly resulting from splice intermediates or alternative splicing.

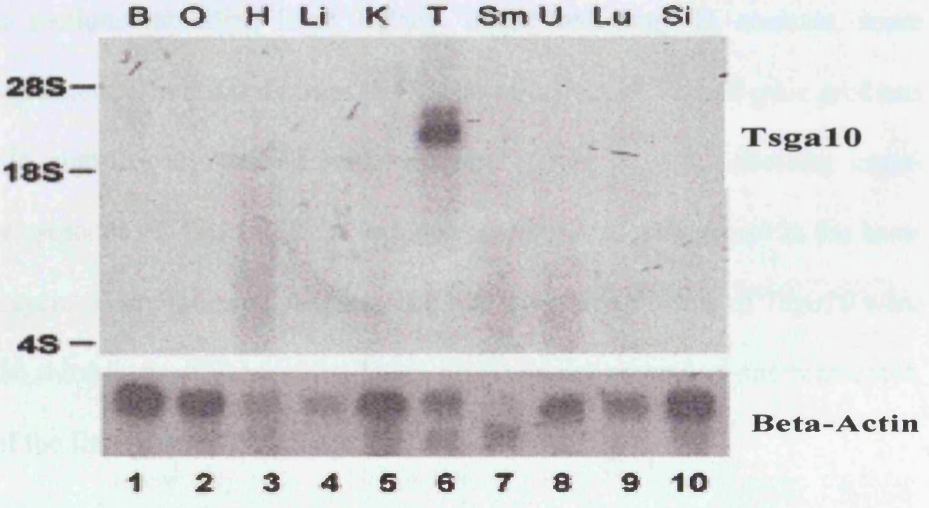
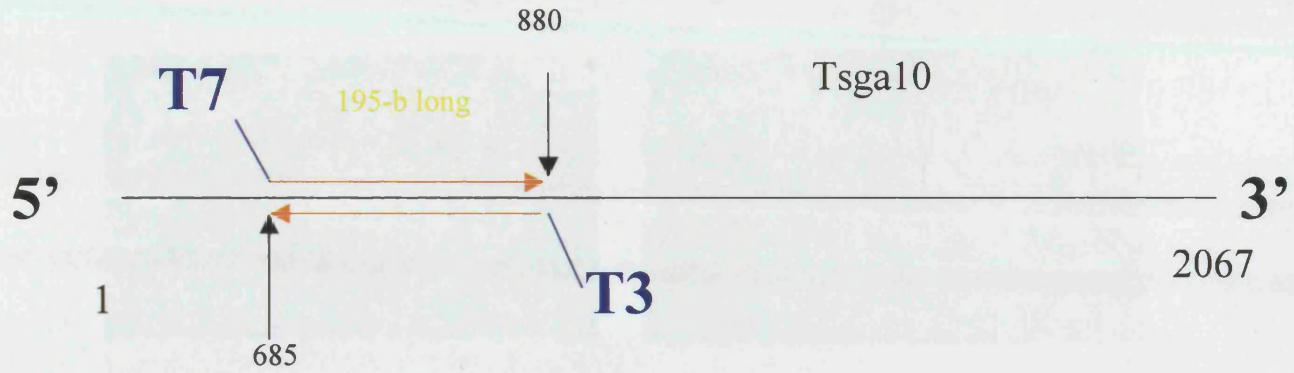


Figure 3.2A- *Tsga10* mRNA is expressed in testis. RNA was isolated from adult mouse tissues and analyzed by Northern blotting using *Tsga10* cDNA as a probe (top) and ß-actin (bottom) as a control. Lanes 1–10: brain (B), ovary (O), heart (H), liver (Li), kidney (K), testis (T), skeletal muscle (Sm), uterus (U), lung (Lu), and small intestine (Si) RNA. Indicated are 28S and 18S ribosomal marker RNAs and 4S tRNAs.

In parallel the in *situ* hybridization technique was used to identify and localize *Tsga10* RNA in testis.

To prepare a labeled RNA probe (riboprobe), a 195-nucleotide DNA sequence was identified which was unlikely to form a stable secondary structure, and oligonucleotides T7Mtc10 and T3Mtc10 (figure 3.2B1) were designed and used to amplify this region from cDNA. 1μg DNA of PCR product was used after purification (with Qiagen PCR purification kit) as the template for RNA synthesis using either T7 or T3 RNA polymerase. 2μl of each riboprobe before and after treatment with DNase I was taken for running in an agarose gel (figure 3.2B2). The remaining (16μl) was stored at -20°C until required.

The probes were labelled with digoxigenin (using the alkaline phosphatase method) and hybridised to mouse tissue sections. *Tsga10* gene products (mRNA) were not detectable in sections including liver, kidney, heart, and lung. In contrast, some expression was detected in brain sections in clusters of cells, and *Tsga10* gene products were found in significantly stained testis sections (figure 3.2.C). Labelling signal indicating the presence of *TSGA10* RNA was detected in most cells except in the basal (peripheral) layers of seminiferous tubules. The highest relative levels of *Tsga10* were detected in the third layer of germ cells. These might be the secondary spermatocytes, the product of the first meiotic division.



Sense

T A A T A C G A C T C A C T A T A G G A C T C
A G C G A C A T C T T G C T A A G A A A A A
A T A T G A G C T A C A G C T T A C T C A G
G A G A A A A T T A T G T G C T T G G A T G
A A A A A A T T G A T A A T T T C A C G A G
G C A A A A T A T T G C A C A G C G A G A
A G A A A T C A G C A T T C T T G G T G C A
A C C C T C A A T G A C C T G G C T A A A G
A A A A G G A A T G C C T G C A A G C G *T*
G T C T G G A T A A A A A G T C T G G A A A
T C A C T C C C A A T T A A

<--T3

Anti-Sense

Figure 3.2B1- Schematic feature and sequence of Riboprobe Synthesis. Oligonucleotides with the T3 and T7 RNA polymerase promoters were designed to amplify a 195-bp section of the *Tsga10* cDNA (nucleotides 685-880), and below the sequence of this region. The primer sequences are underlined, and the T3 and T7 recognition sequences added to these are shown in blue.

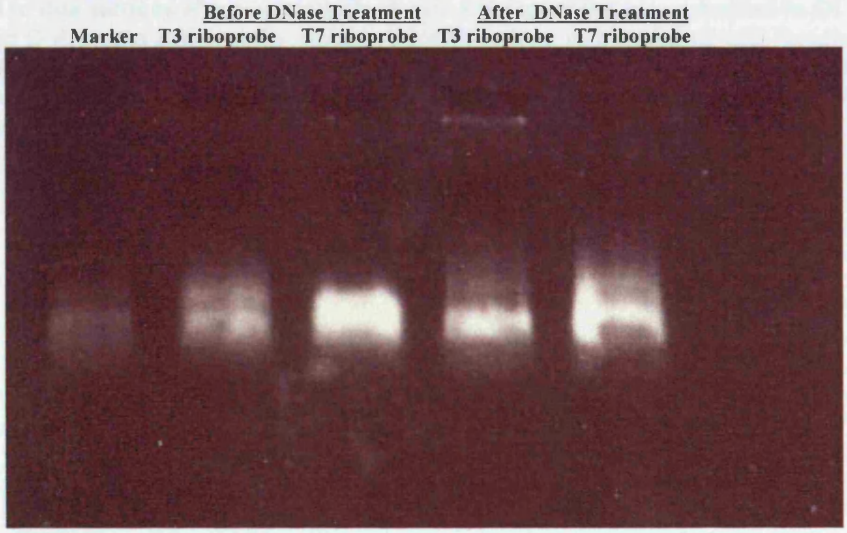


Figure 3.2B2- *Tsga10* sense and antisense riboprobes. 2µl each was run before and after DNase treatment on a DEPC treated gel to check presence and integrity of the RNA. After DNase I treatment any template DNA is removed, and the RNA is still present and full length.

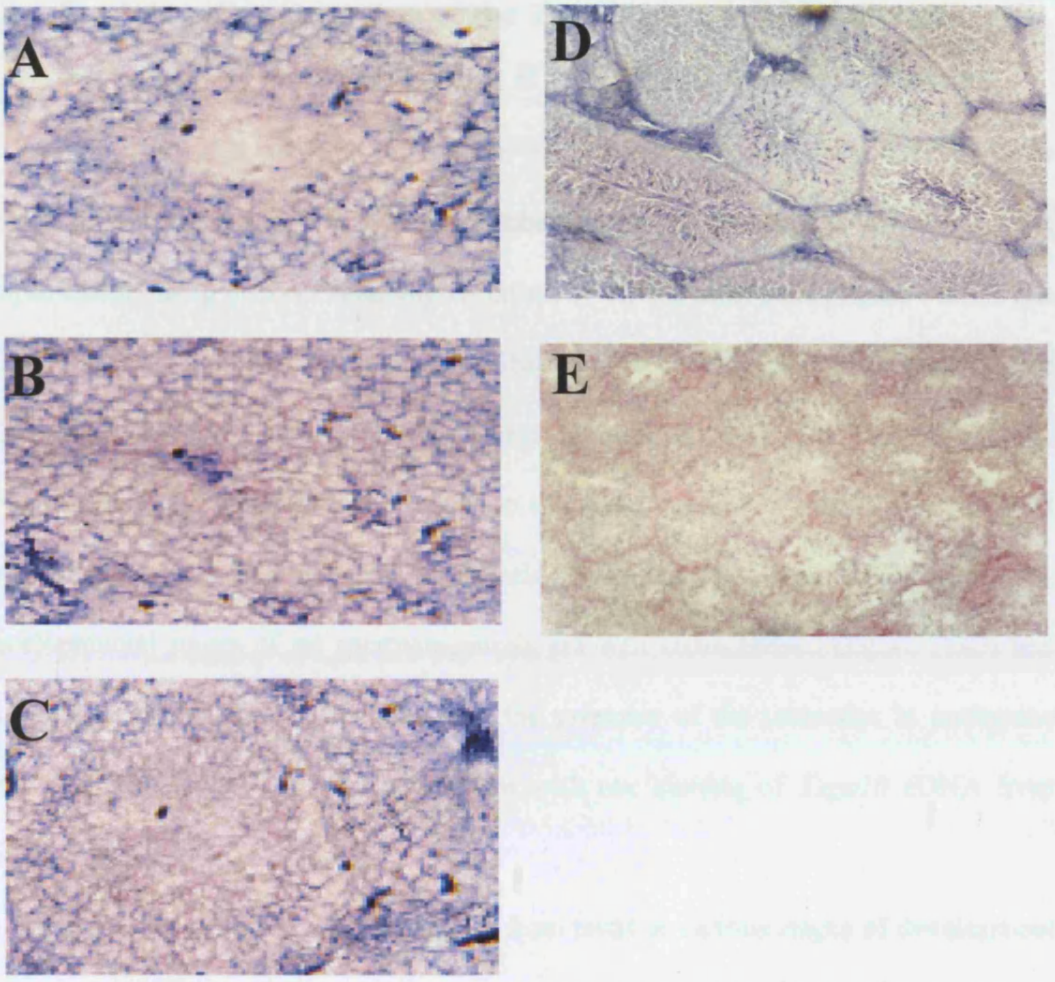


Figure 3.2C- In situ hybridisation, 100 x. T3 and T7 –primed riboprobes to Tsga10 were hybridised to thin sections of mouse testis fixed with 4% formaldehyde-embedded in OCT.
Panel A and D show the antisense T3 probe hybridisation. D is counterstained with hematoxyline and Eosine (H&E), B shows the sense probe and C and E are controls without probe. Hybridisation is seen in panels A and D. Some artificial spots are seen in common between A, B and C, which are caused by the microscope and camera.

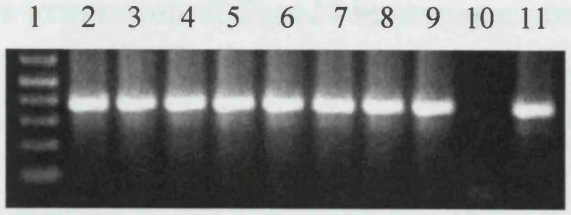
3.3 Studying the expression of the *Tsga10* gene in testis at various stages of development using RT-PCR and Northern blotting

RNA was extracted from rat testis at different developmental stages. The RNA was then converted to cDNA, which was checked for the presence of cDNA by PCR amplification using primers spanning an intron of the rat actin gene (figure 3.3A). The cDNA was then amplified using primers GSP5 and R6H from the *Tsga*10 gene (figure 3.3B); note that mouse and rat *Tsga*10 [accession no. AF530050, XM_237066] are over 95% identical. Rat germ cells were used in this experiment. The results show that the *Tsga10* transcript is first detected in rat testis between Days 15 and 21 after birth. The developmental stages of rat spermatogenesis are well characterised (figure 3.3D) and expression at days 15-21 corresponds to the presence of the transcript in pachytene cells during meiosis. This is in agreement with our cloning of *Tsga10* cDNA from spermatocytes.

In Northern blot analysis, a filter of RNA from testis at various stages of development was hybridised with a PCR product spanning the Tsga10 cDNA. Hybridisation is seen from 21 days after birth onwards (figure 3.3C).

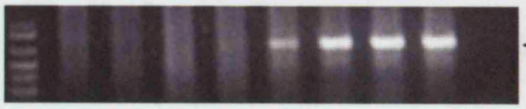
The result of Northern blot hybridisation is in agreement with the results of the RT-PCR.

A-

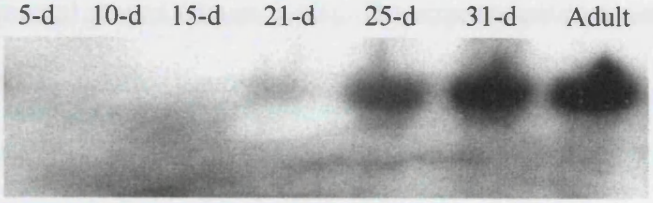


1- Marker
2- Foetal testis
3- Testis (5 days)
4- Testis (10 days)
5- Testis (15 days)
6- Testis (21 days)
7- Testis (25 days)
8- Testis (31 days)
9- Testis (adult)
10- Negative control (water)
11- Positive control

B-



C-



D-

Stage No	Description of developmental stage	Age of rat (days)
I	Spermatogonia and somatic cells only	6-7
II	Initiation of meiosis I-leptotene cells	13-14
III	Appearance of zygotene cells	17-18
IV	Appearance of early pachytene cells	19-20
V	Appearance of late pachytene cells	22-23
VI	Appearance of round spermatids	24-25
VII	Appearance of elongated spermatids	30-31
VIII	Appearance of elongated spermatozoa	36-37

Figure 3.3- *Tsga10* expression in different developmental stages of rat testis.
A- A control RT-PCR amplification using rat β-actin primers spanning an intron (intron II), indicated that all the cDNAs that were tested, gave a product of the correct size (375bp) as shown.
B-RT-PCR using rat oligo dT primed cDNA and human *TSGA10* primers GSP5 and R6H at an annealing temperature of 53 °C. The expected PCR product size (indicated by an arrow) was 632bp.
C- Northern blot of RNA derived from rat testis at various numbers of days after birth ranging from (5 days to adult). The blot was hybridised with the whole mouse cDNA, reamplificated by PCR at 60°C. Hybridisation is seen faintly at 21 days after birth, and strongly thereafter.
D-Developmental schedule of cell types present in rat testis at different stages of development (Malkov et al. 1998).

3.4 The expression of *Tsga10* in mouse embryos at various stages of development, and in embryonic brain

RNA from various developmental stages of mouse embryos was converted to cDNA and used for PCR using primers from the *Tsga*10 and ß actin genes; Primer pairs of MG4 (reverse primer mouse) and R6H (forward primer human, mouse, rat) were used. Either testis-derived cDNA or a cloned *Tsga10* plasmid construct was used as positive controls with *Taq* DNA polymerase. The RT-PCR results show that *Tsga10* is not expressed in embryonic stem cells (figure 3.4), which suggests that it may be turned on sometime between 4.5 and 7.5 dpc. This is confirmed by the RT-PCR of embryos at later developmental stages (figure 3.4B). The expression commenced by 7.5 dpc and contained until newborn (figure 3.4C). Also it was expressed in brain of 11 dpc embryos (figure 3.4B and D). Alternative splicing (isoform) products were observed in whole embryos (figure 3.4D) as well as mature testis (figure 3.4A) but not in the embryo brain RNA (figure 3.4D).

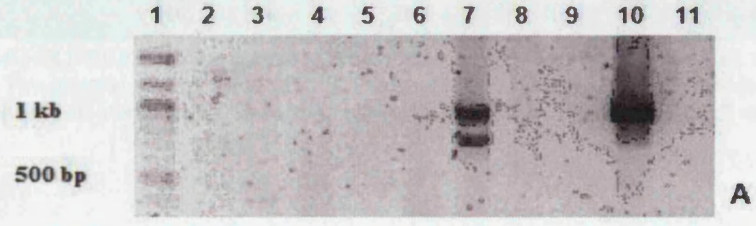


Figure 3.4A- RT-PCR in mice 3 dpc embryos (embryonic stem cells) shows no *Tsga10* expression in embryonic stem –ES- cells but in mature testis with an alternative splicing.
1-Marker
2- ES cells (1) (first division)
3- ES cells (2) (first division) Embryonic stem cells (embryo after about 3 days) from ES cell culture
4- ES cells (1) (fourth division)
5- ES cells (2) (fourth division) ES cells treated with Retinoic acid (makes cells neural differentiation)
6- ES cells (3) (first division)
7- cDNA from mouse testis
8- cDNA from mouse heart
9-cDNA from NIH3T3 cell
10- Positive control from p2 clone, mouse *Tsga10*
11- Negative control

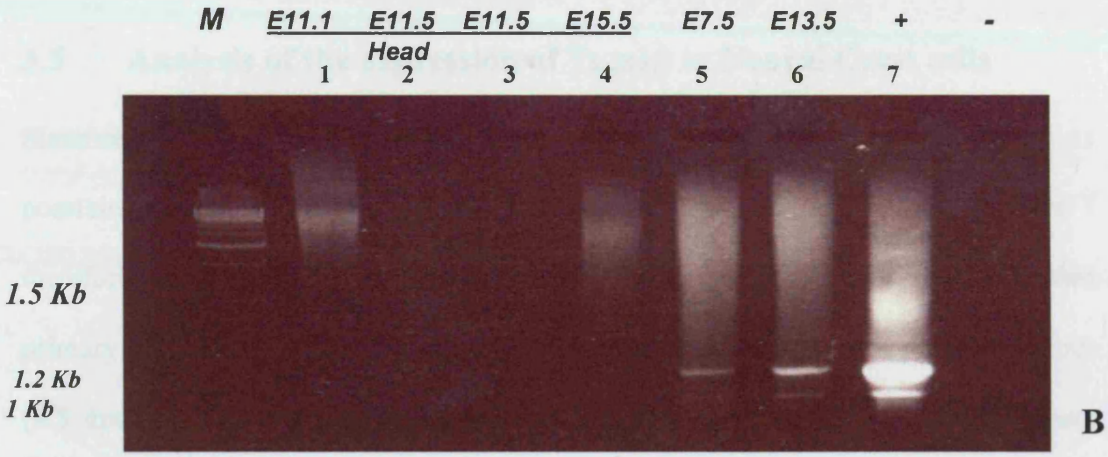


Figure 3.4B- RT-PCR of RNA extracted from mouse embryos using mouse *Tsga10* primers MG4 and R6H at 52°C. Lanes 1-4 RNA is from head of mouse embryos at various days p.c. Each lane contains RNA from two heads. In lanes 5 and 6 RNA is from whole embryos at 7.5 and 13.5 days p.c. The positive control is testis RNA. Negative is water. Expression can be seen in lanes 5 and 6. The two bands represent the alternative splicing products.

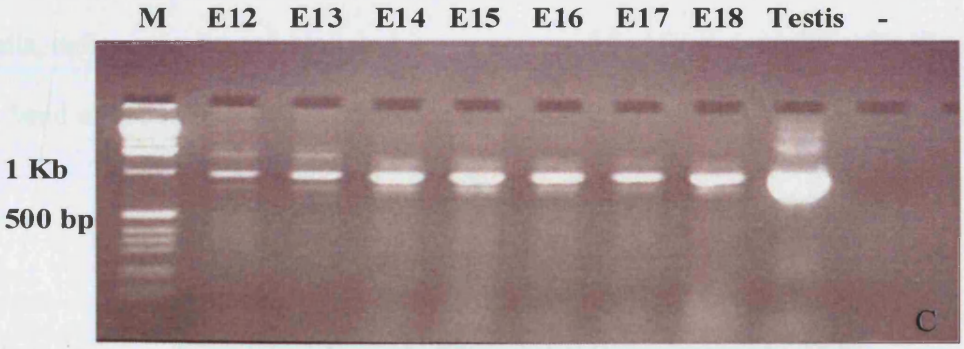


Figure 3.4C- RT-PCR of RNA extracted from mouse embryos at stages from day 12 to newborn. The primers are the same as in B. Positive control is testis. Negative control is water. Expression is seen at all stages. The splice variant can only be seen at stages E12 and E13.

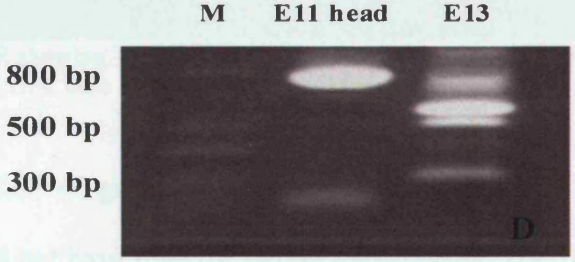


Figure 3.4D- Reamplification and semi-nested PCR of 1 ul from the PCR product obtained in B, from RNA extracted from head at 11 days p.c. and whole embryos at 13 days p.c. Semi-nested primers were R68 and ER18. The primers span the region of alternative splicing. Alternate splice products can be seen in the whole embryo, but not in the head. However, expression of the full-length transcript is seen in the head after reamplification (compare with B), and most likelyrepresents expression in the brain.

3.5 Analysis of the expression of Tsga10 in Neural Crest cells

Since neural crest cells are another example of dividing and differentiating cells, it was possible that *TSGA10* would also be expressed in these cells and their derivatives. I therefore carried out RT-PCR using cDNA which the RNA was extracted from primary cell cultures of neural crest cells obtained from mouse embryo neural tubes (8.5 dpc) cultured for 6 days (a gift from Dr Aldamaria Puliti). A pair of primers (MFE17 and MRE18) in different exons (17 and 18) was used. *Tsga10* expression also was assessed in 13-dpc embryonic guts as a neural crest derivative. RNA extracted from adult mouse brain was used as a control. The mouse neuro2A (derived from neuroblastoma) cell line was also tested. A 169-bp band was obtained from neural crest cells, embryonic gut and brain but it was not amplified from neuro2A cells (figure 3.5). A band of about 500 bp was present in all the samples tested. This band is likely to

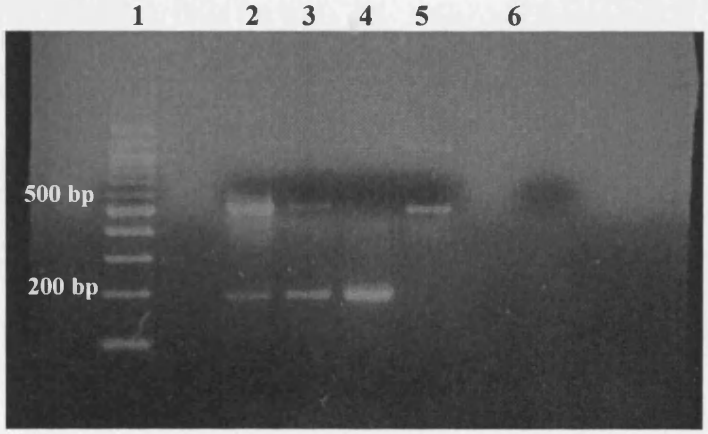


Figure 3.5– RT-PCR showing Tsga10 expression in neural crest cells and the brain. 1) 100 bp ladder 2) E8 Neural crest cells 3) E13 Gut 4) Brain 5) Neuro2A 6) Negative control

represent amplification of genomic DNA contamination, since intron 18 is 304-bp long. However, I did not have time for further experiments. The 186-bp band from the brain was sequenced and shown to correspond exactly to the expected *Tsga10* cDNA

specific fragment. Tsga10 is expressed in the neural crest cells, as it is in the brain. However, no expression is seen in neuro2A (a neuroblastoma cell line). Expression of *Tsga10* is seen in several in malignant tissues (see section 3.6, Tanaka et al., 2004) but not all tumour types.

3.6 *Tsga10* is expressed in the Acute Promyelocytic Leukemia (APL) specific NB4 cell line

There are some ESTs from both the 5' and 3' ends of the *Tsga10* gene in GenBank (Accession nos. AW057728, AI696619, AW591313, BE047007, and BF243403), which were isolated from poorly differentiated adenocarcinoma of the stomach, acute myelogenous leukaemia and germ cell tumors. To test the expression pattern of *Tsga10* gene in an Acute Myeloid Leukemia (AML), acute premyelocytic leukemia (APL) cell line (NB4) was selected and RNA was extracted from it and converted to cDNA. *Tsga10* 5'-end primer pairs of MHF4' (forward primer), MR5' (reverse primer) and *Tsga10* 3'-end primers of MFE17 (forward primer) and MRE18 (reverse primer) were used. Amplification was carried out using DNA Taq polymerase (AB) for 32 cycles, each cycle consisting of 94°C for 30 s, 50°C for 30 seconds and 72°C for 75 seconds. Both experiments showed *Tsga10* expression (figure 3.6) in the acute premyelocytic leukemia (APL) cell line (NB4). A larger band (1kb) is also seen, which is probably due to genomic DNA contamination and indeed, the size of the PCR product is what would be expected with these primers if the amplification included intron 17.

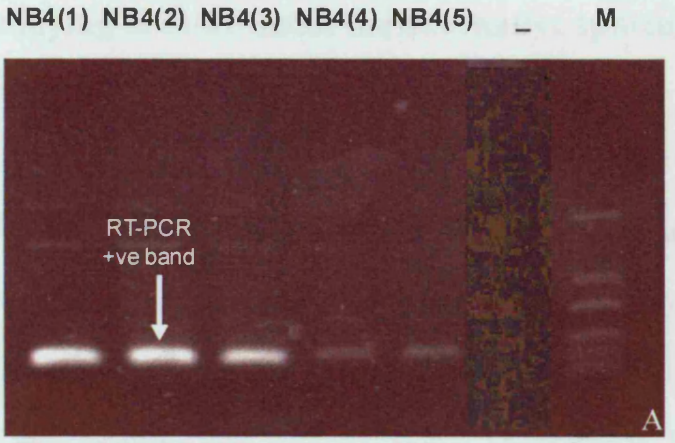


Figure 3.6A- RT-PCR experiment of RNA extracted from 5 different NB4 cell culture plates. Each RNA sample was amplified separately using 5' terminus primers MHF4' and MR5'. The primers span an intron of 475 bp. The RT-PCR product of 200 bp can be seen in all samples.

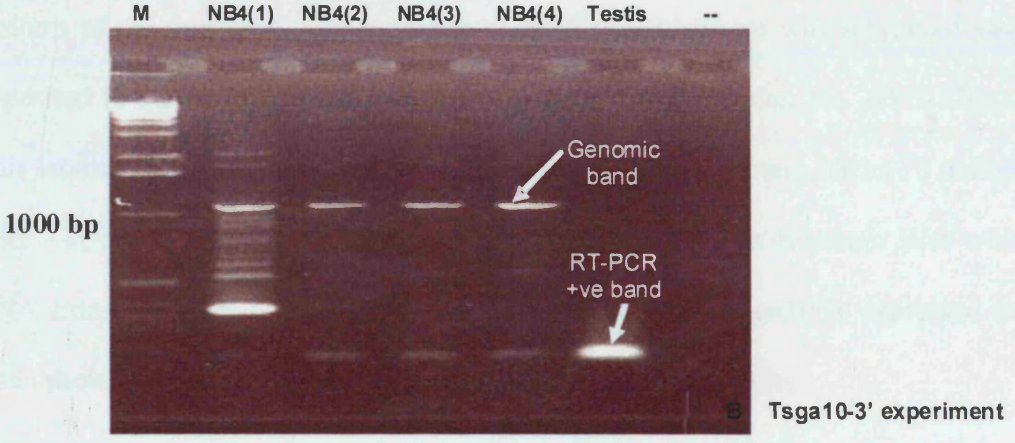


Figure 3.6B- RT-PCR experiment of RNA extracted from 4 different NB4 cell culture plates. Each RNA sample was amplified separately using 3' terminus primers MG4 and MFE17. The primers span an intron of 860 bp, and two products can be seen. The smaller is the result of RNA amplification, and the longer is probably the result of genomic DNA contamination.

3.7 Studying in more detail the alternative splicing of *TSGA10* exon 16

Results shown in sections 3.2 and 3.4 illustrate that isoforms of the *Tsga10* gene transcript (cDNA) exist (e.g. see figures 3.2A, 3.4A). I investigated whether alternative splicing caused these. The PCR reaction using *Tsga10* primer pairs spanning the relevant region -MG4 (reverse primer in exon 12) and R6H (forward primer in exon nineteen) with testis and mouse whole embryo RNA as the template produced a smaller size band in addition to the expected 990 bp band. This small band was reamplified after purification, using the same primers. Sequence analysis confirmed that the whole of exon sixteen might be spliced out to make a shorter isoform of the transcript. The sequence of the *Tsga10* isoform was submitted and deposited in the NCBI GenBank database (BLAST as Accession No. AY618264). This isoform does not cause a frame shift, but results in a protein, lacks 70 amino acids –in cDNA check it lacks 210 bases- (Appendix 8.1). In Northern Blot with RNA extracted from different mouse tissues (figure 3.2), *Tsga10* is expressed in testis showing two transcripts compatible with the RT-PCR results.

Two transcripts were observed in mouse embryo: *Tsga10* alternative splicing (isoform) products were observed in whole embryos as well as mature testis but not in the embryo brain (figure 3.4A,B).

Exon 16 is flanked by two long introns of 17 and 20 kb respectively. Alternative splicing is observed in other genes where exons are flanked by long introns (e.g. *PKC-δ* gene which its first intron to the exon is as long as 17 kb in human), indicating a complexity involved in gene splicing (Suh et al., 2003).

3.8 *Tsga10* gene structure

Previous work from our laboratory and in this thesis has identified *TSGA10* as a testis-specific human gene. It has been shown to be expressed in normal testis, but not in a variety of other tissues. It is expressed in neither foetal testis nor testis of two infertile patients but it is expressed in several human foetal tissues and some primary tumours. *TSGA10* sequence data was deposited with the EMBL/GenBank Data Libraries (Accession No. AF254756). Later a sequence called "NYD-SP7" was submitted by a group in China and was named as a member of the nucleoporin family (without any evidence being shown). *TSGA10* has 99% homology to this sequence. A number of rodent Tsga10 sequence has been submitted to GenBank. Table 3.8A summarises these various sequences.

Table 3.8A- Summary of rodent *Tsga10* sequences submitted in GenBank database.
Human *TSGA10* (H), Mouse *Tsga10* (M), Rat *Tsga10* (R).

Accession number	Name	Species	Deposited by	Homology (%)	Description
AF530050	Tsga10	M	B. Behnam	89% to H	18 exons
AY618264	Tsga10	M	B. Behnam	90% to H	17 exons (spliced version)
AF092091	Tsga10	R	S. Hoyer-Fender	94% to H	Partial cDNA from rat spermatocyte cDNA library
XM_343557	Cp431	R	IMAGE	94% to H and M	16 exons
XM_136734	Cp431	M	IMAGE	90% to H	19 exons, produces hypothetical protein 4933432N21 extends 77bp upstream of AF530050 first exon and extends the open reading frame in a 5' direction by 202 aa
NM_207228	Tsga10	M	RIKEN	90% to H	20 exons, replaces sequence XM_136734
AK030254	Cp431	M	RIKEN	89% to H	19 exons

The 5' ends of my mouse *Tsga10* Sequence and all of the cp431 sequences are aligned in figure 3.8A. This figure shows that upstream of the conserved 'double methionine' start codon (green box), the sequences differ: XM_136734 extends 77 bp upstream into a possible first exon, which extends the ORF in 5' direction by 606 nucleotides. Figure 3.8B illustrates how the different 5' sequences in the different Tsga10 transcripts may represent alternative first exons.

```
MOUSE CP431       GGCGGTCCACGTCGCTACGTGCTCGCTCG------------------------------ 80
MOUSE CP431       GGCGGTCCACGTCGCTACGTGCTCGCTCGGGTGCGCGCGCACGATCCGTAGTCTCCGCGA 176
AF530050          ------------------------------------------------------------
MOUSE Tsga10      GCCGGACGGAAGTGGAGCGTAGGCGCCGGGGAG----------------------GTGG 39
RAT TSGA10        TGCTGTCTACGTCGCTACGTGCTCGCTCG------------------------------ 332

MOUSE CP431       -----------------GGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTT 120
MOUSE CP431       GAGGGCGTGTCGAGACAGGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTT 236
AF530050          -----------------GGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTT 43
MOUSE Tsga10      AAGAAGAGCCCCGCTCAGGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTT 99
RAT TSGA10        -----------------GGTTACCACCAAAGAACATCAGTTTGGTCTCGGGGCTCCAGTT 372
                                  ***** ************************ ************

MOUSE CP431       TCTGAAGCTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGAT 180
MOUSE CP431       TCTGAAGCTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGAT 296
AF530050          TCTGAAGCTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGAT 103
MOUSE Tsga10      TCTGAAGCTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGAT 159
RAT TSGA10        TCTGAAGCTGGAAGAAACCAGAATACTTTCCAACTAGAACAAGAAGGGAGAACCCAAGAT 432
                  ******************************************** ************

MOUSE CP431       AGATTCATCTCTACACTGAAATT-ACAGATTGAAGATCTCAAACAGACAAATCATGACTT 239
MOUSE CP431       AGATTCATCTCTACACTGAAATT-ACAGATTGAAGATCTCAAACAGA-AAATCATGACTT 354
AF530050          AGATTCATCTCTACACTGAAATTTACAGATTGAAGATCTCAAACAGACAAATCATGACTT 163
MOUSE Tsga10      AGATTCATCTCTACACTGAAATT-ACAGATTGAAGATCTCAAACAGACAAATCATGACTT 218
RAT TSGA10        AGAGTCATCTCTACACTGAAATT-ACAGATTGAAGATCTCAAACAGACAAATCATGACCT 491
                  *** ****************** ********************** ********** *

MOUSE CP431       GGAAGAATATGTTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGA 299
MOUSE CP431       GGAAGAATATGTTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGA 414
AF530050          GGAAGAATATGTTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGA 223
MOUSE Tsga10      GGAAGAATATGTTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGA 278
RAT TSGA10        GGAAGAATGTGTTAGGAAACTCTTGGATAGTAAAGAGGTGGTAAGCACTCAAGTAGATGA 551
                  ******** *************************** ********************

MOUSE CP431       CTTAGCCAACCACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGA 359
MOUSE CP431       CTTAGCCAACCACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGA 474
AF530050          CTTAGCCAACCACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACATGCAGA 283
MOUSE Tsga10      CTTAGCCAACCACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGA 338
RAT TSGA10        TCTAACCAACCACAATGAGCATCTTTGTAAAGAGTTGCTTAAACTTGACCAACTAGCAGA 611
                  ** ***************** *************** ***************  *****

MOUSE CP431       GAAATTACAAAAAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGC 419
MOUSE CP431       GAAATTACAAAAAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGC 534
AF530050          GAAATTACAAAAAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGC 343
MOUSE Tsga10      GAAATTACAAAAAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGC 398
RAT TSGA10        GCAATTGCAAAAAGAAAAGAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGC 671
                  * **** *********** ****************************************

MOUSE CP431       AAAGATTGAACTCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAG 479
MOUSE CP431       AAAGATTGAACTCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAG 594
AF530050          AAAGATTGAACTCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAG 403
MOUSE Tsga10      AAAGATTGAACTCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAG 458
RAT TSGA10        AAAGATTGAACTCATTTGCCAGCAAAATAACATAACAGTATTAGAAGATACAATTAAAAA 731
                  ***************************** ********************** ***
```

```
MOUSE CP431       GCTTAAGTCCATAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGA 539
MOUSE CP431       GCTTAAGTCCATAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGA 654
AF530050          GCTTAAGTCTATAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGA 463
MOUSE Tsga10      GCTTAAGTCCATAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGA 518
RAT TSGA10        GCTTAAGTCAATAATTTTAGAGACTGAAAAAGTACAAAATACATCTCCTTCTAGACTTGA 791
                  ********* ******************* * ************************ **

MOUSE CP431       TTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTA-TTATAAGA-CTGAAGCTCAGAAT 597
MOUSE CP431       TTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTA-TTATAAGA-CTGAAGCTCAGAAT 712
AF530050          TTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTAGTTATAAGAACTGAAGCTCAGAAT 523
MOUSE Tsga10      TTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTA-TTATAAGA-CTGAAGCTCAGAAT 576
RAT TSGA10        TTCCTTTGTCAAGACTTTAGAAGCAGACAGAGATTA-TTATAAGG-CTGAAGCTCAGAAT 849
                  ********************************** *******  ************** 

MOUSE CP431       TTAAG-AAAAATGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCG 656
MOUSE CP431       TTAAG-AAAAATGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCG 771
AF530050          TTAAGAAAAAATGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCG 583
MOUSE Tsga10      TTAAG-AAAAATGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCG 635
RAT TSGA10        TTGAG-AAAAATGATGAGAAATAGATCTAAGAGTCCAAGACGCCCATCACCAACTTCTCG 908
                  ** ** ***************************** *********** ******** **
```

Figure 3.8A- Alignment of mouse and rat *Tsga10* 5' end sequences. The first four sequences (including AF530050) are mouse transcripts. Accession numbers of the first and second mouse cp431 in alignment: XM_136734, and AK030254 respectively. The first sequences. Green box is double methionine starting ORF of Tsga10. Red nucleotides are sequencing errors of my sequence (AF530050).

The presence of different sites for an ORF in a gene may result in a different promoter region, which in turn might have a role in regulating the transcriptional activity of that gene. To investigate the possible functional effects of any variation in the 5' flanking region of *Tsga10* or different splicing sites that start its ORF at different sites, it would be an idea for future studies to examine Tsga10 transcriptional activity, perhaps using the luciferase reporter gene assay (see discussion section 7.2).

Regarding conservation, besides human and mouse, for Tsga10 and its longer variant with 202 more amino acids in the 5' upstream sequence, there are ESTs or cDNA sequences for human, mouse, rat, bovine, hamster, zebra fish, chicken, pig and Chimpanzee. At the protein level for Tsga10, there are known amino acid sequences, which are submitted in GenBank for human, mouse and rat. However, for cp431 protein, amino acid sequences are submitted just for mouse (XP_136734), cow (XP_416892), and rat (AAC72234) species.

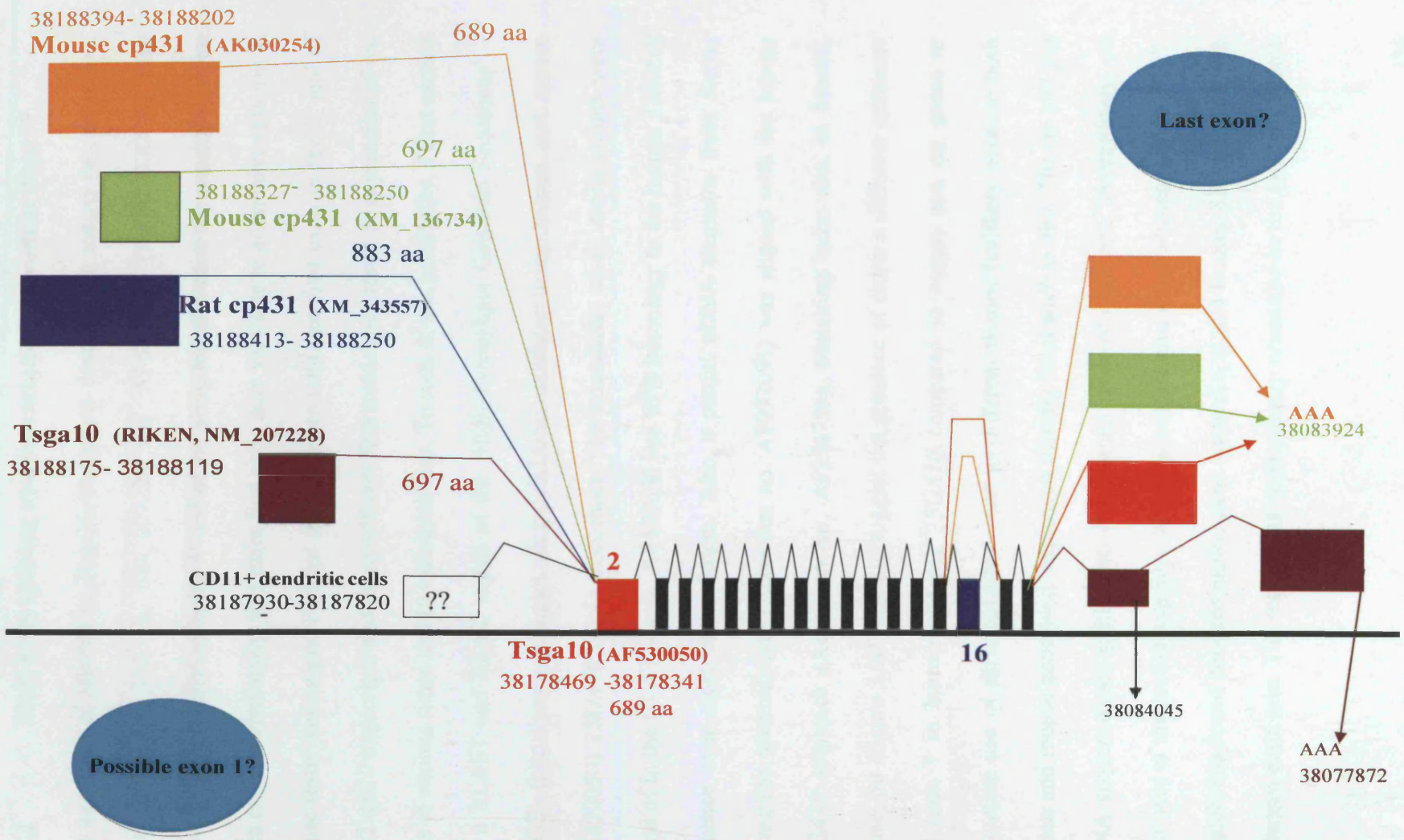


Figure 3.8B- Exons/introns boundaries and schematic structure of *Tsga10* gene and in its isoforms in mouse and rat.

3.8.1 First exon splicing and its variations in rodent *Tsga10*

The existence of alternative Tsga10 transcripts described in mouse (accession no. AK030254, XM_136734, NM_207228 and AF530050) (figure 3.8A,B) may represent polymorphism in the 5' end of the gene, but also raise the possibility of an extra exon. Additionally, this extra first exon may vary due to alternative splicing in mouse testis transcripts. Perhaps there is more than one first exon in the 5' end of the *Tsga10* transcripts and these different first exons may result in expression of the gene at various times and in various tissues. To look at this possibility, I first carried out a BLAST and BLAT search of the 'High-Throughput Genomic Sequences' – HTGS- (http://www.ncbi.nlm.nih.gov/HTGS/) database using human and mouse (AF530050) *TSGA10* cDNA sequences. The boundaries of all exons/introns were found of in the BLAT search. AF530050 had 89% homology to the human *TSGA10* sequence over most of its length. Also a partial cDNA sequence from *Rattus norvegicus* spermatocyte (Accession no. AF092091) was aligned with the human *TSGA10* sequence (Accession no. AF254756). Matching exon one in human, mouse, rat (figure 3.8B) and pig suggest the presence of either a different promoter and exon 1 in human and pig *TSGA10* compared to mouse and rat genes or alternative use of different exon ones in different tissues (unlikely because both human and mouse gene were sequenced directly from adult testis). The rat and pig cDNA sequences are partial but both contain most of exon one. Interestingly, on alignment of the mouse sequence to the human genomic clone sequence AC019097, a match was found between mouse exon one and a region located 8260bp upstream of human exon one. The region was appropriately orientated in the human genome

clone and showed 89% identity to mouse exon one (base 1 to 130). With the exceptions of exons 1 and 19, the positions of the mouse exon / intron junctions were conserved compared to the human gene.

On searching the Ensambl database I found one additional sequence to those found in the BLAST database (ENST000035053). This sequence was most similar to long cp431 and had 21 exons.

The 5' flanking region of the most highly conserved mouse exon one was sequenced from PAC 618H14. It was aligned with the corresponding 5' upstream region of the match site of mouse exon one in the human genomic clone AC019097 and the 5' flanking region of the genuine human *TSGA10* exon one also from AC1019097. 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. There is no significant match to the human exon one 5' region. Now, mouse *Tsga10* four transcripts were aligned with human *TSGA10* sequence using BLAT genome browser. Interestingly, only one of mouse transcripts (NM_207228) was found downstream (99072542) to the first possible human exon one. The upper part of figure 3.8.1A shows human transcripts mapped back onto the human genome and the lower part of it shows mouse and rat transcripts mapped back onto the human genome. Figure 3.8.1B shows mouse transcripts mapped back onto the mouse and human/rat genome respectively. Each transcript is shown in a particular colour.

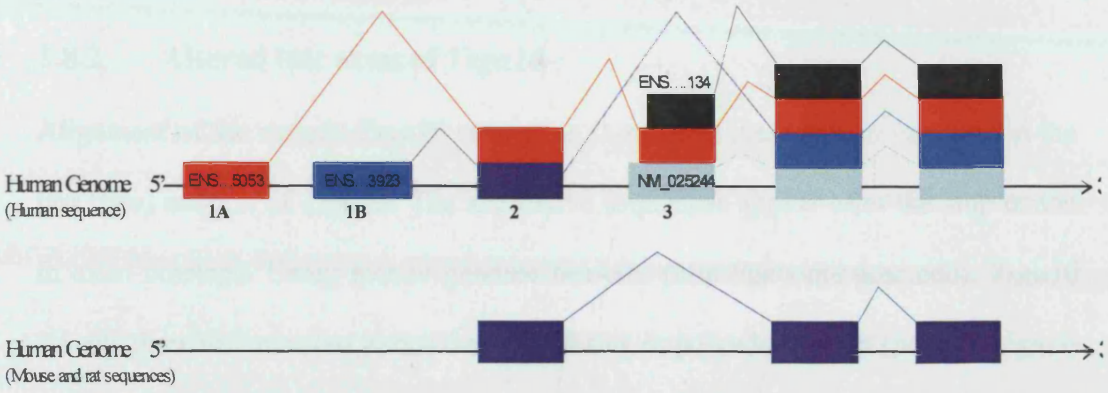


Figure 3.8.1.A- – Schematic location of human *TSGA10* 5' end exons and possible splicing by using *Ensembl*- The fourth exon in this schematic is conserved in human *TSGA10* transcripts but the third one by using mouse and rat *Tsga10* sequences. However, the first exon of hu TSGA10 seems to be Altered in mouse. Lower scheme shows the location of mouse and rat homologues of the 5' exons in human genome.

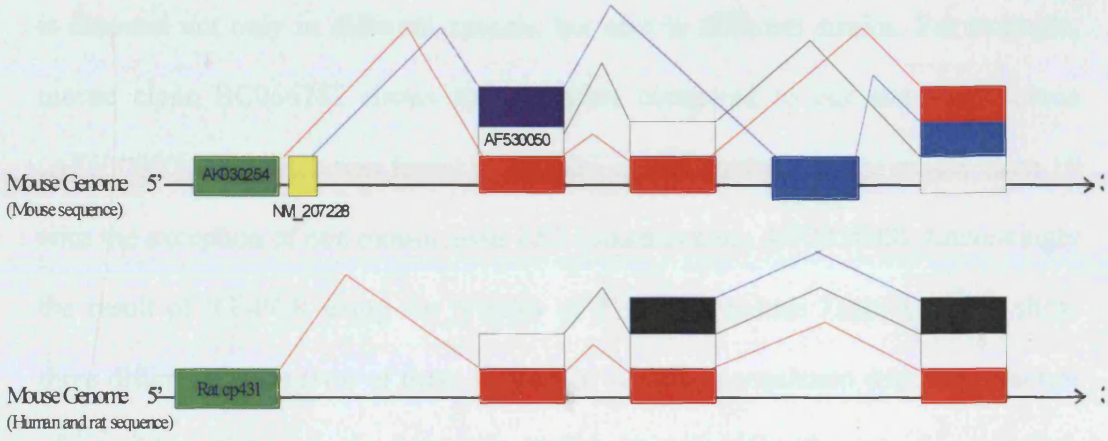


Figure 3.8.1.B – Schematic location of mouse *Tsga10* 5' end first exons and possible splicing by using *Ensembl*- The third exon in this schematic is conserved in mouse *Tsga10* transcripts but the fourth one by using human and the first one in rat *Tsga10* sequences. However, the first ex of human *TSGA10* seems to be altered in mouse. Lower scheme shows the location of rat and human homologues of the 5' end exons in mouse genor

It seems that there is a lot of alternative splicing for the first exon of the *TSGA10/Tsga10* gene, which may be tissue specific. Therefore, there may be alternative promoters that may be tissue specific. Although there are several possible exon one for the gene transcripts in different species, the rodent and human first exons of *TSGA10/Tsga10* gene are different and neither is found in the genome of the other species.

3.8.2 Altered last exon of *Tsga10*

Alignment of the various Tsga10 sequences shows that there is also variation in the last (two) exon(s) of *Tsga10*. The alternative sequences appear after the stop codon in exon nineteen. Using mouse genome browser (http://genome.ucsc.edu), *Tsga10* reveals three different last exons due to differing in polyadenylation (poly A) signals in mouse and rat transcripts (see figure 3.8B). Alternative splicing in exons 19 and 20 may result in three different transcripts. This variation in the last exon of *Tsga10* is detected not only in different species, but also in different strains. For example, mouse clone BC066782 shows this variation compared to our sequenced clone (AF530050). No match was found by BLAST in any database to the mouse exon 19 with the exception of one mouse testis EST (accession no. AV283000). Interestingly the result of RT-PCR using the primers of 3' end of mouse *Tsga10,* which show three different bands (two of them very close in size) is consistent with the presence of possible variation in the last exon of *Tsga10* and might illustrate this variation (see figure 3.4D).

3.8.3 Alternative splicing of *Tsga10* exon 16

As was described previously, exon 16, which is flanked by two long introns of 17 and 20 kb respectively, can be omitted from the mRNA. Alternative splicing is observed in other genes where exons are flanked by long introns, e.g. the human *PKC-δ* has an intron of 17Kb preceeding the first exon (Suh et al., 2003). The first exon of this gene may be spliced out completely without any mutational frame shift.

Considering the sizes of the Tsga10 protein components detected by Western Blotting, exon 16 may not only be spliced out completely resulting in no frame shift, but also may be the end of a transcript, resulting in the 65 KDa protein component detected by the anti-TN antibody (figure 4.2A).

Chapter IV:

Studying the expression of the protein encoded by Tsga10

Looking at the expression patterns of the Tsga10 gene gave an indication where to study expression of the protein. Data of protein expression can give a good indication of the function of a gene, based on its localisation in a tissue, its association with other proteins and its association with particular cellular structures.

I have studied the expression pattern of the TSGA10 protein using Western blotting, immunohistochemistry of tissue sections, and these different methods of subcellular localisation.

4.1 Raising antibody against Tsga10

Initially the Tsga10 amino acid sequence was analyzed in order to choose a region suitable to raise antibodies against. The Tsga10 ORF starting from double methionine (at aminoacid # 1) is conserved in all transcripts encoding the protein. This conserved region of the Tsga10 protein was considered for raising an antibody. Several other factors were considered and came into play when selecting peptides for antibody production. First and foremost were the hydrophilicity, then antigenicity, and finally accessibility. All of these parameters are not independent and take other parameters into account as well. For example, in order for a peptide to be antigenic, it should be fairly hydrophilic (chances of being on the surface are higher). If it is likely to be on the surface then it will be fairly accessible. Since it is always advisable to select more than one region to ensure production of useful antibodies, 3 peptides were selected from the Tsga10 protein sequence. Antibodies to various regions may serve different purposes. A given antibody may be more suitable for blotting, immunoprecipitation, histochemistry, etc. In addition, if both antibodies react with the same band, then we

can be pretty sure that they recognize the right protein. Of course cross-reactivity of these sequences with other closely related members and other desirable features specific to this protein could be considered. Also it was a good idea to study some possible post-translational modification of the protein.

In the Tsga10 protein sequence, the antigen prediction file generated by Peptide Structure (GCG) shows that it is not the best protein to work with due to the lack of proline residues apart from the N-terminus region. However, the N terminus looks as though it is completely exposed and could be worth considering. (Provided there is no reason to think it is chopped off post translationally.) Prolines, R/K and some hydrophobics are good to be included.

The peptides selected for raising polyclonal antibodies against Tsga10 as follows:

1) N-terminal- aa 2-15 + cysteine C-terminally for coupling: H2N - MRNRSKSPRRPSPT- $COOH_2$, very nice prediction, no alpha helix prediction, so the antibodies against them may recognize the native protein. This might be possibly the best peptide. A BLAST search shows the specificity of the selected amino acid sequence to TSGA10 protein.

2) Middle- aa 207-220 + cysteine N-terminally for coupling: H2N - CYENTEKDLSDTQRH- $COOH_2$, good surface probability and antigenic index, predicted rather alpha helical like most of the Tsga10 sequence, so the antibodies against this sequence probably work only in Western Blots. The result of a BLAST search shows the specificity of selected amino acid sequence to TSGA10 protein.

3) C-terminal- aa 676-689 + cysteine N-terminally for coupling: H2N - CHRFPDRGFDRSLEE - COOH (C-terminal). BLAST searches show the specificity of the selected amino acid sequence to TSGA10 protein.

Tsga10 peptides' sequences and their names are shown in table 4.1. There is evidence that the Tsga10 protein may be cleaved into two parts of 27 and 57 KDa (Modarressi et al., 2004, see section 4.2). As far as the antibodies are concerned, both the 'N-terminus' and 'middle' antibodies recognise the 27 KDa N-terminal portion. The sequences of the above peptides are conserved in all three (human, mouse and rat) species.

Table 4.1- Tsga10 peptides' sequences which antibodies raised against them

Name (number)	Peptide name and sequence		Position of amino acids in protein
TN (EP021582)	N-terminus peptide:	MRNRSKSPRRPSPT	2 – 15
TM (EP021583)	Middle peptide:	YENTEKDLSDTQRH	207 - 220
TC (EP0311142)	C-terminus peptide:	HRFPDRGFDRSLEE	676 - 689

After synthesis of the selected peptides, they were injected into rabbits, then the antisera were obtained by bleedings and affinity purified based on the protocol described (materials and methods). Titration of antibodies against Tsga10 peptides was determined by ELISA assay with the peptide as antigen. Of the four immunized rabbits, two generated antibodies that recognized Tsga10 N-terminus peptides TN and TM, and two generated antibodies against the Tsga10 C-terminus peptide TC. As shown in the following titration curves (figure 4.1), the immunization programs were successful, in that antibodies were raised against each of the peptides injected.

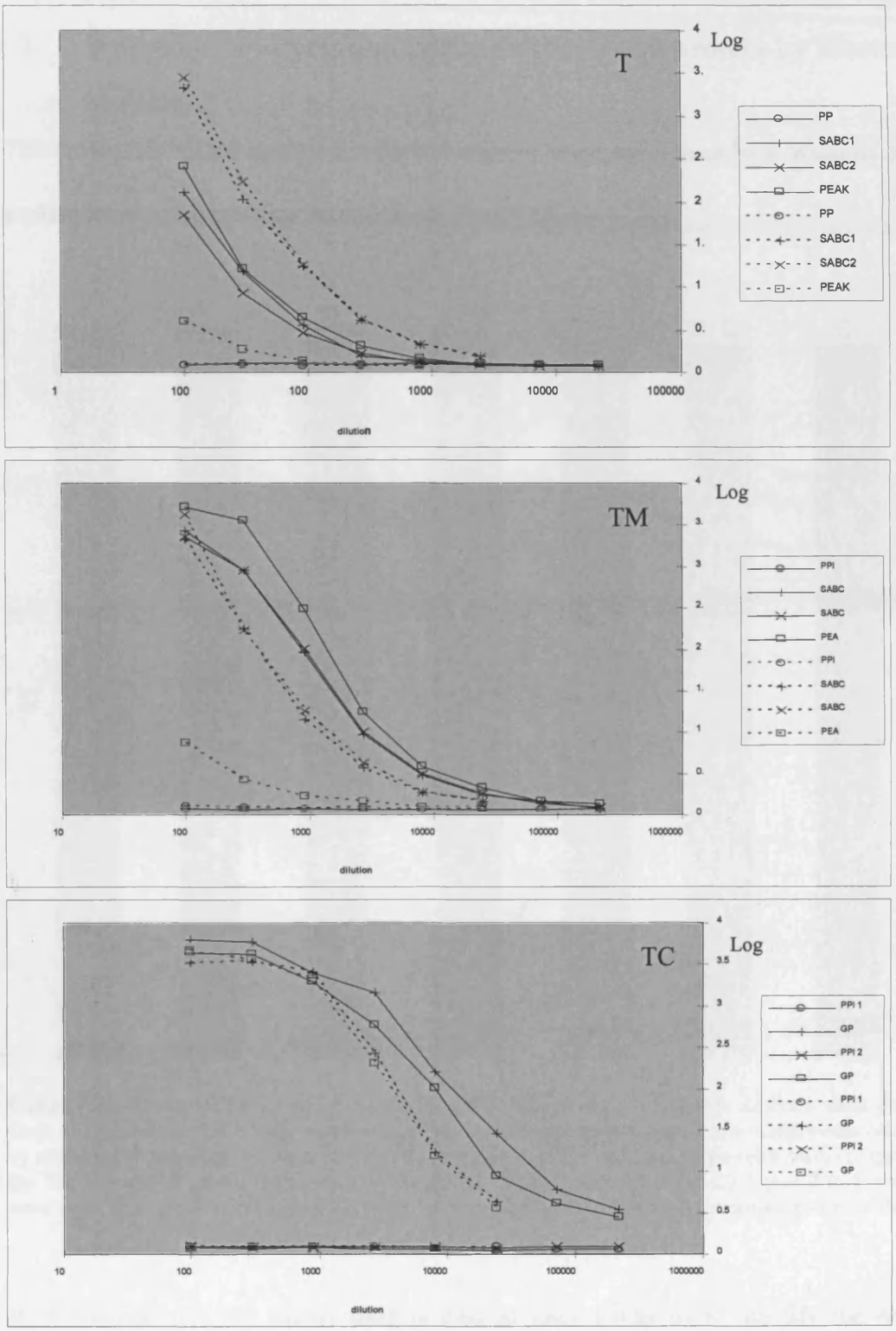


Figure 4.1- ELISA result for Tsga10 rabbit antisera against the N, C-termini peptides: TN (upper), TM (middle) and TC (bottom); Peptide (straight line) and Carrier (discontinuous) PPI= Pre-immune serum, SAB C1/C2 = Serum antibody in Rabbits 1 and 2 (2 rabbits are immunized), PEAK= affinity purified antibody

4.2 Studying the expression pattern of TSGA10 protein by Western blotting

The antibodies raised against the Tsga10 protein were hybridised to a Western blot of protein extracted from adult mouse testis (figure 4.2A).

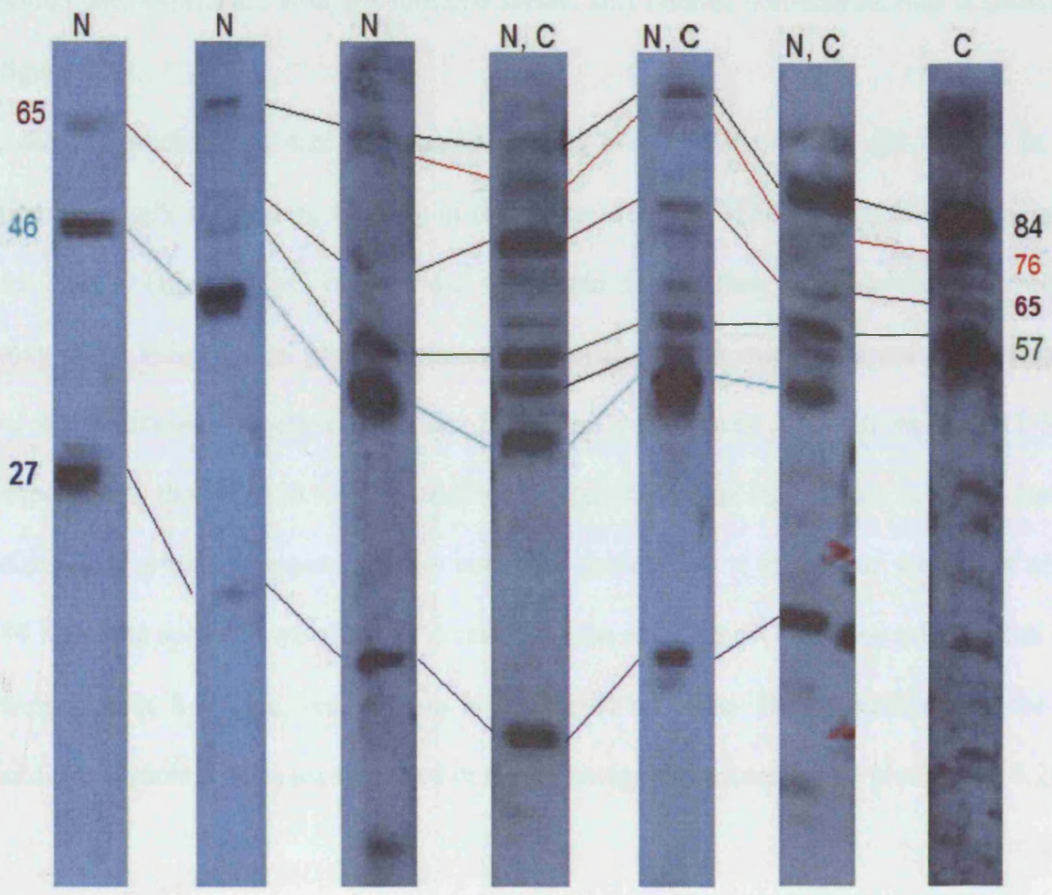


Figure 4.2A- Immunoblotting indicates Tsga10 protein expression in testis with different sizes. Panels of western blots containing protein extracts from testis were probed to illustrate the components recognised by either the N-terminus or C-terminus (N, C) peptides. A different subset of proteins were recognised by the TN, TM and TC antibodies than were recognised by both antibodies (N, C). Lanes 3 (N) and 7 (C) were taken from the blot in figure 4.2B, where a control hybridisation with pre-immune serum is shown.

In this experiment the panels were probed in such a way as to identify the proteins recognised by either the N or C terminus antibodies. Since the antibodies are produced as a mixture, it is necessary to block the blot with the peptide that is not required to

hybridise. In the membranes which were probed with TN antisera (i.e. blocked by TC peptide), there are bands of about 27, 46, 57, 65 and 84 KDs size for the testis protein while in the blot stained by TC antisera (i.e. blocked by TN and middle peptides), 84-KD and 57-KD protein components were prominent. No bands were seen in the blot, which was hybridised with pre-immune serum, this control non-stained blot is shown in figure 4.2B.

Ensembl predicts the size of Tsga10 protein to be 81261.60 Da (81 KDa). In the immuno-study by western blotting in testis, the presence of several protein components for Tsga10 (figure 4.2A) may be due to protein degradation during protein extraction, may be an indication of post-translational modification, or may represent cross reaction of the antibodies to other proteins. However, we already know from the RT-PCR experiments that *Tsga10* mRNA undergoes alternative splicing, so this accounts for the sizes of 2 protein components: the complete protein has a molecular weight of about 84 KDa, the splicing out of exon 16 results in the removal of 70 amino acids which will weigh about 8.5 KDa, resulting in a protein of 46 KDa. Other explanations for the additional protein sizes are explored in the following experiments (see also figure 4.2D).

4.2.1 Studying the expression pattern of Tsga10 proteins in developmental stages of embryogenesis by Western blotting

Using the same hybridisation protocol as the previous experiment, the expression pattern of the Tsga10 protein was studied during embryogenesis as well as in adult testis (figure 4.2B). In the case of the 10.5-dpc mice embryo protein extracts stained by the TN

antibody there are just 57 and 103-KDa prominent protein bands. In the 13.5 dpc mouse embryo there are some bands including 84 and 27 KDa. In the 17.5-dpc-mouse embryo 84 and 57 KDa bands seem to be the prominent protein extracts components. The non-specific bands might be due to further processing of Tsga10 during developmental stages of embryogenesis. In the blot stained by TC antibody just a bright band of 84 KDa protein component is observed in 10.5-dpc embryos. Apart from this 84-KDa protein, a ~30 KDa protein band in 13.5 dpc and the 57 KDa protein band in 17.5 dpc mouse embryo were detected.

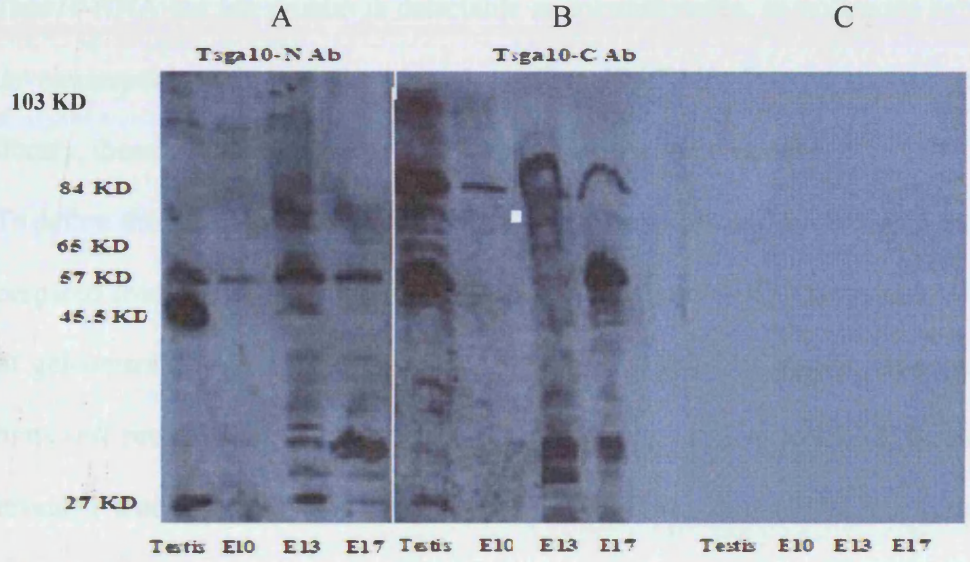


Figure 4.2B- Western blot of protein extracts from embryos and adult testis.
In each panel lane 1 is protein extract from mature testis, lane 2 is extract from 10.5 dpc mouse embryo (E10), lane 3 is from 13.5 dpc mouse embryo (E13) and lane 4, 18.5 dpc mouse embryo (E18). In **A**, the blot was hybridised to the N-terminus antibody (TN). In **B**, it was hybridised with the C-terminus (TC) and in **C** it was hybridised with the pre-immune sera as a control. Different sized proteins are recognised by two ends of the protein, and also at different stages of embryogenesis.
In the blot which was stained by **TN antibody**: there are 6 bands of about 27, 30, 45.5, 57, 76 and ~103 size for the testis protein while just 57 and 84-KD bands are prominent in 10.5 dpc mouse embryo. In 13.5 dpc mouse embryo prominent protein components of 76, 57, 30 and 27 KDa are detected. In 17.5 dpc mouse embryo 76 and 57 bands seem to be detected as main protein components.
In the blot which stained by **TC antibody**: ~103, 84 and 57-KD protein components are stained in testis (although 76 and 65 KDa faint bands were seen as well) while just a bright band of 84-KD (and a faint 103-KDa) protein component is observed in 10.5 dpc. Apart from this 84-KD protein, another 30-KD protein bands in 13.5 dpc mouse embryo and 57-KD protein component in 17.5 dpc mouse embryo are observed.

4.2.2 Examining Tsga10 protein expression in sperm fractions by Western blotting

To further characterize the Tsga10 protein and analyze its temporal expression pattern during spermatogenesis, germ cells were elutriated and analyzed. Three fractions were obtained containing pachytene spermatocytes, round spermatids, and elongating spermatids. TN antisera recognized a 65-KDa band in germ cells. This analysis also showed that Tsga10 protein is most abundant in elongating spermatids (figure 4.2C(A)). A small amount of Tsga10 protein is visible in the extract from round spermatids, which may be due to of the presence of elongating spermatids in that cell fraction. Because *Tsga10* RNA but not protein is detectable in spermatocytes, as evidenced by the present developmental study and the cloning of *Tsga10* cDNA from a spermatocyte cDNA library, these results suggest the possibility of translational control.

To define the location of Tsga10 in cells, nuclear extracts and cytoplasmic extracts were prepared from germ cells. Protein loading was analysed using Coomassie Blue staining of gel-separated proteins (not shown). The same amount of protein was transferred to blots and probed with all three anti-Tsga10 antiserum. The result of immunoblotting revealed that the protein is present in the cytoplasm, but not in the nucleus (figure 4.2C(B), lanes 2 and 3). A control demonstrates that, as expected, the CREM τ transcription factor is predominantly present in the nuclear extract (figure 4.2C(B), lanes 4 and 5). Figure 4.2C (C) shows the results of an experiment designed to examine the association of the Tsga10 protein with different regions of the sperm tail. In the first lane of protein extract from whole sperm tails a small amount of the 65 KDa Tsga10 protein is detectable, but there are an additional 2 proteins of 32 and 27 KDa. Purification of the two sperm tail components outer dense fibers (ODF) and fibrous sheath (FS) shows that

the protein represents the N-terminal portion of the protein since it hybridises to the TN and TM antibodies (figure 4.2C(D)). The TC antibody hybridises to the whole 65 KDa protein but not to the 27 KDa component. It seems that the protein is cleaved post-translationally into two portions, one of which is associated with the fibrous sheath. This will be examined further in section 4.4.3.

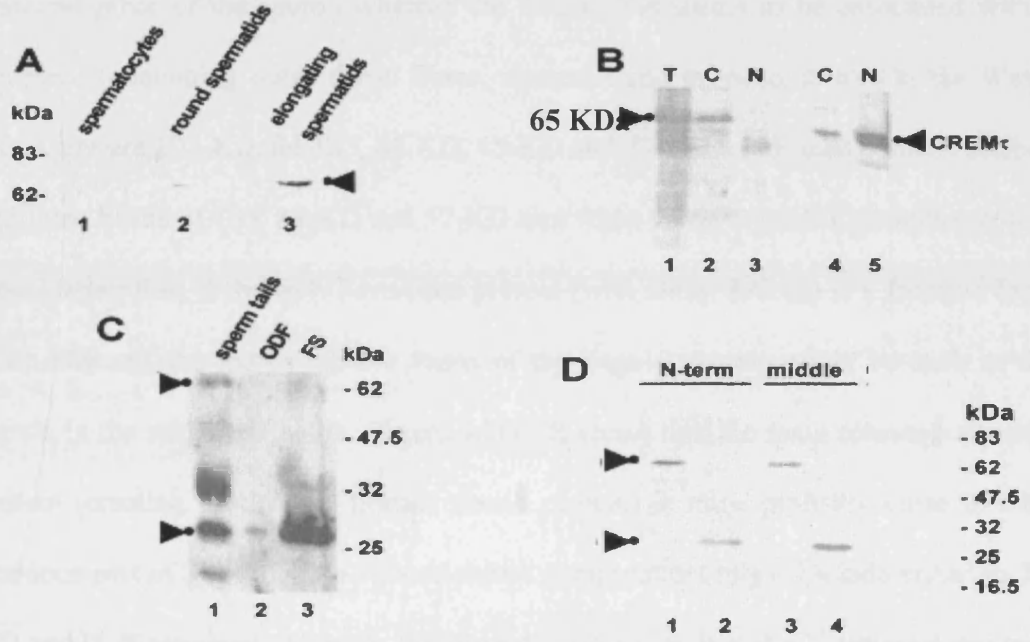


Figure 4.2C- Immunoblot analysis of Tsga10 protein. **A**) Indicated germ cell fractions (lanes 1–3) were obtained from adult rat testes using centrifugal elutriation. Cells were lysed and used for Western blot analysis using TN, TM and TC antiserum. Lanes 1–3: spermatocytes, round spermatids, and elongating spermatids, respectively. Note that 65-kDa Tsga10 is predominantly expressed in elongating spermatids. **B**) Western blot analysis of total testis (T), cytoplasmic (C), and nuclear (N) fractions of testicular germ cells using anti-Tsga10 antiserum raised against all three peptides (lanes 1–3, respectively). Loading of equal amounts of cytoplasmic and nuclear samples was analyzed by Coomassie Blue staining before immunoblotting (not shown). As a control, cytoplasmic and nuclear proteins were also analyzed using anti-CREMτ antibody, which detects the testicular CREMτ transcription factor (lanes 4 and 5, respectively). Note the presence of Tsga10 in cytoplasmic fractions. **C**) Analysis of Tsga10 in sperm tail. Extracts from rat sperm tails (lane 1) and two major components of rat sperm tails, ODF and FS (lane 2 and 3, respectively), were gel separated and used for the immunoblot analysis using anti-Tsga10 antiserum raised against all three peptides. Note that FS, but not in ODF, contain a prominent 27-kDa immunoreactive Tsga10 protein. **D**) Immunoblot analysis of 65-kDa Tsga10 and mature 27-kDa FS Tsga10 was carried out using affinity-purified antibodies against peptides located in three different regions of Tsga10 protein. Lanes 1 and 3: cytoplasm fraction from testicular germ cells. Lanes 2 and 4: FS proteins isolated from sperm tails. Lanes 1 and 2: filters were probed with affinity-purified antibody against a peptide located at the N-terminus (amino acids 2–15: N-term). Lanes 3 and 4: filters were probed with affinity-purified antibody against a peptide located at amino acids 206–219 (middle).

4.2.3 Discussion: Tsga10 protein processing and modification

These experiments studying the localisation of the two halves of the Tsga10 protein strongly indicate that the two halves are independently located in the sperm, and therefore may have different functions. This requires that the protein is modified post-translation by cleavage. The N-terminus is associated with the fibrous sheath in the principal piece of the sperm, whereas the C-terminus seems to be associated with the midpiece (containing outer dense fibres, axoneme and mitochondria). On the Western blot, there are 103-KD, 84-KD, 65-KD, 45-KD and 27-KD bands using anti-N antibody, and three bands of 103, 84-KD and 57-KD size when using Tsga10-C-terminus antisera. Considering this, if the cp431-encoded protein (with about 892 aa) is a form of Tsga10 (with 689 aa), the most probable forms of the Tsga10 protein might be such as those shown in the schematic below (figure 4.2D). It seems that the main cleavage site of the protein (creating the 27-KD fibrous sheath protein) is most probably close to the N-terminus part of Tsga10. This fibrous sheath compartment might include either an 18.5-KD cp431 N-terminus upstream component (202 aa) or just the N terminal regions of Tsga10. The 65-KD protein component, which was detected by the anti-TN antibody, may be the result of translation of a possible alternative splice variant of Tsga10 containing just 15 exons.

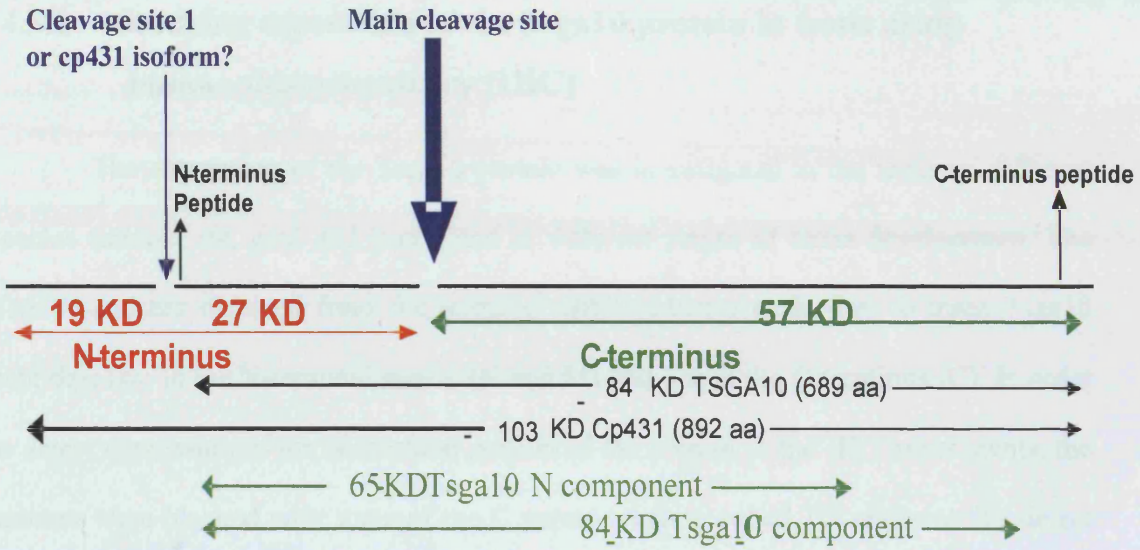


Figure 4.2D- Schematic feature of possible Tsga10 cleavage sites resulting in its post-translational modification. Cleavage may be less likely occurs very close to Tsga10 N-terminus resulting in 19-KD cp431 N-terminus component to be part of 27-KD N-terminus FS protein. More likely, the cleavage occurs in a place of Tsga10 protein where all of 27-KD FS component is the first Tsga10 N-terminus. 65-KD component could be the product of a *Tsga10* transcript containing 15-exon.

4.3.1 Studying expression of the Tsga10 protein in testis using immunohistochemistry (IHC)

The expression of the Tsga10 protein was investigated in the testis of different species (mouse, rat, goat and boar), and in different stages of testis development. The Tsga10 antisera obtained from the injected rabbits contain antibodies to three Tsga10 peptides: two in the N-terminal region (N and M) and one at the C terminus (C). In order to detect expression of the N-terminal portion of the protein in the IHC experiments, the antisera were blocked with some of the C peptide (hereto called TN antisera). To detect the C terminus the antisera were blocked with N and M peptides (TC antisera). As a control the preimmune sera was used. In immunohistochemistry studies of 21-day (immature) mouse testis using TN antisera, we found a limited number of primary spermatocytes expressing the Tsga10 protein N-terminus (figure 4.3.1A). Staining was stronger in mature mouse testis (figure 4.3.1B). The N-terminus of the Tsga10 protein was seen weakly staining a filamentous/fibrous shape in primary spermatocytes, but stronger staining was associated with elongated spermatids at stage V, and mature sperm. Germ cells and spermatogonia were not stained using this antibody. Immunohistochemistry on mature rat testis using the TN antiserum showed staining specifically of elongated sperms. However, when the C terminus was used staining was seen in the cytoplasm of almost all spermatogonial cells within the seminiferous tubules (figure 4.3.1C (A)(B)). These patterns of changes in localisation of the Tsga10 protein during the different stages suggest Tsga10 is physiologically significant in spermatogenesis.

In addition the different locations of the staining of the two halves of the protein may correlate with the N-terminus being associated with the fibrous sheath of more mature sperm. However, the difference in staining pattern observed between anti TN and anti TC may be due to posttranslational modification of the protein (as shown in Western Blotting). Alternatively, it might be due to differential accessibility of the antigens; in other words the N-terminal might be buried in one place and the C-terminal in another.

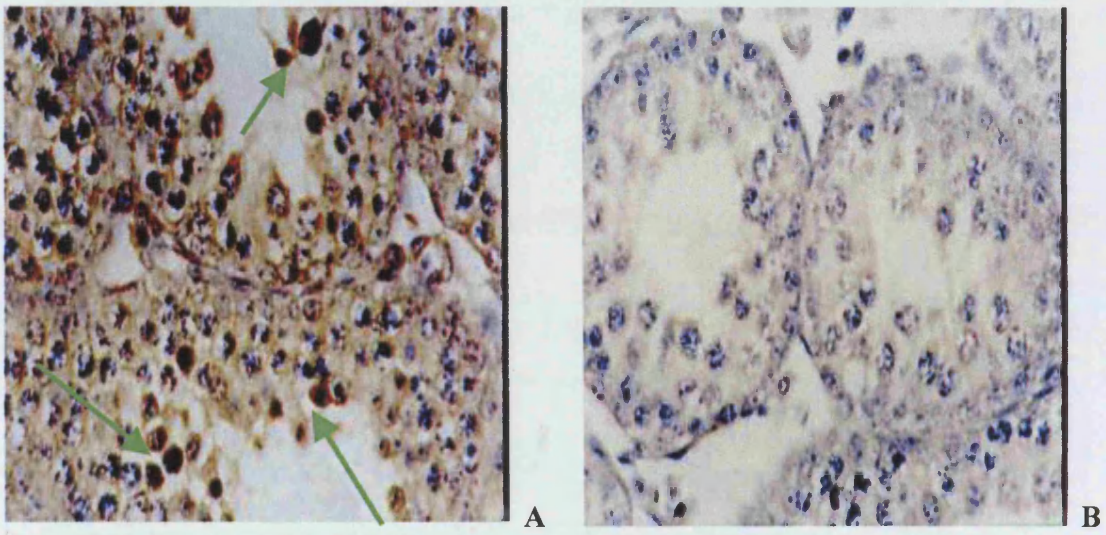


Figure 4.3.1A- Immunostaining of immature mouse testis with the Tsga10 N-terminal peptide. Adjacent sections of a 21 day after birth (immature) mouse testis were hybridised to **(A)** Tsga10 antisera blocked with C terminal peptides, leaving the N terminal peptides available for hybridisation and **(B)** Tsga10 preimmune serum as a control. Positive staining (brown colouration) is seen in a small number of secondary spermatocytes (indicated by green arrows) in (A). No staining is seen in the control experiment (B), just the blue H&E background stain.

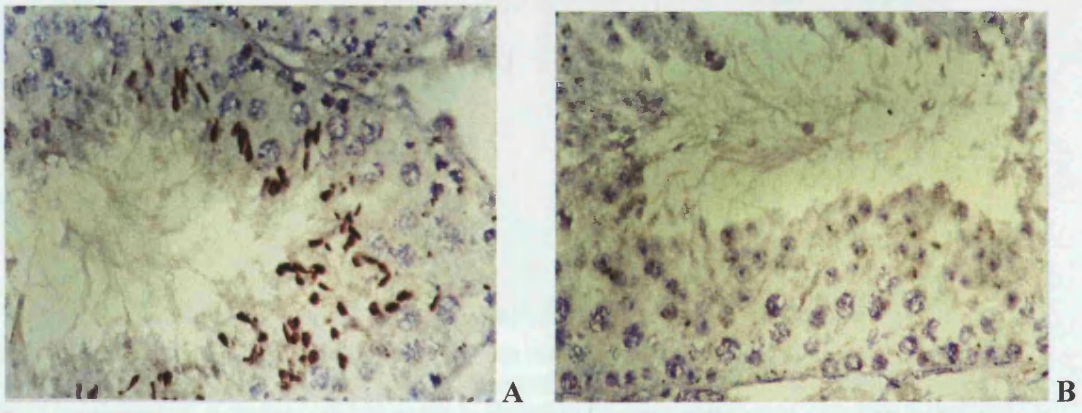


Figure 4.3.1B- Immunostaining of mature mouse testis using the Tsga10 N-terminal peptide. Adjacent sections were stained with **(A)** the TN antiserum and **(B)** the preimmune serum as a control. Strong brown staining can be seen in (A), and is associated with elongated spermatids.

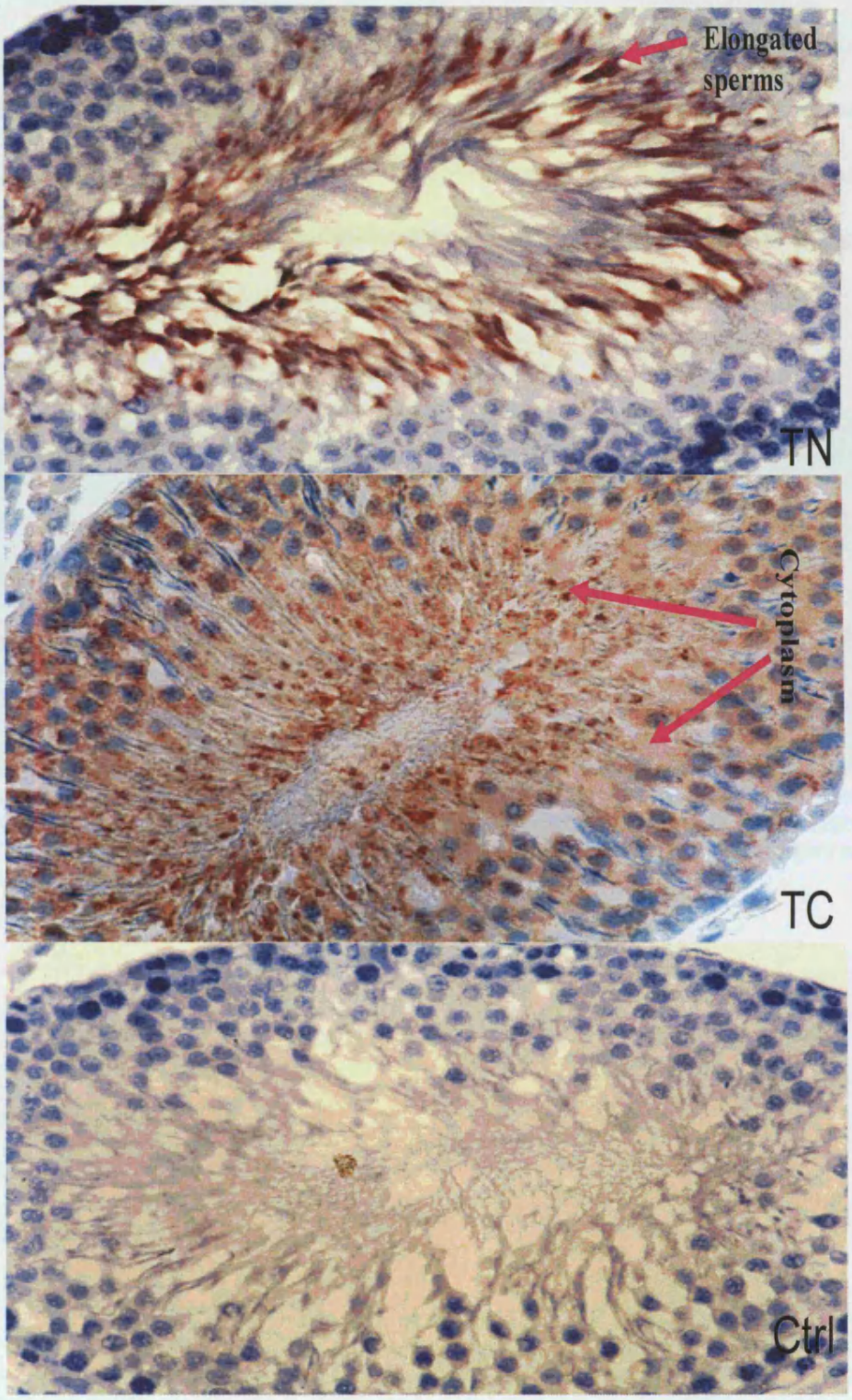


Figure 4.3.1C- Immunostaining of mature rat testis (x400). Adjacent sections of mature rat testis were hybridised to **(A)** the TN antisera, **(B)** the TC antisera and **(C)** pre-immune sera. The N-terminus of the Tsga10 protein is detected predominantly in elongated sperms, whereas the C-terminus is detected in the cytoplasm of almost all spermatogonial cells (B). Only blue background staining is seen in the control experiment (C).

Immunohistochemistry was also carried out on mature pig and goat testes using the TN antisera. Staining was seen in both species. In the pig it seemed that the staining was localised in the seminiferous tubules and epithelium, but it is difficult to say exactly due to structural difference from the slides being prepared from *retes* testis (figure 4.3.1D). The staining in the goat appeared to be associated with elongated spermatids.

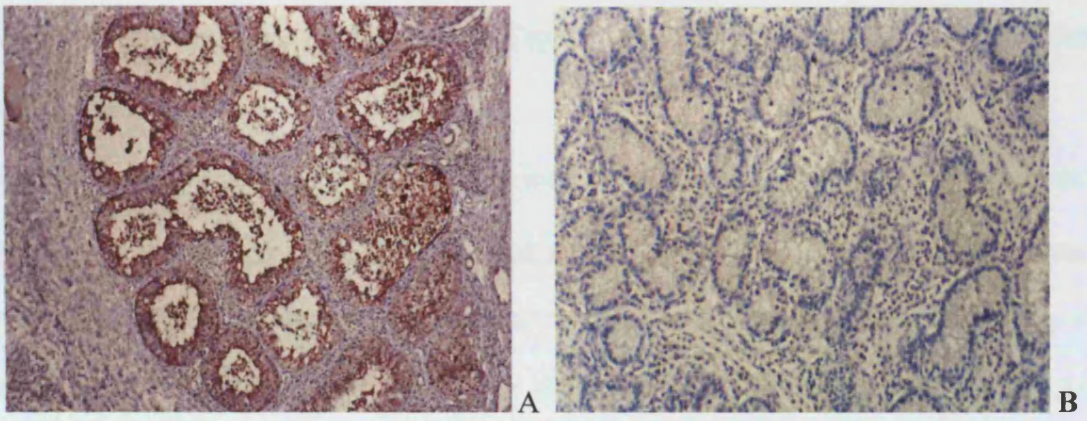


Figure 4.3.1D- Mature pig (boar) testis stained with the Tsga10 antisera. In **(A)** the section is stained with the TN antisera and in **(B)** the section is stained with pre immune serum as a control. Brown staining can be seen in (A) but, due to the morphology of the *retes* testis, it is difficult to say which cell types it is associated with.

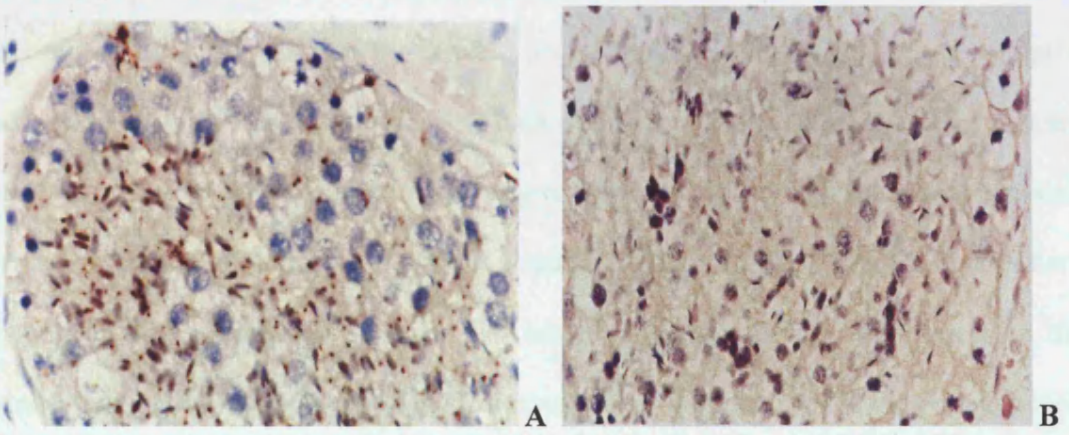


Figure 4.3.1E- Immunostaining of mature goat testis with Tsga10 antisera. In **(A)** the section is stained with the TN antisera and in **(B)** the section is stained with pre immune serum as a control. Brown staining is seen in (A), and may be associated with elongated spermatids, but it is not possible to be certain of this localisation from these sections.

4.3.2 Examining Tsga10 protein expression in brain and glia cells by immunohistochemistry

Having examined the expression pattern of the Tsga10 protein in testis, and seeing that it appears to be spatially and temporally regulated, I continued to look at its expression in other tissues. From the RT-PCR results, it was clear that the RNA is expressed in neural crest cells and in brain (sections 3.4 and 3.5). In addition, on genome database searching, a large number of ESTs homologous to Tsga10 were found in medulla and cerebellum cDNA libraries.

In these series of experiments, using adult mouse brain, whole brains were dissected and serial sections of 10-12 microns including all parts of the brain were obtained. From approximately 100 sections 4 or 5 sections representative of each different brain region were selected and transferred to a single slide ready for hybridisation. Figure 4.3.2A shows transverse sections of the cerebellum and the hippocampus and surrounding areas (cortex and thalamic nuclei) in adult mouse brain. The stained parts of the brain seem to be every cell of glia origin, including the granular layer of the cerebellum, hippocampus, thalamus and subventricular zone. Interestingly, the signals of Tsga10 immunoreactivities were observed using both of TN and TC antisera suggesting that the whole protein is expressed in these dividing neural cells. In the hippocampus there were certain distinct areas that did not express Tsga10, which correlated with the areas where pyramidal neurons are found. Pyramidal neurons are also found in the areas in the cerebellum where Tsga10 was not being expressed. There is staining in the cortex, which is apparent throughout the white matter and is most likely again to represent glial cell origin. Glial cells divide throughout adulthood and are first present in the foetal cortex

towards the end of development. The presence of staining in the foetal brain embryos from day 10.5 dpc onwards indicates Tsga10 protein expression in prenatal brain when active axonal transport is happening. Also Tsga10 protein expression was evident in astrocytes (neuroglia cells) (figure 4.3.2B).

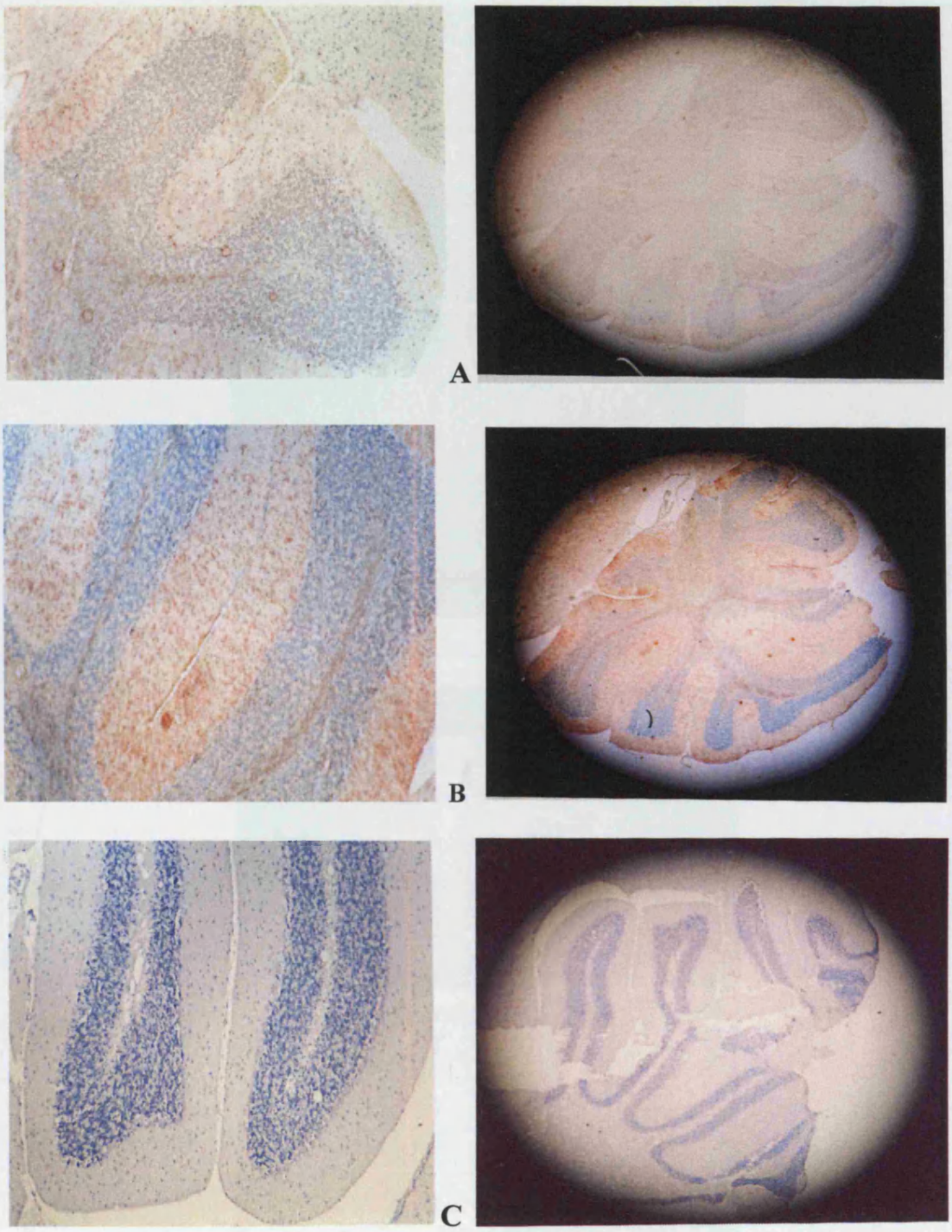


Figure 4.3.2A- Left panel consists of sections of dissected cerebellum, fixed in Bouin's were stained with DAB and counterstained with H&E (x100). The right panel shows the sections of dissected cerebellum at a lower magnification (x10). The sections A were hybridised with the TN antiserum, in (B), TC antiserum and in (C) the control was blocked with N, M, C peptides. Brown staining (DAB) is seen with both N and C peptides in the granular layer, but not in the control experiment (C).

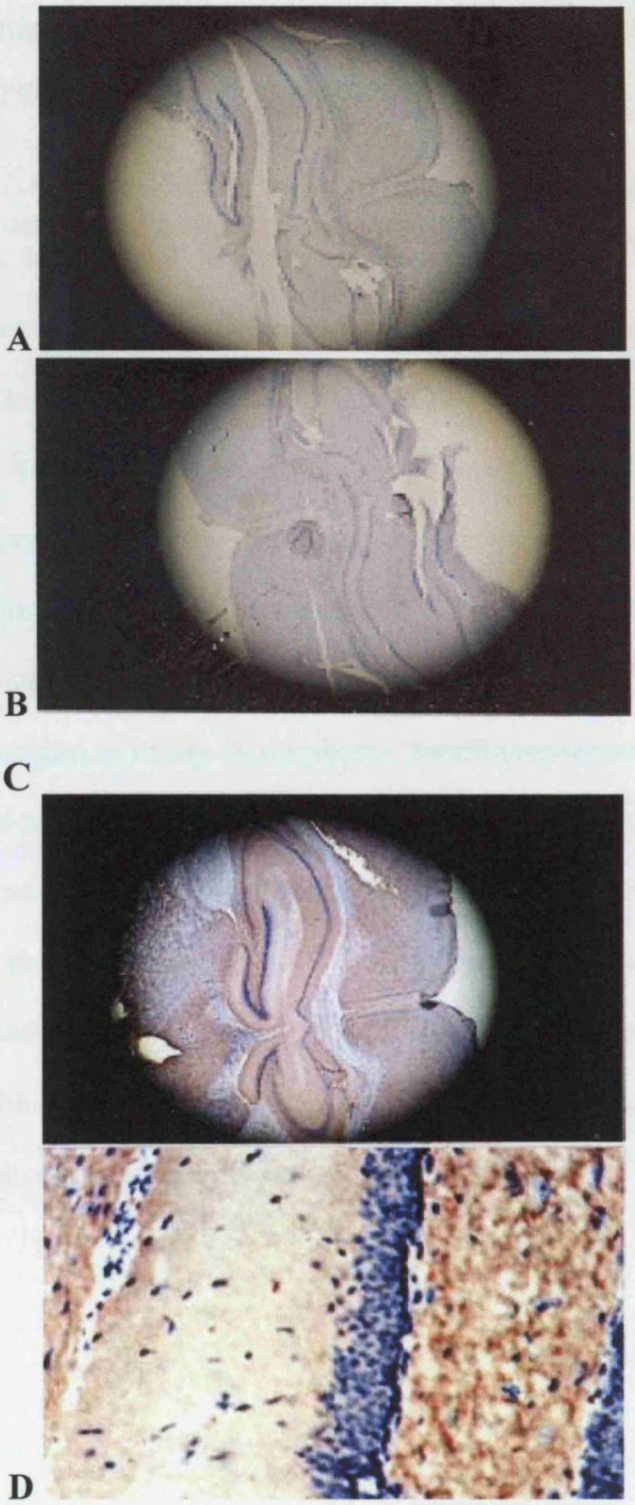


Figure 4.3.2B- Immunostaining of the adult mouse brain subventricular zone. (A) x10 control experiment blocked with all three Tsga10 antibodies prior to hybridisation and staining with DAB and H&E. No brown staining is seen in either the control experiment or when the section is hybridised with the TN antibody (B). When hybridised with the TC antibody brown staining can be seen at 10x magnification (C), and more clearly at high magnification in (D), x400. The staining seems to be localised in the granular layer, which includes astrocytes.

4.3.3 Looking for expression of Tsga10 by Immunostaining whole embryos at various developmental stages

From the RT-PCR experiments, which showed that the *Tsga10* gene is expressed in testis and brain, a high level of expression was also seen in RNA extracted from embryos at different stages of development (section 3.4). Also, the alternative splice variant seemed to appear only at the later stages of embryonic development. It is most likely that Tsga10 is expressed in actively dividing tissues. To verify this hypothesis I examined the expression of the Tsga10 protein in whole embryo sections by immunostaining with the TN and TC antisera.

The following series of figures show mouse embryos at various stages of embryogenesis from day 10 to day 19 (newborn). Paraffin-embedded sections of 10-12 microns hybridised to either the Tsga10 N-terminal (TN) or C-terminal (TC) antibody. For control experiments the antibody mix was blocked with all 3 of the Tsga10 peptides. The slides were stained with DAB, and counterstained with H&E. Brown signal represents hybridisation. The figures are shown in sequence of increasing age and show amplifications of the regions where hybridisation to the probes was seen. A summary of the results is presented following the figures.

Figures 4.3.3A-K: Immunostaining of whole embryo sections at various stages of development. In each case adjacent sections were stained with either the TN, or TC antibody, or, as in controls the antibody mix was blocked with all three Tsga10 peptides. The brown signal represents staining. In each of the following figures is shown the age of the embryo (days post coitus), the magnification, the antibody used, and a brief description of the tissue in which hybridisation is seen.

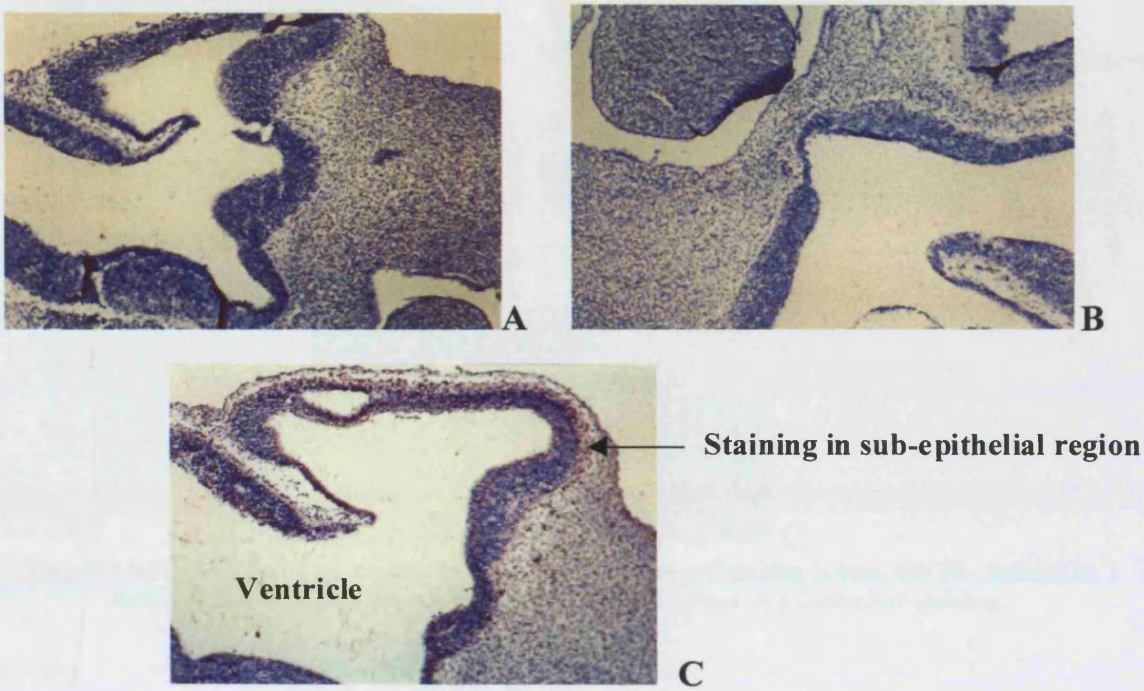


Figure 4.3.3A- 10.5 dpc, Brain, (x100), (A) TN-no staining is seen, (B) Control- no staining is seen, (C) TC-staining is seen in the subepithelial region.

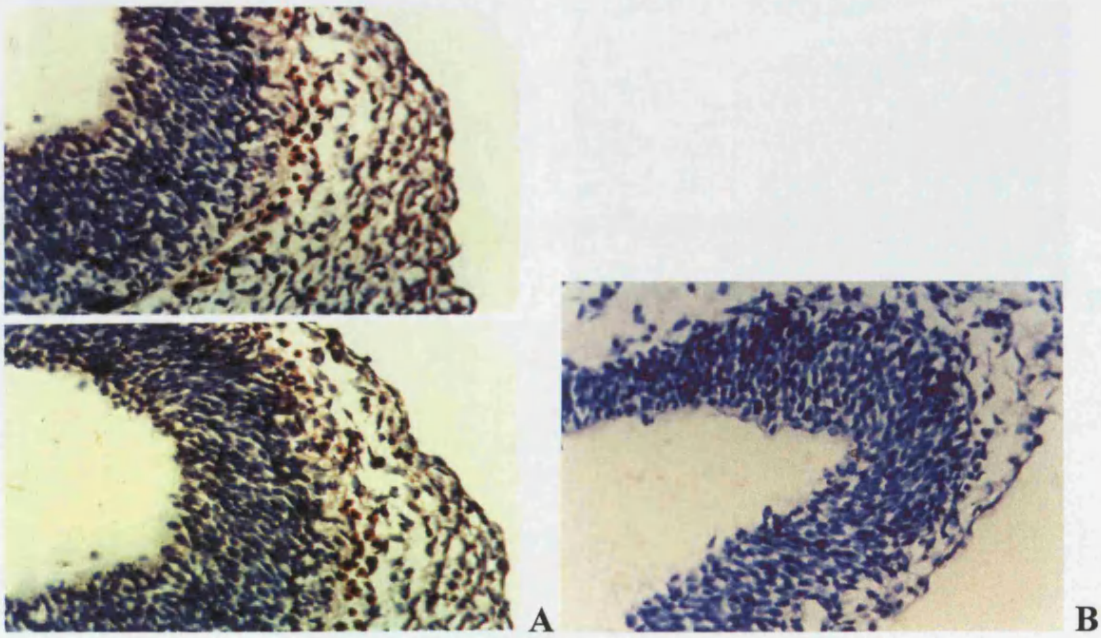


Figure 4.3.3B- 11.5 dpc, Brain, (x400). (A) TC- staining is seen in some non-differentiated cells in the sub-epithelial region of the brain. (B) Control-no staining

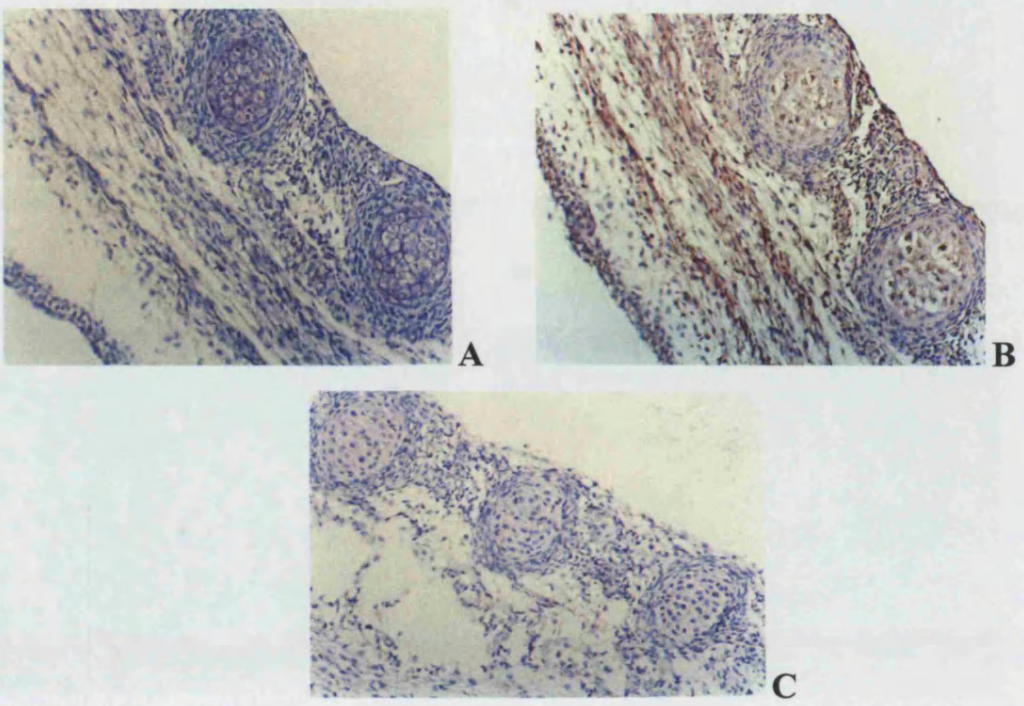


Figure 4.3.3C- 14.5 dpc, Spine, x200. (A) TN- a small amount of staining is seen. (B) TC- staining is localised in the striation of muscle, plus within the vertebrae. (C) Control-no staining.

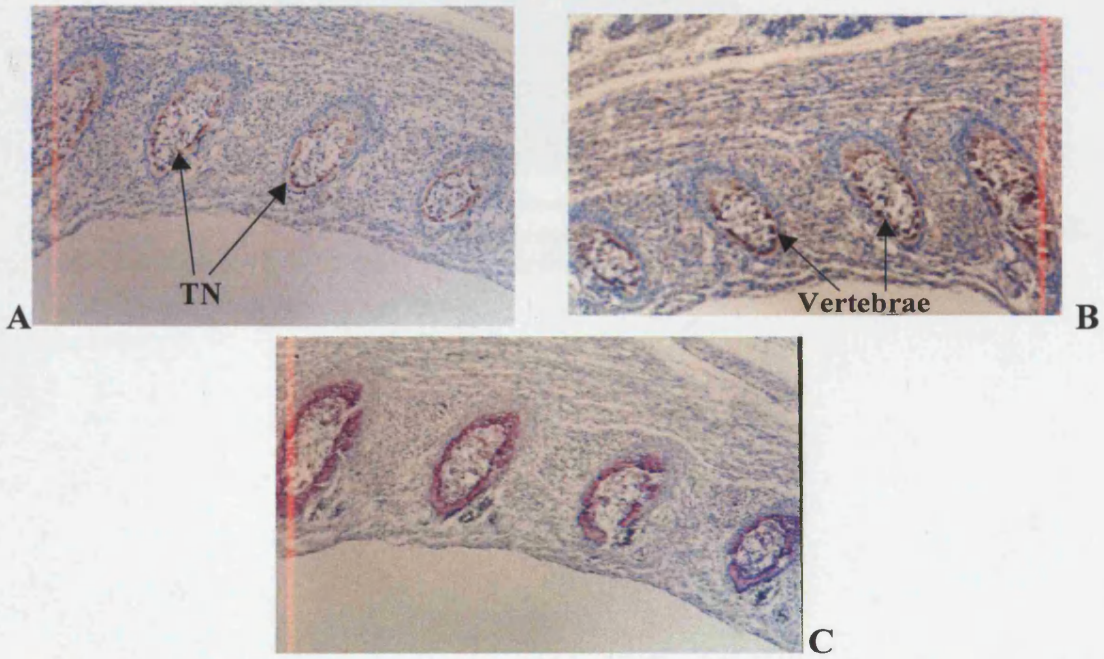


Figure 4.3.3D- 15.5 dpc, Spine, (× 100). (A) TN, (B) TC-with both the TN and TC antibodies intense staining is seen in the vertebrae. (C) Control-no staining

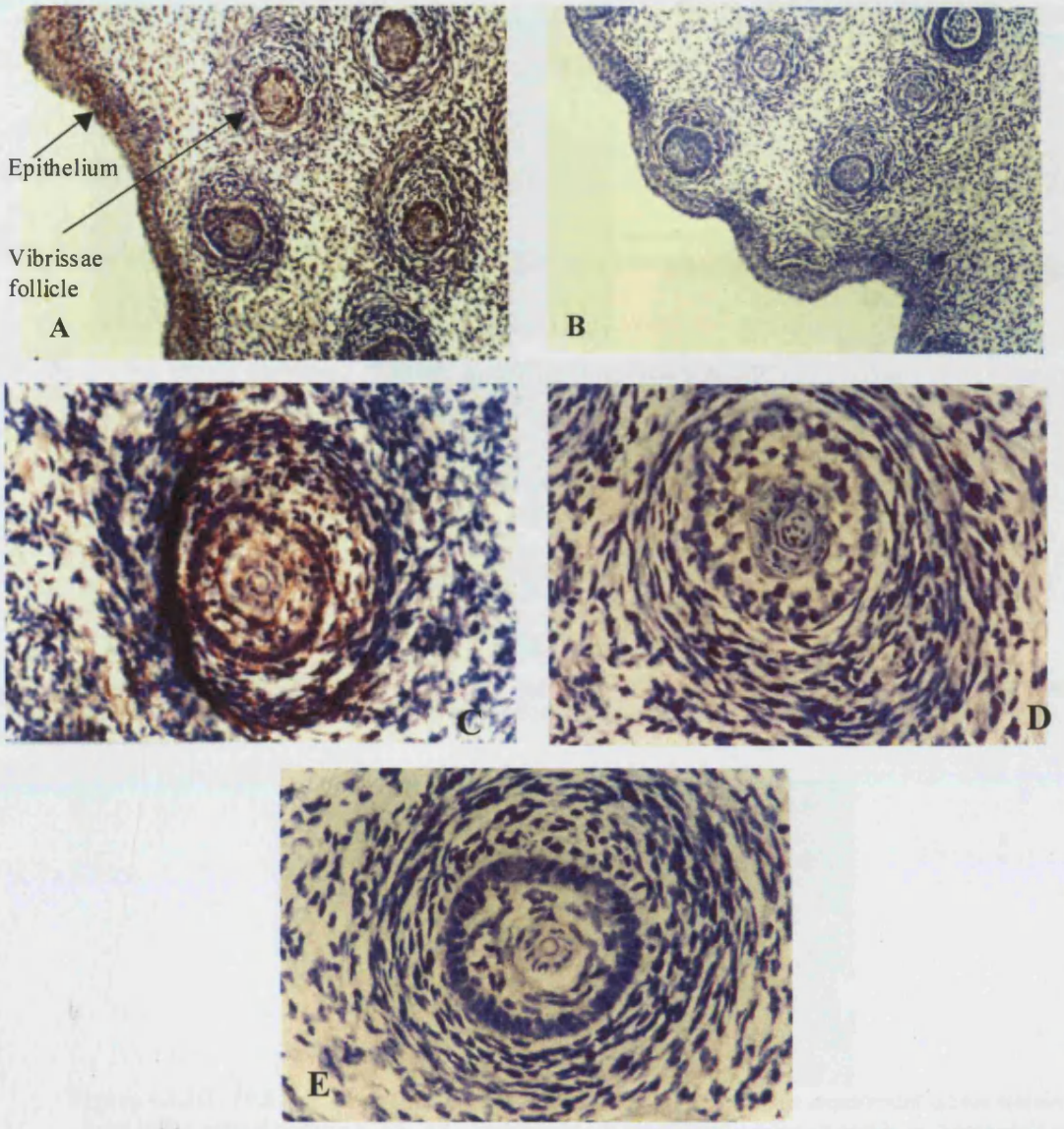


Figure 4.3.3E- 15.5 dpc, Whiskers (Vibrissae follicles), (A-B, x200), (C-E, x630).
(A, C) TN-staining can be seen in the vibrissae follicles and the ectodermal epithelium. (B, D) TC- some faint staining in the follicles can be seen at higher magnification. (E) Control- no staining is seen.

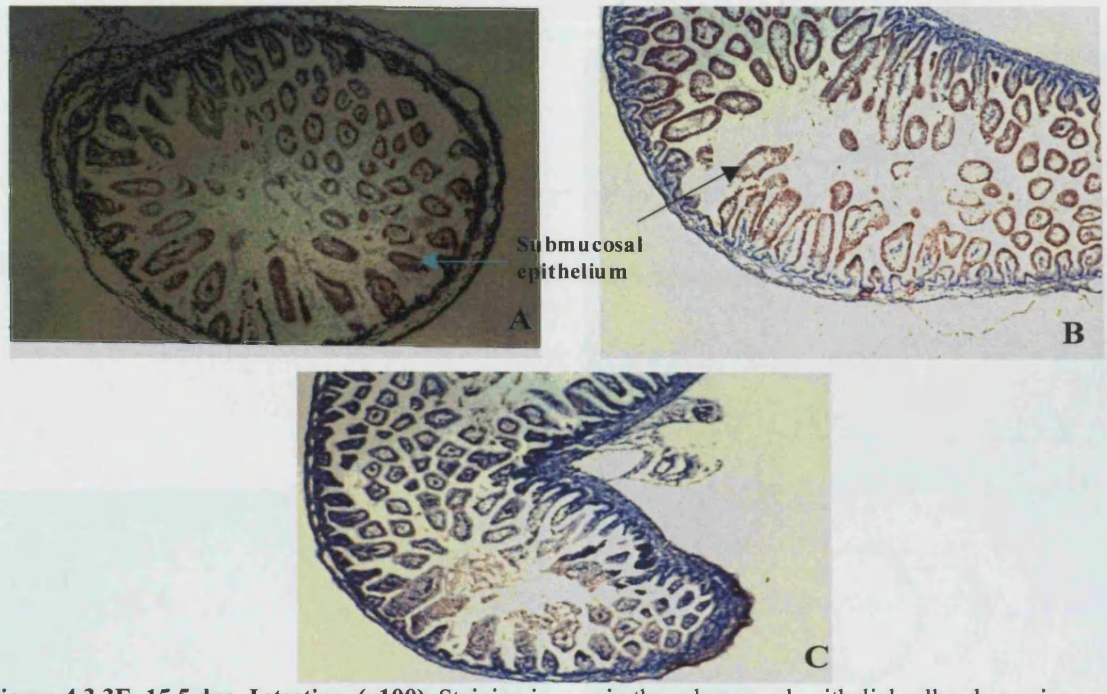


Figure 4.3.3F- 15.5 dpc, Intestine, (x100). Staining is seen in the submucosal epithelial cells when using both the TN and TC antibodies for hybridisation. In the control experiment (C) no staining is seen.

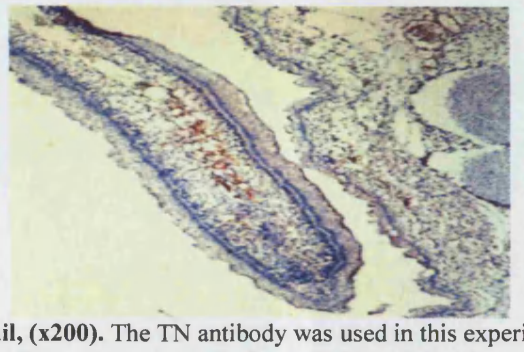


Figure 4.3.3G- 16.5 dpc, tail, (x200). The TN antibody was used in this experiment where staining can be seen in the central portion of the tail - perhaps this is proliferative bone or cartilage. Some staining may also be seen in the lower visceral organs.

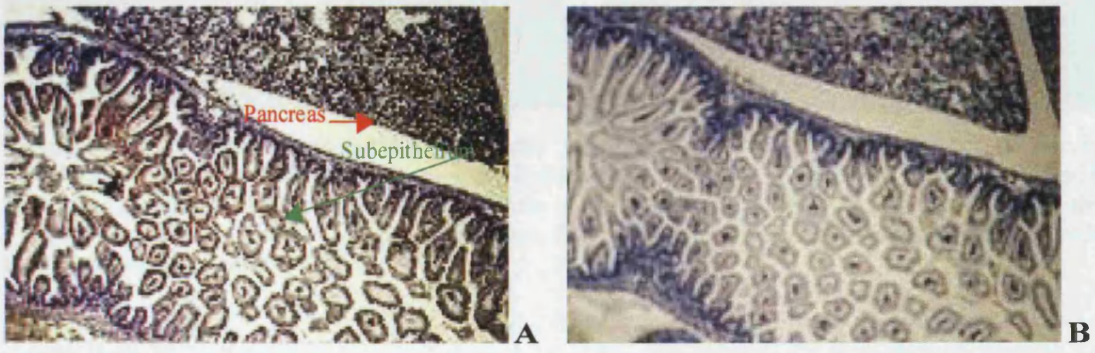


Figure 4.3.3H- 17.5 dpc, Pancreas, (x100). In this experiment (A) the TC antibody hybridizes to the subepithelial layers of the intestinal mucose. These cells are derived from neural crest cells. (B) Control- no staining is seen.

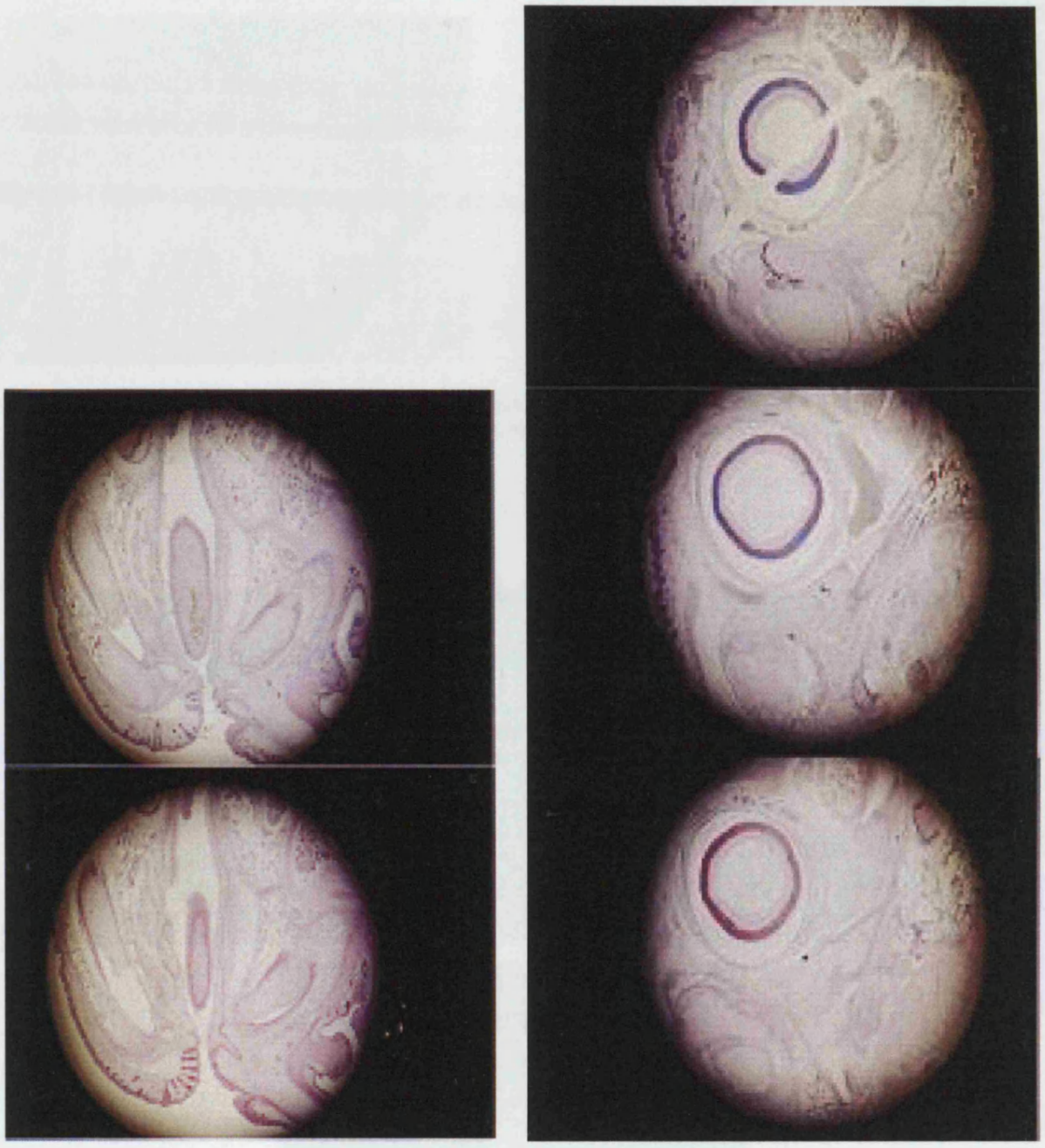


Figure 4.3.3I- 17.5 dpc, Head, (x100). Left panel: Top- Using the TN antibody staining can be seen in the tongue. Bottom- Control, no staining is seen. **Right panel:** Top- The TN antibody also detects Tsga10 expression the muscles of the eyeball. Middle-These muscles are also stained in the experiment using the TC antibody. Bottom- Control, no staining is seen. ESTs have been identified in cDNA libraries made from RNA extracted from eyeballs.

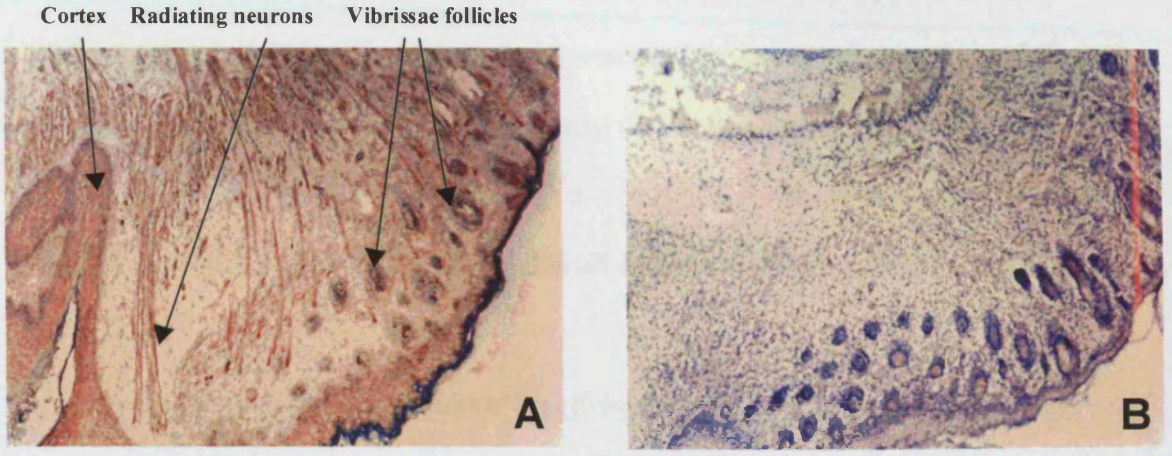


Figure 4.3.3J- New born, Brain, (x100). Staining is seen when using the TC antibody (A), particularly in the brain cortex – seen as a bundle of radiating neurons, in the cortex, and in the vibrissae follicles. In (B) the TN antibody was used, and no staining was seen.

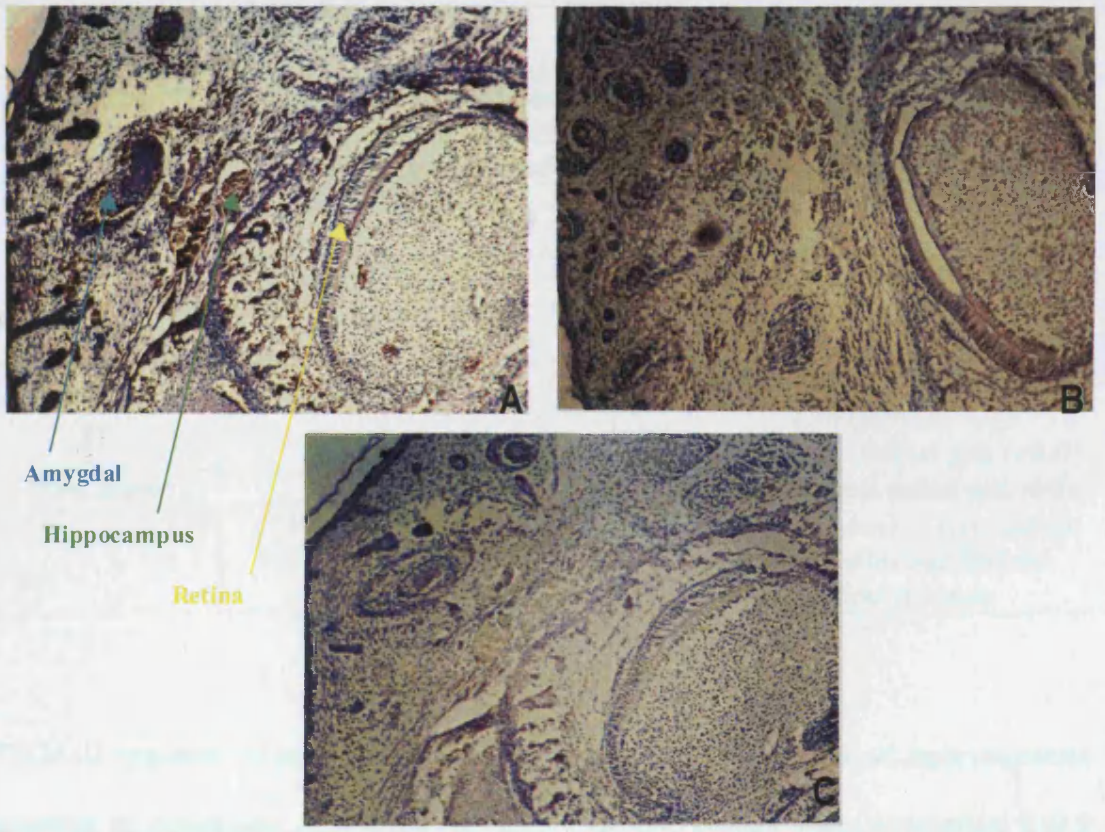


Figure 4.3.3K- New born, Brain, x100. Using the TC antibody (A) produces staining in several regions of the brain including the amygdal region, the hippocampus, the retina and the vibrissae follicles. With the TN antibody (B) the only definite staining is in the vibrissae follicles. In the control experiment (C) no staining is seen.

Table 4.3.3 summarises the observations presented in the previous series of figures (figures 4.3.3 A-K), along with some additional results not shown as figures.

In most cases, once protein expression was detected in an organ or tissue-type by one or other of the Tsga10 antibodies it was seen at all subsequent developmental stages.

Table 4.3.3– Summary of the pattern of Tsga10 protein expression in mouse Embryogenesis

Mouse Embryos	Tsga10-N	Tsga10-C
E10.5	-	(brain) sub-epithelial region
E11.5	-	Non-differentiated cells Brain sub-epithelial cells
E13.5	-	Paravertebral Epithelium
E14.5	-	Vertebrae and Paravertebral regions
E15.5	Primordia of vibrissae-follicle Brain Trunk Vertebrae Epithelium	Vertebrae and Paravertebral regions Intestinal mucosa Eye muscle
E16.5	Tail	-
E17.5	Tongue Eye Muscles	Intestinal mucosa Pancreas Eye Muscles Liver
E18.5	-	Tail Epithelium Vertebrate Eye (Retina) Telencephalon (Radial glia cells?)
New Born	Intestinal mucosa Primordia of vibrissae follicle	Brain (cortex and radial glia cells, subventricular zone) Eye muscle Primordia of vibrissae follicle Intestinal mucosa

TSGA10 appeared to be expressed weakly during the early stages of embryogenesis however its expression increased over time. This was evident when comparing E10.5 and E11.5 embryos, where the C-terminus is expressed in both, but is stronger by stage E11.5 (figures 4.3.3 A-B). In E13.5 embryos, both the N and C-terminus are expressed,

however the signals from the C-terminus are far stronger. The C-terminus seems to be expressed in epithelium and in the paraventricular areas. The N-terminus is not only expressed more weakly but also it is not present in as many areas. The N-terminus may have a structural role in cells (see section 4.4.2).

The C-terminus is expressed in the paraventricular zone; this is significant as the paraventricular zone is a key area for cell proliferation in the developing nervous system. The expression of TSGA10 increases in E14.5 embryos, the C-terminus is now evident in the vertebrae as well as the paraventricular zone and the epithelium. The N-terminus is also weakly expressed in certain areas, but increases significantly by E15.5, when it is expressed strongly in the brain (which reflects the expression in adult mice), the trunk, lower visceral organs, epithelium and primordia of vibrissae follicles. This may mark a transition in the developing embryo from proliferation to differentiation in these areas. The C-terminus is still expressed in the vertebrae although not as strongly as in E14.5 which could be due to the degradation of the protein or a decrease in translation, or a combination of these due to the slowing down of proliferation. The C-terminus is now also expressed in the developing eye. By E16.5 the expression of TSGA10 in the epithelium and vertebra has diminished. The C-terminus is now being expressed in the developing tail bud and the N-terminus is being expressed in the developing digits, this coincides with the time at which the digits re-orientate themselves to be parallel. The expression patterns change again during E17.5. The C-terminus is strongly expressed in the lower visceral organs including the small intestine and the liver whereas the N-terminus is not. Both the N and C termini of Tsga10 protein are expressed in tissue surrounding the eye, this is probably in the muscles of the eyelids, and the time of

expression coincides with the time at which they close. At E18.5 there is no expression of the N-terminus, but the C-terminus is expressed weakly in many areas including the vertebra, tail epithelium, lower visceral organs and the cortex. The newborn mouse expresses high amounts of both the C- and N-terminus in the brain, intestine, primordia of vibrissae follicles and head areas.

The timing of expression of the Tsga10 protein during embryogenesis from E10.5 was similar to that of its message (mRNA) in the mouse embryo and in embryonic brain development as revealed by RT-PCR.

4.3.4 Using immunohistochemistry to examine expression of the Tsga10 protein expressions in cancerous cells (IHC)

Some evidence that TSGA10 might be expressed in cancerous cells was provided when an EST (BE047007) in GenBank isolated from a germ cell tumor was seen to be identical to the *TSGA10* transcript in nucleotides 2378-2981.

I decided to investigate thus pattern of expression further by using the method of immunohistochemistry to identify Tsga10 expression in a non-teratomatous germ cell tumour. The tumour was diagnosed and provided as a gift by Professor Ming-Qing Du (Department of Pathology, UCL, at the moment in Department of Pathology, University of Cambridge).

Using TN antibody a general diffuse staining was observed in the human female germ cell tumour (figure 4.3.4). This may represent expression of the whole TSGA10 protein, and confirm the findings of others (e.g. Tanaka *et al.*, 2004) that TSGA10 is expressed in cancers.

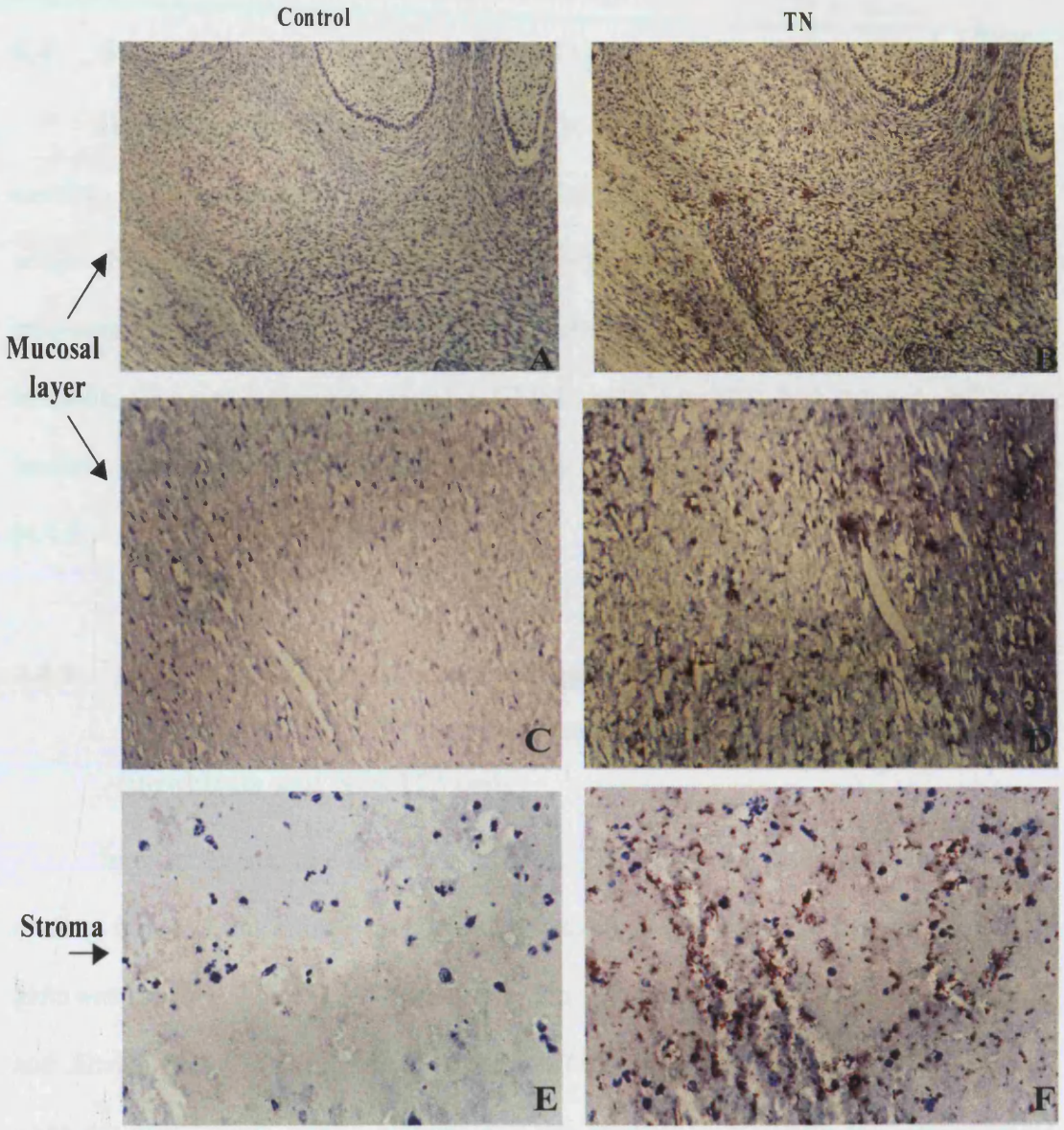


Figure 4.3.4- 10-12-micron sections of a human female (ovarian) germ cell tumour were hybridised to the TN antibody (right hand panel) and were stained with DAB and counterstained with H&E. Adjacent sections were used for the experiment and the controls (left hand panel). For controls the antibody mix was blocked with all three of the Tsga10 peptides. Different parts of the tumour are shown in each of the images. Diffuse staining is seen in the slides of the right hand panel and indicates TSGA10 expression.

4.4 Subcellular localization of the Tsga10 protein

Having examined Tsga10 expression at the protein level by Western blotting, and then in different tissues by immunohistochemistry, I wanted to study its expression within the cell with the hope that the localisation of the protein may give further indication of its function. I studied the subcellualar expression of protein using two methods: first using transfection of a GFP/Tsga10 construct into several cell types (section 4.4.1) and by immunocytochemistry in astrocytes (4.4.2) and sperm cells (4.4.3).

4.4.1 Studying the intracellular localisation of Tsga10 protein by Transfection of GFP/Tsga10 construct in HeLa, hamster fibroblasts and NIH3T3 cells

In order to examine the subcellular localization of the Tsga10 protein, the open reading frame (from double methionine to Stop codon) region of the mouse *Tsga*10 gene was subcloned from cDNA clone P2 into the vector pEGFP-C2 using the *Hind*III and *Eco*RI sites. Two oligonucleotides, "MF4TC10clone" and "MR19TC10clone", were designed which would amplify a 2.1kb fragment containing the open reading frame of the *Tsga10* gene. They were given 5′ extensions containing a *Hind*III site and an *EcoR I* site respectively. After amplification using proofreading DNA polymerase *Pwo*, the 2.1kb fragment was electrophoresed through an agarose gel (figure 4.4.1A), cut out and purified using a Qiagen column. It was then digested with *Eco*RI and *Hind*III and ligated into pEGFP-C2. After transformation of *E. coli*, plating and selection of the colonies containing the insert, its miniprep product was sequenced to

confirm that the cloning of Tsga10 into the vector had not caused a frame-shift mutation (figure 4.4.1(A)B) (map in Appendix).

Transfections of hamster fibroblasts (64063a12), HeLa cells and mouse fibroblasts were carried out. The cell lines were transfected with either the pEGFP-*Tsga10* DNA, which produces a GFP-*Tsga10* fusion protein, or as a control with pEGFP vector DNA that produces Green Fluorescent Protein (GFP) and were fixed and stained with DAPI to visualize nuclei 24 or 48 h after transfection. Transfected mouse fibroblasts (NIH3T3) were also treated with Phalloidin to detect actin fibers.

In hamster fibroblasts, the Tsga10 protein was located in the cytoplasm (figure 4.4.1B) and seemed to be associated with short, thick filaments, and has particularly localized in the perinuclear region. This pattern of localization is also seen when the fusion protein is transfected into HeLa cells (figure 4.4.1C).

Although these cells (fibroblasts and the HeLa cells) are not necessarily those cells that would be expected to express Tsga10- these would more likely be the testis, brain and other actively dividing cells- the fibroblasts and HeLa cells were available and seemed a reasonable cell type in which to try out the experiment. It seems that they have proved to be a reasonable model, since a similar pattern of localization was consistently seen, and this was markedly different from the control experiment. I then continued to carry out a similar experiment in the more relevant cells of sperm and astrocytes.

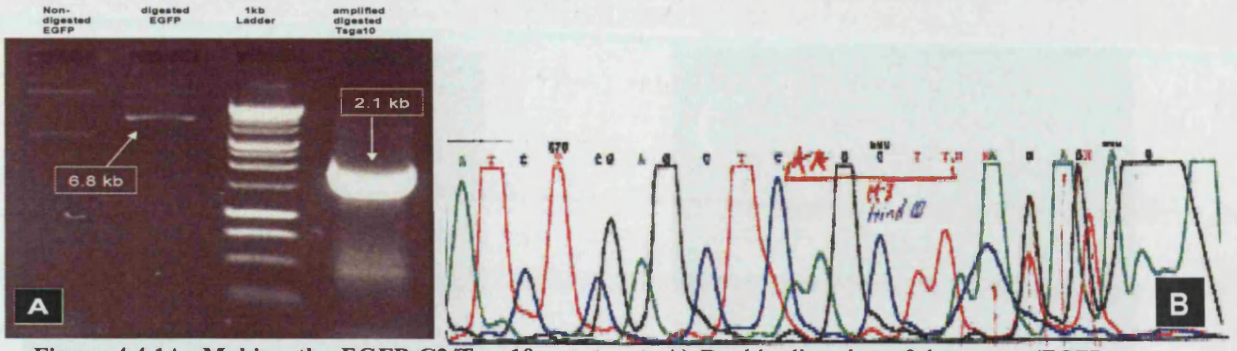


Figure 4.4.1A- Making the EGFP-C2/Tsga10 construct. A) Double digestion of the vector (EGFP-C2) and amplified insert (*Tsga10*) using *Hind III* and *EcoR I* sites. **B)** DNA sequence of the EGFP-C2/Tsga10 construct showing the junction confirming that no frame shift mutation has been created.

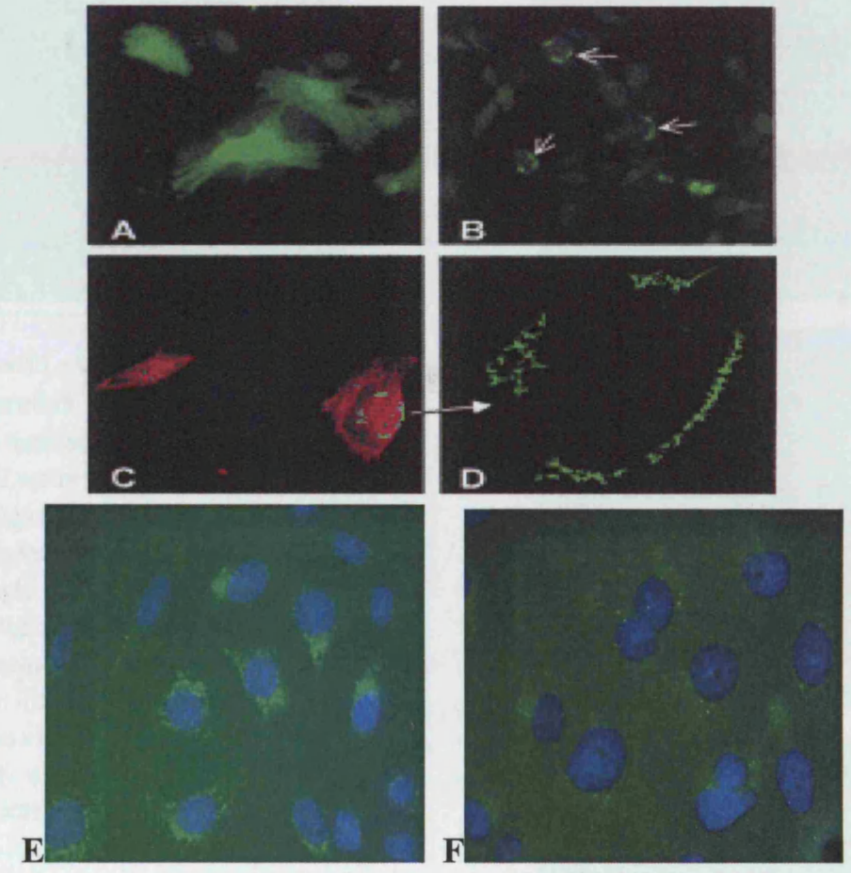


Figure 4.4.1B- A) NIH3T3 mouse fibroblasts were transfected with the pEGFP vector DNA as a control. In this case the GFP can be seen throughout the cell. **B)** The cells were transfected with the Tsga10 construct. In this case the GFP-Tsga10 fusion protein seems to be localized in specific areas of the cell. **C)** Atin was stained using TRITC-coupled Phalloidin (red). The GFP-Tsga10 fusion protein is also seen (green), and this is shown at higher magnification in **D.**
E) Hamster fibroblasts (64063a12) were transfected with the pEGFP-Tsga10 construct (green) and nuclei were stained with DAPI (blue). The fusion protein is highly associated with the nucleus. In the control **(F)** the cells were transfected with the pEGFP-C2 vector. The green colorations is more scattered in this case. Magnifications were (A-C, x20), (D, x100), (E, F, x400).

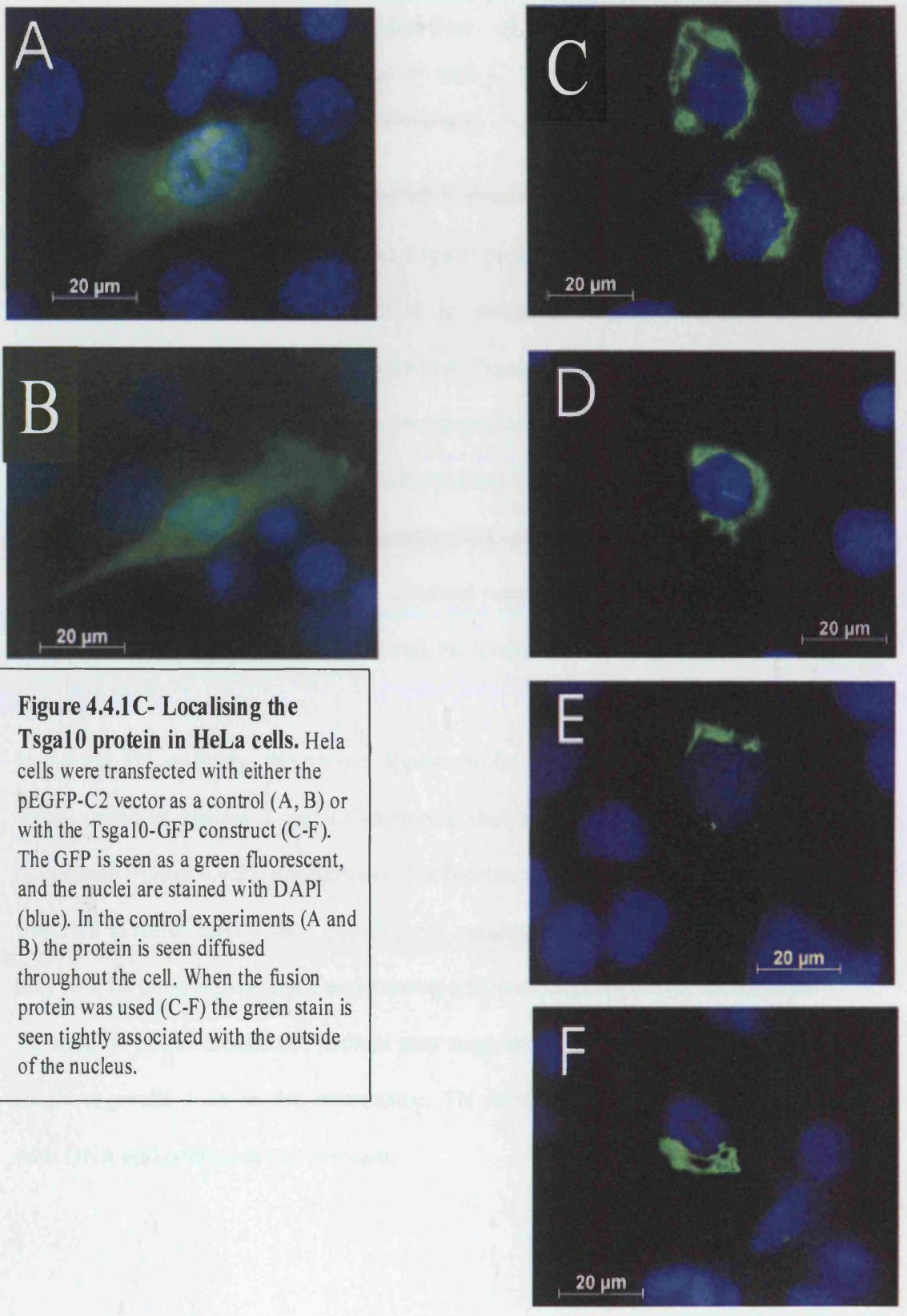


Figure 4.4.1C- Localising the Tsga10 protein in HeLa cells. Hela cells were transfected with either the pEGFP-C2 vector as a control (A, B) or with the Tsga10-GFP construct (C-F). The GFP is seen as a green fluorescent, and the nuclei are stained with DAPI (blue). In the control experiments (A and B) the protein is seen diffused throughout the cell. When the fusion protein was used (C-F) the green stain is seen tightly associated with the outside of the nucleus.

4.4.2 Studying the localisation of the Tsga10 protein within astrocytes using the N and C terminus antibodies, visualised by fluorescence microscopy

In section 4.3.1 immunostaining of the adult mouse brain hybridised to the TN and TC antibodies indicated expression of the Tsga10 protein in astrocytes. Also, some ESTs with Tsga10 sequence were identified in medulla, cerebellum and brain cDNA libraries. Based on this evidence I studied Tsga10 protein expression in astrocytes derived from a man involved in astrocytoma grade III, WHO classification.

In this experiment the astrocytes were hybridised to either the TN or TC antibody, and were visualised using a mouse anti-rabbit FITC-conjugated IgG secondary antibody. When the slides are viewed using a confocal microscope the nuclei are seen as blue (DAPI staining), and the antibody reveals the location of the Tsga10 protein as a green signal.

Using the TN antibody, the protein appears to be localized within the nucleus, while using antibody against Tsga10-C-terminus shows localisation close to the nuclear membrane (figure 4.4.2). Furthermore, the C-terminus seems to be localised in a single structure close to the nucleus, possibly the centrosome or the Golgi. The N-terminus may also be present here, but clearly shows additional signals within the nucleus.

TC specific points around the nucleus may suggest that the protein is associated with a single organelle such as the centrosome. TN inside the nucleus suggests interaction with DNA and control of cell division.

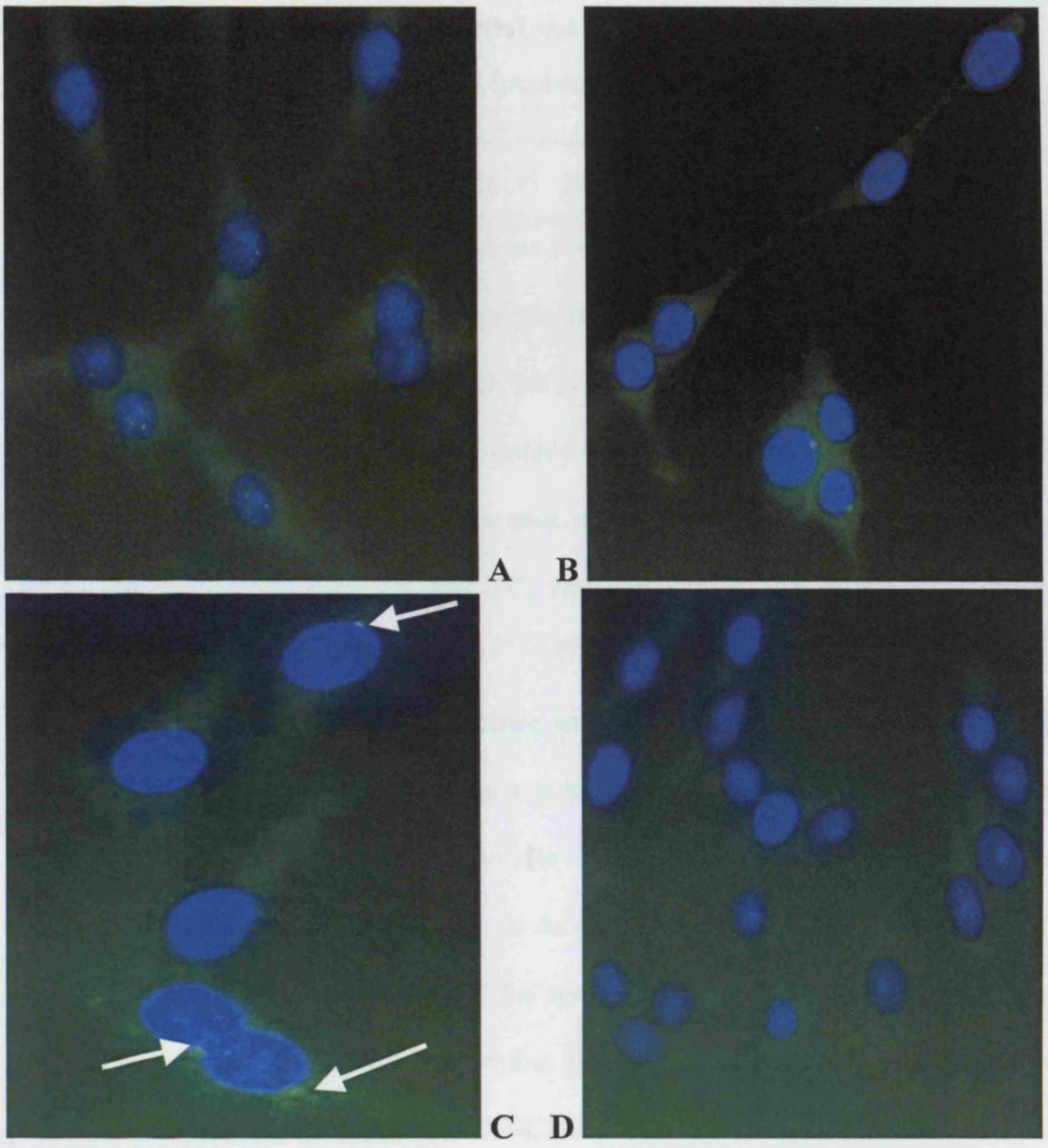


Figure 4.4.2- Immunolocalization of the Tsga10 protein in the astrocytes. Tsga10-N-terminus is localised within the nucleus while Tsga10-C-terminus is localized at nuclear membrane. Each photo contains some astrocytes derived and cultured from an anaplastic astrocytoma tumour (Grade 3, WHO). The photomicrograph is taken at 630X magnification. Mouse anti-rabbit FITC-conjugated IgG used as secondary antibody.

(A) The astrocytes hybridised with the TN antibody;

(B,C) The astrocytes hybridised with TC antibody, in this case the green signal is in a specific perinuclear localisation;

(D) Control, treated with no primary antibody.

4.4.3 Determining the localisation of the Tsga10 protein in sperm by immunocytochemistry and immunofluorescence

Evidence from Western blotting of protein extracts from the two major components of rat sperm tail, the outer dense fibres (ODF) and the fibrous sheath (FS), indicated that in the fibrous sheath a prominent 27-KDa band is seen (figure 4.2C). This 27-KDa fragment seems to represent the N terminus of the Tsga10 protein, which is cleaved from the complete protein. In order to confirm this observation and to refine the localisation indirect immunofluorescence microscopy of epididymal sperm was carried out using antibodies TN and TC, which recognise the N and C terminal portions of the protein respectively.

Using the TN antibody, fluorescence was seen primarily in the principal piece region of the sperm tail, in both mouse and boar (figures 4.4.3A-C). There is some fainter fluorescence in the mid-piece, and it is likely that the Tsga10 N terminal region is expressed here, but it may be difficult to detect due to the mitochondrial sheath and the solenoid structure of this portion of the sperm tail. In the control experiment, the TN antibody did not hybridise to sperm that had been blocked with peptides to both parts of the Tsga10 protein (figures 4.4.3B-C).

Direct immunofluorescence microscopy of epididymal sperm from mouse, cow and boar was also carried out using antiserum TC. The result shows that C-terminus component of Tsga10 is present in the midpiece of the tail but not the principal piece (figures 4.4.3D-F). Sperm incubated with all of the blocked primary antibodies did not show any staining and non-specific secondary antibody binding to sperm was also undetectable.

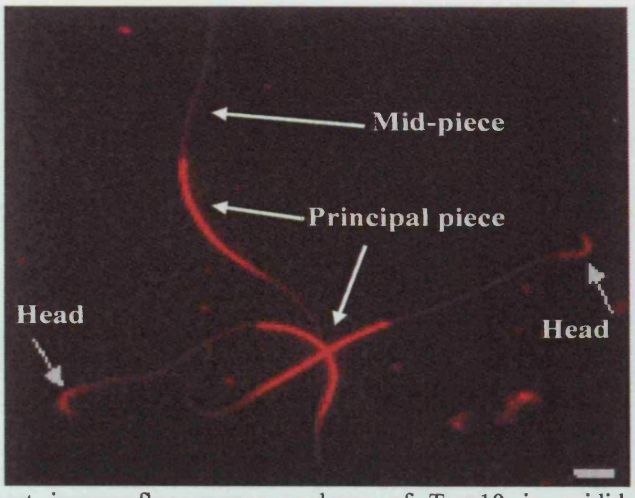


Figure 4.4.3A- Indirect immunofluorescence analyses of Tsga10 in epididymal **mouse** sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and hybridised with anti-TN antiserum followed by secondary antibody conjugated to Cy3. The principal piece of sperm tails, which is distal to the midpiece, strongly fluoresced (red). Sperm heads, showing very weak staining, are indicated with arrows. Bar = 20 μm

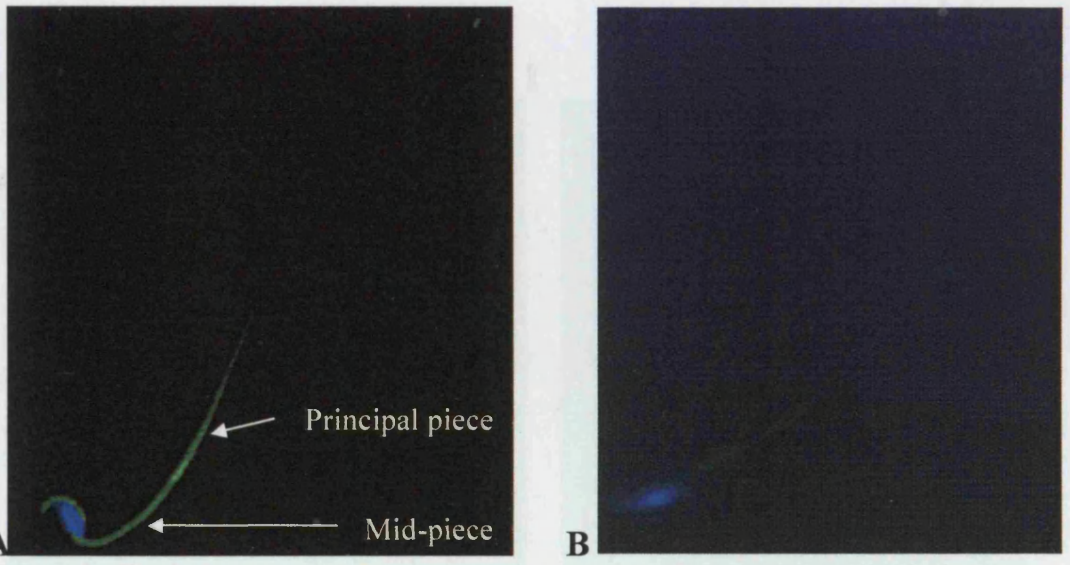


Figure 4.4.3B- Indirect immunofluorescence analysis of Tsga10 in epididymal **mouse** sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and: **A)** stained with anti-TN antiserum followed by secondary antibody conjugated to FITC. The principal piece of sperm tails, which is distal to the mid piece, fluoresced strongly (green). Sperm heads and midpiece, showed weak staining;
B) Control (blocked with TN, TM, TC peptides) (× 400) sperm heads appear blue due to DAPI staining of nuclei

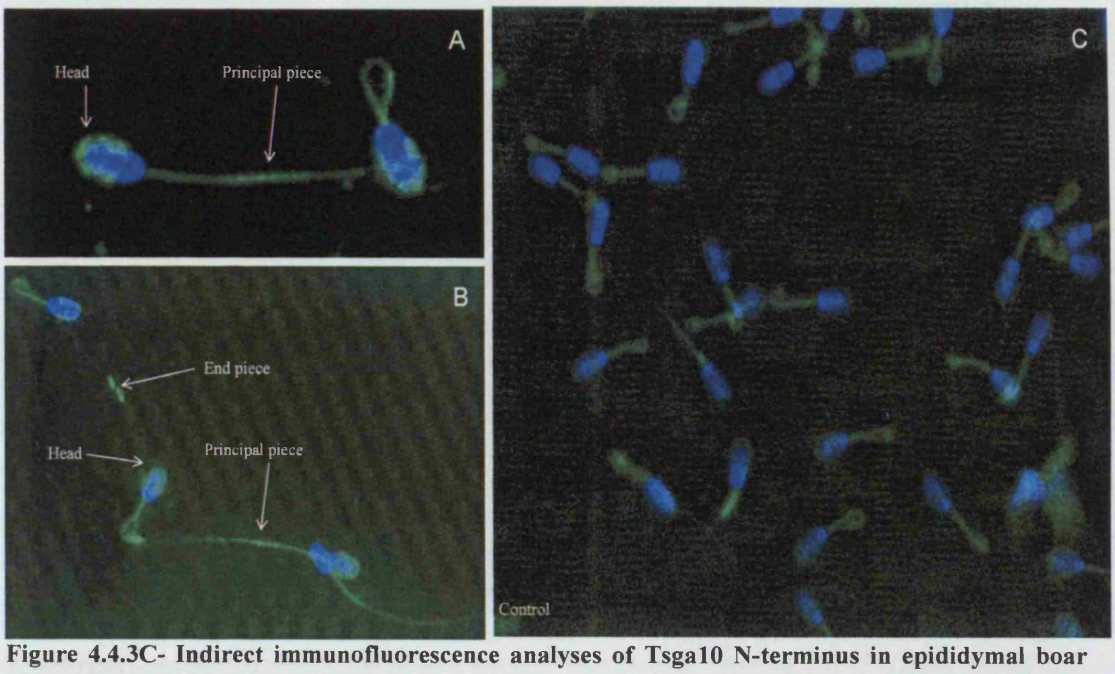


Figure 4.4.3C- Indirect immunofluorescence analyses of Tsga10 N-terminus in epididymal boar sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and: **A,B)** stained with TN antiserum followed by secondary antibody conjugated to FITC. The principal of sperm tails, which is distal to the mid-piece, as well as sperm head and end piece of sperm tail strongly fluoresced. **C)** Control, sperm heads appear blue due to DAPI staining of nuclei, A(×630) B,C (×400)

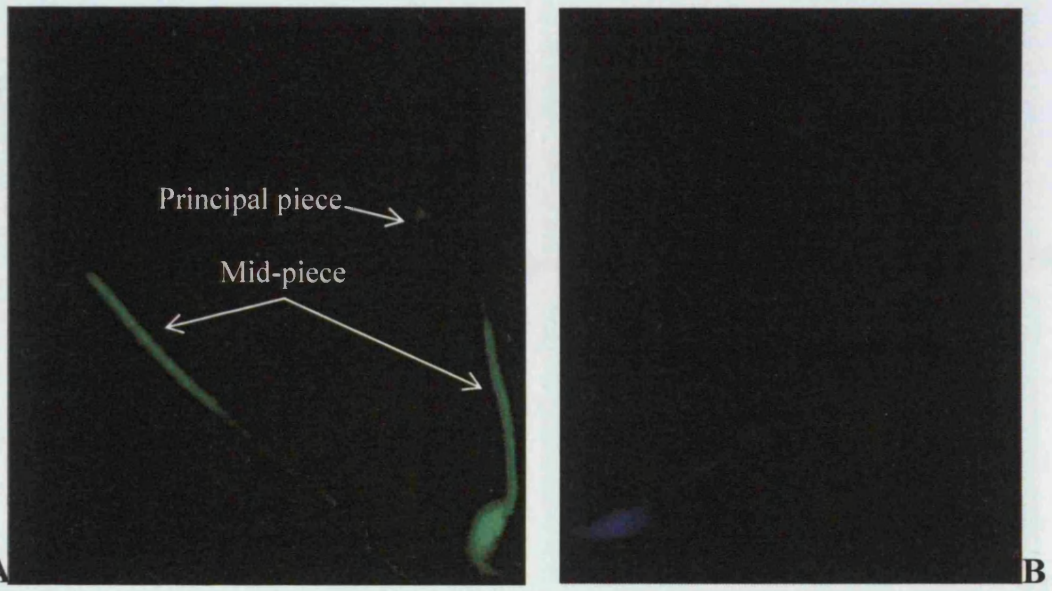


Figure 4.4.3D- Indirect immunofluorescence analysis of Tsga10 in epididymal **mouse** sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and: **A)** stained with TC antiserum followed by secondary antibody conjugated to FITC. The midpiece of sperm tails, which is proximal to the principal piece, strongly fluoresced (green). Sperm heads show weak staining;
B) Control, sperm heads appear blue due to DAPI staining of nuclei (TN, TM, TC blocked) (× 630)

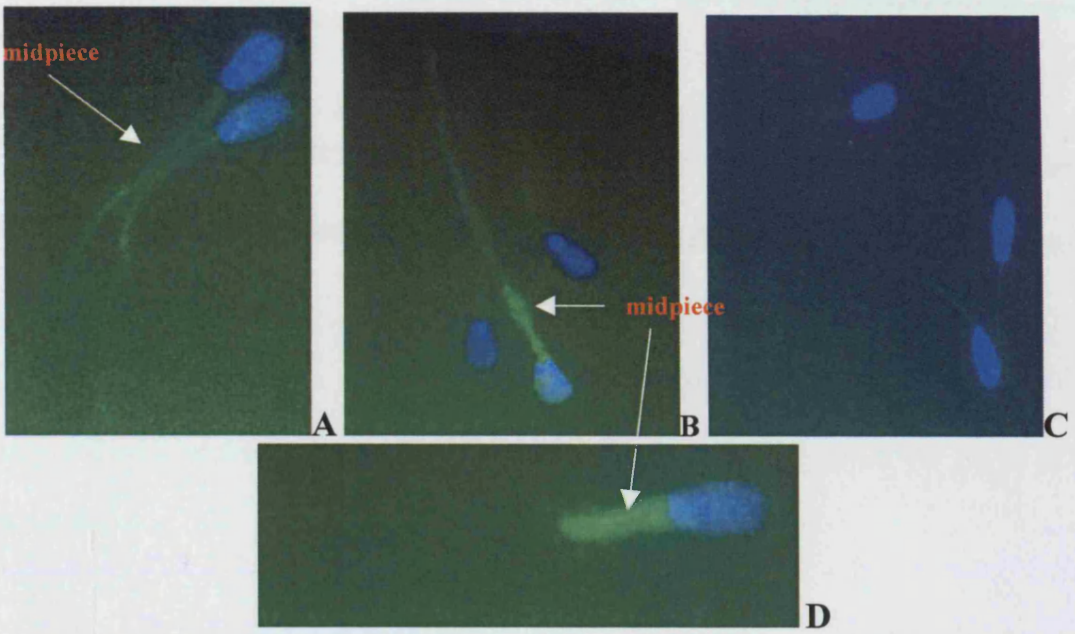


Figure 4.4.3E- Indirect immunofluorescence analyses of Tsga10 in epididymal **bovine** sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and: **A,B,D)** stained with TC antiserum followed by secondary antibody conjugated to FITC. The midpiece of sperm tails, which is proximal to the principal piece, strongly fluoresced. Sperm heads show weak staining;
C) Control (blocked with C, M, N peptides), sperm heads appear blue due to DAPI staining of nuclei

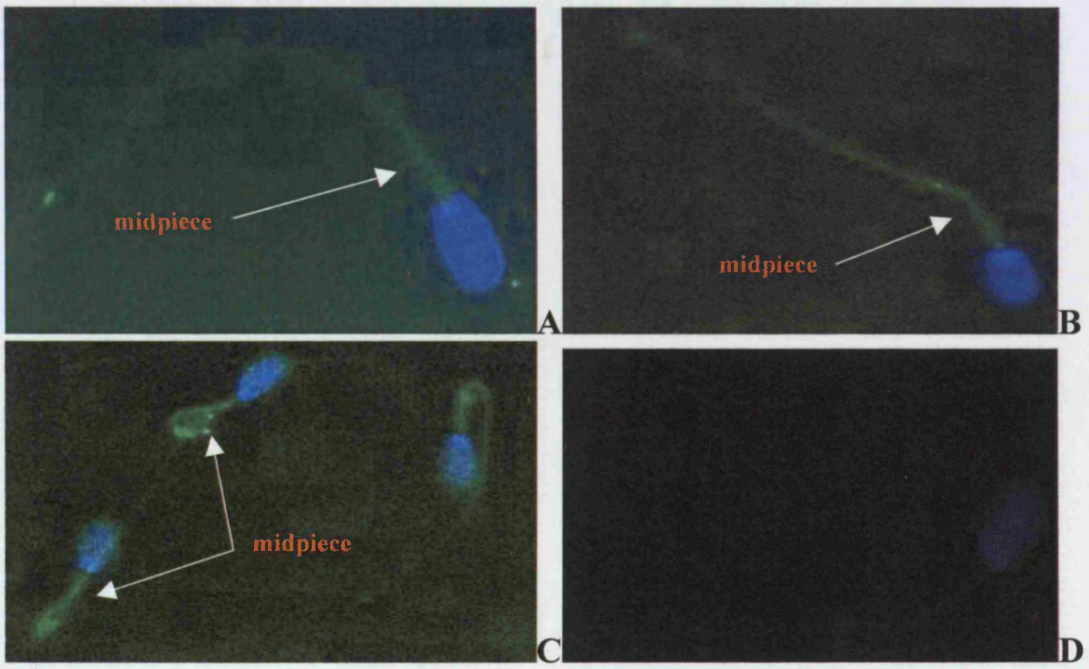


Figure 4.4.3F- Indirect immunofluorescence analyses of Tsga10 C-terminus in epididymal **boar** sperm. Spermatozoa were fixed on poly-L-lysine coated glass cover slips and: **A,B,C)** stained with TC antiserum followed by secondary antibody conjugated to FITC. The midpiece of sperm tails, which is proximal to the principal piece, strongly fluoresced. Sperm heads show weak staining; **D)** Control, sperm heads appear blue due to DAPI staining of nuclei

Chapter V:

Examining TSGA10 protein interaction using the Yeast Two Hybrid assay

In the previous experiments overall observations about the behaviour of the Tsga10 protein have come to light. The two halves of the protein seem to behave differently in the cell: in sperm the N terminus is associated with the fibrous sheath in the principal piece of the sperm tail, whereas the C terminus is associated with the midpiece (section 4.3.3), in astrocytes the N terminal region appears to be located within the nucleus, whereas the C terminal portion is distinctly perinuclear. A way to try and understand the reasons for the locations of the protein, and therefore to be able to make predictions about function, is to investigate which other proteins Tsga10 might interact with. This can be achieved using the yeast two-hybrid system (Zhu and Hannon, 2000). The experiment involves using a plasmid construct containing the cDNA encoding the protein of interest (the 'bait' Tsga10), and introducing (by yeast mating, or co-transformation) this into yeast cells along with plasmid constructs containing 'prey' sequences, in this case a rat testis cDNA library. Where the bait protein interacts with a prey protein, the yeast genes encoding selectable markers within the constructs can be transcribed, and the yeast colony will survive on selection media (see methods section 2.5).

Making the bait construct- "YTHM" and "MR19TC10clone" oligonucleotides were used to amplify a 2.1kb fragment containing the entire open reading frame of the *Tsga10* gene using pGFP-C2/*Tsga10* as a template. These two oligonucleotides were given 5′ extensions containing an *NcoI* site and an *EcoR I* site respectively.

After amplification using proofreading DNA polymerase *Pwo,* and digestion with *Nco*I and *Eco*RI the 2.1kb fragment was electrophoresed through an agarose gel and purified

using a Qiagen column. It was then ligated into the bait vector pGBKT7 which contains the GAL4 binding domain promoter; T7, ADH1, a multiple cloning site and TRP1 in the cassette (see Appendix for its map). The vector had been previously digested with the same two restriction enzymes. The construct of the bait plasmid was sequenced to confirm its proper junction to make sure that no from shift mutations had been created and then the yeast strain AH109 was transformed by it with a high efficiency ($\sim \times 10^8$ colonies per μg DNA). Then pretransformed rat testis (in pGAD424 plasmid) was used to transform yeast strain AH109 containing the bait construct.

Yeast Transformation- Since the transformation could be performed so efficiently, it was reasonable to transform AH109 yeast strain already containing the bait plasmid with the library plasmid. An aliquot of the transformed yeast was plated onto SD/-Trp/-Leu plates to select for the presence of both plasmids, and surviving colonies were counted.

Table 5A- Yeast clones with library and bait plasmid counted after co-transformation in yeast strain AH109

Date of counting	Approximate number of colonies
09.10.03 (1:2000 diluted control)	213
09.10.03	400,000

$$\text{Transformation efficiency} = \frac{\text{no. colonies x total suspension volume (μl)}}{\text{Volume plated(μl) x dilution factor x amount DNA(μg)}}$$

$$\approx \frac{210 \times 2000}{100 \times 0.005 \times 0.1}$$

$$\approx 8 \times 10^6 \text{ colonies per μg DNA}$$

This is about ten times less efficient than the single plasmid transformation. This a good reason for doing the mating experiment as well, however it is still quite efficient.

The remainder of the transformation was plated on SD/-Trp/-Leu/-His/-Ade, to select for a two hybrid reaction. Seventeen yeast colonies were obtained after a 2-week incubation.

PCR of plasmid mini-preps using primers pGAD 3' and pGAD 5' confirmed the presence of the library plasmid in these colonies (results not shown) and showed the size of the inserts. Only four out of the seventeen screened colonies, looked like they may have contained an insert. These colonies, M1, B1, S2 and S3 were re-screened by PCR again using the primers pGAD 3' and pGAD 5' on a mini-prep product from the colonise pGADT7-T was used as a template for the positive control, and no template in the negative control. The PCR products were then run on an agarose gel alongside the Gibco 1 Kb DNA ladder (figure 5A).

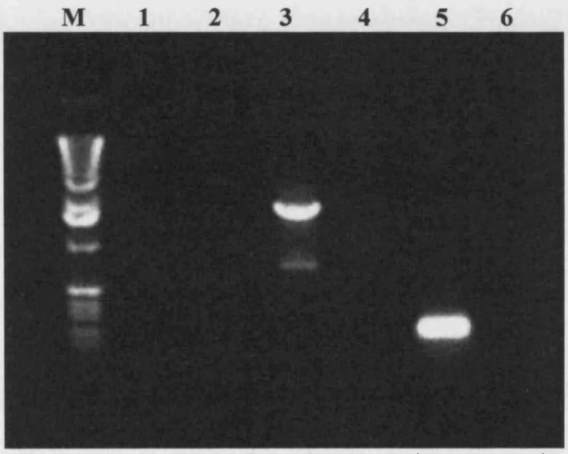


Figure 5A- PCR products of plasmids extracted from yeast colonies that survived the selection procedure. In this research, using primers pGAD3' and pGAD5', only are plasmid (S3 lane 3) seems to contain one insert, and maybe two. The inserts are 2 Kb and 750 bp in size. Lanes 1 (M1), 2 (S2), 3 (S3), 4 (B1), 5 (positive control), 6 (negative control).

The only colony in which there appears to be a plasmid with an insert is S3, which has two, one around 750 bp, and one around 2,000 bp. These bands were labelled S3b (big band), and S3s (small band).

The S3 colony could contain 2 populations of cells, each containing a different plasmid. The plasmids were isolated from the yeast cells and transformed into *E. coli* DH5α. This

enabled the two types of plasmid to be separated as each *E. coli* cell can only take up one plasmid. Several individual *E. coli* colonies were then picked and grown up in liquid culture. A sample of this was re-plated, while DNA was prepared from the rest of each culture using the mini-prep method. The mini-prep DNA then underwent PCR using the same primers as above. Both plasmids were rescued and representative colonies were then stored as a glycerol stocks at -70°C and were sequenced. **S3b** showed 100% identity to an anonymous transcript from *Rattus norvegicus* described as "similar to 'hypothetical protein FLJ32880'". This transcript had been found in testis cDNA libraries, maturing and mature sperm, and pancreas. **S3s** showed a high degree of similarity to *Mus musculus* 'phosphorylase kinase alpha2 (Phaka2)', which is expressed in a range of tissues, including the adult testis (Table 5B).

Yeast mating

In order to obtain more yeast colonies containing plasmids with inserts I carried out the two-hybrid experiment again using the yeast mating method, rather than co-transformation. First it was necessary to transform the yeast strain Y187 with the rat testis library in pGAD424. Then Y187 colonies growing on SD/-Leu were counted. To be sure that every clone in the testis cDNA library had a reasonable chance of being rescued, at least three times as many colonies were sought, as there were clones in the original library. As about 200,000 independent recombinant clones were in the testis library, I aimed to rescue at least 600,000 colonies in Y187 (See table 5.1A).

Around 4,500,000 colonies containing the library plasmid were obtained. This is sufficient to expect that each independent cDNA clone be represented many times. Both

Table 5B- Rat testis library clones counted after transformation in yeast strain Y187

Date of counting	Approximate number of colonies on plate
01.10.03	90,000
03.10.03	160,000
03.10.03	600,000
06.10.03	2,000,000
06.10.03	2,000,000
06.10.03 (1:2000 diluted)	2,400

large and small colonies were seen.

To estimate transformation efficiency, yeast colonies were counted on the control plate (from the 06.10.03) by using the formula:

$$\text{Transformation efficiency} = \frac{\text{no. colonies x total suspension volume(µl)}}{\text{Volume plated(µl) x dilution factor x amount DNA(µg)}}$$

$$\approx \frac{2400 \times 2000}{100 \times 0.005 \times 0.1}$$

$$\approx 9.6 \times 10^7 \text{ colonies per µg DNA}$$

This transformation was about 10 times as efficient as the co-transformation method.

Mating between the bait strain, AH109 and the library strain, Y187 was then carried out according to the protocol outlined in the methods section (2.19). As a control, Y187 containing pGADT7-T (library control plasmid) and AH109 containing pGBKT7-53 (bait control plasmid) were also grown up in overnight culture. The Y187 containing the library control plasmid was mated to both the AH109 containing the bait control plasmid and to AH109 containing the bait plasmid itself. The diploid cells produced by the mating were then plated out on SD-Trp/-Leu, to select for the presence of both plasmids

and to permit an estimate of mating efficiency, and also on SD-Trp/-Leu/-His/-Ade. The results are shown in Table 5C.

Table 5C- Transformation efficiency and bait positive control by yeast mating

Mating	Y187 containing library control (pGADT7-T) plasmid mated to AH109 containing:	Media	Approximate number of colonies
A	Bait control (pGBKT7-p53)	SD-Trp/-Leu	~ 5,000
B	Bait control (pGBKT7-p53)	SD-Trp/-Leu/-His/-Ade	~ 3,000
C	Bait (pGBKT7-Tsga10)	SD-Trp/-Leu	~ 5,000
D	Bait (pGBKT7-Tsga10)	SD-Trp/-Leu/-His/-Ade	0

The cells from mating A that grew on the double selection media demonstrated that the *TRP1* and *LEU2* marker genes were both active in the cells, in other words, the co-transformation was successful. A similar number of colonies were obtained under the double selection from mating C, showing that the bait strain-control strain mating was also successful, thus the method works with the experimental (Tsga10) bait plasmid.

Only cells from mating B grew on the quadruple selection media. The two hybrid proteins encoded by the control plasmids should interact to activate the *HIS3* and *ADE2* selectable genes, and thus be capable of growing on -Trp/-Leu/-His/-Ade drop-out media. The large T antigen and the Tsga10 protein however should not interact; no colony growth is seen from mating D, under the quadruple selection. This shows that the selection for the two-hybrid interaction worked.

Results of the mating experiment between strains AH109 containing the bait plasmid (Tsga10) and Y187 containing the library plasmid: The bait strain AH109 containing pGBKT7-Tsga10 was mated to the library strain Y187 containing the rat cDNA library in pGAD424. The resulting diploid cells were plated onto twenty-six plates containing more than two hundreds yeast colonies grown. This is because the protein interaction has to take place in order to switch on genes that are necessary for growth on PCR analysis, three colonies grew on selection media: L1 (insert approximately 150 bp), L3 (insert approximately 2,000 bp) and Y1 (insert approximately 750 bp) (figure 5B). Further PCR reactions of yeast colonies from this experiment are shown in the following figures (5C and 5D).

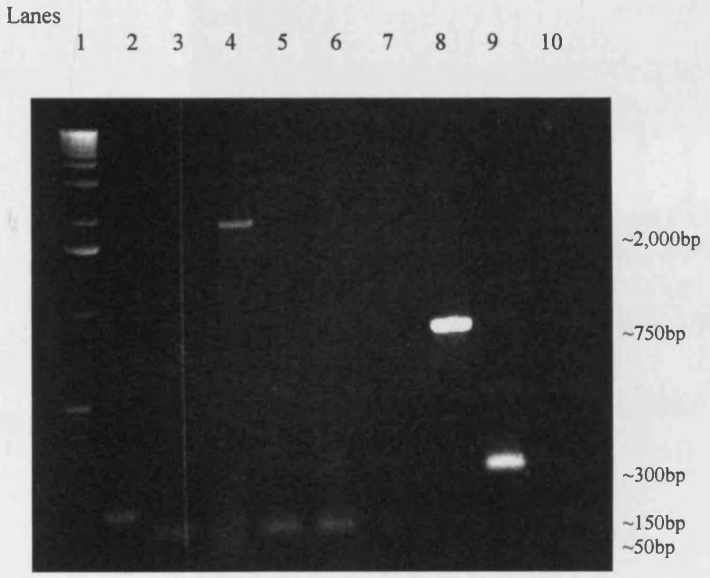


Figure 5B- Electrophoresis of PCR products, from DNA mini-prep of colonies produced by yeast mating- DNA was prepared from the colonies using the mini-prep method. PCR reactions were then carried out using the primers pGAD 3' and pGAD 5'. The PCR products were then run on the electrophoresis gel alongside the Gibco 1 Kb DNA ladder. Lanes are: (2) L1, (3) L2, (4) L3, (5) Z1, (6) Z2, (7) Z3, (8) Y1, (9) positive control, (10) negative control.

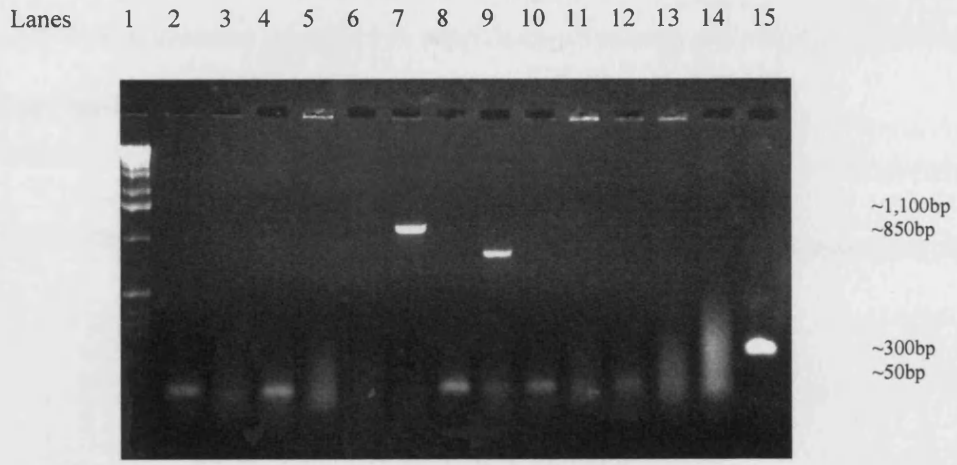


Figure 5C- In this PCR experiment from yeast colonies, two appear to contain plasmids with inserts, these are A'1 (insert approximately 1,100 bp) and A'6 (insert approximately 850 bp). Lanes are (2) A1, (3) A2, (4) A3, (5) A4, (6) A5, (7) A'1, (8) A'2, (9) A'6, (10) C1, (11) D1, (12) D2, (13) D3, (14) D5, (15) positive control.

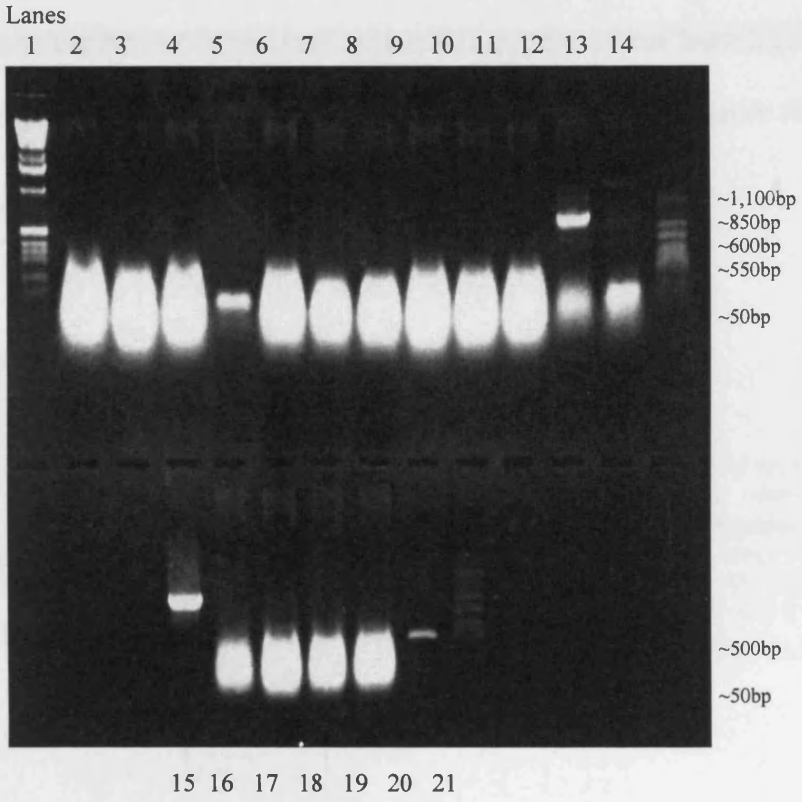


Figure 5D- In this PCR reaction, three of the yeast colonies seem to contain plasmids with inserts, these are F2 (approximately 850 bp and 600 bp), F3 (approximately 1,100 and 550 bp) and F5 (approximately 500 bp). Lanes are (1) Kb ladder. (2) C2, (3) D4, (4) D7, (5) D9, (6) E1, (7) E2, (8) E3, (9) E4, (10) E5, (11) F1, (12) F2, (13) F3, (14) F4 . (15) F5, (16) F6, (17) G2, (18) G3, (19) G4, (20) positive control, (21) 100-bp ladder.

In total only 9 yeast colonies remained as candidates containing potentially interesting inserts. They are listed in Table 5D.

Table 5D

Colony name	Approximate insert size (base pairs)
L1	150
L3	2,000
Y1	750
U14	850
A'1	1,100
A'6	850
F2	600 and 850
F3	1,100 and 550
F5	500

Electrophoresis was carried out in which all of the PCR products from colonies obtained by mating containing sizeable inserts were run alongside each other (see figure 5E).

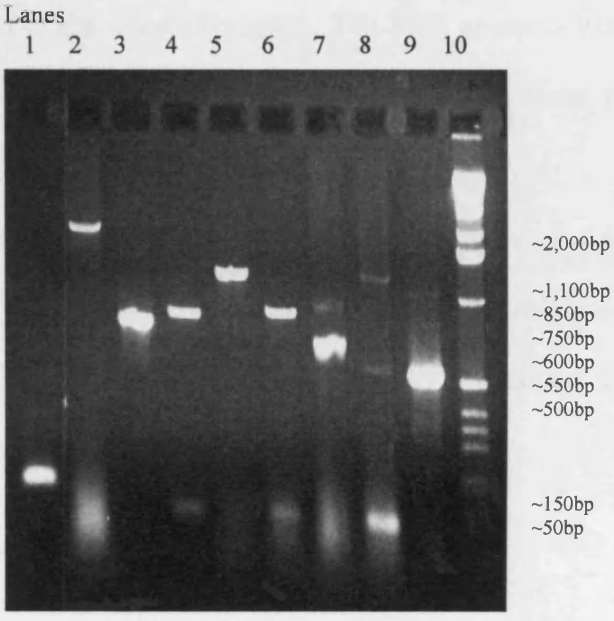


Figure 5E- PCR of each of the 9 clones containing inserts already shown in figures 5B-D, illustrating the range of insert sizes. Lanes are: (1) L1, (2) L3, (3) Y1, (4) U14, (5) A'1, (6) A'6, (7) F2, (8) F3, (9) F5, (10) 1Kb ladder.

Three individual colonies were picked from each of the re-plated colonies screened. Plasmid DNA was isolated from them using the mini-prep method. The plasmids were transformed into *E. coli* DH5α. Where there were two plasmid types, i.e. in colonies F2, F3 and S2, this enabled the two types of plasmid to be separated as each *E. coli* cell can only take up one plasmid. Several individual *E. coli* colonies from each transformation were then picked and grown up in liquid culture. A sample of this was re-plated, while DNA was prepared from the rest of each culture using the mini-prep method. The mini-prep DNA then underwent PCR using the same primers as above. Colonies containing appropriate bands (seen in the electrophoresis of the PCR products) were then stored as a glycerol stock at -70°C to be sent for sequencing (see below). The plasmid DNA was isolated from the *E. coli* cells using the mini-prep method, and used as a template for PCR. The PCR reactions were carried out using the primers pGAD 3' and pGAD 5'. pGADT7-T was used as a template for the positive control. No DNA template was added to the negative control. The PCR products were then run on the electrophoresis gel alongside the Gibco 1 Kb DNA ladder to check the inserts. A sample of each PCR product was sent for sequencing.

All of the nucleotide sequences reads are shown in Appendix 8.1. NCBI basic alignment search tool (BLAST) was used to identify the origin of the cloned inserts by aligning the sequences obtained with those stored in its database. The results of this search are shown in tables 5E.

Table 5E- Bioinformatic characteristics of yeast colonies containing candidate prey insert interacting to bait Tsga10 protein by transformation (S3b) and mating (the rest of them)

Colony name	Insert size (bp)	Length of sequence (bp)	Significant alignment found in BLAST search	Accession number	Further information about gene
L1	150	121	pGAD242 cloning vector	U07647.1 GI:464015	
L3	2,000	~ 850	Rattus norvegicus similar to hypothetical protein FLJ32880	--	
Y1	750	716	Mus musculus Tripartite motif protein 37 (Trim37)	BC059070.1 GI:37589511	Ubiquitously expressed
U14	850	780	Mus musculus Hypothetical protein MGC23918	XM_343490.1 GI:34866381	Ubiquitously expressed
A'1	1,100	866	Mus musculus Transcription factor-like protein MORF-related gene 15 protein (MRG15)	XM_222925.2 GI:34880963	Expressed in a large range of tissues, including the adult testes.
A'6	850	830	Mus musculus Cytochrome c-1 (Cyc1)	XM_216944.2 GI:34866852	Ubiquitously expressed mitochondrial protein
F2	600	468	Mouse peroxisomal biogenesis factor 3 (Pex3)	BC054252.1 GI:32450617	Ubiquitously expressed
F3	~ 1,100	701	Rattus norvegicus Sperm outer dense fibre major protein 2 (Odf2) Cenexin 1	AF162756.2 GI:17388905	A molecular scaffolding protein with a link between centriole maturation and sperm tail protein
F5	~ 500	439	Mus musculus 1500002M01 gene	BC004785.1 GI:13435875	Expressed in a range of tissues, including the adult testes.
S3b	~ 2,000	832	Rattus norvegicus similar to hypothetical protein FLJ32880	XM_341991.1 GI: 34861446	Found in cDNA libraries of testes, maturing and mature sperm, and pancreas
S3s	750	1,250	Chimeric clone: First portion: *Mus Musculus* Phosphorylase kinase alpha2 (Phka2) Second portion: *Mus musculus* BAC clone RP24-233P3	AL928806.14 GI:33569179 AC122445.3 GI:25046668	Expressed in a range of tissues, including the adult testes

Clone L1 showed homology to a part of the cloning vector and thus was probably a false positive. The inserts of Y1, U14, A'6 and F2 are clones containing ubiquitously expressed sequences, i.e. they are expressed in every cell of the body. Those that have been characterised appear to have housekeeping-related functions. It seems unlikely that any of these, except CytC-1 (A'6) would actually interact with the Tsga10 protein in the sperm tail, as it is such a specific structure. These interactions may have been false positives, or may have been chance interactions between proteins that do not normally come into contact with one another. The same was probably true of the genes from which the inserts cloned in A'1, F5 and S3s are derived although these genes are known to be expressed in the testis. The three inserts that looked the most promising were those found in clones F3, L3 and S3b particularly since clones L3 and S3b contained the same sequence found in testis and sperm cDNA libraries, and F3 contains the sequence encoding a protein associated with outer dense fibres.

5.1 Confirming the interaction of the Tsga10 protein with 3 candidate proteins: FLJ32880 hypothetical protein, Cyc-1 and odf2

This confirmation experiment was needed to eliminate the possibility that some mutation in the yeast cells in the previous co-transformation and mating experiments had permitted colony growth in the absence of a genuine two-hybrid interaction. Cells from the bait strain stock (containing the pGBKT7-Tsga10 plasmid alone), were co-transformed with the plasmids isolated from the colonies in which an interaction was previously detected. The plasmids were grown up in *E. coli* prior to the transformation. The resulting transformation suspension was then plated out on both double (-Trp/-Leu) selective media, and quadruple (-Trp/-Leu/-His/-Ade) selective media. A sample of AH109 that did not contain the pGBKT7-Tsga10 plasmid was transformed and also plated out, as well as controls on single (-Leu), and triple (-Trp/-Leu/-Ade) selective media. See Table 5.1A.

Table 5.1A- Repeat transformation and screening of plasmids containing candidate proteins interacting with Tsga10

Colony name	AH109/-Leu	AH109/-Leu/ -His/-Ade	AH109 with pGBKT7-Tsga10/ -Leu/-Trp	AH109 with pGBKT7-Tsga10/-Leu/ -Trp/-His/-Leu
L1	Colony growth	No growth	Colony growth	No growth
L3	**Colony growth**	**No growth**	**Colony growth**	**Colony growth**
Y1	Colony growth	No growth	Colony growth	No growth
U14	Colony growth	No growth	Colony growth	No growth
A'1	Colony growth	No growth	Colony growth	No growth
A'6	**Colony growth**	**No growth**	**Colony growth**	**Colony growth**
F2	Colony growth	No growth	Colony growth	No growth
F3	**Colony growth**	**No growth**	**Colony growth**	**Colony growth**
F5	Colony growth	No growth	Colony growth	No growth
S3b	**Colony growth**	**No growth**	**Colony growth**	**Colony growth**
S3s	Colony growth	No growth	Colony growth	No growth

The cells transformed with the library plasmid alone grew on SD/-Leu; this showed that the library plasmid conferred on the AH109 *leucine auxotrophs* the ability to grow without leucine, as it should do. Similarly, when the AH109 cells were co-transformed, the ability to grow on media without *tryptophan* was conferred. This showed that the plasmid selections were working. No cells grew in the absence of the bait plasmid when under selection for Adenine and Histidine as well as Leucine. This provided evidence that the proteins encoded by the library plasmids were not capable of inducing expression from the *His* and *Ade* promoters on their own. There were no colonies found to be growing on the plates containing cells transformed with the plasmids isolated from L1, Y1, U14, and A'1, F2, F5 and S3s. This means that the selectable genes downstream of the His promoter, were not activated, indicating that no interaction was taking place between the protein encoded by the library plasmid, and *Tsga10*. These must have been false positives. However, there was growth on the plates containing cells transformed with the plasmids isolated from L3, A'6, F3 and S3b (see figure 5.1A).

Using the yeast two-hybrid assay I found three proteins that interact with Tsga10. The first is "L3 or S3b": a hypothetical protein, **FLJ32880**, which had been found in human spermatids, medulla and pancreas. The second is "F3": an outer dense fibre protein encoded by the **Odf2** gene, which is expressed in rat and human testis, epididymis and foetal brain. The third is "A'6": **Cytochrome C1**, which is a subunit of an electron-transfer chain in the mitochondria.

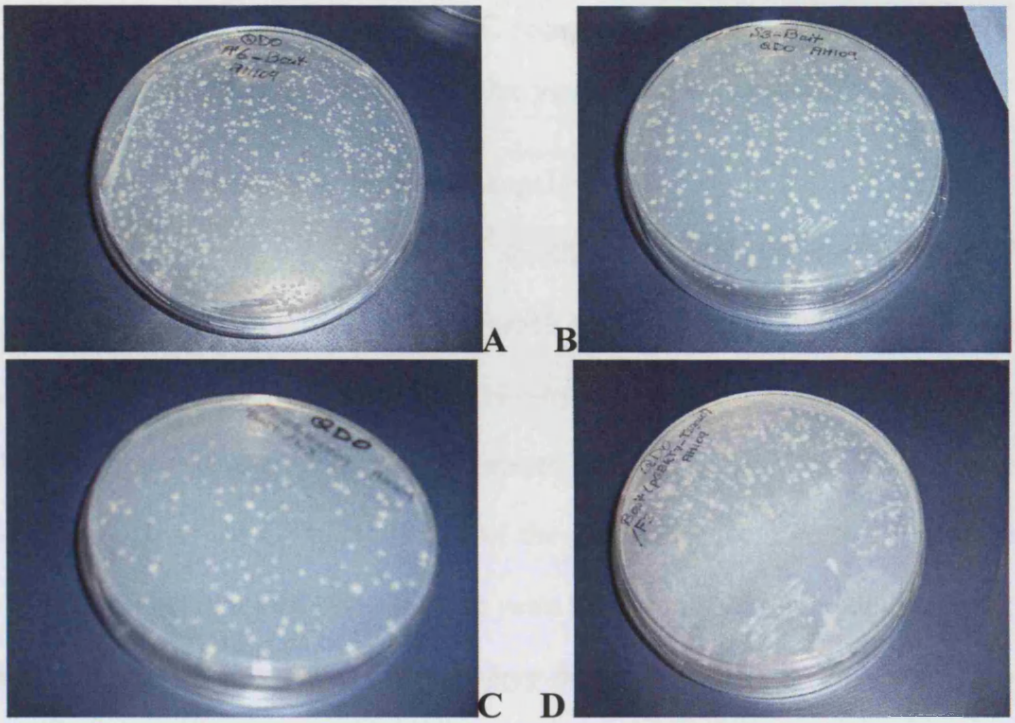


Figure 5.1A- Co-transformation of bait construct with candidate interacting proteins in AH109 yeast strain. This experiment confirmed possible interaction of the Tsga10 protein (encoded by the bait construct) with proteins encoded by A'6 (A), F3 (B), L3 (C) and S3 (D) constructs.

5.2 Determining whether the N or C terminus of the Tsga10 interact with the proteins identified by the yeast two hybrid assay

In previous sections I have shown the Tsga10 protein may be cleaved, and its N-terminus localizes in sperm tails as a fibrous sheath protein. The focus of this part of project was to find out which half of the Tsga10 protein interacts with the proteins identified above by performing separate yeast two-hybrid assays with each portion of the gene. The gene was divided into two fragments by amplifying with PCR primers designed to amplify either the 5′ or 3′ half of the gene sequence. The two fragments were then used to construct bait plasmids. The yeast two-hybrid screens were carried out by co-transforming the yeast with bait and prey plasmids and by mating yeast cells containing the bait plasmid with those carrying the prey plasmid. The yeast cells were plated onto medium, which could select those cells in which a protein-protein interaction had taken place. As before, the experiment was carried out, once using co-transformation of yeast strain AH109, and once using mating between strains AH109 and Y187 in order to introduce the bait and prey plasmid combinations into the same yeast cell.

Making the Constructs- Figure 5.2A shows Tsga10 5′ N-terminus and 3′ C-terminus halves of Tsga10 amplified with primers A-D (see methods section 2.19.2). The product of this PCR reaction were then purified by ethanol precipitation and digested with restriction enzymes *EcoRI* and *Hind III* (figure 5.2B).

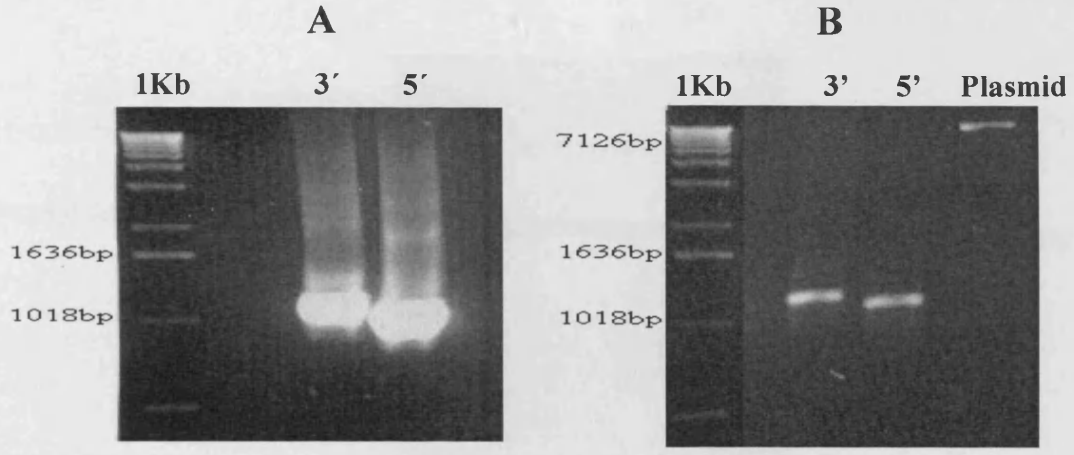


Figure 5.2- A) Agarose gel (1.3%) electrophoresis of purified PCR products of Tsga10 N/C-portions **B)** Electrophoresis of the products from (A) with *EcoR*I and *NcoI.* The marker is 1 Kb ladder

The wide bands present on the gel (figure 5.2A) were produced by amplification of the two portions of *Tsga10*. The 5′ band can be seen to be approximately 1000bp which corresponds to its predicted length of 1079bp. The band produced by the C-terminal half (3') is about 1100bp in size, which corresponds to its predicted length of 1142bp. By comparing the brightness of the bands to those produced by the ladder, the amount of DNA estimated to be in each reaction was 300ng/µl. Figure 5.2B shows the DNA purified from the restriction double digests (*Eco*RI and *NcoI*). The digested fragments were predicted to be 1059bp (5′) and 1137bp (3′) in length due to their ends having been trimmed by the restriction enzymes (*Eco*RI and *NcoI*). The band produced by the plasmid corresponds to the size of pGBKT7, which is 7.3kb in length. Once the DNA had been cleaved by the same restriction enzymes, two *Tsga10* fragments were ligated into the other plasmid. Figure 5.2C shows the two portions of the *Tsga10* gene ligated into the pGBKT7 plasmid:

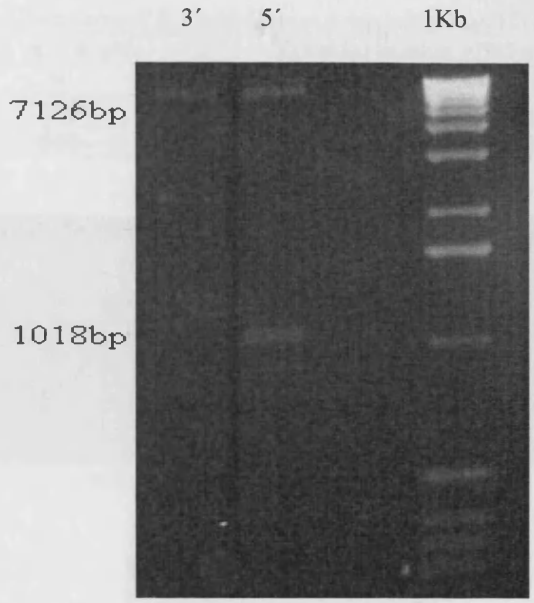


Figure 5.2C- Electrophoresis of ligated DNA of Tsga10 5'/3'-portions (2µl ligation reaction, 1µl Loading buffer, 2µl 1Kb ladder. Run 2hours at 40v)

The gel shows that the ligation reactions were partially successful. A faint band produced by unligated 5′ fragments can still be seen at about 1000bp. The larger bands of approximately 8000bp represent the plasmid ligated with each of the two portions of *Tsga10*. Then plasmids containing the two portions of *Tsga10* were transformed into competent *E. coli* cells. A positive control was included with cells transformed with 1µl of the empty plasmid (diluted to give 1ng/µl). The agar contained kanamycin because the pGBKT7 plasmid is kanamycin resistant; therefore, only cells containing the plasmid would be able to grow on the plates (figure 5.2D). The transformation efficiency was calculated using the following equation and the results are shown in table 5.2A:

$$\frac{\text{No. of colonies}}{\text{1ng pGBKT7}} \times \frac{10^6}{1000} \times \frac{\text{1000µl (total vol. transformation)}}{\text{200µl (vol. transformation plated)}}$$

Table 5.2A- Transformed *E. Coli* colonies containing Tsga10-5'/3' portions

Plate	No. of colonies	Transformation efficiency (cfu/µg plasmid)
5′	80	4×10^5
3′	100	5×10^5
pGBKT7	600	3×10^6

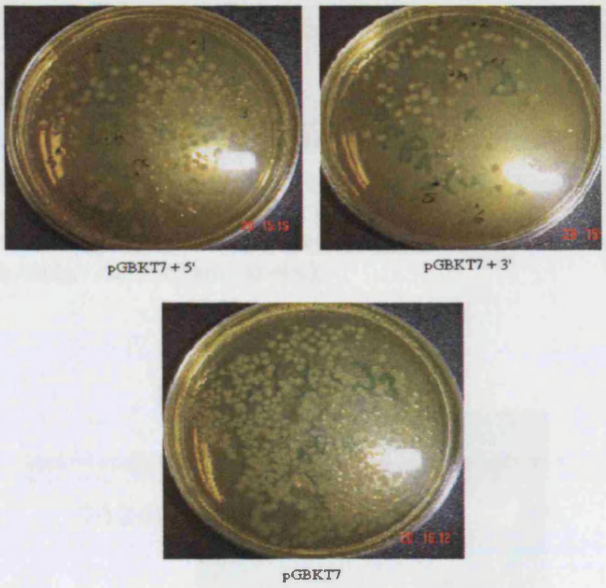


Figure 5.2D- The *E. coli* colonies which were transformed with the plasmids (Tsga10-5'/pGBKT7, Tsga10-3'/pGBKT7) to make bait constructs

The colonies numbered in the photographs in figure 5.2D were tested by PCR with primers pGAD424-5' and pGAD424-3' to see if they contained the expected inserts. A negative control was included for each set of primers with water replacing the DNA template. The results are shown in figure 5.2E. The gel shows that three of the *E. coli* colonies transformed with pGBKT7 and the 5′ fragment took up plasmids containing the insert, and one colony took up the 3' plasmid and insert. The very small fragments at the bottom of the gel were produced by unused primers. Two colonies were selected for DNA miniprep and yeast transformation, one containing the 5′ portion of *Tsga10* and one containing the 3′ portion (figure 5.2F).

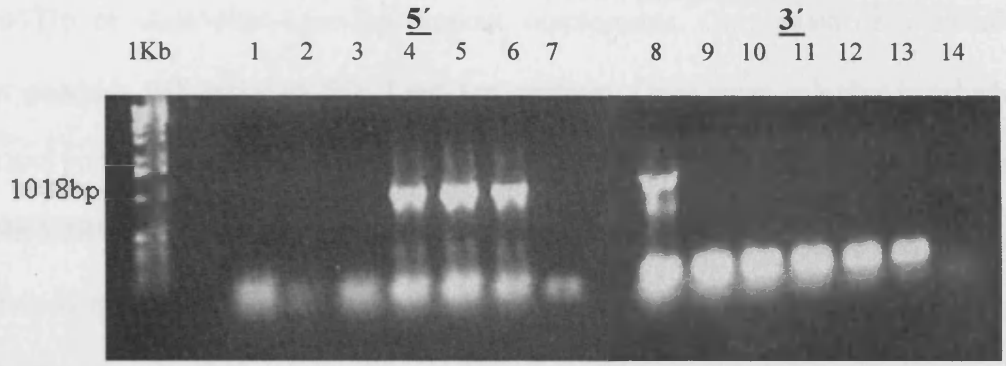


Figure 5.2E- Electrophoresis of PCR products of Tsga10 5' (lanes 1-6) and 3' (lanes 8-13) ends ligated into the pGBKT7 vector and transformed into *E. Coli*. Primer pairs were (YTHM, B) for 5' and (C, MR19TC10clone) for 3' portions. Lanes 7 and 14 contained water instead of DNA. (2µl PCR product, 1µl Loading buffer, 2µl 1Kb ladder. Run for 1hour at 40v).

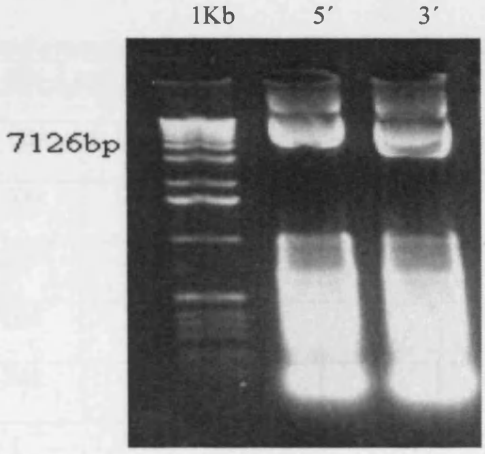


Figure 5.2F- Electrophoresis of the plasmid miniprep products of Tsga10-5' and Tsga10-3' (5µl miniprep, 1µl Loading buffer, 3µl 1Kb. Run for 1hour at 40v)

Co-transformation of yeast- The *Saccharomyces cerevisiae* yeast strain, AH109, was co-transformed with plasmid containing either of the two portions of *Tsga10* (bait) and one of the three different genes (prey) encoding proteins previously shown to interact with the intact Tsga10 protein. The gene fragments used in the prey plasmid were A′6 (*Cyc1*), F3 (*Odf2*), L3 and S3b (both like *FLJ32880*). Parallel controls were included; cotransformation with pGBKT7-Tsga10 (the complete *Tsga10* gene) and pGBKT7 (empty plasmid). The co-transformed yeasts were plated onto SD medium with either

-Leu/-Trp or -Ade/-His/–Leu/-Trp dropout supplements. Only yeast cells containing both plasmids will grow on SD/–Leu/-Trp medium. Only yeast colonies in which the bait and prey proteins have interacted will grow on SD/-Ade/-His/–Leu/-Trp medium. Table 5.2B shows the results of the yeast co-transformation. The plates marked * also contained many small red colonies produced by yeast cells which were able to grow without expressing the reporter genes and are not evidence of an interaction between two proteins. The red colour indicated that the yeast still had a '-Ade' phenotype. In results 100 colonies or more are counted.

Table 5.2B- Co-transformed yeast colonies containing Tsga10 and Tsga10/5', Tsga10/3' portions

AH109	**SD/–Leu/-Trp**				**SD/-Ade/-His/–Leu/-Trp**			
	A'6	**F3**	**L3**	**S3**	**A'6**	**F3**	**L3**	**S3**
5'	**88**	300	200	17	16	14	200	10
3'	51	300	300	2	0	3*	0	10
Tsga10	25	200	200	0	17	100	100	0
pGBKT7	23	300	300	4	0*	19*	0*	2*

The SD/-Leu/-Trp plates were set up to select yeast cells, which had been transformed with both plasmids. It was therefore expected that colonies would grow on all these plates, provided the co-transformation procedures were successful. Colonies were found on all the plates except where the yeast cells had been transformed with *Tsga10* and S3. However, few colonies were found on any of the S3 plates indicating that the transformation efficiency of the plasmid containing this gene was poor. The transformation efficiency of the plasmid containing the A'6 gene was also reduced, although to a lesser extent. These results were probably caused by a low concentration of plasmid being used for the co-transformations.

In co-transformation with pGBKT7 vector, no colonies were expected to grow on the SD/-Ade/-His/-Leu/-Trp plates. However, small numbers of colonies were found on two of the plates, those with the F3 and S3 genes. It may be possible that the activation domain in the prey plasmid was able to activate transcription without a DNA-binding domain. However, as this had not been seen with the same plasmids in previous experiments and most of the colonies were very small and red, they were ignored. Interactions were expected to take place between the Tsga10 protein and the proteins transcribed from all four bait plasmids as these results had already been seen in previous yeast two-hybrid screens. The expected results were observed for proteins L3 and F3, however, few colonies with the A′6 protein and no colonies with S3 grew. Again this was likely to be due to a low concentration of plasmid being used.

The results show a strong interaction between the N-terminal portion of the Tsga10 protein and L3 and probably with S3 (given the small numbers of colonies overall with S3). Weaker interactions also took place between this part of Tsga10 and A′6 and F3. The C-terminal portion of Tsga10 showed only a weak interaction with F3 though the reaction with S3 was comparable to that of the N-terminal portion. As such a small number of colonies containing the 3′ portion of *Tsga10* and F3 grew, it is possible that this result is a false positive. Therefore, it is likely that interactions between Tsga10 and other proteins take place primarily in the N-terminal portion but to a lesser extent may also involve the C-terminal portion of the protein.

This experiment could not be taken any further because of lack of time. However, it is currently being repeated and extended by a third year undergraduate project student.

Testing the two halves of Tsga10 interactions using the yeast-mating method;

Transformation of Yeast strains for yeast mating experiment- Transformations were carried out for use in a yeast mating. AH109 was transformed with the two portions of *Tsga10* and a second yeast strain, Y187, was transformed with the three genes known to interact with *Tsga10*. The transformed AH109 yeast cells were plated onto SD/–Trp medium and the transformed Y187 yeast plated onto SD/–Leu medium. As controls, the transformed yeast cells were also plated onto SD/–Leu/-Trp medium. The purpose of the controls was to check that the yeast strains were only transformed with one plasmid. The results of the transformations are shown in table 5.2C.

Table 5.2C- Transformation of candidate proteins and the N and C termini of Tsga10 into yeast Y187

	SD/-Leu	SD/-Trp	SD/–Leu/-Trp
Y187 + A'6	100		-
Y187 + F3	100		-
Y187 + L3	100		-
Y187 + S3	100		-
AH109 + 5'		300	-
AH109 + 3'		300	-

As expected, no colonies were able to grow on the SD/–Leu/-Trp plates as no yeast cells contained both the bait and prey plasmids. Colonies were found on the other selection plates indicating that the yeast strains successfully took up the plasmids they were transformed with. More colonies from the AH109 strain grew than from the Y187 strain. This may be due higher plasmid concentrations being used to transform this strain. It is also possible that the Y187 yeast cells were not as competent as the AH109 cells.

PCR of colonies- Samples of the yeast colonies were tested by PCR to check for the presence of the two portions of *Tsga10* (figures 2.19.2A-B). A negative control was included with water replacing the DNA. Figure 5.2G shows the results of the PCR reactions:

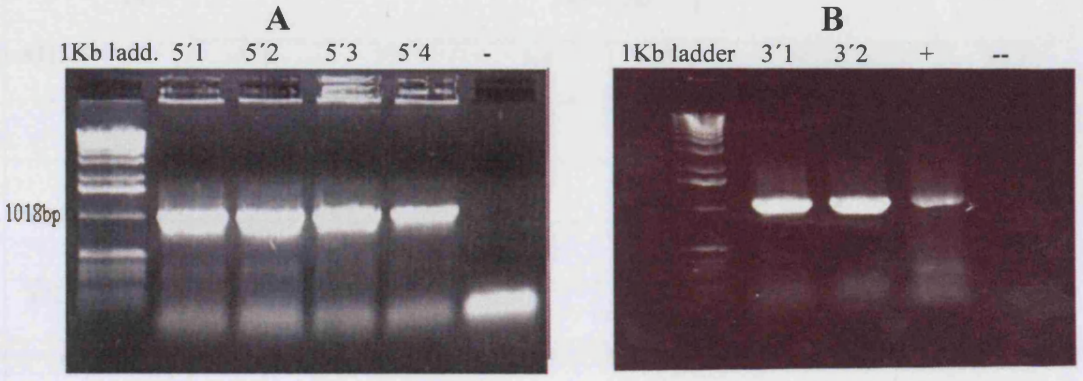


Figure 5.2G- Electrophoresis of PCR product of Tsga10 5' (A) and 3' (B) ends expressed in yeast colonies. Primers were (YTHM, B) and (C, MR19TC10clone) (25µl PCR product, 5µl Loading buffer, 15µl 1Kb ladder. Run at 40v)

All four samples of DNA amplified from the yeast colonies transformed with the 5′ portion of *Tsga10* have produced bands of about 1000bp, which corresponds to the size of the 5′ fragment. Also the correct size DNA was amplified from the samples transformed with the 3′ portion of Tsga10. Small size bands on the gel were produced.

Yeast mating- The yeast strains AH109 (containing the two portions of *Tsga10*) and Y187 (containing the three genes known to produce proteins which interact with Tsga10) were mated to confirm the protein-protein interactions which took place in the yeast co-transformations. Control matings were also included. The control bait plasmid used was pGBKT7-53 and the control prey plasmid was pGADT7-T. The bait strains were also mated with untransformed Y187 cells and the prey strains were mated with untransformed AH109 cells and with pGBKT7 containing the full *Tsga10* gene. The

results of the yeast matings plated onto SD/-Leu/-Trp plates are shown in table 5.2D. This plating demonstrates whether there are two plasmids present in the yeast cells and indicates that the mating had worked.

Table 5.2D- The number of colonies grown by Yeast mating of Y187 strains (containing candidate interacting proteins) to AH109 strains (containing bait constructs)

Y187 / AH109	SD/-Leu/-Trp					
	A'6	F3	L3	S3	pGADT7-T	Y187
5'	500	500	500	500	500	0
3'	500	500	500	500	500	0
Tsga10	500	500	500	500	500	0
pGBKT7-53	0	0	0	0	14	0
AH109	0	23	0	0	0	0

The untransformed AH109 and Y187 yeast cells when transformed with any other strain should have been unable to grow on the -Leu/-Trp plates, as they should have had only one of the *LEU2* or *TRP1* genes. However, a small number of colonies were found on the plates containing normal AH109 cells and Y187 cells with F3 (23 colonies). This result was probably caused by contamination of the plate with yeast cells containing a bait plasmid.

All the yeast cells containing the *p53* gene and the prey plasmids should have grown on the -Leu/-Trp plates. Only those cells with *p53* and the *T-antigen* gene formed colonies due to known interaction between *p53* and *T-antigene* (provided by Clontech. Kit).

Colonies were found on all the plates where part or all of Tsga10 was present as well as the four proteins with which it interacts. This indicates that the yeast cells successfully mated to produce progeny cells containing the bait and prey plasmids.

The results of the platings onto SD/-Ad/-His/-Leu/-Trp are shown in table 5.2E. This illustrates where an interaction has taken place between the bait and prey plasmids.

The untransformed AH109 and Y187 yeast cells when mated with any other strain should have been unable to grow on the '-Leu/-Trp/-Ade/-His' plates as they should have had only one of the *LEU2* or *TRP1* genes. Yeast cells containing the *p53* gene were also unable to produce proteins, which could interact with the products of A'6, F3, L3 or S3. Therefore, no colonies were formed. The protein products of *p53* and the *T-antigen* gene are known to interact strongly yet only a small number of colonies were found.

Table 5.2E- Yeast mating of Y187 strains (containing candidate interacting proteins) to AH109 (containing bait constructs) plus controls (see text)

Y187 / AH109	SD/-Ade/-His/-Leu/-Trp					
	A'6 (Cyt C1)	F3 (Odf2)	L3 (FLJ32880)	S3 (FLJ32880)	pGADT7-T	Y187
Tsga10-N	0	0	100	100	0	0
Tsga10-C	0	0	80	60	0	0
Tsga10	200	100	100	100	0	0
pGBKT7-53	0	0	0	0	50	0
AH109	0	0	0	0	0	0

As was expected, strong interactions took place between the full Tsga10 protein and the proteins known to interact with it. Strong interactions between the N-terminal portion of the protein and L3 and S3 (both FLJ32880) also took place, as well as these proteins and the C-terminal portion of Tsga10. These results show that the interaction between FLJ32880 and Tsga10 takes place in both N-terminal and C-terminal portions of Tsga10. Although the full Tsga10 protein interacted strongly with CYC1 and ODF2, there were

no interactions between the separate portions of Tsga10 and these proteins. This suggests that interactions with these proteins may require the presence of the whole Tsga10 protein.

Table 5.2F- Comparison between the results of yeast mating and transformation using Tsga10 and its N, C termini fragments as bait constructs

	Yeast Cotransformation					*Yeast Mating*			
	L3	**S3**	**A'6**	**F3**		**L3**	**S3**	**A'6**	**F3**
Tsga10	+	_	+	+		+	+	++	+
N terminus	++	+ (few)	+ (few)	+ (few)		+	+	_	_
C terminus	_	+ (few)	_	_		+	+	_	_

The most conclusive results were obtained from the yeast mating experiment rather than the co-transformation (table 5.2F). In this case both the complete Tsga10 protein and its two proteins interact with L3 and S3, which makes since these clones both contain the FLJ32880 gene.

To test any possibility of Tsga10 interaction (the whole gene or the N and/or C-termini) with the other major outer dense fiber protein Odf1, its interaction with Odf1 protein was assayed (by using pGAD424-Odf1 plasmid -containing activating domain- which was provided by Prof. Frans Van der Hoorn as a gift) both by yeast mating and co-transformation. However, no significant interaction was found (data not shown).

Chapter VI:

Further examination of the *Tsga10ip (FLJ32880)* gene and its predicted protein

This gene is located on human chromosome 11(11q13.1) (accession no. AK057442, BAB71488, NP_689975), mouse chromosome 19 (XM_355133) and rat chromosome 1 (1q43) (NM_001004278, AY690667, XM_341991). There are eight predicted exons in the human and mouse FLJ32880 cDNA sequence, and the protein encoded by the gene is predicted (by Ensembl) to be a nuclear protein (84% probability). The sequences of mouse and rat (genes that encodes) FLJ32880 hypothetical protein have only 70% homology to human sequence (table 6A). This homology was much greater for Tsga10 (90-94%), and the low homology for FLJ32880 hypothetical protein seems to be due to some gaps in the sequences (even between rat and mouse, see table 6.2.3B). Figure 6A shows the alignment of the human, mouse and rat predicted protein sequences. In human the gene is expressed just in **round spermatids** and **testis** (TESTI2004122 Clone), and transcripts are found in the dbEST in NCBI GenBank, which are restricted to the **embryo, medulla, heart, thymus and pancreas** (1810013N20). Searches using *UniGene* have identified two protein similarities for FLJ32880 hypothetical protein: one to an "unconventional myosin-15" (A59266) and one to "PRP3-mouse proline-rich protein MP-3" (P05143). Also the mouse FLJ32880 hypothetical protein is shown to have 29% similarity to human TBP-associated factor 4 (a TATA box binding protein) (NP_003176.1).

A number of *FLJ32880* neighbouring genes is listed in table 4.5 using MapViewer, BLAST (http://www.ncbi.nlm.nih.gov/entrez) which seem to have a role in the cell cycle.

CLUSTAL W (1.82) multiple sequence alignment (FLJ32880 hypothetical protein)

```
Mouse       ----------MEMAERAVPELAG--TCHFLSTKSRALGDIQKNR------CLGSGDSLQS 42
Rat         ------MLNTDQLLVRAISERPRKSTEPLVPGTPTGLFSLLSNISPEEQGRLGSGDSLQS 54
Human       MGQDTDMLNTYQQLVRTPSVRPGQDVRLQAPGTRTGLLKLLSTVSQDKQGCLGSGDGVPN 60
                       :   *: .  .   .  . .* .: ..        *****.: .

Mouse       QSCQLQRSYSAGQTTKKERKARRRNKKGRGSAEAEDLFSPPSRKPSFPFQWAWESFIIDG 102
Rat         QSCQQQRSYSAGQTTKKERKPRRRNKKGRGSAEAEDLFSSP-RKPSFPFQWAWESFIIDG 113
Human       QDLQ-QRPQSSRQTAKKDRKPRGQSKKGQGSEESEDHFPLLPRKPSFPFQWAWESIATDV 119
            *. * **. *: **:**:**.* :.***:** *:** *.   *************:  *

Mouse       QALFQSSSSMAMGHRPLLLPPAAPQCKSRHKSVANLSEDLRACHKMEEQSVGRRHHLGGW 162
Rat         QALLQSGSSVAVGHRSLLFPPAAPQCKTRHKSVANLSEDLRACHKSEVQNLGRRYQPGAW 173
Human       RAVLQP-SSPTPGHQALPMPSSFSQRQSRRKSTANLPEAHGCCWKTEAQNLKARQQLGAW 178
            :*::*. ** : **:.* :*.: .* ::*:**.***.*   .* * * *.:  * : *.*

Mouse       ANFPLPLSKVESQGLDQPS-FWLSGK--GSGSESEDVLEIEGQNSEEAEKSLSSGELPQL 219
Rat         ANLSLPLGKAESQGLERPT-FWSTGK--GSGSECEDVSEVEGQNADEAEKSLSTGELPQL 230
Human       GGVSIPTGKGE-LGSEPPSGLQLPGRRPGSGSASDKQVQLQSLGAEEAERGLSSGVLPQR 237
            ....:* .* *  * : *: :  .*:  **** .:.  :::. .::***:.**:* ***

Mouse       PGQSLILEEELISEAMEE----EHSDPHKGKGSSVNKGGNSGEKGSEEGELQSHNQGSSS 275
Rat         PGQGLTLEEELISEVMEE----EEHN--RRKGSSVNKGRNSGEKGSEEGELQSHNQGSSS 284
Human       PRRGSISEEEQFSEATEEAEEGEHRTPCRRRAGCQRKGQISGEEASDEGEVQGQSQGSSP 297
            * :.   *** :**. **    *.    : :... .**  ***:.*:***:*.:.****.

Mouse       SFNSLRKPQKGISRAKDLKGPWDLERLHRQLQEELECGPQKQTWKALRAAVQASARNKKT 335
Rat         NSNSLRKSPKGTSGAKEFKGPWDLERLHRQLQEELESGPQKQTWKALRAAVQASARNRKT 344
Human       SFNNLRRRQWRKTRAKELQGPWDLEKLHRQLQRDLDCGPQKLPWKTLRAAFQASKRNGKA 357
            . *.**:     : **:::******:******.:*:.**** .**:****.*** ** *:

Mouse       PILGDDESFLSANFPNRTFHKRQEATR---------------------------------- 362
Rat         PVTGEEESFLTANFPNRTFHKRQEATRTWALRDTYFFPQSLTSGPHAPPCYLGLPRAPDP 404
Human       YASGYDETFVSANLPNRTFHKRQEATR---------------------------------- 384
               * :*:*::**:**************

Mouse       ---------NLLRAWEQQQVKERQQAEVRRAREQQVQQQVARCLAAYTSGGNRGTLAPQR 413
Rat         HGHPPFLTRNLLQAWEQQQLKEKQQAEMRRAREQQVQQQVARCLAAYTPGGNRGTLGPQR 464
Human       ---------SLLQAWERQRQEERQQAELRRARTQHVQRQVAHCLAAYAPRGSRGPGAAQR 435
                     .**:***:*: :*:****:**** *:**:***:*****:. *.**. ..**

Mouse       KLEELRRKERQRFAEYQAELQGIQHRVQARPFLFQQAMQTNARLTANRRFSQVLSALGVD 473
Rat         KLEELRRKERQRFAEYQAELQGIQHRVQARPFLFQQAIRP-------------------- 504
Human       KLEELRRQERQRFAEYQAELQGIQHRVQARPFLFQQAMQANARLTVTRRFSQVLSALGLD 495
            *******:******************************::.

Mouse       EDQLLAEAGNAESTPRKHRSNRSLRAEMEPSSQSPPKTEPTSSQPGRRPSPTLDPDYSPR 543
Rat         ------MPGSQQTGASPR-CCQHWEWMRNSCWLRQAMQRASPGNTGATGHLE-------- 549
Human       EEQLLSEAGKVDREGTPRKPRSHRSMGVRMEHSPQRPPRTEPTGSQPDRHYNPSLDPECS 555
                  ... :  .:..    .
```

Figure 6A- Alignment of the predicted protein sequences encoded by human, mouse and rat FLJ32880 gene. Antibodies were raised against the regions marked in red (section 6.2). The underlined sequence is a possible PDZ (post-synaptic density) domain. Asterisks indicate exactly identical amino acid, while ':' indicates similar amino acids and '.' Shows the ones with less similarity.

Table 6A- UniGene comparison of the FLJ32880 sequence between human, mouse and rat.

UniGene		
H.sapiens	Hs.350671	hypothetical protein FLJ32880
M.musculus	Mm.329659	RIKEN cDNA 1810013N20 gene
R.norvegicus	Rn.56956	Rattus norvegicus similar to hypothetical protein FLJ32880 (LOC361707), mRNA

Species	Gene	aa%ID	nt%ID
H.sapiens	**FLJ32880**		
vs. M.musculus	1810013N20Rik	57.5	70.4
vs. R.norvegicus	LOC361707	57.7	70.4
M.musculus	**1810013N20Rik**		
vs. H.sapiens	FLJ32880	57.5	70.4
vs. R.norvegicus	LOC361707	84.6	90.5
R.norvegicus	**LOC361707**		
vs. H.sapiens	FLJ32880	57.7	70.4
vs. M.musculus	1810013N20Rik	84.6	90.5

My nucleotide sequence of the rat FLJ32880 protein was longer than the one, which had been found in NCBI GenBank, so my sequence was deposited in the NCBI GenBank database as accession number AY690667. Also the 'HUGO Gene Nomenclature Committee' (http://www.gene.ucl.ac.uk/nomenclature) agreed to name it '***TSGA10IP***' which stands for the gene coding ***TSGA10 I***nteracting ***P***rotein.

6.1 Studying the expression of *Tsga10ip* gene transcripts (mRNA) in embryogenesis

Since *Tsga10* expression is found during developmental stages of embryogenesis as well as developmental stages of spermatogenesis, investigation into the expression of the gene that encodes Tsga10-interacting protein (*Tsga10ip*), may be interesting. It may indicate a role for Tsga10/Tsga10ip in all actively dividing and differentiating cells. For RT-PCR using cDNA from various developmental stages of mouse embryos, a pair of mouse homologue *Tsga10ip* gene-specific forward (FLJ32880F) and reverse (FLJ32880R) primers in different exons were used (9.6-kb genomic and 475-bp cDNA size). The result shows the presence of its transcript in embryonic developmental stages from 14 dpc onwards (figure 6B). This may indicate a role in embryogenesis.

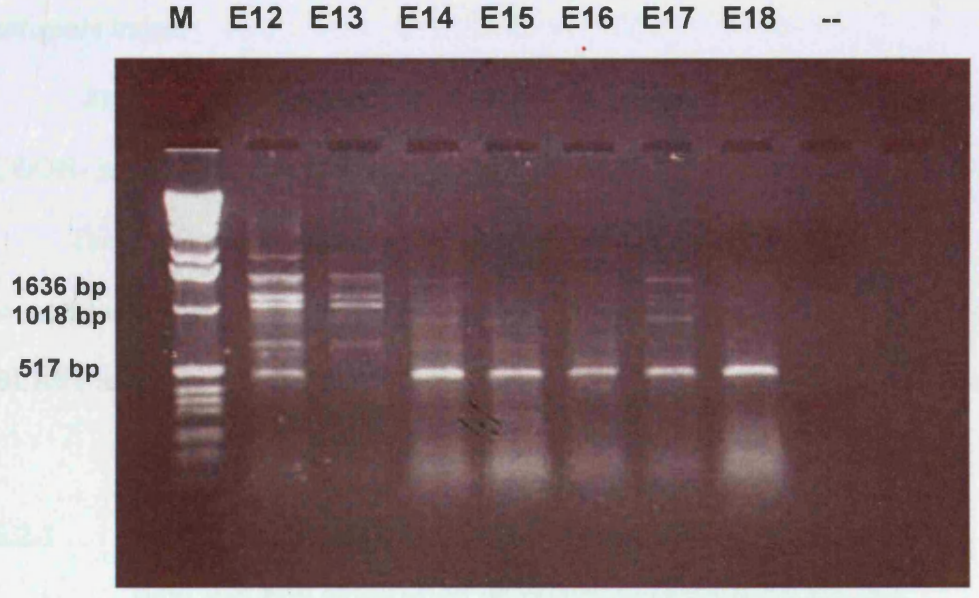


Figure 6B- Electrophoresis of RT-PCR showing *Tsga10ip* gene expression in developmental stages of mouse embryogenesis (E12 – E18). PCR products are seen in lanes E14-E18. Bands in lanes E12 and E13 are probably the result of non-specific amplification. The control lane (-) has water instead of DNA. Lane M is 100Kb ladder.

6.2 Raising antibody against mouse homologue Tsga10ip protein, and studying its localisation in sperm by immunofluorescence

Bioinformatic analysis of Tsga10ip protein to target peptides for raising antibody- In the case of the Tsga10ip protein, it was hard to find good sequences that were homologous in all three species (human, rat and mice). However, the following sequences were selected that were completely identical in mouse and rat:

1) Amino acid sequence 56-79 (N-terminus): NH2-RRNKKGRGSAEAEDL-COOH- this region is very good in terms of antigenicity. The hydrophilicity index as is the probability of a surface location.

2) Amino acid sequence 245–260 (Middle): NH2-KGSSVNKGRNSGEKG-COOH - is comparable to the first peptide; it also has a good hydrophilicity and antigenic index.

3) Amino acid sequence 679–691 (C-terminus): NH2-KLEELRRKERQRFAE-COOH- is also a very good choice in terms of hydrophilicity and surface probability.

The sequences those were high in as many parameters as possible were selected. In addition, these sequences were checked for specificity to the Tsga10ip protein using BLAST sequence homology. No significant homology to other proteins was seen.

6.2.1 Raising polyclonal antibodies against Tsga10ip selected peptides and evaluation of relative amounts by ELISA

Two immunized rabbits generated antibodies that recognized the Tsga10ip protein candidate peptides EP040254, EP040255 and EP040256 based on the protocol described in materials and methods. Titration of these antibodies was determined by

ELISA assay using the peptide as antigen. As shown in the following graphs (see figures 6C (A-B)) the immunization programs were properly done with a good level of antibodies consistent with the results from the results obtained subsequently for protein assays (immunocytochemistry and immunostaining).

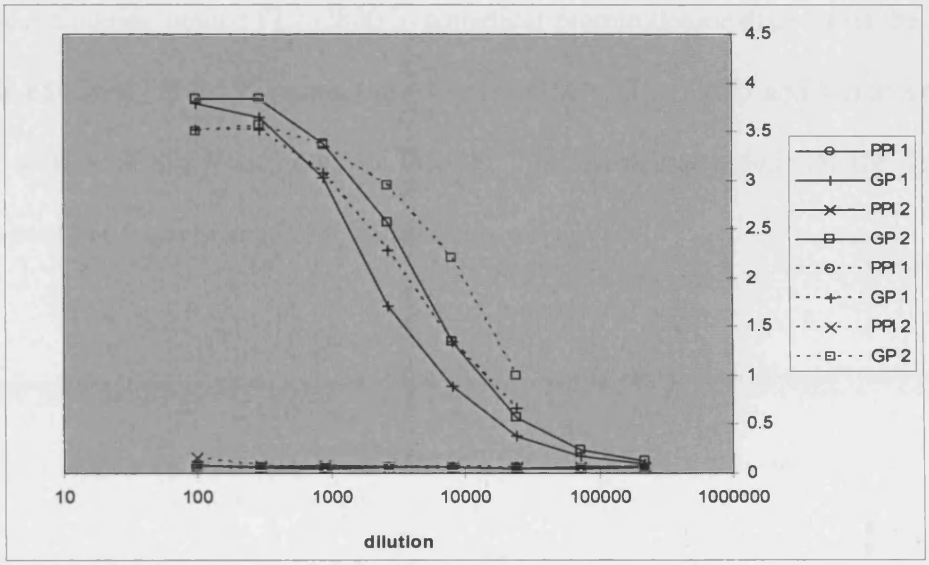


Figure 6C(A)- ELISA result for Tsga10ip protein rabbit antisera against the N-terminus peptide (continuous line) and Carrier (discontinuous). PPI= Pre-immune serum, GP 1/2 = final bleeding antibody

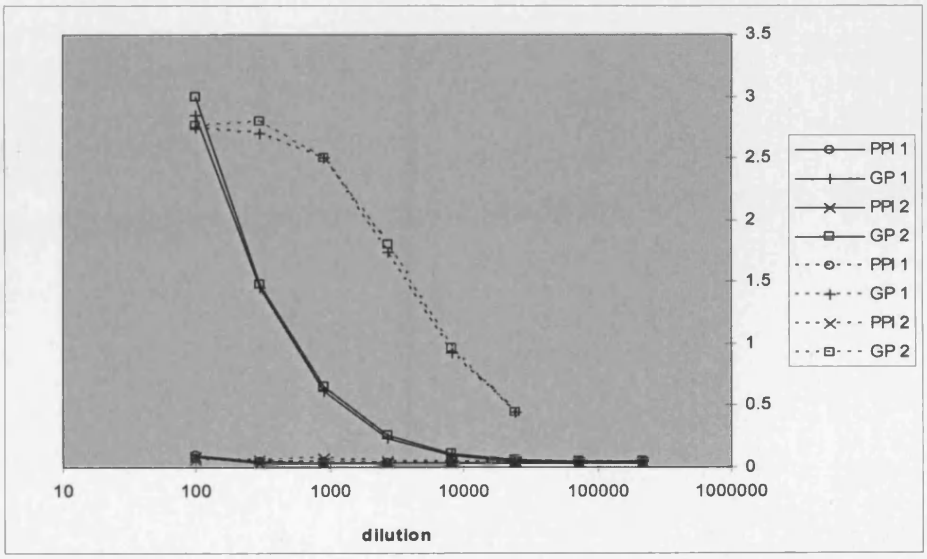


Figure 6C(B)- ELISA result for Tsga10ip protein rabbit antisera against the C-terminus peptide; Peptide (continuous line) and Carrier (discontinuous). PPI= Pre-immune serum, GP 1/2 = final bleeding antibody

6.2.2 Studying the location of Tsga10ip protein in sperm

The antibodies described above were used to analyse the localisation of the Tsga10ip protein pattern in mature sperm by immunofluorescence. Epididymal mouse and bovine sperm were prefixed (in 4% PFA) and smeared on microscope slides. Hybridisation with the antibodies against FLJ32880 hypothetical protein demonstrated that the protein is localized mainly to the principal piece of sperm tails (figures 6D and 6E) in a manner similar to that of the N-terminus of Tsga10. This result suggested that the Tsga10ip protein could be a component of the fibrous sheath (FS).

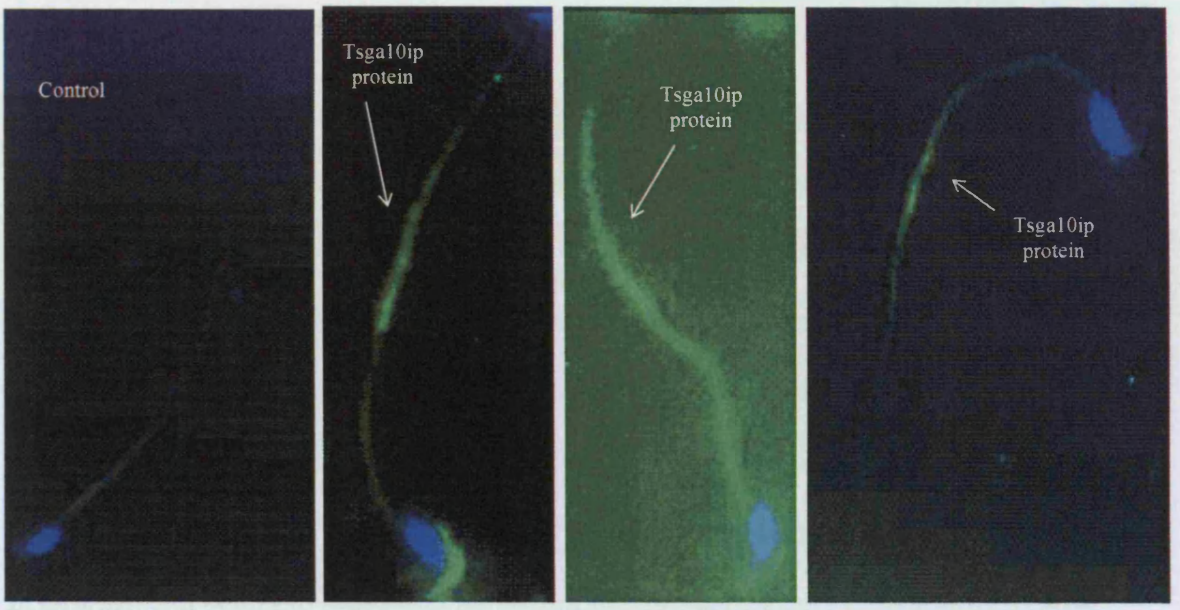


Figure 6D- Indirect immunofluorescence analyses of Tsga10ip in epididymal mouse sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass cover slips and incubated with anti-Tsga10ip antiserum followed by a secondary antibody conjugated to FITC. The principal piece of sperm tails, which is distal to the midpiece, strongly fluoresced (green). Sperm heads show very weak fluorescence. The control (left panel) are seen as blue areas due to DAPI staining.

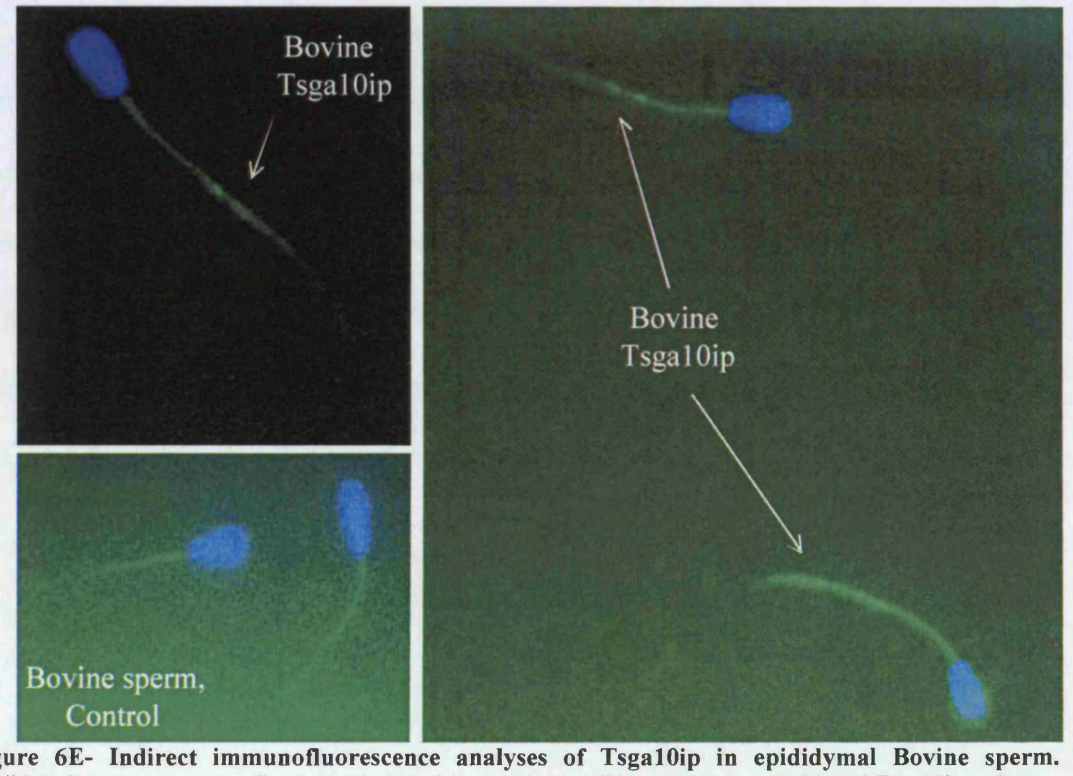


Figure 6E- Indirect immunofluorescence analyses of Tsga10ip in epididymal Bovine sperm. Epididymal spermatozoa were fixed on poly-L-lysine coated glass slides and incubated with anti-Tsga10ip antiserum followed by a secondary antibody conjugated to FITC. The principal piece of sperm tails fluoresced more strongly than the midpiece, although the midpiece does seem to be fluorescing to some degree. The control (left panel) are seen as blue areas due to DAPI staining.

6.2.3 Discussing the function of Tsga10ip

We found an interaction between Tsga10 and Tsga10ip proteins by yeast two hybrid assays. To show *in vitro* protein interaction between Tsga10ip and Tsga10 proteins, immunoprecipitation by protein G agarose beads and analysis by SDS-PAGE of the precipitants were used several times. However, for unknown reasons it failed and remains to be answer in the future. However, *in vivo* information and Tsga10ip localization in principal piece of sperm tail led to the renaming of FLJ32880 hypothetical protein to Tsga10 interacting protein (Tsga10ip). UniGene is not very helpful when it comes to predicting the role of this protein through its similarity to other proteins. UniGene predicts human TSGA10IP has 27% similarity to a mouse proline-rich protein (PRP3; MP-3, sp: P05143) as well as to a human unconventional myosin 15. Also mouse Tsga10ip is shown to have 29% similarity to human TBP-associated factor 4 (a TATA box binding protein) (NP_003176.1). By searching for Tsga10ip homology by Pfam, VP4 and FliS are come up as candidate proteins, which may have similar motif to it (see table 6B). VP4 is a viral protease (Jagadish et *al.,* 1988). The large RNA segment of birnaviruses codes for a polyprotein (N-VP2-VP4-VP3-C). FliS is a flagellar protein encoded by the FliD operon and is transcribed in conjunction with FliD and FliT although no function is known for it.

Table 6B- Tsga10ip protein Pfam study
(http://www.sanger.ac.uk/cgi-bin/Pfam/), fs = fragment domain

Domain	Start	End	Evalue	Mode
Birna_VP4	123	137	0.49	fs
FliS	413	426	0.64	fs

Although it seems that *Tsga10ip* has eight exons in mouse and rat, it is not highly conserved (see table 6C). Similarly the TSGA10IP amino acid sequence is not as well conserved as TSGA10 in mouse, rat and human.

Table 6C- *Tsga10ip* exon/intron boundaries using rat genome browser (http://genome.ucsc.edu)

	Rat ***Tsga10ip*** **NM_001004278, AY690667**	**Mouse** ***Tsga10ip*** **XM_355133**
1		5191551 - 5191515
2	5191485 - 5191356	
3		5191304 – 5191265
4	5190190 - 5190045	
5	5189962 – 5189259	5189962 – 5189259
6	5189085 – 5188938	5189085 – 5188938
7	5183592 – 5183306	
8		5183476 – 5183306
9	5180036 – 5179037	5180036 – 5179037
10	5179750 – 5179632	5179750 – 5179632
11	5179220 - 5179070	5179220 - 5179070

It is likely that Tsga10ip also plays a role in the movement of the sperm tail.

TSGA10 and Tsga10ip (FLJ32880 hypothetical) proteins show the following similarities:

- Both of them (or at least a part of them) are localized in the principal piece and head of sperm (fibrous sheath proteins).
- Tsga10 protein has similarities to myosin heavy chain tail and ERM (Ezrin/ radixin/merlin that contains a polyproline region between the helical and C-

terminal domains). Similarly, Tsga10ip has a similarity (27% predicted by UniGene) to an unconventional myosin 15. This reflects the presence of a coiled coil domain in Tsga10ip. Also Tsga10ip is similar (27% predicted by UniGene) to a mouse proline-rich protein (MP-3 which is aligned with human TSGA10IP). A fibrous sheath specific isoform of glyceraldehyde 3-phosphate dehydrogenase (GAPD-S) is reported to have a proline rich amino acid domain as well, which is important and critical for sperm motility (Bunch et al., 1998). More interestingly and recently, a proline-rich region is detected for 'enkurin' protein (a novel calmodulin and TRPC - transient receptor potential-canonical- cation channel binding protein in sperm) that contains predicted ligand sequences for SH3 domain proteins, including the SH3 domain of the p85 regulatory subunit of 1-phosphatidylinositol-3-kinase (PI3-K) (Sutton et al., 2004).

- Both of these proteins are expressed in glia cell lines derived from astrocytoma.
- Both of these proteins are expressed during embryogenesis and astrocytes (I showed by RT-PCR and ICC, see sections 6.1).

The pattern of co-expression and co-localisation of these two structural proteins suggest that they may share functions other than in the sperm tail. One potential pathway could be cell cycle regulation and mitosis (since Tsga10 and Tsga10ip are expressed in some tumours). Another candidate pathway may be cell differentiation (because of their expression during embryogenesis –organogenesis- and development). Expression of Tsga10/Tsga10ip in glia cells indicates their role in neurogenesis and synaptogenesis. Of course, neurogenesis and neuroproliferation require both cell division and

differentiation. These proteins are also likely to be present in the structure of cilia and play a dual role (structural and scaffolding) in it.

Structural proteins in sperm (such as odf2) can have some particular functions as well. Consistent with this description, we found interactions between TSGA10IP/ODF2 and TSGA10, all of them as structural proteins as well as roles in cell division.

Genes in the vicinity of *TSGA10/Tsga10* and *TSGA10IP/Tsga10ip* have almost similar functions, although in different chromosomes.

The following proteins of known function are near to the Tsga10ip gene (mapped on 11q13.1 in human and 1q43 in rat): mitogen-activated protein kinase kinase kinase 3, basophilic leukemia expressed protein BLES03, squamous cell carcinoma antigen, cation chanel, kinesin light chain 2 orthologous, breast cancer metastatic suppressor 1, Solute carrier nucleotide transport, oral cancer overexpressed protein 2, and a protein similar to sperm ion channel are the proteins encoded by genes close to *Tsga10ip/TSGA10IP* on chromosome 11 (table 6D).

Tsga10ip/TSGA10IP is located on or very close to the D11S913 locus of human 11q13 –candidate region for multiple endocrine neoplasia, type I (MEN1)- (Williams et al., 1997). Also it may be a member of neighboring cluster genes that bind to TATA-box and regulate transcription complex. It may be consistent with the presence of potential transcription factor binding site in 2 Kb upstream of *TSGA10/Tsga10* which does not need TATA box for transcription. Transcriptional corepressor DRAP1 (encoded by a gene located within 25 Kb of *Tsga10ip*) has a very specific role in regulation of nodal activity during embryogenesis (Iratni et al., 2002). While Tsga10ip is 30% identical to a TATA-binding protein, the DRAP1 interacts with DR1 that in turn interacts with

TATA-binding protein. This could be reinforced by the finding (RT-PCR) (figure 6.1) of *Tsga10ip* expression during mouse embryogenesis.

Table 6D- *TSGA10IP* (mapped on 11q13.1 in human and 1q43 in rat) and genes within 500 Kb

Name	Location Human, Rat, Mouse	Distance	Role / Expression
Squamous cell carcinoma Antigen recognized by T-cell 1 (SART1)	**H 11q13.1 R 1q43 M 19qA**	**2 Kb**	**Squamous cell carcinoma new tumour marker (Cromer et al., 2004)**
DR1-associated protein 1 (DRAP1) (Negative cofactor 2-alpha)	**H 11q13 R 1q43 M 19qA**	**25 kb**	**Facilitates the maintenance of a differentiated state by silencing specific genes, a corepressor of the transcription that interacts with DR1. DR1 interacts with TATA-binding protein**
Basophilic leukemia (BLES03)	**H 11q13 R 1q43 M 19qA**	**30 Kb**	
Cation channel, sperm associated 1	**H 11q12.1 R 1q43 M 19qA**	**55 Kb**	**Required for evoked Ca^{2+} entry and control of flagellar function in sperm (Carlson et al., 2003)**
Mitogen-activated protein kinase kinase kinase 11	**H 11q13.1-11q13.3 R 1q43 M 19qA**	**300 Kb**	**Required for cell proliferation, both mitogen and cytokine activation of B-Raf, JNK (MAPK8), ERK (MAPK3), and p38(MAPK14). Phosphorylate Rb1 and stimulate kRas2 (Chadee and Kyriakis, 2004)**
Kinesin light chain 2 ortholog	**H 11q13.1 R 1q43 M 19qA**	**300 Kb**	**Kinesin structure, target of 14-3-3 (a key regulator in cell signaling pathway mediated by protein phosphorylation (Ichimura et al., 2002)**
RAB1B	**H 11q13 R 1q43 M 19qA**	**300 Kb**	**Member RAS oncogene family**
Breast cancer metastatic suppressor 1	**H 11q13-11q13.2 R 1q43 M 19qA**	**400 Kb**	**Human breast carcinoma metastasis suppressor gene**
Solute carrier nucleotide transport	**H 11q13**	**400 Kb**	**Induced by immediate-early response (IER) genes following serum or growth factor stimulation of cells, within 80 kb of the D11S913 locus of human 11q13 –candidate region for multiple endocrine neoplasia, type I (MEN1) (Williams et al., 1997)**

Chapter VII:

Discussion

What could be the function of TSGA10?

The experimental data presented in the previous results chapters has indicated several functions for the Tsga10 protein. In this section I will examine further evidence for these possible functions.

7.1 The expression pattern of *Tsga10*

My experiments demonstrated the expression of the Tsga10 gene and its encoded protein in the testis where it was localised in the sperm tail. There was also expression in developmental stages of mouse embryos, some cancerous cells and

A Virtual Northern search among EST data from libraries and SAGE data (http://cgap.nci.nih.gov) revealed that the *Tsga10* gene is predominantly expressed in testis but possibly also at very low levels in pancreatic islet, skin, prostate, muscle, lung and brain. Microarray expression data deposited in genome browser (http://genome.ucsc.edu, GNF Expression Atlas 2 Data from U133A and GNF1H Chips) show a very high expression level for *TSGA10* in testis, indeed it is one of the most abundant transcripts.

However, UniGene also shows that *Tsga10* transcripts have been isolated from many tissues (including uterus, melanocyte, testis, brain, kidney, liver, spleen, medulla, primary Lung Epithelial Cells and placenta as well as some tumours including germ cell tumor, leiomyosarcoma, glioblastoma, acute myelogenous leukemia, well-differentiated endometrial adenocarcinoma, and poorly differentiated adenocarcinoma with signet ring cell features). The pattern of the gene expression in mouse, (UniGene Cluster Mm .332756), shows the sources of cDNAs from some normal tissues including cerebellum,

bone marrow, thyroid, brain, testis, E9.5 whole embryo, round spermatids, pooled from multiple mice, cecum, eyeball, 18-day preleptotene spermatocytes as well as hematopoietic Stem Cell (Lin-/c-Kit-/Sca-1) line. In rat, (UniGene Cluster Rn .30052), it is originated from testis. The gene is highly conserved in mammalian species, and a search using PBLAST with the *Tsga10* predicted protein sequence identifies the *mud* gene in *Drosophila Melanogaster*, which is involved in the control of neuroblast proliferation and encodes a coiled-coil protein with 21% homology to TSGA10.

7.2 Potential transcriptional control regions in *Tsga10*

Areas of homology upstream of the first exon may indicate areas that are important recognition sites for transcription factors. The Human *TSGA10* start position is known by 5' RACE. A Computer search for transcription factor binding sites in *TSGA10* (using the program TESS at http://searchlauncher.bcm.tmc.edu/seq-search/gene-search.html) showed a match with myocyte-specific enhancer factor 2 (MEF2), TTTTAAATAG at position –32 nucleotide upstream of human *TSGA10*. Also the search showed a region at -129 to –159 that is present in many sequences and can be a potential promoter site. There is neither conventional TATA nor CCAAT box in the expected positions (-30 to –70). However, a CAAT/GT box (named epidermal growth factor response element 'EGFRE' in the mouse lactoferrin gene) is placed at –21. It activates GKLF in human endometrial carcinoma (RL95-2) transfected cells. Also, gut-enriched Krueppel-like factor (GKLF) (a zinc finger domain transcription factor) binds to CAAT/GT box of mouse lactoferrin gene (Shi *et al.*, 1999). It shows a series of temporal changes during embryogenesis and binds GC/GT-rich sequences. Potential binding sites within 300 bp

of the 5' flanking region of short *TSGA10* transcript found by TESS (http://searchlauncher.bcm. tmc.edu/seq-search/gene-search.html) show five binding sequences for SRY, PU.I, DEF, STE12 and HiNF-A transcription factors. These transcription factors mainly mediate in sex differentiation and spermatogenesis, cell cycle, division or immune cell responses. SRY, C/EBPbeta and EBP were shown by TESS in 338 bp 5' upstream of mouse *Tsga10* (AF530050) transcript as well. A Search of the 2 kb upstream genomic sequence of different *Tsga10* transcripts for transcription factor binding sites using TESS and Matinspector V2.2 (http://www.gsf.de/biodv/matinspector.html) revealed multiple putative transcription factor binding sites that were homologous to the known consensus-sequence motifs, including more than 20 for octamer-binding factor 1 (which plays an important role in germ cell differentiation as well as embryogenesis, embryonic brain development, and oncogenesis), 6 for 'sex hormone receptor', 6 for neurogenin 1/3 (involved in neurogenesis)', 6 for 'tumour suppressor p53 and wilm's', and 3 for 'SOX5' binding sites. Putative transcription factor binding sites in the upstream promoter region of the mouse protein kinase C (*PKC*) gene suggest its involvement in the same processes as TSGA10, including spermatogenesis (SRY), embryogenesis (GATA), development (MYOD, GATA), brain generation (RORA), oxidative environment (NRF) and oncogenesis (p53, AML, NFκB) (Suh et al., 2003). A similar pattern of potential transcription factor binding sites was found in the 2 kb upstream of the TSGA10 gene including GATA, GAGA, SOX, and p53.

Transcription factor binding sites with high homology are summarized in tables 7.2A and 7.2B. The diversity of the regulatory elements in *Tsga10*, if functional, may reflect

the multitude of pathways to which *Tsga10* contributes. A description of AP-2, SP1, C/EBPbeta, MZF-1 and GKLF transcription factors which are closer and frequent in the upstream of both Tsga10 transcripts, are discussed below and description of some other predicted factors is in the appendix. **AP-2** is an activator protein (also known as AP-2alpha; activator protein 2; Ker-1; AP-2A; AP2; KER1), which its C-terminal part interacts with the bHLH-ZIP (helix-loop-helix-ZIP) domain of c-Myc, but not Mad (mitotic arrest deficient). Also it is an activator that controls keratinocyte-specific gene expression, and negatively regulates trans-activation exerted by c-Myc with high-level expression causes "self-interference" (e. g. in N-ras transformed cells). It appears to be

Table 7.2A- Summarized result of 2 kb upstream genomic sequence of short *Tsga10* (AF530050) transcript (sites 38180468-38178468 of mouse genome) for transcription factor binding sites (using the program TESS at http://searchlauncher.bcm.tmc.edu/seq-search/gene-search.html)

Factor	Beg	Sns	Len	Sequence
T01806 p53	-1805	N	10	TGACATGCCT
T01591 P	-1755	N	9	ACCAACCAG
T00644 Oct-1A, 1B, 1C, 2.1, 2.3. 2.4	-1737	N, R	10	ATTCAAATGA
T00702 PU.1	-1634	R	10	CAGGAACTGA
T01806 p53	-1519	N	10	AGACTTGCCT
T00395 Hb	-1468	R	10	TTTTTTTTTG
T01806 p53	-1124	R, N	10	AGACATGTCT
T00395 Hb	-1040	N	10	GAATAAAAAA
T00196 Dl	-865	N	11	TGGGAAAAACA
T01427 p300	-709	R	14	GGCAAACTCCCTGA
T02637 RAV1	-647	R	12	GCCTCAGGTGTG
T01481 Pbx-1a	-584	R	9	TTGATTGGT
T00196 Dl	-573	R	11	TGGTTTTCTGT
_00000 V$PADS_C	-562	N	9	AGTGGTCTC
T00063 Bcd	-465	N, R	8	GGGATTAT
T00273 Evi-1	-394	N	11	TGATTAGATAG
T00996 SRY	-338	R	12	TCTTTTGTTTTG
T00996 SRY	-333	R	12	TGTTTTGTTTTT
T00794 TBP, TFIID	-326	R	10	TTTTTAAAGA
T00581 C/EBPbeta	-80	N	13	GTTTTCAAAACTT

N: Normal, R: Reverse, Black: highest match, Blue: high match, Red: lowest match

Table 7.2B- Summarized result of ~2 kb upstream genomic sequence of long *Tsga10* (AK030254, NM_207228) transcripts (sites 38190413-38188413 of mouse genome) for transcription factor binding sites (using TESS at http://searchlauncher.bcm.tmc.edu/seq-search/gene-search.html). [GKLF: gut-enriched Krueppel-like factor]

Factor	Beg	Sns	Len	Sequence
T01156 AP-1	-1866	R	11	TGTGAGTCACC
T00321 GCN4	-1864	R	10	TGAGTCACCT
T02450 GKLF	-1753	R	14	CCCTTTTCCTTTTT
T02450 GKLF	-1752	R	14	CCTTTTCCTTTTTC
T02450 GKLF	-1739	R	14	CCTTTCCTTTCCTT
T02450 GKLF	-1734	R	14	CCTTTCCTTTCCTT
T02450 GKLF	-1729	R	14	CCTTTCCTTTCCTT
T02450 GKLF	-1724	R	14	CCTTTCCTTTCCTT
T02450 GKLF	-1719	R	14	CCTTTCCTTTCCTT
T02450 GKLF	-1714	R	14	CCTTTCCTTCCCTT
T02450 GKLF	-1705	R	14	CCCTTCCCTTTCCT
T02450 GKLF	-1700	R	14	CCCTTTCCTTTTTC
T02450 GKLF	-1671	R	14	CCTCCCTTTCCTTT
T02450 GKLF	-1668	R	14	CCCTTTCCTTTTTC
T02450 GKLF	-1667	R	14	CCTTTCCTTTTTCC
T00759 Sp1	-1467	R	12	AGACCCCTCCCC
T00505 MEF-2	-1452	N	11	TAAAAATGACA
T01614 Skn-1	-1428	R	12	CTAATGATAATG
T00267 GATA-1	-1391	R	10	TTTTATCAGT
T01157 Oct-1	-1342	R	14	GAAATGCAAATTAT
T00644 Oct-1A, 1B, 1C, 2.1, 2.3, 2.4	-1339	N, R	10	ATGCAAATTA
T01614 Skn-1	-1299	N	12	CATCATCATGAA
T00120 CF2-II	-1016	R	9	TACATATAC
T00107 C/EBPalpha	-949	N	9	ATTTGGAAA
T00772 STE12	-826	N	7	ATGAAAC
T00529 MZF-1	-815	N	13	AGGTGAGGGGGAA
T00755 Sp1	-732	R	12	AAGAGGGGGGAT
T02637 RAV1	-701	N	12	AAACAACAGAAA
T00456 Kr	-620	N	10	ACAGGGTTAA
T00725 REB1	-193	N	8	TCACCCGG
T01467 deltaEF1	-149	N	11	CCCCACCTGAA
T00035 AP-2	-59	N	11	CGCCTGGGGCT
	Tsga10 (NM_207228)			
T00561 NF-muE1	41	R	12	TCGGCCATCTCG
T00755 Sp1	140	N	12	GAGGCGGGCCGG
T00755 Sp1	190	R	12	GAGAGGGCGTGT

N: Normal, R: Reverse, Black: highest match, Blue: high match, Red: lowest match

involved in gene responses to PKA- and PKC-mediated signals (Imagawa et al. 1987).

Sp1 has 3 zinc finger motifs and interacts with BPV-1, bovine papilloma virus type 1. It has some role in interaction of nuclear factors with the upstream region of the alpha-subunit gene of muscle acetylcholine receptor and in muscle differentiation and denervation (Piette et al., 1989). Interestingly, Sp1 activates transcription without enhancing DNA-binding activity of the TATA box factor (Schmidt et. al, 1989).

C/EBPbeta (also known as IL-6DBP; NF-M; NF-IL6; LAP1; LAP; H-APF-2; CRP2; AGP/EBP; ANF-2) has very relaxed DNA-binding specificity, may even bind to AP-1 sites and may heterodimerize with either c-Fos or c-Jun (Hsu et al., 1994). Its leucine zipper motif may also interact with the Rel-domain of NF-kappaB1 (p50) thus causing synergistic effects of both factors (LeClair et al., 1992; Matsusaka et al., 1993). It is a transcriptional activator *in vitro* and *in vivo* (Akira et al., 1990) which is involved in acute phase reactions, inflammation and haematopoiesis and induced by IL-1, IL-6 and LPS (bacterial lypopolysacharides). It is a master regulator of the acute-phase response and regulates constitutive as well as IL-1-induced transcription of IL-6 at the IL-6 promoter. It acts antagonistically to the p53 tumor suppressor (Margulies and Seghal, 1993) and its phosphorylation by p21(ras)-dependent MAP kinases enhances transcriptional activation (Nakajima et al., 1993).

MZF-1: it has a glycine-rich structural feature with myeloid cell specificity and a negative regulatory effect on CD34 and c-myb (Perrotti et al., 1995). It is possibly involved in hematopoietic development (Hromas et al., 1991).

GKLF: gut-enriched Kruppel-like factor (GKLF) amino acid sequence contains three tandem zinc fingers that are related to members of the Kruppel family of transcription

factors. GKLF is localized to the cell nucleus and is found in high levels in growth-arrested cells and is nearly undetectable in cells that are in the exponential phase of proliferation. In the mouse, *GKLF* expression is most abundant in the colon, followed by the **testis**, lung, and small intestine. *GKLF* is enriched in epithelial cells located in the middle to upper crypt region of the colonic mucosa. Taken together, GKLF is potentially a negative regulator of cell growth in tissues such as the gut mucosa and testis, where cell proliferation is intimately coupled to growth arrest and differentiation.

The range of potential transcription factor binding sites in Tsga10 indicates sites that would be consistent with functions of the gene in spermatogenesis or cell cycle. For example, the SRY/SOX binding sites may be important in male-specific expression and C/EBPbeta in expression in the testis during spermatogenesis, and also in mature sperm. Sites for GATA, p53, SOX, NFκB and GKLF may be significant in terms of the possible role of Tsga10 in cell division for example in neural crest (SOX) and embryogenesis (GKLF).

The experiment to test which transcription factor-binding sites may actually be used would be to locate the promoter of the gene and assess the effect of various transcription factors on it. For this purpose, a construct of the promoter with a luciferase reporter gene would be transfected into a cell line to achieve a base line expression level. Then co-expression with target transcription factors that may regulate the gene and its comparison with the base line can indicate which transcription factor up-regulate the promoters. Mutagenesis assays of the promoter by disruption of transcription factor recognition sites can also give good information.

7.3 Considering the function of Tsga10 protein by studying motifs in the protein sequence, and homologies to other proteins

The predicted protein sequence of human TSGA10 protein (Accession number AAH28366) consists of 698 amino acids compared to 689 amino acids in the mouse Tsga10 protein (Accession number AAN01136). Protein sequences in databases were searched using the PBLAST program at NCBI. As I described previously, there are two long cDNA sequences in the database for the gene, *Tsga10* (Accession no. AF530050, AK030254) and a predicted sequence similar to mouse *cp431* (Accession no. XM_136734). Comparison of these two long 5'UTRs (consisting of more than 620 nucleotides) suggests 4 nucleotides less in the 5'UTR of *Tsga10* (AF530050), which results in an ORF of 202-amino acids more in its predicted protein. A search using Tsga10 showed 92% identity to TSGA10 and 95% homology to the predicted protein translated from the *Rattus norvegicus* cDNA named *cp431*. Interestingly, the PBLAST search also showed 95% homology of Tsga10 to an unnamed protein product derived from *Macaca fascicularis* testis (Accession no.: BAB84031) but with 237 amino acids less due to possible alternative splicing in the protein (aa #267-688). It also gave a 51% identity over the whole length of a protein derived from *Xenopus laevis* embryo in stage 17/19 (Accession no.: AAH48022) and a 40% similarity to a human protein of unknown function (KIAA0635) predicted from a brain cDNA sequence (Accession no. AAH62951). In searching the database (using PBLAST) with Tsga10 protein sequence, putative conserved domains named "Smc (Structural Maintenance Chromosome, SMC)" and "MAD (mitotic arrest deficient)" have been detected (figure 7.3A, 7.3B). Tsga10 and both the Smc and MAD domains can be considered as homologues of the

"Myosin class II heavy chain tail" domain (Cytoskeleton). The myosin molecule is a multi-subunit complex made up of two heavy chains and four light chains. It is a fundamental contractile protein found in all eukaryote cell types. The members of the myosin protein family consist of the coiled-coil myosin heavy chain tail region. Many coiled-coil proteins showed about 20% homology to Tsga10.

Different parts of homologous proteins to SMC have conserved similarity in leucin and glutamate (E) sites. All SMC proteins have a bipartite ATPase domain (de Jager et al., 2004). Interestingly Tsga10 has a '*bipartite nuclear localization signal*' motif (table 7.3A) in amino acids 113-127 (RRXXXXXXXXXXRXK) (Munoz-Fontela et al., 2003), which is conserved highly in proteins involved in chromosome segregation via sister chromatid cohesion mechanism, Jones and Sgouros, 2001).

At least a "Leucine zipper" like domain seems to be conserved in Tsga10 protein as well as in other proteins. Since the helix-loop-helix (HLH) domain of myosin tail, is the most important part of myosin to be dimerized and for protein interaction, Tsga10 protein as a structural protein may interact via it with other proteins. Also, Tsga10 leucine zipper motif (Tsga10 spiral motif in figure 7.3E) may also interact with the Odf2 protein by dimerization (making a dimerized HLH) via its leucine zipper motif, then resulting in a synergistic effect on spindle formation and centrosome stimulation. Interestingly, Ensembl (by Markov CLustering) shows a cluster gene (for *TSGA10*) coding for a centrosomal protein named *cep135*. Since leucine zippers act as dimerization motifs in the bZIP family of transcription factors, as does the myosin tail protein it will be interesting to investigate the possibility of Tsga10 self-interaction *in vivo*.

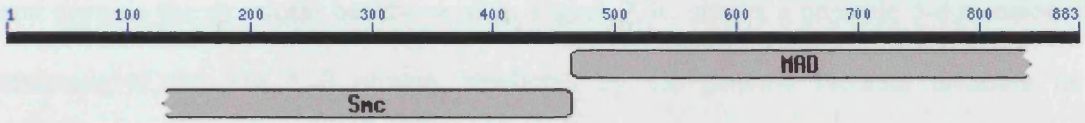


Figure 7.3A- Putative conserved domains comparison in mouse Tsga10 (cp431) by PBLAST (http://www.ncbi.nlm.nih.gov/Structure/cdd)

```
Tsga10:  24    ELLKSTARDREELKCMLEKYERHLAEIQGNVKVLTSERDKTFLLYEQAQEEIARLRREMM  83
SMC    : 253   EELEELQEELEEAEKEIEELKSELEELREELEELQEELLELKEEIEELEGEISLLRERLE  312

Tsga10:  84    KSCKSPKSTTAHAILRRVETERDVAFTDLRRMTTERDSLRERLKIAQETAFNEKAHLEQR  143
SMC    : 313   ELENELEELEER--LEELKEKIEALKEELEERETLLEELEQLLAELEEAKEELEEKLSAL  370

Tsga10:  144   IEELECTVHNLDDERMEQMANMTLMKETITTVEKEMKSLARKAMDTESELGRQKAENNSL  203
SMC    : 371   LEELEELFEALREELAELEAELAEIRNELEELKREIESLEERLERLSERLEDLKEELKEL  430

Tsga10:  204   RLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQREEISILGATL  263
SMC    : 431   -------EAELEELQTELEELNEELEELEEQLEELRDRLKELERELAELQEELQRLEKEL  483

Tsga10:  501   KVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIEMMENELDSARSEIELLRSQMTNE  560
SMC    : 718   LEELKRELA-------ALEEELEQLQSRLEELEEELEELEEELEELQERLEELEEELESL  770


Tsga10:  176   EKEMKSLARKAMDTESELGRQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQL-TQEK  234
MAD    : 132   LTDLKQLEESEAKAESEAAEAKERSKQLEQKLDKLSLRTDAQSKKLQKEKEDQQADAKES  191

Tsga10:  235   IMCLDEKIDNFTRQNIAQREEISILGATLNDLAKEKECLQACLDKKSENIASLGESLAMK  294
MAD    : 192   ISKLKNQLSEMQLRAMNSETELKLLEKELEDLKEQLEELQKELAEAEQKLQSLKASQAER  251

Tsga10:  295   EKTISGMKNIIAEMEQASRQSTEALIMCEQ-----DISRMRRQLDETNDELGQIARERDI  349
MAD    : 252   ADNEQLIKHLEEELKLYEQDAEVVKSMKEQLQRLPELERELEQLREENKKLKSSHENNEL  311

Tsga10:  350   LAHENDNLQ----------EQFAKVKQENQALSKKLN----------------DTHNEL  382
MAD    : 312   LKEELEDLQRKLERAEKMRSKLADLELENEKLEAELKSWENLLQNIGLNLRTPEDISRRI  371

Tsga10:  383   SDIKQKVQDTNLEVNKLKNILKSEESENRQIMEQLRKANEDAENWENKARQTEAENNTLK  442
MAD    : 372   VELQKEELILTEKNGSLTSDAKNLKTANQQLQLERQQALAEILELKKKLETLKRLNRRLQ  431

Tsga10:  443   LELITAEAEGNRLKEKVDALNREVEQHLNAERSYKSQIATLHKSLVKMEEELQKVQFEKV  502
MAD    : 432   RQLSLLTKERDLLRAILDSYDSE-NTETEASNQLTRRLEEAEDMVQLVDSYKAKMEAQLK  490

Tsga10:  503   SALADLSSTRELCIKLDSSKELLNRQLVAKDQE--------IEMMENELDSARSEIELLR  554
MAD    : 491   ELEDELGGQKQRAETLEKELKLLKEQLSSNEREALNALRLKNESLERERDRLRSEKALLE  550

Tsga10:  555   SQM  557
MAD    : 551   MKL  553
```

Figure 7.3B- **Top**: alignment of Tsga10 sequence to its homologue domain Smc (consensus) sequence, possible leucine zipper domain (in red) and hypothetical motif (-X-Ser/Thr-X-Val/Leu-COOH which) may be involved in Tsga10 cleavage. **Bottom:** alignment of Tsga10 sequence to its homologue domain MAD (consensus) sequence.

The coiled-coil domain and some other known motifs among the Tsga10 protein sequence are listed in table 7.3A. The coiled-coil is composed of the tail from two molecules of myosin. These can then assemble into the macromolecular thick filament

and provide the structural backbone of it. Figure 7.3C shows a possible 3-dimensional structure of the TSGA10 protein, predicted by the genome browser database (at http://genome.ucsc.edu/), based on the significant homology of TSGA10 to myosin II heavy chain (fused to alpha-actinin 3). Tsga10 self-interaction (such as myosin tail dimerisation), can also be tested via making Tsga10 bait and prey constructs in a yeast two hybrid assay.

Tsga10 protein shows (see figure 7.3D) a region with low compositional complexity (RKMMRNRSKSPRRPSPTSRAAN) and another region of scattered coiled-coils along all the rest of the amino acid sequence. Coiled-coil sequences are often involved in protein-protein interactions (Beck and Brodsky 1998; Lupas 1997).

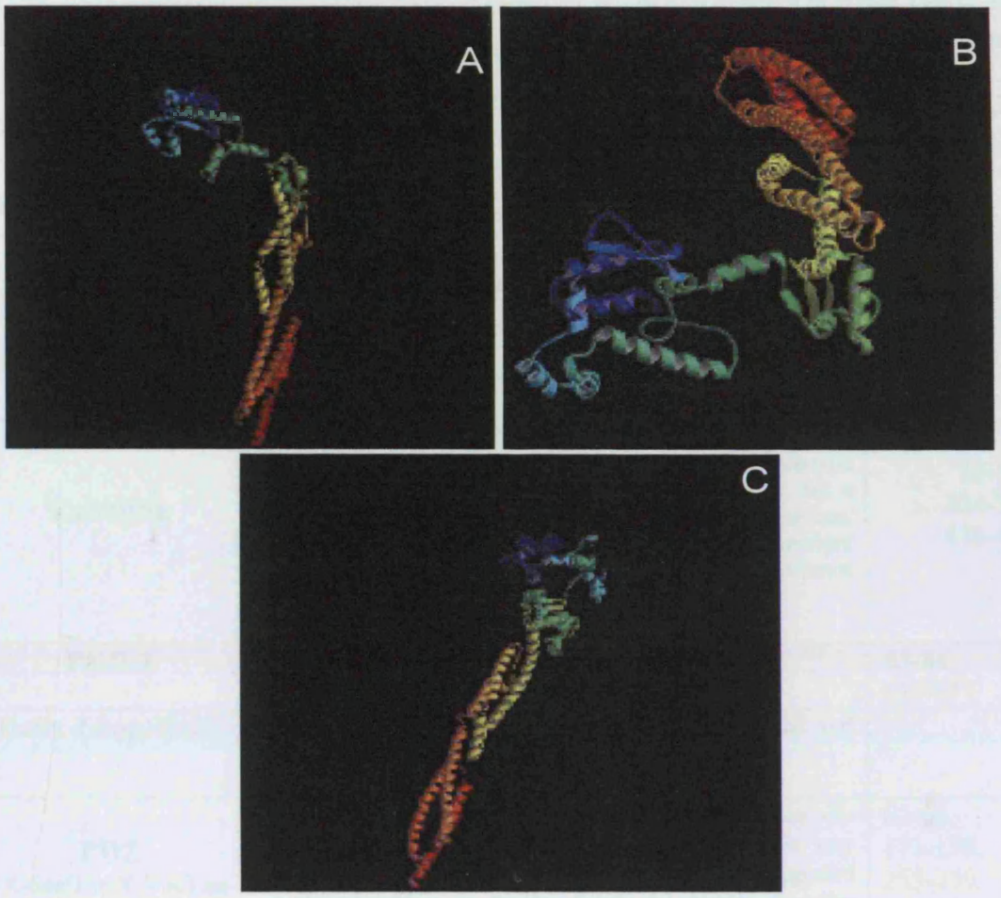


Figure 7.3C- Predicted Comparative 3D Structure of TSGA10 by ModBase (http://genome.ucsc.edu /cgi-bin/hgGene?org=Human&hgg_gene=BC028366&hgg_chrom=none&hgsid=37747339&db=hg17).
A) Front view **B)** Top view **C)** Side view

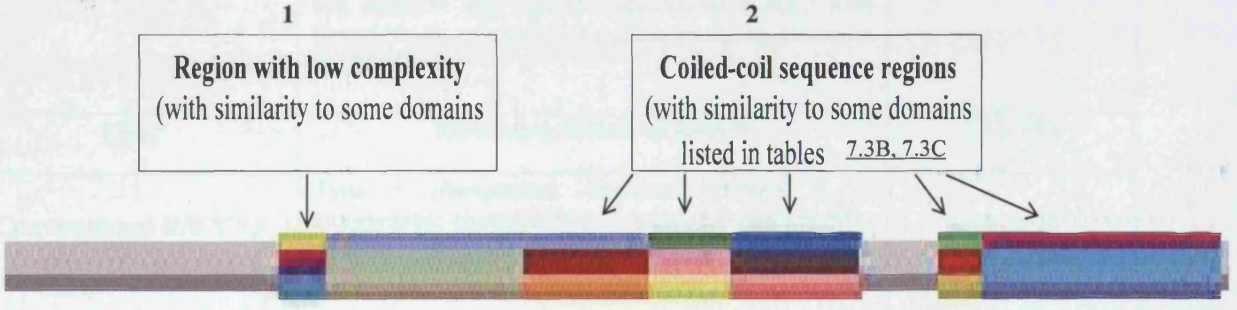


Figure 7.3D- shows Domain within Tsga10 (cp431) amino acid sequence. Different regions (in different colors) show different domains.

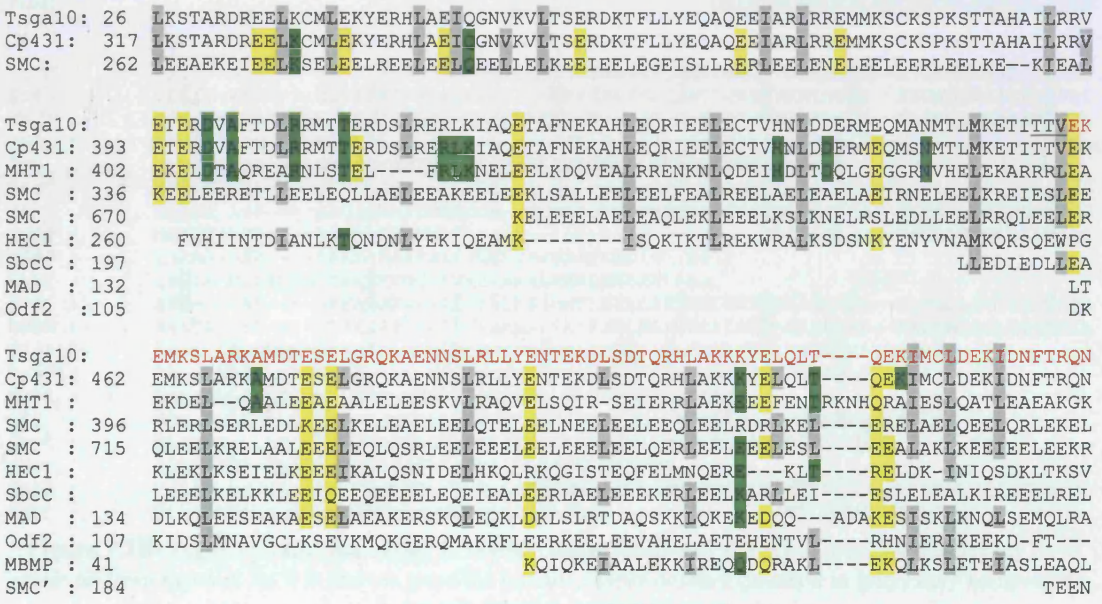
Table 7.3A- Occurrence and position of known conserved motifs and domains in mouse Tsga10 protein

Motif / domain	Description	# Amino acid
Coiled-coil	Coiled-coil myosin heavy chain tail is composed of the tail from two molecules of myosin which can then assemble into the macromolecular thick filament. Coiled-coil domain often involved in protein-protein interactions	26-88, 112-164, 172-199, 228-248, 263-283, 319-500, 522-570, 613-633
Keratin-I	Found in the manchette, the insertion site of the microtubular mantle. The manchette is a highly dynamic structure providing microtubular tracks to structural proteins participating in the sperm tail development (Kierszenbaum et al., 2002)	322-345 476-502
Rabaptin	A coiled-coil protein that interacts with the GTP form of the small GTPase Rab5, a potent regulator of endocytic transport (Stenmark et al., 1995). Mainly cytosolic, but a fraction co-localises with Rab5 to early endosomes. Demonstrating functional similarities with other members of the Ras superfamily. Rabaptin-5 is a Rab effector required for membrane docking and fusion	33-51 324-344 476-498
PKD-2	Play a role in ion channel and in ciliary beating	65-84, 384-401, 535-553
Helix-Loop-Helix	Most important part of myosin to be dimerized and for protein interaction	
PDZ (-X-Ser/Thr-X-Val/Leu-COOH)	Bind to a C-terminal motif with the sequence -X-Ser/Thr-X-Val/Leu-COOH, where X represents any residue. The presence of a PDZ domain suggested that rhophilin (a highly expressed sperm flagellar protein in Rho kinase pathway) works as an adaptor molecule	62-66, 108-112, 171-175, 189-193, 255-259, 390-394, 484-488, 501-505, 619-623
LXXLL	Mediate in interactions of CBP [cAMP-response element-binding protein (CREB)-binding protein = a general co-integrator of various signalling pathways and interacts with lots of transcription factors] with ligand-binding domains of nuclear receptors (Klein et al., 2004)	567-571
LSK	RII binding domain of AKAPs	371-373
Conventional R/KVXF	Type I phosphatase interacts through it, a recognizable characteristic of many phosphatase PP1-targeting subunit in AKAP350 (a centrosomal anchoring protein)	496-499
ESEIQLLK	MAPKKK highly conserved motif in several embryo and non embryo ESTs	589-596
Bipartite nuclear localization signal (RRXXXXXXXXXXRXK)	Necessary and sufficient for nuclear import of the proteins; All SMC proteins are predicted to have a bipartite ATPase domain (de Jager et al., 2004)	113-127

From table 7.3A it can be seen that there are a large number of coiled coil and PDZ domains in the Tsga10 sequence. These are seen significant and consistent in terms of the function of the Tsga10 protein. Many different genes and molecules are related to myosin tail (e.g. giantin in Golgi) and probably all of them conserved a helix-loop-helix domain. Tsga10 protein, as well as other proteins, might have evolved from myosin heavy chain tail (figure 7.3E). In Tsga10, it is likely that these coiled coil domains allow Tsga10 to interact with other proteins, and in self-dimerisation. The PDZ domains (highly expressed in a sperm flagellar protein) act as an adaptor molecule. In Tsga10 they may mediate in Rho kinase pathway.

Other domains are also found:

The ***PKD2*** gene is highly expressed in ovary, foetal and adult kidney, testis, and small intestine. Polycystin-1 requires this protein for its proper function via its C-terminus, and all mutations between exons 1 and 11 result in nonfunctional polycystin-2 because



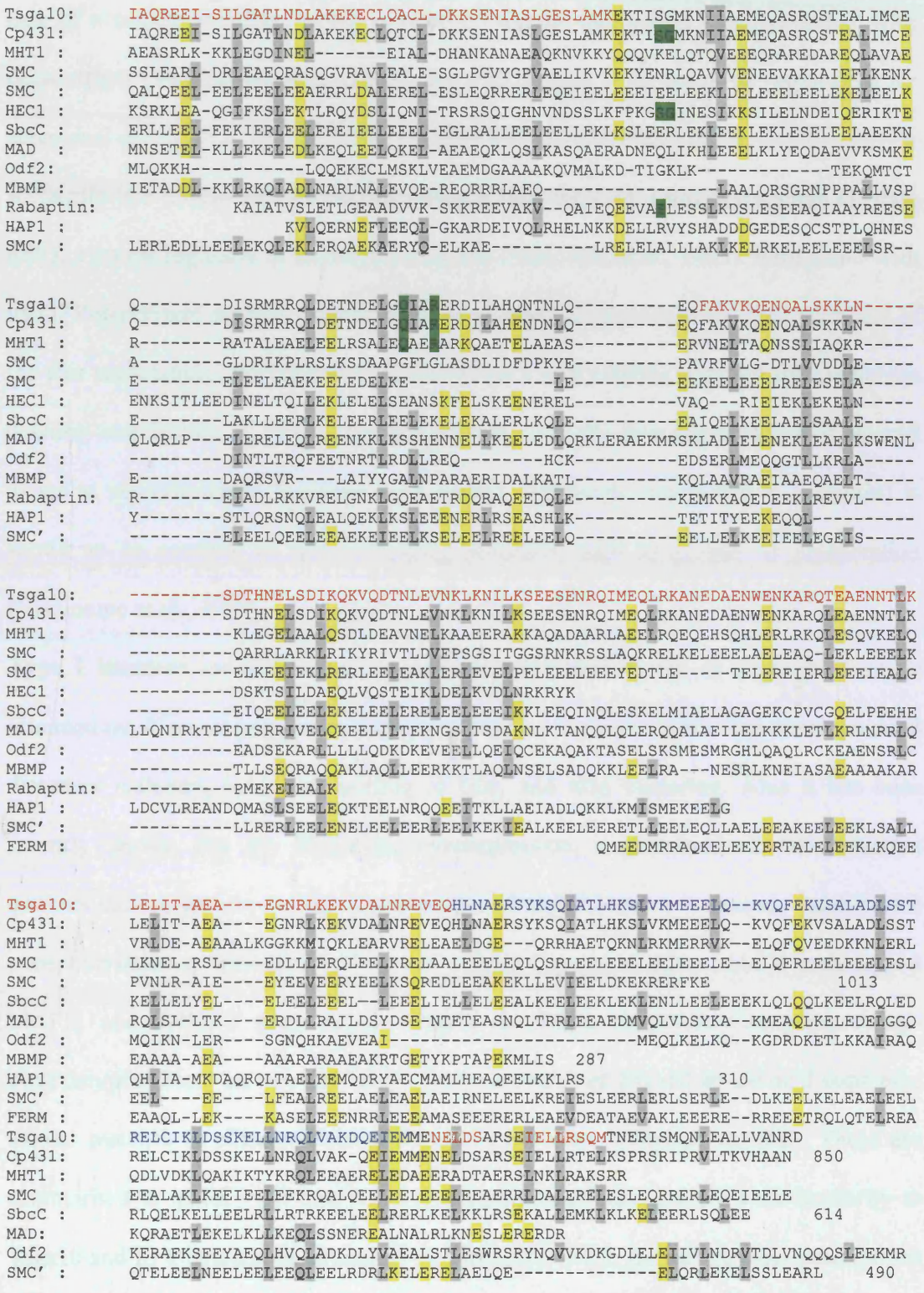


Figure 7.3E- Tsga10 protein homology to myosin heavy chain tail and other proteins which may have evolved from myosin. As it is shown, possible leucine zipper domain (marked in gray) may be involved in protein function and interaction

lack of a calcium-binding, which is required for the interaction of polycystin-2 with polycystin-1 (Veldhuisen et *al.,* 1997). The *PKD2* gene is shown to be responsible for autosomal dominant polycystic kidney disease (Mochizuki et al., 1996).

Rabaptin is a coiled-coil protein that interacts with the GTP form of the small GTPase Rab5, a potent regulator of endocytic transport (Stenmark et al., 1995). Rabaptin-5 with its GTP-dependent manner, demonstrates functional similarities with other members of the Ras superfamily. Rabaptin-5 is shown to be a Rab effector required for membrane docking and fusion. It seems that this protein may regulate regeneration of injured axons at neuromuscular junctions as a neurite outgrowth. Neurocrescin (rabaptin) is shown to be secreted in active dividing processes such as growth or proliferation (Nishimune et al., 1997).

Type I keratins are differentiation markers which are known as a group of acidic intermediate filament proteins and associated with a number of inherited developmental disorders including baldness, beading of hair, and skin blistering. Also it has been recently shown that by PKCalpha overexpression, expressions of differentiation markers such as keratin 1 (MK1) and keratin 10 (MK10) were increased, and ERK1/2 phosphorylation is concurrently induced without change of other MAPK such as p38 MAPK and JNK1/2 (Seo et al., 2004). In Pfam (http://www.sanger.ac.uk/cgi-bin/Pfam/pfamblast_server) and NCBI database search of Tsga10 amino acid sequence, similar matches to other proteins are found using other database searches. These are summarised in tables 7.3B and 7.3C. A description of each protein with similarity to Tsga10 and its alignment is written in the Appendix 8.2.2. Below are brief descriptions of two proteins with high homology to Tsga10 protein, and to myosin heavy chain).

Table 7.3B- Conserved domains of rodent Tsga10 (cp431) protein (892 aa in mouse)

Similar Family Proteins	Start	End	Homology %
Chromosome segregation ATPase	-160	178	29.1
Chromosome segregation ATPase	25	612	49.3
Myosin tail, Myosin class II heavy chain	23	554	60.2
Membrane-bound metallopeptidase	208	477	58.8
HEC1, involved in chromosome segregation	98	498	45.2
Mitotic checkpoint protein	175	556	57.1
ERM (Ezrin/radixin/moesin)	278	517	27.9
SbcC (ATPase involved in DNA repair)	25	534	46
Hook protein	105	578	44.5
Intermediate filament protein	301	576	49.3

Homologous proteins to Tsga10 (Tsga10 starts at aa # 203) by NCBI GenBank (http://www.ncbi.nlm.nih.gov/entrez/viewer) and Ensembl (www.ensembl.org) databases. Each homologous protein was aligned with Tsga10 by usin 'Genebee' (http://www.genebee.msu.su/).

Table 7.3C- Potential matches for Tsga10 892 amino acids residues using Pfam.

Domain	Start	End	Evalue	Pfam ID	Mode
Myosin_tail_1	-177	563	0.00054	PF01576	ls
EB1	-66	-27	0.27	PF03271	ls
E-MAP-115	13	186	0.37	PF05672	ls
MAD	106	688	0.078	PF05557	ls
ERM	223	502	0.47	PF00769	ls
Prefoldin	238	346	0.39	PF02996	ls
Rabaptin	259	386	0.27	PF03528	ls
Filament	345	628	0.035	PF00038	ls
Tropomyosin	355	385	0.032	PF00261	fs

Known domains with E-values below the cut off (Evalue < 0.5) are considered.
Is = complete domain, fs = fragment domain (http://www.sanger.ac.uk/cgi-bin/Pfam/)

Mitotic checkpoint protein family- consists of several eukaryotic MAD (**m**itotic **a**rrest **d**eficient) proteins. The mitotic spindle checkpoint monitors proper attachment of the bipolar spindle to the kinetochores of aligned sister chromatids and causes cell cycle arrest in prometaphase when failures occur. Several mitotic spindle checkpoint components including Mad1-dependent complex (Mad2, Mad3, Bub3 and Cdc20) have been identified.

Membrane bound metallopeptidase domain- there is a rat protein with 22.8% (not very high but considerable) similarity to Tsga10, encoded by the gene in rat chromosome 9q21 and human chromosome 2 (synteny to the same place of Tsga10) named "Ece11" (endothelin converting enzyme-like 1) that plays a role (Kato et al., 2002) in damage-induced neuronal metallo-endopeptidase and is involved in neuroprotection. Interestingly it is a membrane-bound metallopeptidase with molecular function of hydrolase activity. It may mediate neuropeptide processing and enhances the expression and the activity of antioxidant enzymes such as Cu/Zn-superoxide dismutase (SOD), Mn-SOD, and glutathione peroxidase under oxidative stress. Its expression is in neurons of the central nervous system (CNS) and peripheral nervous system. Its transcription is increased by nerve growth factor (NGF) deprivation and by leukemia inhibitory factor (LIF) but not by other growth factors and cytokines; mRNA is increased in the dorsal root ganglion (DRG) neurons by sciatic nerve injury. Intracellular signalling cascade, neuropeptide signalling pathway, proteolysis and peptidolysis are some of its biological functions. More interestingly it is a cellular component integrating into membrane and plasma membrane and plays a role in pathogenesis of respiratory insufficiency (Schweizer et al., 1999).

Bioinformatic analysis of the 84-kD protein also revealed the presence of a region with significant similarity to the **ERM** (**E**zrin/**r**adixin/**m**oesin) domain (aa # 223-502). The similarity of domains to ERM (as a key membrane-cytoskeleton linker) may indicate a role for Tsga10 in the formation of the fibrous sheath (FS) during spermiogenesis. A CD44-ERM attachment complex is involved in migration of embryonic fibroblasts, activated macrophages, and metastatic cells. It has been

demonstrated that the function of the NF2/ERM/4.1 superfamily is to maintain cell integrity, motility and differentiation. Some of these proteins are involved in human diseases (Surace et al., 2004), and suppression of the expression of all ERM proteins with antisense oligonucleotides in cultured fibroblasts/epithelial cells blocked membrane ruffling and motility (Bornhauser et al., 2003). It has been speculated that ERM proteins regulate cell motility by linking cortical F-actin to the plasma membrane in different cell types (Olsson et al., 2000) and may have an identical function in tumor-cell invasion (Kosako et al., 2000): EhM2, which contains the ERM domain, is expressed in high-metastatic but not in low-metastatic K-1735 murine melanoma cells and regulate the metastatic potential of melanoma cells by affecting cytoskeletal–membrane interactions (Shimizu et al., 2000). It is also expressed in the liver, lung, kidney, and testis and in 7 to 17 day mouse embryos. Tsga10, which may have functional myosin/ERM sequences, is also expressed in foetal tissues and several cancer cells including cancers originating from bone marrow, thyroid, germ cell, stomach and uterus. It has been shown that ezrin protein suppression results in reduction of activation of Rho kinase, CD44 and MAPK phosphorylation.

Since *Tsga10* has similarity to a membrane-bound metallopeptidase and an ezrin/radixin/moesin (ERM) protein domain, it may have some contribution in cell trafficking via the **'nexin'** protein family. These proteins have an endocytic pathway function to internalise extracellular components and participate in diverse signal-transduction pathways. SNX9 one member of the nexin family binds preferentially to the precursor, but not the processed forms, of the metalloproteases through the interaction of its SH3 domain with Pro-rich sequences that are present in the

cytoplasmic tails of the metalloproteases. SNX17 has an ezrin/radixin/moesin (ERM) protein domain 4.1: four-point-one, ezrin/radixin/moesin, FERM and is also associated with the cytoskeleton (Worby et al., 2002). It is indicated that FERM domains of ERM proteins evolved in response to multi-cellularity, rather than as cytoskeletal proteins while ERM proteins are involved in microfilament-membrane attachment and Rho-signal-transduction pathways. The FERM domain also binds to Rho pathway signalling molecules while an F-actin binding site is present in the carboxy terminus of ERM proteins (Bretscher et al., 2002). Since the connection between the cell membrane and actin cytoskeleton is required for metastatic cells to engage their microenvironment, and for axonal transport, TSGA10 may be an indicator of metastasis (marker) in addition to being a tumour marker (recently proposed as a tumour marker by Tanaka et *al.*, 2004). The location of the ERM domain in the Tsga10 protein is downstream of the proposed Tsga10-N-terminus 27-KD cleavage site. ERM members are phosphorylated at a C-terminal threonine through kinases such as Rho kinase/PKC, and at N-terminal tyrosines through receptor tyrosine kinases such as growth factors (probably GDNF which is shown to be expressed during spermatogenesis), cytokines and c-MET (Krieg et al., 1992; Crepaldi et al., 1997). It is shown that IL-1β activates ERM family members (e.g. Tsga10) by inducing phosphorylation, allowing them to form a bridge between F-actin and integral membrane proteins at the plamalemma (Mangeat et al., 1999). Then activated ERM family members associate with integral membrane proteins directly (e.g. the sodium-hydrogen exchanger isoform 1, CD44, membrane-bound Tsga10) and indirectly via the scaffolding proteins (e.g. ODF2) (John et al., 2004).

ERM proteins are identified as targets of a single regulatory mechanism, a signalling pathway mediated by GTPases of the Rho family (Amano et al., 2000).

PBLAST at NCBI GenBank (table 7.3B) showed homology between Tsga10 and chromosome segregation ATPase, myosin heavy chain, membrane-bound metallopeptidase, hook protein and polyamine modulated factor 1 binding protein 1 which is expressed in sperm tail (Ohuchi et al., 2001). In the Gene Ontology database (www.ensembl.org/) several functions for the TSGA10/Tsga10 protein are given:

GO:0005741: Mitochondrial outer membrane **GO:0007283: Spermatogenesis**

GO:0015288: Porin activity GO:0016021: Integral to membrane

TSGA10 protein in other species- Other proteins with homology to TSGA10 are found in species in addition to mouse and rat, these are summarized in table 7.3D. Some proteins with high similarity to Tsga10 have known functions. These include "Rab6-interacting protein 2" that regulates intracellular transport at the level of the Golgi complex, "Restin" (Reed-Steinberg cell-expressed intermediate filament) and "CLIPs" (cytoplasmic linker proteins). These proteins are involved in the local regulation of microtubule dynamics in response to positional cues.

Rab6 protein is a member of '*Ras-like GTPase*' family and '*Rab6-interacting protein 2*' (Rab6IP2) is involved in endosome-to-TG [trans-Golgi] network and participates in assembly of a multiprotein complex including PI3-kinase (Monier et al., 2002). GTP-Ras proteins bind to and activate a number of effector molecules including protein kinase B (PKB), protein kinase C (PKC) and lipid kinase PI3-K (Vojtek and Der, 1998). Rab6IP2 is expressed mostly in brain cytosol and is recruited on to Golgi membranes in

Table 7.3D– UniGene study shows *Tsga10* sequence homology to some known functional genes and proteins http://www.ncbi.nlm.nih.gov/UniGene/clust.cgi?ORG=Mm&CID=332756

Mus musculus	ref:NP_444434.1 - Rab6-interacting protein 2	20.58 % / 710 aa
R.norvegicus:	ref:NP_113933.1 - restin	20.41 % / 702 aa
A.thaliana:	ref:NP_176681.1 - hypothetical protein	22.37 % / 707 aa
C.elegans:	ref:NP_505094.1 - myosin	20.44 % / 709 aa
D.melanogaster:	pir:T13030 - T13030 microtubule binding protein D-CLIP-190	19.78 % / 730 aa
S. cerevisiae:	sp:P25386 - USO1_YEAST Intracellular protein transport protein USO1	21.07 % / 714 aa

a Rab6:GTP-dependent manner. Apart from Rab6, a melanosomal membrane protein - Rab27- is another Rab protein, which is expressed particularly in testis and recruits myosin Va to the melanosome surface or links to myosin through a synaptotagmin-like protein (Slac2). Also Tsga10ip neighbouring gene encodes synaptotagmin protein. Slac2 also displays an actin-binding site and is highly expressed in brain, lung, and testis itself (Fukuda and Kuroda, 2002). Null expression of myosin Va, Rab27 and Slac2 results in a failure of melonosome transfer to keratinocytes and partial albinism of hair and skin (Seabra et al., 2002; Wu et al., 2002). Proteins in non-mammals, which are conserved with 20 percent homology, may be more important as evolutionarily related proteins. This includes a "Myosin heavy chain family member" which has ATP binding and nuclear localization domains, motor activities and localises in membrane. Another protein is "Microtubule binding protein D-CLIP-190" in *Drosophila melanogaster.* Myosin is associated with it in neurons and at the posterior pole of Drosophila embryos. "Intracellular transport protein USO1" in yeast is the other similar protein to Tsga10 which plays a role as a cytoskeletal component in the protein transport from the ER to the later secretory compartments (Golgi complex).

7.4 What does localisation of the protein tell us about function?

7.4.1 Tsga10 in the fibrous sheath and its role in male fertility

After completing a range of immunohistochemistry and immunofluorescent (IF) experiments it was clear that TSGA10 is expressed in testis (throughout sperm maturation), and is localized in the sperm tail. This was consistent with the gene expression data (obtained by RT-PCR and Northern Blot analysis) that TSGA10 was found throughout sperm development. *Tsga10* mRNA is detectable from the pachytene stage of meiosis but protein expression is not seen until the elongating spermatid stage. So a delay between *Tsga10* transcription and its encoded protein expression is predicted. Tsga10 whole protein was localised in seminiferous tubules and a 27-KD product was detected in the fibrous sheath component of sperm tails. Therefore post-translational processing of the protein happens during spermiogenesis. The 27 KD form of Tsga10 associated with the fibrous sheath of the sperm tail, is likely to play a role in sperm motility.

It is not clear whether myosin is present within the sperm tail. However, there is some evidence for the possibility of myosin involvement in sperm function (Ashizawa et al., 1995; Kierszenbaum et al., 2003; Weil et al., 1995). A number of unconventional myosins, including myosin-X, contain the Myosin Tail Homology 4 and ERM domains (Berg et al., 2000). Interestingly Tsga10 subcellular localization and its pattern within the cell showed filament formation which is consistent with the above description and it may suggest the presence of a putative functional myosin/ERM domain in elongating spermatids.

Recent models suggest that FS proteins can also serve as scaffolds for protein complexes involved in regulating signal transduction processes (Miki et al., 2002; Eddy et al., 2003). It was suggested that components of the Rho signaling pathway associated with the FS might be involved in regulating flagellar function. These include the Rho-binding protein rhophilin (Nakamura et al., 1999) and ropporin, a spermatogenic cell-specific protein that binds to rhophilin and to the FS (Fujita et al., 2000). Rhophilin is highly expressed in testis and is localized specifically in sperm flagella. The presence of a PDZ domain suggested that rhophilin works as an adaptor molecule (Peck et al., 2002). PDZ domains bind to a C-terminal motif with the sequence -X-Ser/Thr-X-Val/Leu-COOH (figure 7.3A), where X represents any residue. This motif (PDZ) is present nine times among Tsga10 amino acid sequence (aa numbers 62-66, 108-112, 171-175, 189-193, 255-259, 390-394, 484-488, 501-505, and 619-623), while its presence in the region between aa numbers 255 and 259 may be involved in cleavage of Tsga10 (see section 4.2). It is therefore possible that the cleavage site of Tsga10 is between aa 259 and 260. Further investigation will confirm whether Tsga10 is involved in regulating signal transduction and/or can bind rhophilin via measurement of that in the environment containing Tsga10 FS protein. This pathway is known as a main regulator of cell shape, migration, and motility (Schmitz et al., 2000; Riento and Ridley, 2003). Tsga10 has an apparent ERM domain and I have shown it to be expressed in astrocytes. ERM family proteins deactivate GTPase-Rho kinase (ROCK) pathway in the presence of IL-1β. It has been shown that the deactivation of the Rho GTPase-Rho kinase (ROCK) pathway by IL-1β regulates the phenotype of astrocytes (John et al., 2004). Deactivation of this pathway affects cellular morphology and migration via

effects on F-actin and its interactions with focal adhesions, nonmuscle myosin, and microvillar adapter proteins of the ezrin-radixin-moesin (ERM) family. Hypothetically, Tsga10 is most likely phosphorylated in Rho-ROCK axis in parallel with the proposed role of ROCK in ERM phosphorylation (Matsui et al., 1998). If we can consider sperm maturation and migration in testis as an active natural model of metastasis, then there might be a common Rho-GTPase-Tsga10 pathway in cancer metastasis and sperm motility.

Again considering the possible biological activity of TSGA10 (bearing in mind its interaction with Cyt C1, which is one of the major proteins involved in ATP synthesis in sperm mitochondria), Tsga10 could be a protein that plays an important role in the mechanism by which ATP (as sperm motility energy supplement) can be shuttled from the mitochondrial sheath (MS) to the rest of the flagellum.

During spermiogenesis, microtubules are twisted round the axoneme in the sperm tail. This process localises mitochondria to the tail so that large amounts of ATP can be synthesised in order for the tail to move. The fibrous sheath extends from the mitochondria, gathered around the axoneme, to the end of the sperm tail, which is further evidence for a connection between Tsga10 and Cyt C1. The fibrous sheath surrounds the outer dense fibres, which in turn surround the axoneme in the midpiece of the sperm tail. These fibres, of which one type is encoded by *outer dense fibre 2* (*ODF2*), are also involved in sperm motility.

A function of the fibrous sheath (FS) is to protect sperm from oxidative stress that could interfere with sperm motility and DNA damage has been proposed (Fulcher et al, 1995). In the case of any defect in mitochondrial oxidative phosphorylation, sperm still

produce ATP (in lower amounts) and sperm motility is still present (although reduced) (Narisawa et al, 2002). This is achieved via a glycolysis reaction compartmentalized within the FS/principal piece (Turner 2003). Tsga10 protein also interacts with HIF1α (Dr. D.M. Katschinski, Martin-Luther-Universität Halle, personal communication). HIF1α, Cyt C1, and TSGA10 may play a role in glycolysis in the principal piece to supply the fuel for the tail while oxidative stress via its interaction with HIF1α and Cyt C1. HIF1 α is a transcription factor that enables tumors to adapt to hypoxic environments by activating survival genes. Interestingly, the HIF1α protein is among the list of known flagellar proteins. The *HIF1*α gene is frequently overexpressed in colon tumors (Zhong et al, 1999), and regulation often occurs at the level of protein degradation; however, studies also document the regulation of HIF1α mRNA levels (Lukahev et al, 2001, Caniggia et al, 2002).

It is shown that in absence of an A-kinase anchor protein -AKAP4- (the major mouse sperm fibrous sheath protein), a major reduction or loss of other FS proteins occurs resulting in severely reduced sperm motility (Carrera et al, 1994 and Miki et al, 2002). A possible case of this effect is thought to be an alteration in protein kinase A (PKA) distribution resulting in either significant structural changes or complete disruption of other FS proteins (Miki et al., 2002). It is shown that AKAPs tether protein kinase A (PK-A) to the FS. The presence of at least two identified FS components as members of the A-kinase anchoring protein (AKAP) family (AKAP3, AKAP4), implies that "TSGA10 & TSGA10IP" proteins may also be members of the AKAP family. Considering Tsga10 similarity to ERM and consistent with it, there is some evidence that 'Ezrin' protein is an AKAP as well (Dransfield et al., 1997). Interestingly human

and mouse pro-AKAP82 proteins are proteolytically processed in a similar manner to Tsga10. The C-terminal part of pro-AKAP82 remains in the fibrous sheath. Most of known FS proteins (AKAP4 which is isolated as more than half of FS proteins, AKAP3 and TAKAP-80) have anchoring sites for cAMP-dependent protein kinase (Eddy et al., 2003). PK-A activates a downstream tyrosine kinase(s) whose targets are primarily located in the flagellum (Leclerc et al, 1996; Si and Olds-Clarke, 2000). To date, just a few number known target proteins for PK-A phosphorylation have been identified in sperm (Tash and Bracho, 1998) including axonemal dynein (Tash 1989). TSGA10 protein with a high number of phosphorylation sites (nine PK-A among forty six in total) could be one of the protein targets for PK-A phosphorylation. AKAP4 recruits PKA to the fibrous sheath to facilitate neighbouring proteins that regulate flagellar function to be phosphorylated (Miki et al., 2002). In that case, TSGA10 could be considered as another member of the AKAP fibrous sheath proteins. In this case, possible roles for TSGA10 are broadened to include serving as scaffolds for protein complexes involved in regulating signal transduction processes (Edwards and Scott, 2000; Colledge and Scott, 2000; Colledge et al., 2000; Dodge and Scott, 2000). Interestingly in agreement of our hypotheses of TSGA10 as an AKAP as well as its role in Rho signalling pathway and to link both of them together, it was recently suggested that Rho signalling pathway components associated with the fibrous sheath via AKAP to regulate flagellar function (Carr et al., 2001); including Rho-binding protein rhophilin (Nakamura et al., 1999) and ropporin that binds to rhophilin and fibrous sheath (Fujita et al., 2000). Also AKAP-associated sperm protein and ropporin have

similar regions to the AKAP-binding domain of PKA and bind to the PKA-anchoring domain of AKAP3 (Carr et al., 2001).

All of the above findings suggest that the TSGA10 protein has both functional and structural roles in the fibrous sheath.

Finally, a defect in TSGA10 may be one of the causes of a known sperm motility disease, dysplasia of the fibrous sheath (DFS). To date all studies have failed to identify problems with any FS known genes or proteins (e.g. AKAP4) in this disease. Common characteristics of this disease include marked disorganization of the FS as well as distortion of the microtubule doublet and axonemal (Chemes et al., 1987).

It is thought that, complete formation of the fibrous sheath occurs near the end of spermiogenesis in mouse (Sakai et al., 1986) and rat (Irons and Clermont, 1982), and it is after the appearance of the axonemal ribs and concurrent with the AKAP4 incorporation into the FS (Miki et al., 2002). DFS might be caused by a mutation in a gene affecting FS formation rather than in the AKAP4 gene, in DFS there are disorganised, thickened FS; (Turner et al., 2001). Therefore, because of the gene expression pattern and localisation of the protein, mutations in the *TSGA10* could result in DFS aethiology.

Usher syndrome, is an autosomally inherited subtype of 'retinitis pigmentosa' and is characterised by infertility, retinal degeneration and congenital deafness. It is possible that mutations of the *TSGA10* gene might cause a form of Usher syndrome (type IB) since defects in a myosin molecule (VIIa) are known to cause this disease (Gibson et al, 1995; Hasson et al, 1995; Weil et al, 1995).

7.4.2 Other functions of Tsga10 in the sperm tail

Apart from the localisation of the Tsga10 protein in the fibrous sheath, it is possible that Tsga10 may play a role in mitosis. This function has been proposed for Odf2 protein (Nakagawa et. al, 2001) as well. It seems the pathway by intracellular cAMP that activates PKA may play a fundamental role here via AKAP family members in sperm; Odf2 is co-localized with β-tubulin (Nakagawa et al, 2001): tubulin is identified as a binding partner of an AKAP (AKAP84/AKAP121), that binds and target protein kinase A (PKA) to the outer mitochondrial membrane. It concentrates on mitochondria in interphase and on mitotic spindles during M phase (Cardone et al, 2002). S-AKAP84/AKAP121 complexes localize signalling enzymes to mitochondrial and cytoskletal components in sperm (Cardone et al, 2002). So, the FS complex of TSGA10/Odf2/Tsga10ip may play a similar role in sperm mitochondria outer membranes via the mitochondrial Cyt C1 protein: TSGA10 has predicted porin activity (Gene Ontology, Ensembl) in mitochondrial outer membrane and I have shown that it interacts with Cyt C1. A proteolytic activity in the Tsga10 complex might possibly mediate assembly and degradation processes of membrane protein complexes and thereby exert a key function in the maintenance of mitochondrial membrane integrity such as other mitochondrial members of ATP dependent metallopeptidases (AAA) family (Arlt et al., 1996). Similarly a membrane-embedded metallopeptidase, Oma1, as a quality control system component (including a proteolytic ATP-dependent AAA protease) is known to maintain the inner membrane of mitochondria by removal of misfolded membrane proteins (Kaser et al., 2003).

Protein kinase C (PKC) and an inositol dephosphorylation enzyme (Inositol polyphosphate-4-phosphatase type I, that participates in cell growth downstream of transcription factor GATA-1 and catalyzes the hydrolysis of the 4-position phosphate of phosphatidylinositol 3 -PI3-) are encoded by genes which lay within 300-400 kb of *Tsga10*. It has been shown that PKC is upregulated specifically during the spermatid stage of spermatogenesis (Um et al., 1995). It has been recently shown that inhibition of PKC signal transduction results in an activation of matrix metallopeptidase (Woo et al., 2004). So TSGA10 protein (which has similarity to membrane-bound metallopeptidases), may be a candidate for playing a role in the PKC pathway during spermatogenesis. PKC and phosphatidylinositol 3-kinase (PI3-kinase) are associated with development and hyperactivation of sperm motility (via *Ras*). After thrombin activation, the 4-phosphatase is translocated to the actin cytoskeleton along with PI 3-kinase, which requires both integrin engagement and aggregation (Munday et al., 1999). Meanwhile a considerable fraction of PKC is associated with PI3-kinase and Ras in cauda sperm plasma membrane. Also, activation of PI3-kinase by GTP-Ras can activate PKC as a potential signal transduction pathway in spermatozoa (NagDas et al., 2002); then PKC phosphorylate flagellar proteins (such as Tsga10 with its high number of phosphorylation sites) on serine and threonine residues or activate other signalling elements to function in the regulation of sperm motility.

Tsga10 has about fifty potential phosphorylation kinase sites including nineteen for PKC, and nine for PKB, which is also most widely associated with the PI3K signalling pathway (Vanhaesebroeck and Alessi 2000). The Tsga10 protein may play a role in polarising the sperm tail in a pathway via PI3-K (Haugh et al., 2000). Tsga10 has a

similarity to microtubule binding protein D-CLIP-190 (cytoplasmic linker protein 190 and/or its mouse homologue [CLIP-170]). The distribution of *CLIP*-associated proteins (CLASPs) is mediated by PI3-kinase. CLIP-170 and CLIP-115 contain two microtubule-binding domains. Microtubules are essential for chromosome segregation in mitosis and for organelle movement and positioning in interphase cells (Akhmanova et al., 2001). Hereby, Tsga10 similarities to mitotic checkpoint protein and to chromosome segregation domain are interesting. Most of the CLASPs are involved in the regulation of microtubule dynamics, represented by the hippocampus cDNA clones (in consistent with our immunostaining results) and mainly detected in the region of the Golgi complex. All of these findings suggest Tsga10 protein as a candidate member of trans-Golgi CLASPs.

Coordination of sperm motility requires the interaction of the cytoskeletal filament systems in the sperm tail. In addition, these systems are involved in the transport of cellular components in the cell. Interestingly, a molecule that mediates coordination between the actin and microtubule cytoskeletons is a homologue of cytoplasmic linker protein (CLIP)-170 that interacts with a class VI unconventional (95F) myosin. Consistent with our findings, it is shown that this unconventional myosin and D-CLIP-190 are coexpressed in a number of tissues during embryogenesis and axonal process of neurons in *Drosophila* (Lantz and Miller, 1998). Tsga10 is similar to both myosin and a homologue of microtubule-binding protein, and these two proteins are associated with the nervous system and the posterior pole of *Drosophila* embryo, where both microtubule and actin dependent processes are known to be important. TSGA10 may link the actin and microtubule cytoskeletons (dynein in sperm). This hypothesis is

reinforced in sperm by the evidence that CLIP-170 is a member of a protein family that includes DP-150^{Glued}, which is known to be a part of a large complex, dynactin. The dynactin complex stimulates dynein-mediated motility of vesicles *in vitro* (Schroer, 1996). An interaction is reported (by Ensembl) between Cep135 (a coiled-coiled centrosome protein which shows 45% homology to *TSGA10*) and dynein/dynactin complex. While centrosomes ensure fidelity of chromosome segregation during mitosis by guiding microtubule nucleation towards the formation of a bipolar spindle, this Cep135/dynein complex is shown to be necessary in assembly and maintenance of both microtubule and centrosome functionalities (Uetake et al., 2004) where Tsga10 may play a role. Also similar to Tsga10 overexpression in the cell, a pattern of filamentous polymers induced by Cep135 overexpression is seen (Ryu et al., 2000).

Spag4 is another ODF-interacting spermatid-specific protein localized to two microtubule-containing spermatid structures; to the manchette and to the axoneme in elongating spermatids and epididymal sperm (Shao et al., 1999). Like *TSGA10*, *SPAG4* is expressed in testis and pancreas as well as being upregulated in a wide range of neoplastic tissues (Kennedy et al., 2003). The similarity between the localisation of Tsga10 and Spag4, and the interaction of both with 'outer dense fibres' (Odf) proteins proteins suggest that TSGA10 may be another component of the manchette.

7.4.3 Further evidence for role in microtubules

By consideration of dimerized HLH domain of Tsga10 protein, its possible role in microtubule movement regulation in sperm tail while interacting with Odf2 (which is colocalized with β-tubulin), Tsga10 dimer walking along microtubule filaments such as kinesin (Hoenger et al., 2000) can be proposed. That is why similarly, kinesin is shown with two motor domains and the alpha-helical neck forming a coiled-coil (Sack et al., 1997; Kozielski et al., 1997) to walk on the tubulin protofilament towards the plus-end of microtubules (right side of figure 7.4.1).

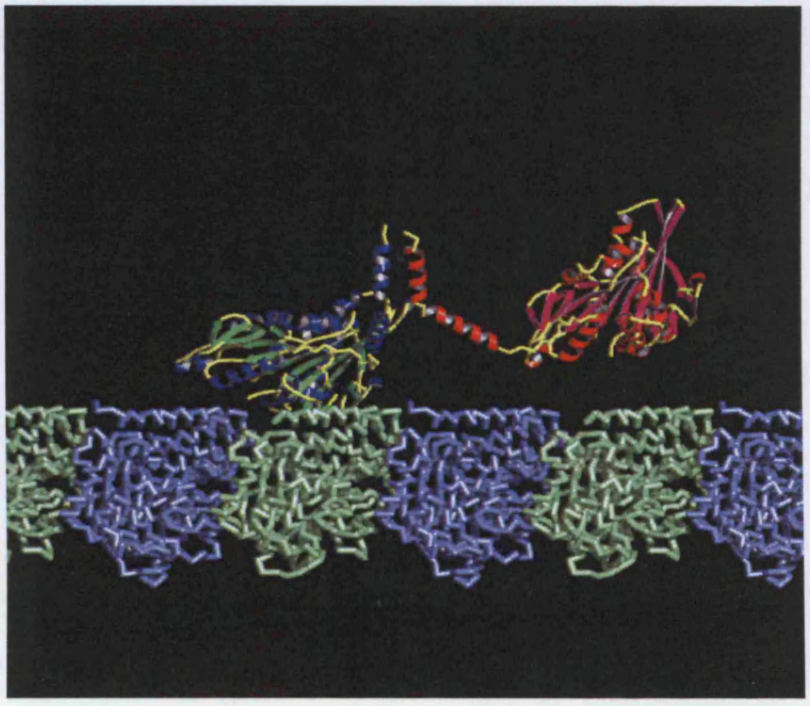


Figure 7.4.1- Kinesin; A candidate model for Tsga10 protein function. The motor domains attach to successive beta-tubulin subunits and the two heads move in a non-equivalent fashion in order to avoid twisting the stalk. alpha-Tubulin is pale blue, beta-Tubulin is pale green. The upper (C-terminal) part of the coiled-coil neck is shown to be permanently connected since it has a high coiled-coil potential, similar to a leucine zipper. The lower part of the neck is shown to open and close reversibly in order to allow the heads to detach, move, and reattach. This is coupled to a reorientation of the linker region (yellow) between the neck helix and the motor domain. The picture is from the web site: http://www.mpasmb-hamburg.mpg.de/ktdock.

Finally a dual microtubule-based and actin-based motor protein has already been suggested to mobilize vesicle cargos by intramanchette transport in developing spermatids (Kierszenbaum et al., 2003). Tsga10 protein may give a good description for the reason of these two transport systems or a link between them.

7.5 Tsga10 protein expression in tissues other than sperm

From the IHC studies I have shown that Tsga10 is expressed in a range of tissues in the embryo. It is possible that the gene is involved in patterning morphogenetic change in the vertebrate embryo. Its role could be in events occurring at the embryonic cell surface that involve cell migration, signaling and probably adhesion. There is some variance in the expression of the two forms of TSGA10 (namely the N Terminus and the C-Terminus), whereby in some cases it appeared that both the N terminus and the C-terminus were being expressed which suggests that the protein is present in its whole form and has not been post-translationally modified. In other instances only one terminus was being expressed, the differential expression of the termini is nicely illustrated in figure 4.3C of mature rat testis where the N-terminus is clearly only being expressed in the sperm midpiece, whereas the C-terminus is being expressed in all germ line cells. In these cases the protein has been post translationally modified into a smaller 27kDa protein. It is likely that the different forms of the protein have specific roles with the roles linked to dividing or differentiating cells.

Tsga10 protein expression in the primordia of vibrissae follicles is reminiscent of a similar pattern of expression of some proteins in testis and the olfactory system. For example, such 'pheromone receptors' are expressed in the accessory olfactory organs as well as spermatids and have an important role in sperm maturation and/or migration (Tatsura et al., 2001). In both cases this may reflect a connection with the presence of cilia.

7.6 The role of the Tsga10 protein in neurogenesis, cell cycle regulation and cancer

Neurogenesis: the mammalian hippocampus is a highly plastic brain structure in which new neurons are generated throughout adulthood (Jessberger and Kempermann, 2003). The mechanical problem of accommodating the dividing neurons has been overcome by the discovery that neural stem cells exist throughout life in the adult brain and can renew and give rise to new neurons, astrocytes, and oligodendrocytes, just as in the developing brain. This was first shown from the subventricular zone (Reynolds and Weiss, 1992; Richards et al., 1992) and then in the dentate gyrus of the hippocampus (Gage et al., 1995; Palmer et al., 1997), and in most structures of the brain examined (Palmer et al., 1995; Shihabuddin et al., 2000), even where no neurons existed (Palmer et al., 1999; Kondo and Raff, 2000).

Tsga10 protein is localized in some part of embryonic and mature brain where it may be involved either in the process of axonal transport (synaptogenesis) during embryogenesis or glia cells (active neurogenesis). Interestingly the cilium is seen within the astrocytes as well as in the brain (named as ependymal dynein). Actin rearrangement is crucial in axonal transport and synaptogenesis via PKC pathway while PKC expression level in brain newborn mice is elevated (Peters et al., 1999) indicating its role in brain development. Also, protein kinase C (PKC) plays an important role in neurogenesis and synaptogenesis in particular in some parts that need axonal transport while PKC activation is triggered by contact of astrocytes via actin rearrangement (Hama et al., 2004). It is shown that D-AKAP1 colocalizes with a PKA subunit at the postsynaptic membrane of the vertebrate neuromuscular junction (but not in the

presynaptic region) (Perkins et al., 2001). Astrocytes participate actively in synaptogenesis, and adult neurogenesis occurs in two regions of the adult brain: subventricular zone (SVZ) and the hippocampus subgranular zone (Hama et al., 2004) (exactly where we got the signals of Tsga10 expression). In parallel, it is well shown that myosin regulatory interacting protein (MIR, a member of ERM protein family) is localized specifically to neurons in cerebellum and hippocampus (Olsson et al., 2000). Mature hippocampus astrocytes regulate neurogenesis by instructing stem cells to adopt a neural fate. Tsga10 may play a role in this common pathway (because of its similarity to myosin domain) based on its expression in the part of brain with active synaptogenesis and neurogenesis.

Other areas of current project consist of the mechanisms of neural crest cell migration because of Tsga10 expression in some components of it including melanocyte, pancreatic islet cell, intestinal mucosa, primordial vibrassea follicles (Nosrat et al., 2004) and brain (cranial neural crest cells). Tsga10 expression in intestinal mucosa from E15 to newborn is compatible with the time when submucosal ganglia are first detected around E15.5 (Jiang et *al.*, 2003).

The expression of a gene that is expressed predominantly in the testis and in the mature rat brain is at first thought puzzling, as it was previously thought that cell division is very rare in the adult brain. However, recent research supports the idea of select pockets of continually dividing cells in the adult brain (A. Alvarez-Buylla et al 2001). This would explain the correlation as to why so many testis specific genes are also expressed in the brain. (I.A. Muslimov et al 2002, S. Suzu et al 2002, J.Dalmau et al 1999). The cell types that are widely accepted as continually dividing in the adult brain are granular

cells (found in the cerebellum and hippocampus) and glial cells (found throughout the central nervous system especially in the cortex). Our results showed that TSGA10 is being expressed strongly in certain areas of the cerebellum and hippocampus as well as being weaklier expressed in the cortex. The signals in the cortex have a striated pattern, which reflects the morphology of glial cells (G.W. Kreutzberg 1996). Only the N-terminus is expressed in the cortex and hippocampus whereas the N- and C-terminus were both expressed in the cerebellum. It is possible that this differential pattern of TSGA10 expression in the cerebellum could be due to a different morphology of the cell types expressing the protein, thus they could be using TSGA10 for alternative purposes. Only certain layers of the cerebellum expressed TSGA10, namely the granular layer and the white matter both of which contain continually dividing cells. In the hippocampus there were certain distinct areas that did not express TSGA10. The areas where there was no staining in the hippocampus correlated with the areas where pyramidal neurons were found. Pyramidal neurons were also found in the areas in the cerebellum where TSGA10 was not being expressed.

There was staining in the cortex, which was apparent throughout the white matter and was most likely also from glial cells. Glial cells divide throughout adulthood and are first present in the foetal cortex towards the end of development. The presence of staining in the foetal cortex of E15.5 embryos is further evidence that support the hypothesis that it is glial cells that are expressing TSGA10 in the adult cortex. There was a striking difference in the staining between certain areas. For example the corpus collosum (the neural tract which joins the left and right cortical hemispheres) and fimbria are clearly unstained in contrast to the other areas of cortex. Consistent with this

finding, the presence of Tsga10 N-terminus within the glia nuclei, and Tsga10-C-terminus in perinuclear region, may mean the whole Tsga10 protein (without any cleavage) is localized in a structure adjacent to nuclear membrane but after processing, its N-terminus enters the astrocytes nuclei. Considering the perinuclear localization (by IF and subcellular localization using GFP construct) of the Tsga10 protein, and its homology to a centrosomal protein 'cep135', Tsga10 is most likely a protein which is associated with the centrosome as a microtubule organizing structure in flagellae or related to Golgi apparatus.

As described earlier, Gene Ontology predicts that Tsga10 may be a 'porin'. Evidence in appoint of this hypothesis might be found from the fact that Tsga10 binds to 'Cyt C1', and that Tsga10 is a component of 'voltage-dependent anion-selective channel' proteins (VDACs). Interestingly, VDACs form pores in the outer mitochondrial membrane of eukaryotes as well as in brain postsynaptic membranes (Al-jamal 2004) and Cytochrome C is a component of the isolated VDAC complex (Vyssokikh et al, 2004).

Recently, a correlation between overexpression of 'Glial cell line-derived neurotrophic factor' (GDNF), RET (GDNF receptor) and testicular seminoma is confirmed. Also, GDNF expression is increased during testis development and peaks on postnatal day 7 in rat testis (Devouassoux-Shisheboran et al., 2003) as well as it is identified as a candidate regulator of germinal stem cell (GSC) proliferation and differentiation (Ymogida et al., 2003; Tadokoro et al., 2002). GDNF signal transducing receptor complex is composed of RET (Rearranged during Transfection) receptor tyrosine kinase and a glycosylphosphatidylinositol-linked co-receptor, GFRα1 (Saarma & Sariola 1999). GDNF is expressed in Sertoli cells while its receptors (RET and GFRα1) in

spermatogonia (Meng et al., 2000). We show TSGA10 protein expression specifically in glial cells of brain and in neural crest originated cells of embryos.

Cell cycle regulation and cancer- Additionally, GDNF and its receptors (RET, GFRα 1) are expressed in sperm as well as primordial of vibrassea follicles (Nosrat et *al.*, 2004) while RET proto-oncogene is known as a neural crest marker (Tanaka et al., 2004). Alternatively, Tsga10 expression in vibrassea follicles may be described by the fact that hair follicles have been recently shown to be local reservoirs of skin mast cell precursors (Kumamoto et al., 2003). There are evidences for RET and *c-kit* proto-oncogenes collaboration in the tumours with neural crest origin (Kato et al., 2004). The c-kit receptor tyrosine kinase belongs to PDGF/CSF-1/c-kit receptor subfamily. The proto-oncogene C-KIT encodes a tyrosine kinase receptor for SCF (Stem Cell Factor) and expressed in PGC (Primordial Germ Cells) (Devouassoux-Shisheboran et al., 2003). The coincidence of sites of expressions of TSGA10 and of GDNF and its receptors suggests that TSGA10 could be a component in an unclear pathway (involving GDNF, RET/GFRα1 and C-KIT), which spermatogonia enter for maturation and differentiation. Interestingly, in the context of Tsga10 expression in the tissues with neural crest origin, apart from some ESTs from pancreatic islet cells in BLAST for *Tsga10, FLJ32880* is shown to be expressed in pancreas as well as spermatid cells.

Modarressi (and co-aouthors 2001; 2004) have shown expression of TSGA10 in tumour cells. This is confirmed and extended in this thesis (chapters 3 section 3.6 and chapter 4 section 4.3.3) where I have shown *Tsga10* expression in premyeloid leukemia and the protein expression in human germ cell tumour. In addition Tanaka (et al., 2004)

detected Tsga10 protein as a testis-cancer tumour marker in some cancers including hepatocellular carcinoma and melanoma. However, it is not clear why TSGA10 is expressed in these tumours.

Hypothetically, all the cells that express TSGA10 (astrocytoma, acute myeloid leukemia, germ cell tumour, primordia of vibrassea follicles, malignant melanoma and others) originate from a common tissue they are either among totipotential germ cell or pluripotential stem cells or neural crest origin. Interestingly, Tsga10 protein expression is shown in primary cultures of neural crest cells (NCC) of 8-day mouse embryos by IHC (personal communication with Aldamaria Puliti, Università di Pisa, Italy).

Tsga10 expression in the embryonic brain and its localization in the perinuclear membrane, bearing in mind the similarity of Tsga10 to membrane-bound metallopeptidase while being expressed in some cancers may raise the possibility of a role for Tsga10 in number of processes such as axonal transport during brain development via Rho-ROCK-ERM axis (described in section 7.4.1), as well as a function in the cell membrane to afford spatial regulation of Rho family GTPase signaling through ERM proteins during metastasis as has been recently shown for the LyGDI protein (Ota et al., 2004). The possibility of a role for Tsga10 in metastasis is reinforced by its similarity to ERM (and PKC-matrix metalloproteinase interaction described formerly), as it is shown that degradation of extracellular matrix mediated by matrix metalloproteinases (MMPs) is necessary during tumor invasion and metastasis (Stetler-Stevenson 2001). ADAMs (A Disintegrin And Metalloprotease) – a newly described family of proteins- have important roles in adhesion and migration of cranial neural crest cells, the latter of which give rise to a number of important structures in the

head (Alfandari et al., 2001). Similarly, 'integrins' are a large family of transmembrane receptors that bind to components of the extracellular matrix (ECM) and to other cell surface receptors. One of the most significant features of integrins is that they serve as transmembrane links between ECM molecules and the actin cytoskeleton. Integrins have been shown to participate in both outside-in and inside-out signaling events that influence cell adhesion, cytoskeletal organization, and gene expression (Marsden et al., 2003; Davidson et al., 2002; Smith et al., 2002).

There is some other evidence for a parallel expression of *TSGA10* and *c-kit* proto-oncogene as well. In melanogenesis, *c-kit* 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 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 function as cell-associated molecules. SCF/c-kit is involved in different functions in the testis, including cell adhesion, cellular proliferation, PGC migration and anti-apoptotic actions. 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 important signalling cascades in tumourigenesis and cell division. There is some circumstantial evidence supporting the *TSGA10* link to *c-kit:*

a) Similarity of the site for phosphorylation by the tyrosine kinase in *c-kit* to one of *TSGA10* predicted protein sites (ENST00000288135).

***TSGA10* phosphorylation predicted sites:**

Name	Pos	Context	Score	Pred
Sequence	43	MLEKYERHL	0.711	*Y*
Sequence	653	RRQ**NYS**SNA	0.655	*Y*
Sequence	658	SSN**AYH**MSS	0.701	*Y*
Sequence	695	ENLCYRDF-	0.487	.

***c-kit* phosphorylation predicted sites:**

2 EVT**NYS**LKG

3 VK**RAYH**RLC

b) TSGA10 is expressed in melanocytes (EST accession no. N33757).

c) Expression of both proto-oncogene *c-kit* in a broad spectrum of human cancers (Matsuda *et al.*, 1993) and *TSGA10* in at least six different tumours (three from EST database and others by using RT-PCR and immunohistochemistry).

d) Both are expressed in spermatocytes (although *c-kit* expressed elsewhere too).

Tsga10 is introduced as a mitotic spindle checkpoint protein in yeast and higher eukaryotic probably via Mad1, Mad2, Mad3, Bub3 and Cdc20 complex by gene ontology survey in BLAST (http://www.ensembl.org/Homo_sapiens/geneview?gene=ENSG00000135951). In agreement to it, MAPK member proteins family encoded by the genes in neighbouring of both *Tsga10* and *Tsgaip* (*FLJ32880*) genes have been found; 'mitogen-activated

protein kinase kinase kinase kinase 4' in neighbouring *Tsga10* and 'mitogen-activated protein kinase kinase kinase 3' just downstream to *Tsga10ip* (11q13.1-q13.3). Again bioinformatically, an-eight aminoacid identity to mitogen activating protein kinase (MAP kinase kinase kinase (kinase3) sequence (ESEIQLLK) in *Tsga10* C-terminus is found, which is highly conserved in ESTs originating from tissues including: whole eye embryo, four-cell-Embryo, normal thyroid, male adrenal gland, newborn kidney, blastocyst, 7.5-dpc whole embryo, tumors and infiltrating ductal carcinoma. Considering above facts and Tsga10 (and Tsga10ip) protein expression in hipoccampal neurons and our hypothesis of Tsga10 role in Rho-ROCK-ERM axis (described previously), Tsga10/TSGA10 is more likely to be a member of MAPK family (most probably JNK subgroup among its four known -ERK, p38, JNK, ERK5- subfamily) that regulates the signaling via scaffolding mechanism (Tanoue and Nishida, 2002, 2003). That's why, it is shown that JNK-interacting protein family members (JIP1, JIP2 and JIP3) are well-known scaffold proteins which are required for stress-induced activation of JNK in hippocampal neurons via interaction with the kinesin light chain that has a role in axonal cargoes (Whitmarsh et al., 2001; Verhey et al., 2001). Also they are reported to associate with RhoGEF (Meyer et al., 1999; Stockinger et al., 2000). Also, it is shown that 'tumour suppressor RASSF1A' plays a critical role in mitotic progression and tumor suppression by controlling APC-Cdc20 activity.

Tsga10 may play an active role in the cell cycle via Raf kinases. It is shown that Raf kinase inhibitory protein inhibits beta-cell proliferation (Zhang et al., 2004) while Tsga10 is expressed strongly in pancreatic islet cells. Generally, tubulin (which is colocalized with Odf2) glutamylation is shown to be an important posttranslational

modification in microtubule interaction with structural and motor microtubule-associated proteins (MAPs) in brain and axonemes. More specifically it is shown that tubulin glutamylation can be involved in the regulation of the intracellular traffic including chromosomes and in the regulation of ciliary and flagellar motility (Kann et al., 2003).

7.7 Tsga10 and cilia

Different types of cilia including 'ependymal' cilia in the brain, 'respiratory' cilia in lung epithelium, 'nodal' cilia in testis (sperm tail) and embryo, 'renal' cilia and 'connecting cilia' are known. The results of the experiments in this project point to the presence of Tsga10 protein wherever there is a flagellum or a cilium. Tsga10 is expressed in sperm tail, olfactory and lung epithelium (which have cilia), some parts of brain (most likely these containing glia cells), astrocytes (which have cilia), and embryo. Afzelius proposed the presence of ciliary beating in the embryo in 1974 for the first time. It has been recently shown that localized dynein in E7.5 node (in the pole of embryo) is the source of nodal flow which in turn is necessary for left/right asymmetry during embryogenesis (Ibanez-Tallon et al., 2003, 2004). The nodal flow is leftward, *situs solitus* and if rightward *situs inversus*. Also it is shown that the ventral cells of the embryo node are a monociliated nonproliferative layer that can induce the second axis of nodal flow which is necessary for left/right assembly. They then make the embryonal groove of neural crest with contribution of dorsal ectodermal proliferative cells (Schneider and Brueckner 2000). The pattern of Tsga10 protein expression and localization during embryogenesis (in the parts which originated from neural crest) as well as in neural crest derived cells reinforce the hypothesis of Tsga10 role in (embryo) nodal cilia. PKD-2 (as a motif of Tsga10) which is a cation channel required for normal left/right development, localizes to this embryo monocilia node (Barr et al., 2001), is colocalized with α-tubulin (Portmann and Emmons 2004), and plays a role in flow-induced Ca^{2+} signalling (Kaletta et al., 2003). In this project Tsga10 expression is shown in olfactory epithelium which contains cilia microtubule organization and any

disruption in this structure results in anosmia (Kulaga et al., 2004). Also the time of *Tsga10* expression in early development (sometime between 4.5 and 7.5 dpc) is comparable to *Dnah5*, which its role in left/right assembly and its specific time of expression (E7.5) is confirmed (Olbrich et al., 2002). Mutations in mouse *Dnah5*/Dnah5 cause primary ciliary dyskinesia. This gene is expressed in ependymal cells from which the ependymal flow originates. Ependymal flow is necessary for the aqueduct not to be closed; its closure results in hydrocephalus. Nephronophthisis proteins (NPHP1, NPHP2, NPHP3, NPHP4 and NPHP5) are members of a protein family of cilia in which 'nodal' expression during embryogenesis has been shown (Olbrich et al., 2003). *NPHP1* is mapped in 2q13 in the region of *TSGA10*- and is necessary for cerebella development (Tsga10 is also expressed in cerebellum) (Parisi et al., 2004) and is colocalized with kinesin (personal communication). Problems in cilia structure may result in some diseases including primary ciliary dyskinesia. Clinical manifestations of this disease are hydrocephalus, retinitis pigmentosa, infertility and respiratory infection. The pattern of Tsga10 expression is consistent with all of the tissues involved in this disease.

It would be interesting to examine patients with this disease for aberrant expression of *TSGA10*. Also it would be interesting to study the roles of cilia in sperm, both in normal conditions and the pathological consequences caused by their dysfunction. Sperm cilia can be considered as required components for assembly, ion transport, providing energy source and motility. It will be interesting to study whether sperm cilia are the sites of the initial events of sperm signal transduction, and to study the activity of ciliary adenylate cyclases (ACs) (in particular the effects of calcium concentration). Also it is interesting to find the proteins, which may have a role for directional movement inside

the female reproductive tract and are part of a complex or a signalling pathway involved in directional cues necessary for recognition and entry into the female storage organs.

To check Tsga10 and other sperm tail (fibrous sheath) proteins for phosphorylation by activated PKA via intracellular cAMP (and being a member of AKAPs family) as well as their possible porin activity will be interesting. Also possible subcellular colocalization of centrosome (less likely Golgi) and micrortubule-associated Tsga10, Tsga10ip (and AKAPs) as well as their specificity for anchoring protein containing tethering domains for RI & RII-alpha could be nice to understand other proteins' possible role in the PKA phosphorylation pathway (also by kinase A activity of transfected cells and/or an assay for cations/anions of cells environment or western blot). Tsga10 (with its KVXF motif through which type I phosphatase interacts) could be recognized by PP1-targeting subunits such as AKAP350. It is also interesting to note the homology of Tsga10 to microtubule interacting protein in Drosophila and the neighbouring MAP2. Also, it is logical to study flagellar function by finding out which components of the Rho signaling pathway (rhophilin and ropporin) are associated with the fibrous sheath (and AKAPs) as scaffolds for signal transduction proteins. For example both Tsga10 and Tsga10ip as FS proteins (and possible AKAPs) have PDZ domains (rhophilin is an adaptor for it) in the site of its possible cleavage. I am interested to find out more sperm tail proteins which play a role in Ras(/Raf) GTPase Rho kinase pathway. Better understanding this pathway in sperm flagellar, can lead us to understand better the biological process(es) of cancer metastasis since sperm could be the most motile (metastatic! because it is close to undifferentiated germ line as well) normal cell in the body.

Conclusion

In conclusion, there is strong evidence that Tsga10 has a significant role in the development and function of the sperm tail. This role may be both as a structural component of the flagellum, and perhaps also in terms of protein trafficking and energy production. We know that the transcript is absent in two patients with non-obstructive infertility, and it would be logical to think that Tsga10 may be mutated or deleted in case of infertility involving lack of sperm motility such as dysplasia of fibrous sheath.

There seem to be consistent findings of Tsga10 expression in dividing cells, both in embryogenesis, and in adult dividing cells such as astrocytes. In these cases the Tsga1o protein may be functioning as a component of microtubules and/or manchette assembly.

Over-expression of Tsga10 in cancer cells mimics the expression pattern of a number of other genes (e.g. AKAPs) that are expressed in sperm and are over-expressed in some tumours. These so-called cancer-testis antigens can be indicators of the prognosis of these tumours, and may be useful target antigens for immunotherapy.

References

Abou-Haila A., Tulsiani D. R. P., 2000, Mammalian Sperm Acrosome: Formation, Contents, and Function, Archives of Biochemistry and Biophysics, 379 (2): 173-182.

Agarwala, S., Ragsdale, C. (2002). A role for midbrain arcs in nucleogenesis. Development 129(24), 5779-5788

Agarwala, S., Sanders, T.A., Ragsdale, C.W. (2002). Sonic hedgehog control of size and shape in midbrain pattern formation. Science 291, 2147-2150.

Akhmanova A, Hoogenraad CC, Drabek K, Stepanova T, Dortland B, Verkerk T, Vermeulen W, Burgering BM, De Zeeuw CI, Grosveld F, Galjart N.: Clasps are CLIP-115 and -170 associating proteins involved in the regional regulation of microtubule dynamics in motile fibroblasts. Cell. 2001 Mar 23;104(6):923-35.

Akira S., Isshiki H., Sugita T., Tanabe O., Kinoshita S., Nishio Y., Nakajima T., Hirano T., Kishimoto T.: A nuclear factor for IL-6 expression (NF-IL6) is a member of a C/EBP family. EMBO J. 9:1897-1906 (1990).

Alfandari D., Cousin H., Gaultier A., Smith KM., White JM., Darribère T., and DeSimone DW. 2001. Xenopus ADAM13 is a metalloprotease required for cranial neural crest-cell migration. Current Biology. 11, 918-930.

Allan DJ.: Radiation-induced apoptosis: its role in a MADCaT (mitosis-apoptosis-differentiation-calcium toxicity) scheme of cytotoxicity mechanisms.
Int J Radiat Biol. 1992 Aug;62(2):145-52. No abstract available.

Amano M, Fukata Y, Kaibuchi K: Regulation and functions of Rho-associated kinase. Exp Cell Res. 2000 Nov 25;261(1):44-51. Review.

Amano M, Fukata Y, Shimokawa H, Kaibuchi K: Purification and in vitro activity of Rho-associated kinase. Methods Enzymol. 2000;325:149-55.

Arlt H, Tauer R, Feldmann H, Neupert W, Langer T: The YTA10-12 complex, an AAA protease with chaperone-like activity in the inner membrane of mitochondria. Cell. 1996 Jun 14;85(6):875-85.

Ashizawa K, Wishart GJ, Hashimoto K, Tsuzuki Y. Dephosphorylation of a 30-kDa protein of fowl spermatozoa by the addition of myosin light chain kinase substrate peptide inhibits the flagellar motility. Biochem. Biophys. Res. Commun. 1995; 215: 706-712.

Ashizawa K, Wishart GJ, Nishinakama K, Sakamoto T, Tsuzuki Y. Regulatory mechanisms of fowl sperm motility: possible role of endogenous myosin light chain kinase-like protein. J. Reprod. Fertil. 1995; 104: 141-148.

Balhorn R: A model for the structure of chromatin in mammalian sperm. J Cell Biol. 1982 May;93(2):298-305.

Barr MM, DeModena J, Braun D, Nguyen CQ, Hall DH, Sternberg PW: The Caenorhabditis elegans autosomal dominant polycystic kidney disease gene homologs lov-1 and pkd-2 act in the same pathway. Curr Biol. 2001 Sep 4;11(17):1341-6.

Barth, K.A., Wilson, S.W. (1995). Expression of zebrafish Nkx2.2 is influenced by sonic hedgehog/ vertebrate hedgehog-1 and demarcates a zone of neuronal differentiation in the embryonic forebrain. Development 121, 1755-1768.

Beck K, Brodsky B: Supercoiled protein motifs: the collagen triple-helix and the alpha-helical coiled coil. J Struct Biol. 1998;122(1-2):17-29.

Behringer RR, Finegold MJ, Cate RL: Mullerian-inhibiting substance function during mammalian sexual development. Cell. 1994 Nov 4;79(3):415-25.

Bendixen C, Gangloff S, Rothstein R: A yeast mating-selection scheme for detection of protein-protein interactions. Nucleic Acids Res. 1994 May 11;22(9):1778-9.

Berg JS, Derfler BH, Pennisi CM, Corey DP, Cheney RE. Myosin-X, a novel myosin with pleckstrin homology domains, associates with regions of dynamic actin. J. Cell Sci. 2000; 113 Pt 19: 3439-3451.

Bhasin S, de Kretser DM, Baker HW: Clinical review 64: Pathophysiology and natural history of male infertility. J Clin Endocrinol Metab. 1994 Dec;79(6):1525-9.

Bhasin S, Berman N, Swerdloff RS: Follicle-stimulating hormone (FSH) escape during chronic gonadotropin-releasing hormone (GnRH) agonist and testosterone treatment. J Androl. 1994 Sep-Oct;15(5):386-91.

Bianco P, Riminucci M, Gronthos S, Robey PG: Bone marrow stromal stem cells: nature, biology, and potential applications. Stem Cells. 2001;19(3):180-92.

Bijur GN, Jope RS.: Rapid accumulation of Akt in mitochondria following phosphatidylinositol 3-kinase activation. J Neurochem. 2003 Dec;87(6):1427-35.

Boitani C, Geremia R, Rossi R, Monesi V: Electrophoretic pattern of polypeptide synthesis in spermatocytes and spermatids of the mouse. Cell Differ. 1980 Feb;9(1):41-9.

Braun RE: Temporal control of protein synthesis during spermatogenesis. Int J Androl. 2000;23 Suppl 2:92-4.

Braun RE: Post-transcriptional control and male infertility. Results Probl Cell Differ. 2000;28:111-29.

Berg JS, Derfler BH, Pennisi CM, Corey DP, Cheney RE. Myosin-X, a novel myosin with pleckstrin homology domains, associates with regions of dynamic actin. J Cell Sci 2000 113:pt 193439-3451.

Bornhauser BC, Johansson C, Lindholm D: Functional activities and cellular localization of the ezrin, radixin, moesin (ERM) and RING zinc finger domains in MIR. FEBS Lett. 2003 Oct 9;553(1-2):195-9.

Bornhauser BC, Olsson PA, Lindholm D: MSAP is a novel MIR-interacting protein that enhances neurite outgrowth and increases myosin regulatory light chain. J Biol Chem. 2003 Sep 12;278(37):35412-20.

Bretscher A, Edwards K, Fehon RG. ERM proteins and merlin: integrators at the cell cortex. Nat Rev Mol Cell Biol 2002 3:586-599

Breckler J, Au K, Cheng J, Hasson T, Burnside B: Novel myosin VI isoform is abundantly expressed in retina. Exp Eye Res. 2000 Jan;70(1):121-34.

Bretscher A, Edwards K, Fehon RG. ERM proteins and merlin: integrators at the cell cortex. Nat. Rev. Mol. Cell Biol. 2002; 3: 586-599.

Briscoe, J., Sussel, L., Serup, P., Hartigan-O'Connor, D., Jessell, T.M., Rubenstein, J.L., Ericson, J. (1999). Homeobox gene Nkx2.2 and specification of neuronal identity by graded Sonic hedgehog signalling. Nature 398, 622-627.

Briscoe, J., Chen, Y., Jessell, T.M., Struhl, G. (2001). A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in neural tube. Mol. Cell. 7, 1279-1291.

Brown PR, Miki K, Harper DB, Eddy EM: A-kinase anchoring protein 4 binding proteins in the fibrous sheath of the sperm flagellum. Biol Reprod. 2003 Jun;68(6):2241-8. Epub 2003 Jan 22.

Bruch J, Wilda M, Teigler-Schlegel A, Harbott J, Borkhardt A, Metzler M. Occurrence of an MLL/LAF4 fusion gene caused by the insertion ins(11;2)(q23;q11.2q11.2) in an infant with acute lymphoblastic leukemia. Genes Chromosomes. Cancer 2003; 37: 106-109.

Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA: Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis. Biol Reprod. 1998 Mar;58(3):834-41.

Caniggia, I. & Winter, J. L.: Hypoxia inducible factor-1: oxygen regulation of trophoblast differentiation in normal and pre-exlamptic pregnancies. Placenta 23:S47-S57 2002

Carninci P, Hayashizaki Y: High-efficiency full-length cDNA cloning. Methods Enzymol. 1999;303:19-44.

Cardone L, de Cristofaro T, Affaitati A, Garbi C, Ginsberg MD, Saviano M, Varrone S, Rubin CS, Gottesman ME, Avvedimento EV, Feliciello A: A-kinase anchor protein 84/121 are targeted to mitochondria and mitotic spindles by overlapping amino-terminal motifs. J Mol Biol. 2002 Jul 12;320(3):663-75.

Carlson AE, Westenbroek RE, Quill T, Ren D, Clapham DE, Hille B, Garbers DL, Babcock DF: CatSper1 required for evoked Ca2+ entry and control of flagellar function in sperm. Proc Natl Acad Sci U S A. 2003 Dec 9;100(25):14864-8.

Carr I, Clegg EJ, Meek GA: Sertoli cells as phagocytes: an electron microscopic study. J Anat. 1968 Mar;102(3):501-9.

Carr DW, Fujita A, Stentz CL, Liberty GA, Olson GE, Narumiya S: Identification of sperm-specific proteins that interact with A-kinase anchoring proteins in a manner similar to the type II regulatory subunit of PKA. J Biol Chem. 2001 May 18;276(20):17332-8.

Carrera A, Gerton GL, Moss SB: The major fibrous sheath polypeptide of mouse sperm: structural and functional similarities to the A-kinase anchoring proteins. Dev Biol. 1994 Sep;165(1):272-84.

Carrera AC, Borlado LR, Roberts TM, Martinez C.: Tyrosine kinase specific motif at subdomain VIII does not confer specificity for tyrosine. Biochem Biophys Res Commun. 1994 Dec 15;205(2):1114-20.

Catalano RD, Hillhouse EW, Vlad M: Developmental expression and characterization of FS39, a testis complementary DNA encoding an intermediate filament-related protein of the sperm fibrous sheath. Biol Reprod. 2001 Jul;65(1):277-87.

Chabot B, Stephenson DA, Chapman VM, Besmer P, Bernstein A: The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. Nature. 1988 Sep 1;335(6185):88-9.

Chada M, Prusa R, Bronsky J, Kotaska K, Sidlova K, Pechova M, Lisa L: Inhibin B, follicle stimulating hormone, luteinizing hormone and testosterone during childhood and puberty in males: changes in serum concentrations in relation to age and stage of puberty. Physiol Res. 2003;52(1):45-51.

Chadee, D. N.; Kyriakis, J. M: MLK3 is required for mitogen activation of B-Raf, ERK and cell proliferation. *Nature Cell Biol.* 6: 770-776, 2004.

Chai NN, Zhou H, Hernandez J, Najmabadi H, Bhasin S, Yen PH: Structure and organization of the RBMY genes on the human Y chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. Genomics. 1998 Apr 15;49(2):283-9.

Chiarini-Garcia H., Russel L. D., 2001 a, high-Resolution Light Microscopic Characterisation of Mouse Spermatogonia, Biology of Reproduction, 65 (4): 1170-1178.

Chiarini-Garcia H., Hornick J. R., Griswold M. D., Russel L. D., 2001 b, Distribution of Type A Spermatogonia in the Mouse is not Random, Biology of Reproduction, 65 (4): 1179-1185.

Chemes HE, Brugo S, Zanchetti F, Carrere C, Lavieri JC: Dysplasia of the fibrous sheath: an ultrastructural defect of human spermatozoa associated with sperm immotility and primary sterility. Fertil Steril. 1987 Oct;48(4):664-9.

Chemes HE, Carizza C, Scarinci F, Brugo S, Neuspiller N, Schwarsztein L: Lack of a head in human spermatozoa from sterile patients: a syndrome associated with impaired fertilization. Fertil Steril. 1987 Feb;47(2):310-6.
Chu G, Sharp PA: SV40 DNA transfection of cells in suspension: nalysis of efficiency of transcription and translation of T-antigen. Gene. 1981 Mar;13(2):197-202.

Clark AT, Firozi K, Justice MJ: Mutations in a novel locus on mouse chromosome 11 resulting in male infertility associated with defects in microtubule assembly and sperm tail function. Biol Reprod. 2004 May;70(5):1317-24.

Clark AM, Maguire SM, Griswold MD: Accumulation of clusterin/sulfated glycoprotein-2 in degenerating pachytene spermatocytes of adult rats treated with methoxyacetic acid. Biol Reprod. 1997 Oct;57(4):837-46.

Clark AM, Griswold MD: Expression of clusterin/sulfated glycoprotein-2 under conditions of heat stress in rat Sertoli cells and a mouse Sertoli cell line. J Androl. 1997 May-Jun;18(3):257-63.

Clermont Y, Oko R, Hermo L: Immunocytochemical localization of proteins utilized in the formation of outer dense fibers and fibrous sheath in rat spermatids: an electron microscope study. Anat Rec. 1990 Aug;227(4):447-57.

Cobb MH, Goldsmith EJ: How MAP kinases are regulated. J Biol Chem. 1995 Jun 23;270(25):14843-6.

Cobbe N, Heck MM: The evolution of SMC proteins: phylogenetic analysis and structural implications. Mol Biol Evol. 2004 Feb;21(2):332-47.

Colledge M, Dean RA, Scott GK, Langeberg LK, Huganir RL, Scott JD: Targeting of PKA to glutamate receptors through a MAGUK-AKAP complex. Neuron. 2000 Jul;27(1):107-19.

Colledge M, Scott JD: AKAPs: from structure to function. Trends Cell Biol. 1999 Jun;9(6):216-21.

Cooke HJ, Elliott DJ.: RNA-binding proteins and human male infertility. Trends Genet. 1997 Mar;13(3):87-9. Review. No abstract available.

Creemers LB, Meng X, den Ouden K, van Pelt AM, Izadyar F, Santoro M, Sariola H, deRooij DG. Transplantation of germ cells from glial cell line-derived neurotrophic factor-overexpressing mice to host testes depleted of endogenous spermatogenesis by fractionated irradiation. Biol Reprod 2002 66:1579-1584

Crepaldi T, Gautreau A, Comoglio PM, Louvard D, Arpin M: Ezrin is an effector of hepatocyte growth factor-mediated migration and morphogenesis in epithelial cells. J Cell Biol. 1997 Jul 28;138(2):423-34.

Cromer A, Carles A, Millon R, Ganguli G, Chalmel F, Lemaire F, Young J, Dembele D, Thibault C, Muller D, Poch O, Abecassis J, Wasylyk B: Identification of genes associated with tumorigenesis and metastatic potential of hypopharyngeal cancer by microarray analysis. Oncogene. 2004 Apr 1;23(14):2484-98.

Crossley, P.H., Martinez, S., Martin, G.R. (1996). Midbrain development induced by FGF8 in the chick embryo. Nature 380, 66-68.

D'Souza UM, Russ C, Tahir E, Mill J, McGuffin P, Asherson PJ, Craig IW: Functional effects of a tandem duplication polymorphism in the 5'flanking region of the DRD4 gene. Biol Psychiatry. 2004 Nov 1;56(9):691-7.

Davidson, L. A., Hoffstrom, B.G., Keller, R. and DeSimone, D.W. 2002. Mesendoderm migration and mantle closure during Xenopus laevis gastrulation: combined roles for integrin a5b1, fibronectin and tissue geometry. Dev. Biol. 242: 109-129.

De Felici M, Dolci S, Pesce M: Cellular and molecular aspects of mouse primordial germ cell migration and proliferation in culture. Int J Dev Biol. 1992 Jun;36(2):205-13.

de Jager M, Trujillo KM, Sung P, Hopfner KP, Carney JP, Tainer JA, Connelly JC, Leach DR, Kanaar R, Wyman C: Differential arrangements of conserved building blocks among homologs of the Rad50/Mre11 DNA repair protein complex.

J Mol Biol. 2004 Jun 11;339(4):937-49.

Devouassoux-Shisheboran M, Mauduit C, Tabone E, Droz JP, Benahmed M: Growth regulatory factors and signalling proteins in testicular germ cell tumours. APMIS. 2003 Jan;111(1):212-24.

Diviani D, Scott JD: AKAP signaling complexes at the cytoskeleton. J Cell Sci. 2001 Apr;114(Pt 8):1431-7. Review.

Dix DJ, Allen JW, Collins BW, Mori C, Nakamura N, Poorman-Allen P, Goulding EH, Eddy EM: Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. Proc Natl Acad Sci U S A. 1996 Apr 16;93(8):3264-8.

Dix DJ, Rosario-Herrle M, Gotoh H, Mori C, Goulding EH, Barrett CV, Eddy EM: Developmentally regulated expression of Hsp70-2 and a Hsp70-2/lacZ transgene during spermatogenesis. Dev Biol. 1996 Mar 15;174(2):310-21.

Dix DJ, Allen JW, Collins BW, Mori C, Nakamura N, Poorman-Allen P, Goulding EH, Eddy EM.: Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. Proc Natl Acad Sci USA. 1996 Apr 16;93(8):3264-8.

Dodge K, Scott JD: AKAP79 and the evolution of the AKAP model. FEBS Lett. 2000 Jun 30;476(1-2):58-61.

Doenecke D, Albig W, Bouterfa H, Drabent B: Organization and expression of H1 histone and H1 replacement histone genes. J Cell Biochem. 1994 Apr;54(4):423-31.

Doniarch, T. (1995). Basic FGF as an inducer of anteroposterior neural pattern. Cell 83, 1067-1070.

Dransfield DT, Bradford AJ, Smith J, Martin M, Roy C, Mangeat PH, Goldenring JR: Ezrin is a cyclic AMP-dependent protein kinase anchoring protein. EMBO J. 1997 Jan 2;16(1):35-43.

Dupe, V., and Lumsden, A. (2001). Hindbraimn patterning involves graded responces to Retinoic Acid signalling. Development 128(12), 2199-2208.

Durston, A.J., Timmermans, J.P., Hage, W.J., Hendricks, H.F., de Vries, N., Heideveld, M., Nieuwkoop, P.D. (1989). Retinoic acid causes an anteroposterior transformation in the developing central nervous system. Nature 340, 140-144.

Dutcher SK: The tubulin fraternity: alpha to eta. Curr Opin Cell Biol. 2001 Feb;13(1):49-54. Review.

Dwyer ND, Adler CE, Crump JG, L'Etoile ND, Bargmann CI: Polarized dendritic transport and the AP-1 mu1 clathrin adaptor UNC-101 localize odorant receptors to olfactory cilia. Neuron. 2001 Aug 2;31(2):277-87.

Eddy EM, Toshimori K, O'Brien DA: Fibrous sheath of mammalian spermatozoa. Microsc Res Tech. 2003 May 1;61(1):103-15.

Eddy EM.: HSP70-2 heat-shock protein of mouse spermatogenic cells. J Exp Zool. 1998 Sep-Oct 1;282(1-2):261-71.

Eddy EM: Role of heat shock protein HSP70-2 in spermatogenesis. Rev Reprod. 1999 Jan;4(1):23-30.

Edwards AS, Scott JD: A-kinase anchoring proteins: protein kinase A and beyond. Curr Opin Cell Biol. 2000 Apr;12(2):217-21.

El-Alfy M, Moshonas D, Morales CR, Oko R: Molecular cloning and developmental expression of the major fibrous sheath protein (FS 75) of rat sperm. J Androl. 1999 Mar-Apr;20(2):307-18.

Electronic Annotations Based on Mouse Homologs (MGI)

Elliott DJ, Cooke HJ: The molecular genetics of male infertility. Bioessays. 1997 Sep;19(9):801-9.

Ericson, J., Muhr, J., Placzek, M., Lints, T., Jessell, T.M., Edlund, T. (1995). Sonic hedgehog induces the differentiation of ventral forebrain neurons: a common signal for ventral patterning within the neural tube. Cell 81, 747-756.

Fawcett DW, Phillips DM: The fine structure and development of the neck region of the mammalian spermatozoon. Anat Rec. 1969 Oct;165(2):153-64. No abstract available.

Fawcett DW: The mammalian spermatozoon. *Dev Biol.* 1975;44:394-436

Ferlin A, Moro E, Onisto M, Toscano E, Bettella A, Foresta C: Absence of testicular DAZ gene expression in idiopathic severe testiculopathies. Hum Reprod. 1999 Sep;14(9):2286-92.

Ferlin A, Moro E, Garolla A, Foresta C: Human male infertility and Y chromosome deletions: role of the AZF-candidate genes DAZ, RBM and DFFRY. Hum Reprod. 1999 Jul;14(7):1710-6.

Foley, A.C., and Stern, C.D. (2001). Evolution of vertebrate forebrain: How many different mechanisms? J. Anatomy 199, 35-52.

Fujita A, Nakamura K, Kato T, Watanabe N, Ishizaki T, Kimura K, Mizoguchi A, Narumiya S: Ropporin, a sperm-specific binding protein of rhophilin, that is localized in
the fibrous sheath of sperm flagella. J Cell Sci. 2000 Jan;113 (Pt 1):103-12

Fukuda M, Kuroda TS: Slac2-c (synaptotagmin-like protein homologue lacking C2 domains-c), a novel linker protein that interacts with Rab27, myosin Va/VIIa, and actin. J Biol Chem. 2002 Nov 8;277(45):43096-103.

Fukuda M, Kuroda TS, Mikoshiba K: Slac2-a/melanophilin, the missing link between Rab27 and myosin Va: implications of a tripartite protein complex for melanosome transport. J Biol Chem. 2002 Apr 5;277(14):12432-6.

Fulcher KD, Welch JE, Klapper DG, O'Brien DA, Eddy EM.: Identification of a unique mu-class glutathione S-transferase in mouse spermatogenic cells. Mol Reprod Dev. 1995 Dec;42(4):415-24.

Gilbert S. F., 2003, Developmental Biology seventh edition, Sinauer Associates Inc.

Gillett GT, McConville CM, Byrd PJ, Stankovic T, Taylor AM, Hunt DM, West LF, Fox MF, Povey S, Benham FJ. Irradiation hybrids for human chromosome 11: characterization and use for generating region-specific markers in 11q14-q23. Genomics 1993 15:332-341

Godsave, S.F., Koster, C.H., Getahun, A., Mathu, M., Hooiveld, M., van der Wees, J., Hendriks, J., Durston, A.J. (1998). Graded retinoic responses in the developing hindbrain. Dev. Dyn. 213(1), 39-49.

Goldberg E: Transcriptional regulatory strategies in male germ cells. J Androl. 1996 Nov-Dec;17(6):628-32.

Gonzy G, Pokholkova GV, Peronnet F, Mugat B, Demakova OV, Kotlikova IV, Lepesant JA, Zhimulev IF. Isolation and characterization of novel mutations of the Broad-Complex, a key regulatory gene of ecdysone induction in Drosophila melanogaster. Insect Biochem. Mol. Biol. 2002; 32: 121-132

Goud B: How Rab proteins link motors to membranes. Nat Cell Biol. 2002 Apr;4(4):E77-8.

Griswold MD: The central role of Sertoli cells in spermatogenesis. Semin Cell Dev Biol. 1998 Aug;9(4):411-6.

Griswold DE, Hillegass LM, Breton JJ, Esser KM, Adams JL: Differentiation in vivo of classical non-steroidal antiinflammatory drugs from cytokine suppressive

antiinflammatory drugs and other pharmacological classes using mouse tumour necrosis factor alpha production. Drugs Exp Clin Res. 1993;19(6):243-8

Gu W, Kwon YK, Hecht NB: In 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. 1996 Jun;44(2):141-5

Gu W, Hecht NR.: Translation of a testis-specific Cu/Zn superoxide dismutase (SOD-1) mRNA is regulated by a 65-kilodalton protein which binds to its 5' untranslated region. Mol Cell Biol. 1996 Aug;16(8):4535-43.

Gurden, J.B., Dyson, S., and St Johnston, D. (1998). Cells' perception of position in a concentration gradient. Cell 95, 159-162.

Gurden, J.B. and Bourillot, P-Y. (2001). Morphogen gradient interpretation. Nature 413, 797-803.

Guan *et al*, 2000). Mushroom body defect, a gene involved in the control of neuroblast proliferation in Drosophila, encodes a coiled-coil protein. J Proc. Natl. Acad. Sci. U.S.A. 97 (14), 8122-8127 (2000)

Hama H, Hara C, Yamaguchi K, Miyawaki A.: PKC signaling mediates global enhancement of excitatory synaptogenesis in neurons triggered by local contact with astrocytes. Neuron. 2004 Feb 5;41(3):405-15.

Hasson T, Heintzelman MB, Santos-Sacchi J, Corey DP, Mooseker MS: Expression in cochlea and retina of myosin VIIa, the gene product defective in Usher syndrome type 1B. Proc Natl Acad Sci U S A. 1995 Oct 10;92(21):9815-9.

Hecht NB: The making of a spermatozoon: a molecular perspective. Dev Genet. 1995;16(2):95-103.

Hecht NB: Molecular mechanisms of male germ cell differentiation. Bioessays. 1998 Jul;20(7):555-61.

Hecht NB, Distel RJ, Yelick PC, Tanhauser SM, Driscoll CE, Goldberg E, Tung KS: Localization of a highly divergent mammalian testicular alpha tubulin that is not detectable in brain. Mol Cell Biol. 1988 Feb;8(2):996-1000.

Herman JG, Baylin SB: Promoter-region hypermethylation and gene silencing in human cancer. Curr Top Microbiol Immunol. 2000;249:35-54.

Hiwatari M, Taki T, Taketani T, Taniwaki M, Sugita K, Okuya M, Eguchi M, Ida K, Hayashi Y. Fusion of an AF4-related gene, LAF4, to MLL in childhood acute lymphoblastic leukemia with t(2;11)(q11;q23). Oncogene 2003; 22: 2851-2855.

Holstein A. F., Schulze W., Davidoff M., 2003, Understanding spermatogenesis is prerequisite for treatment, Reproductive biology and Endocrinology, 1: 107.

Hoog C: 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. 1995 Oct;39(5):719-26.

Hromas R, Collins SJ, Hickstein D, Raskind W, Deaven LL, O'Hara P, Hagen FS, Kaushansky K: A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. J Biol Chem. 1991 Aug 5;266(22):14183-7.

http://www.ncbi.nlm.nih.gov/entrez/query

http://www.ncbi.nlm.nih.gov/geo/index.cgi?qb=pro

Hsu W., Kerppola T. K., Chen P.-L., Curran T., Chen-Kiang S: Fos and Jun repress transcription activation by NF-IL6 through association at the basic zipper region. Mol. Cell. Biol. 14:268-276 (1994).

Huhtaniemi I: Fetal testis--a very special endocrine organ. Eur J Endocrinol. 1994 Jan;130(1):25-31.

Ibanez-Tallon I, Pagenstecher A, Fliegauf M, Olbrich H, Kispert A, Ketelsen UP, North A, Heintz N, Omran H: Dysfunction of axonemal dynein heavy chain Mdnah5 inhibits ependymal flow and reveals a novel mechanism for hydrocephalus formation. Hum Mol Genet. 2004 Sep 15;13(18):2133-41.

Ibanez-Tallon I, Heintz N, Omran H: To beat or not to beat: roles of cilia in development and disease. Hum Mol Genet. 2003 Apr 1;12 Spec No 1:R27-35.

Ibanez-Tallon I, Gorokhova S, Heintz N: Loss of function of axonemal dynein Mdnah5 causes primary ciliary dyskinesia and hydrocephalus. Hum Mol Genet. 2002 Mar 15;11(6):715-21.

Imagawa M, Chiu R, Karin M: Transcription factor AP-2 mediates induction by two different signal transduction pathways: Protein kinase C and cAMP.
Cell 51:251-260 1987

Iratni, R.; Yan, Y.-T.; Chen, C.; Ding, J.; Zhang, Y.; Price, S. M.; Reinberg, D.; Shen, M. M. : Inhibition of excess Nodal signaling during mouse gastrulation by the transcriptional corepressor DRAP1. *Science* 298: 1996-1999, 2002.

Irons MJ, Clermont Y: Kinetics of fibrous sheath formation in the rat spermatid.
Am J Anat. 1982 Oct;165(2):121-30.

Irons MJ, Clermont Y: Formation of the outer dense fibers during spermiogenesis in the rat. Anat Rec. 1982 Apr;202(4):463-71.

Jagadish MN, Staton VJ, Hudson PJ, Azad AA: Birnavirus precursor polyprotein is processed in Escherichia coli by its own virus-encoded polypeptide. J Virol 1988;62:1084-1087.

Jessberger S, Kempermann G.: Adult-born hippocampal neurons mature into activity-dependent responsiveness. Eur J Neurosci. 2003 Nov;18(10):2707-12.

Jiang et *al.*, Dev Biol 2003; 258:364-84

Jimenez R, Burgos M: Mammalian sex determination: joining pieces of the genetic puzzle. Bioessays. 1998 Sep;20(9):696-9.

John GR, Chen L, Rivieccio MA, Melendez-Vasquez CV, Hartley A, Brosnan CF: Interleukin-1beta induces a reactive astroglial phenotype via deactivation of the Rho GTPase-Rock axis. J Neurosci. 2004 Mar 17;24(11):2837-45.

Johnson A.D. & Gomes W.R. (1977). The testis. In: Advances in physiology, biochemistry and function. Academic Press. New York.

Jones MH, Davey PM, Aplin H, Affara NA: Expression analysis, genomic structure, and mapping to 7q31 of the human sperm adhesion molecule gene SPAM1. Genomics. 1995 Oct 10;29(3):796-800.

Jones S, Sgouros J: The cohesin complex: sequence homologies, interaction networks and shared motifs. Genome Biol. 2001;2(3):RESEARCH0009.

Kaletta T, Van der Craen M, Van Geel A, Dewulf N, Bogaert T, Branden M, King KV, Buechner M, Barstead R, Hyink D, Li HP, Geng L, Burrow C, Wilson P: Towards understanding the polycystins. Nephron Exp Nephrol. 2003 Jan;93(1):e9-17.

Kaser M, Kambacheld M, Kisters-Woike B, Langer T: Oma1, a novel membrane-bound metallopeptidase in mitochondria with activities overlapping with the m-AAA protease. J Biol Chem. 2003 Nov 21;278(47):46414-23.

Kann ML, Soues S, Levilliers N, Fouquet JP: Glutamylated tubulin: diversity of expression and distribution of isoforms. Cell Motil Cytoskeleton. 2003 May;55(1):14-25.

Kato R, et al.: Damage-induced neuronal endopeptidase (DINE/ECEL) expression is regulated by leukemia inhibitory factor and deprivation of nerve growth factor in rat sensory ganglia after nerve injury, J Neurosci 2002 Nov 1;22(21):9410-8.

Kato A, Nagata Y, Todokoro K: Delta-tubulin is a component of intercellular bridges and both the early and mature perinuclear rings during spermatogenesis. Dev Biol. 2004 May 1;269(1):196-205.

Kerr JB: Ultrastructure of the seminiferous epithelium and intertubular tissue of the human testis. J Electron Microsc Tech. 1991 Oct;19(2):215-40.

Khanna C, Wan X, Bose S, Cassaday R, Olomu O, Mendoza A, Yeung C, Gorlick R, Hewitt SM, Helman LJ: The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. Nat Med. 2004 Feb;10(2):182-6. Epub 2004 Jan 04.

Kissel H, Timokhina I, Hardy MP, Rothschild G, Tajima Y, Soares V, Angeles M, Whitlow SR, Manova K, Besmer P: Point mutation in kit receptor tyrosine kinase reveals essential roles for kit signaling in spermatogenesis and oogenesis without affecting other kit responses. EMBO J. 2000 Mar 15;19(6):1312-26.

Kierszenbaum AL, Rivkin E, Tres LL: The actin-based motor myosin Va is a component of the acroplaxome, an acrosome-nuclear envelope junctional plate, and of manchette-associated vesicles. Cytogenet Genome Res. 2003;103(3-4):337-44.

Kierszenbaum AL, Tres LL: Bypassing natural sperm selection during fertilization: the azh mutant offspring experience and the alternative of spermiogenesis in vitro. Mol Cell Endocrinol. 2002 Feb 22;187(1-2):133-8.

Kobayashi Y, Watanabe M, Okada Y, Sawa H, Takai H, Nakanishi M, Kawase Y, Suzuki H, Nagashima K, Ikeda K, Motoyama N: Hydrocephalus, situs inversus, chronic sinusitis, and male infertility in DNA polymerase lambda-deficient mice: possible implication for the pathogenesis of immotile cilia syndrome. Mol Cell Biol. 2002 Apr;22(8):2769-76.

Kobayashi A, Miyake T, Kawaichi M, Kokubo T: Mutations in the histone fold domain of the TAF12 gene show synthetic lethality with the TAF1 gene lacking the TAF N-terminal domain (TAND) by different mechanisms from those in the SPT15 gene encoding the TATA box-binding protein (TBP). Nucleic Acids Res. 2003 Feb 15;31(4):1261-74.

Kosako H, Yoshida T, Matsumura F, Ishizaki T, Narumiya S, Inagaki M: Rho-kinase/ROCK is involved in cytokinesis through the phosphorylation of myosin light chain and not ezrin/radixin/moesin proteins at the cleavage furrow. Oncogene. 2000 Dec 7;19(52):6059-64.

Krieg J, Hunter T: Identification of the two major epidermal growth factor-induced tyrosine phosphorylation sites in the microvillar core protein ezrin. J Biol Chem. 1992 Sep 25;267(27):19258-65.

Krausz C, McElreavey K: Y chromosome and male infertility. Front Biosci. 1999 Jan 15;4:E1-8. Kwon , 1993 Prm1 prm2

Kulaga HM, Leitch CC, Eichers ER, Badano JL, Lesemann A, Hoskins BE, Lupski JR, Beales PL, Reed RR, Katsanis N: Loss of BBS proteins causes anosmia in humans and defects in olfactory cilia structure and function in the mouse. Nat Genet. 2004 Sep;36(9):994-8. Epub 2004 Aug 22.

Kumamoto T, Shalhevet D, Matsue H, Mummert ME, Ward BR, Jester JV, Takashima A: Hair follicles serve as local reservoirs of skin mast cell precursors. Blood. 2003 Sep 1;102(5):1654-60. Epub 2003 May 08

Lahbib-Mansais Y, Barbosa A, Yerle M, Parma P, Milan D, Pailhoux E, Gellin J, Cotinot C: Mapping in pig of genes involved in sexual differentiation: AMH, WT1, FTZF1, SOX2, SOX9, AHC, and placental and embryonic CYP19. Cytogenet Cell Genet. 1997;76(1-2):109-14.

Lamb RF, Roy C, Diefenbach TJ, Vinters HV, Johnson MW, Jay DG, Hall A: The TSC1 tumour suppressor hamartin regulates cell adhesion through ERM proteins and the GTPase Rho. Nat Cell Biol. 2000 May;2(5):281-7.

Langford GM: ER transport on actin filaments in squid giant axon: implications for signal transduction at synapse. FASEB J. 1999 Dec;13 Suppl 2:S248-50.

Lantz,V.A. and Miller,K.G.: A class VI unconventional myosin is associated with a homologue of a microtubule-binding protein, cytoplasmic linker protein-170, in neurons and at the posterior pole of Drosophila embryos. J. Cell Biol. 140 (4), 897-910 (1998).

LeClair et al., 1992

Leclerc P, de Lamirande E, Gagnon C: Cyclic adenosine 3',5'monophosphate-dependent regulation of protein tyrosine phosphorylation in relation to human sperm capacitation and motility. Biol Reprod. 1996 Sep;55(3):684-92.

Leclerc P, Gagnon C: Phosphorylation of Triton X-100 soluble and insoluble protein substrates in a demembranated/reactivated human sperm model. Mol Reprod Dev. 1996 Jun;44(2):200-11.

Lee K, Haugen HS, Clegg CH, Braun RE.: Premature translation of protamine 1 mRNA causes precocious nuclear condensation and arrests spermatid differentiation in mice. Proc Natl Acad Sci U S A. 1995 Dec 19;92(26):12451-5.

Lewis SA, Cowan NJ: Complex regulation and functional versatility of mammalian alpha- and beta-tubulin isotypes during the differentiation of testis and muscle cells. J Cell Biol. 1988 Jun;106(6):2023-33.

Liem, K.F., Jr., Tremml, G., Roelink, H., Jessell, T.M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell 82, 969-979.

Lukahev, D., Caldwell, C., Ohta, A., Chen, P. & Sitkovsky, M: Differential regulation of two alternatively spliced isoforms of hypoxia-inducible factor-1α in activated T lymphocytes. J. Biol. Chem. 276:48754-48763 2001

Luo X, Ikeda Y, Parker KL.: A cell-specific nuclear receptor is essential for adrenal and gonadal development and sexual differentiation. Cell. 1994 May 20;77(4):481-90.

Lupas A: Predicting coiled-coil regions in proteins. Curr Opin Struct Biol. 1997 Jun;7(3):388-93.

Ma C, Staudt LM: LAF-4 encodes a lymphoid nuclear protein with transactivation potential that is homologous to AF-4, the gene fused to MLL in t(4;11) leukemias. Blood. 1996 Jan 15;87(2):734-45.

Maden, M., Gale, E., Kostetskii, I., Zilc, M.H. (1996). Vitamin A-deficient quail embryos have only half a hindbrain and other neural defects. Curr. Biol. 6, 417-426.

Mahadevaiah SK, Odorisio T, Elliott DJ, Rattigan A, Szot M, Laval SH, Washburn LL, McCarrey JR, Cattanach BM, Lovell-Badge R, Burgoyne PS: Mouse homologues of the human AZF candidate gene RBM are expressed in spermatogonia and spermatids, and map to a Y chromosome deletion interval associated with a high incidence of sperm abnormalities. Hum Mol Genet. 1998 Apr;7(4):715-27.

Marigo, V. and Tabin, C.J. (1996). Regulation of patched by sonic hedgehog in the developing neural tube. Proc. Natl. Acad. Sci. USA 93, 9346-9351.

Marin, R., Puelles, L. (1994), Patterning of the embryonic avian midbrain after experimental inversions: a polarizing activity from the isthmus. Dev. Biol. 163, 19-37.

McHugh B, Krause SA, Yu B, Deans AM, Heasman S, McLaughlin P, Heck MM: Invadolysin: a novel, conserved metalloprotease links mitotic structural rearrangements with cell migration. J Cell Biol. 2004 Nov 22;167(4):673-86.

Marsden, M. and DeSimone, D.W. 2003. Integrin-ECM Interactions Regulate Cadherin-Dependent Cell Adhesion and Are Required for Convergent Extension in Xenopus. Current Biology 13, 1182-1191.

Margulies L., Sehgal P. B.: Modulation of the human interleukin-6 promoter (IL-6) and transcription factor C/EBPbeta (NF-IL-6) activity by p53 species. J. Biol. Chem. 268:15096-15100 (1993).

Martinez, S., Wassef, M., Alvarado-Mallart, R-M.: Induction of a mesencephalic phenotype in the 2-day-old chick prosencephalon is preceeded by the early expression of the homeobox gene. 1991 En. Neuron 6, 971-981.

Massague J.: TGF-beta signal transduction. Annu Rev Biochem. 1998;67:753-91.

Matsusaka T., Fujikawa K., Nishio Y., Mukaida N., Matsushima K., Kishimoto T., Akira S.: Transcription factors NF-IL6 and NF-kappaB synergistically activate transcription of the inflammatory cytokines interleukin 6 and interleukin 8. Proc. Natl. Acad. Sci. USA 90:10193-10197 (1993).

Mazeyrat S, Saut N, Sargent CA, Grimmond S, Longepied G, Ehrmann IE, Ellis PS, Greenfield A, Affara NA, Mitchell MJ: The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. Hum Mol Genet. 1998 Oct;7(11):1713-24.

Meeks JJ, Weiss J, Jameson JL: Dax1 is required for testis determination. Nat Genet. 2003 May;34(1):32-3.

Meeks JJ, Russell TA, Jeffs B, Huhtaniemi I, Weiss J, Jameson JL: Leydig cell-specific expression of DAX1 improves fertility of the Dax1-deficient mouse. Biol Reprod. 2003 Jul;69(1):154-60. Epub 2003 Feb 19.

Meeks JJ, Crawford SE, Russell TA, Morohashi K, Weiss J, Jameson JL: Dax1 regulates testis cord organization during gonadal differentiation. Development. 2003 Mar;130(5):1029-36.

Meng X, Lindahl M, Hyvönen ME, Parvinen M, deRooij DG, Hess MW, Raatikainen-Ahokas A, Sainio K, Rauvala H, Lakso M, Pichel JG, Westphal H, Saarma M, Sariola H.: Regulation of cell fate decision of undifferentiated spermatogonia by GDNF. Science 2000 287:1489-1493

Meng X, deRooij DG, Westerdahl K, Saarma M, Sariola H. Promotion of seminomatous tumors by targeted overexpression of glial cell line-derived neurotrophic factor in mouse testis. Cancer Res 2001 61:3267-3271.

Meyer D, Liu A, Margolis B: Interaction of c-Jun amino-terminal kinase interacting protein-1 with p190 rhoGEF and its localization in differentiated neurons. J Biol Chem. 1999 Dec 3;274(49):35113-8.

Migliaccio A, Di Domenico M, Castoria G, de Falco A, Bontempo P, Nola E, Auricchio F: Tyrosine kinase/p21ras/MAP-kinase pathway activation by estradiol-receptor complex in MCF-7 cells. EMBO J. 1996 Mar 15;15(6):1292-300.

Miki K, Willis WD, Brown PR, Goulding EH, Fulcher KD, Eddy EM: Targeted disruption of the Akap4 gene causes defects in sperm flagellum and motility. Dev Biol. 2002 Aug 15;248(2):331-42.

Minamiguchi H, Kimura T, Urata Y, Miyazaki H, Bamba T, Abe T, Sonoda Y.: Simultaneous signalling through c-mpl, c-kit and CXCR4 enhances the proliferation and differentiation of human megakaryocyte progenitors: possible roles of the PI3-K, PKC and MAPK pathways. Br J Haematol. 2001 Oct;115(1):175-85.

Mochizuki T. , Wu G. , Hayashi T. , Xenophontos S.L. , Veldhuisen B. , Saris J.J. , Reynolds D.M. , Cai Y. , Gabow P.A. , Pierides A. , Kimberling W.J. , Breuning M.H. , Deltas C.C. , Peters D.J. , Somlo S.: PKD2, a gene for polycystic kidney disease that encodes an integral membrane protein. Science 272: 1339- 1342 (1996)

Monesi V: Synthetic activities during spermatogenesis in the mouse RNA and protein. Exp Cell Res. 1965 Aug;39(1):197-224.

Modarressi MH, Cameron J, Taylor KE, Wolfe J.: Identification and characterisation of a novel gene, TSGA10, expressed in testis. Gene. 2001 Jan 10;262(1-2):249-55.

Modarressi MH, Behnam B, Cheng M, Taylor KE, Wolfe J, Van Der Hoorn FA: Tsga10 Encodes a 65-Kilodalton Protein That Is Processed to the 27-Kilodalton Fibrous Sheath Protein. Biology of Reproduction 608–615 (2004)

Monier,S., Jollivet,F., Janoueix-Lerosey,I., Johannes,L. and Goud,B.: Characterization of novel Rab6-interacting proteins involved in endosome-to-TGN transport. J Traffic 3 (4), 289-297 (2002)

Morales CR, Hecht NB: Poly(A)+ ribonucleic acids are enriched in spermatocyte nuclei but not in chromatoid bodies in the rat testis. Biol Reprod. 1994 Feb;50(2):309-19.

Morrison DK: KSR: a MAPK scaffold of the Ras pathway? J Cell Sci. 2001 May;114(Pt 9):1609-12.

Moura AA, Erickson BH: Testicular development, histology, and hormone profiles in three yearling angus bulls with spermatogenic arrest. Theriogenology. 2001 Apr 15;55(7):1469-88.

Munday AD, Norris FA, Caldwell KK, Brown S, Majerus PW, Mitchell CA: The inositol polyphosphate 4-phosphatase forms a complex with phosphatidylinositol 3-kinase in human platelet cytosol. Proc Natl Acad Sci U S A. 1999 Mar 30;96(7):3640-5.

Munoz-Fontela C, Rodriguez E, Nombela C, Arroyo J, Rivas C: Characterization of the bipartite nuclear localization signal of protein LANA2 from Kaposi's sarcoma-associated herpesvirus. Biochem J. 2003 Sep 1;374(Pt 2):545-50.

Nachtigal MW, Hirokawa Y, Enyeart-VanHouten DL, Flanagan JN, Hammer GD, Ingraham HA.: Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. Cell. 1998 May 1;93(3):445-54.

NagDas SK, Winfrey VP, Olson GE.: Identification of ras and its downstream signaling elements and their potential role in hamster sperm motility. Biol Reprod. 2002 Oct;67(4):1058-66.

Nadeau OW, Carlson GM, Gogol EP: A Ca(2+)-dependent global conformational change in the 3D structure of phosphorylase kinase obtained from electron microscopy. Structure (Camb). 2002 Jan;10(1):23-32.

Nakagawa Y, Yamane Y, Okanoue T, Tsukita S, Tsukita S.: Outer dense fiber 2 is a widespread centrosome scaffold component preferentially associated with mother centrioles: its identification from isolated centrosomes. Mol Biol Cell. 2001 Jun;12(6):1687-97.

Nakajima,H., Hirata,A., Ogawa,Y., Yonehara,T., Yoda,K. and Yamasaki,M.: A cytoskeleton-related gene, uso1, is required for intracellular protein transport in Saccharomyces cerevisiae. J. Cell Biol. 113 (2), 245-260 (1991)

Nakajima T., Kinoshita S., Sasagawa T., Sasaki K., Naruto M., Kishimoto T., Akira S.: Phosphorylation at threonine-235 by ras-dependent mitogen-activated protein kinase cascade is essential for transcription factor NF-IL6. Proc. Natl. Acad. Sci. USA 90:2207-2211 (1993).

Nakamura N, Oshiro N, Fukata Y, Amano M, Fukata M, Kuroda S, Matsuura Y, Leung T, Lim L, Kaibuchi K: Phosphorylation of ERM proteins at filopodia induced by Cdc42. Genes Cells. 2000 Jul;5(7):571-81.

Nakamura K, Fujita A, Murata T, Watanabe G, Mori C, Fujita J, Watanabe N, Ishizaki T, Yoshida O, Narumiya S: Rhophilin, a small GTPase Rho-binding protein, is abundantly expressed in the mouse testis and localized in the principal piece of the sperm tail. FEBS Lett. 1999 Feb 19;445(1):9-13.

Nantel F, Sassone-Corsi P.: CREM: A transcriptional master switch during the spermatogenesis differentiation program Front Biosci. 1996 Sep 1;1:d266-9.

Nantel F, Monaco L, Foulkes NS, Masquilier D, LeMeur M, Henriksen K, Dierich A, Parvinen M, Sassone-Corsi P.: Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. Nature. 1996 Mar 14;380(6570):159-62.

Narisawa S, Hecht NB, Goldberg E, Boatright KM, Reed JC, Millan JL.: Testis-specific cytochrome c-null mice produce functional sperm but undergo early testicular atrophy. Mol Cell Biol. 2002 Aug;22(15):5554-62.

Nayernia K, Nieter S, Kremling H, Oberwinkler H, Engel W.: Functional and molecular characterization of the transcriptional regulatory region of the proacrosin gene. J Biol Chem. 1994 Dec 23;269(51):32181-6.

Nayernia K, Reim K, Oberwinkler H, Engel W.: Diploid expression and translational regulation of rat acrosin gene. Biochem Biophys Res Commun. 1994 Jul 15;202(1):88-93.

Nayernia K, Nieter S, Kremling H, Oberwinkler H, Engel W: Functional and molecular characterization of the transcriptional regulatory region of the proacrosin gene. J Biol Chem. 1994 Dec 23;269(51):32181-6.

Nguyen A, Burack WR, Stock JL, Kortum R, Chaika OV, Afkarian M, Muller WJ, Murphy KM, Morrison DK, Lewis RE, McNeish J, Shaw AS: Kinase suppressor of Ras (KSR) is a scaffold which facilitates mitogen-activated protein kinase activation in vivo.
Mol Cell Biol. 2002 May;22(9):3035-45.

Nishimune H, Uyeda A, Nogawa M, Fujimori K, Taguchi T.: Neurocrescin: a novel neurite-outgrowth factor secreted by muscle after denervation. Neuroreport. 1997 Nov 10;8(16):3649-54.

Nosrat IV, Smith CA, Mullally P, Olson L, Nosrat CA: Dental pulp cells provide neurotrophic support for dopaminergic neurons and differentiate into neurons in vitro; implications for tissue engineering and repair in the nervous system. Eur J Neurosci. 2004 May;19(9):2388-98.

Ohuchi J, Arai T, Kon Y, Asano A, Yamauchi H, Watanabe T: Characterization of a novel gene, sperm-tail-associated protein (Stap), in mouse post-meiotic testicular germ cells. Mol Reprod Dev. 2001 Aug;59(4):350-8.

Olbrich H, Haffner K, Kispert A, Volkel A, Volz A, Sasmaz G, Reinhardt R, Hennig S, Lehrach H, Konietzko N, Zariwala M, Noone PG, Knowles M, Mitchison HM, Meeks M, Chung EM, Hildebrandt F, Sudbrak R, Omran H: Mutations in DNAH5 cause primary ciliary dyskinesia and randomization of left-right asymmetry. Nat Genet. 2002 Feb;30(2):143-4.

Olbrich H, Fliegauf M, Hoefele J, Kispert A, Otto E, Volz A, Wolf MT, Sasmaz G, Trauer U, Reinhardt R, Sudbrak R, Antignac C, Gretz N, Walz G, Schermer B, Benzing T, Hildebrandt F, Omran H: Mutations in a novel gene, NPHP3, cause adolescent nephronophthisis, tapeto-retinal degeneration and hepatic fibrosis. Nat Genet. 2003 Aug;34(4):455-9.

Okabe M, Ikawa M, Ashkenas J: Male infertility and the genetics of spermatogenesis. Am J Hum Genet. 1998 Jun;62(6):1274-81.

Oko R, Clermont Y: Light microscopic immunocytochemical study of fibrous sheath and outer dense fiber formation in the rat spermatid. Anat Rec. 1989 Sep;225(1):46-55.

Oko R: Comparative analysis of proteins from the fibrous sheath and outer dense fibers of rat spermatozoa. Biol Reprod 1988 39:169-182

Olsson PA, Bornhauser BC, Korhonen L, Lindholm D: Neuronal expression of the ERM-like protein MIR in rat brain and its localization to human chromosome 6. Biochem Biophys Res Commun. 2000 Dec 29;279(3):879-83.

Ota T, Maeda M, Suto S, Tatsuka M: LyGDI functions in cancer metastasis by anchoring Rho proteins to the cell membrane. Mol Carcinog. 2004 Apr;39(4):206-20.

Papalopulu, N., Lovell-Badge, R., Krumlauf, R. (1991). The expression of murine Hox-2 genes is dependant on the differentiation pathway and displays a collinear sensitivity to retinoic acid in F9 cells and Xenopus embryos. Nucleic Acids Res. 19, 5497-5506.

Parisi MA, Bennett CL, Eckert ML, Dobyns WB, Gleeson JG, Shaw DW, McDonald R, Eddy A, Chance PF, Glass IA: The NPHP1 gene deletion associated with juvenile nephronophthisis is present in a subset of individuals with Joubert syndrome. Am J Hum Genet. 2004 Jul;75(1):82-91. Epub 2004 May 11.

Peri A, Krausz C, Cioppi F, Granchi S, Forti G, Francavilla S, Serio M: Cyclic adenosine 3',5'-monophosphate-responsive element modulator gene expression in germ cells of normo- and oligoazoospermic men. J Clin Endocrinol Metab. 1998 Oct;83(10):3722-6.

Perkins GA, Wang L, Huang LJ, Humphries K, Yao VJ, Martone M, Deerinck TJ, Barraclough DM, Violin JD, Smith D, Newton A, Scott JD, Taylor SS, Ellisman MH: PKA, PKC, and AKAP localization in and around the neuromuscular junction. BMC Neurosci. 2001;2(1):17.

Perrotti D, Melotti P, Skorski T, Casella I, Peschle C, Calabretta B: Overexpression of the zinc finger protein MZF1 inhibits hematopoietic development from embryonic stem cells: correlation with negative regulation of CD34 and c-myb promoter activity. Mol Cell Biol. 1995 Nov;15(11):6075-87.

Piette J, Klarsfeld A, Changeux J: Interaction of nuclear factors with the upstream region of the alpha-subunit gene of chicken muscle acetylcholine receptor: variations with muscle differentiation and denervation. EMBO J. 8:687-694 (1989).

Peters CA, Maizels ET, Hunzicker-Dunn M: Activation of PKC delta in the rat corpus luteum during pregnancy. Potential role of prolactin signaling. J Biol Chem. 1999 Dec 24;274(52):37499-505.

Pritchard-Jones K, Fleming S.: Cell types expressing the Wilms' tumour gene (WT1) in Wilms' tumours: implications for tumour histogenesis. Oncogene. 1991 Dec;6(12):2211-20.

Portman DS, Emmons SW: Identification of C. elegans sensory ray genes using whole-genome expression profiling. Dev Biol. 2004 Jun 15;270(2):499-512.

Qi Z, Tang D, Matsuura I, Lee KY, Zhu X, Huang QQ, Wang JH: Regulatory properties of neuronal cdc2-like kinase. Mol Cell Biochem. 1995 Aug-Sep;149-150:35-9.

Ralt D, Manor M, Cohen-Dayag A, Tur-Kaspa I, Ben-Shlomo I, Makler A, Yuli I, Dor J, Blumberg S, Mashiach S, et al.: Chemotaxis and chemokinesis of human spermatozoa to follicular factors. Biol Reprod. 1994 Apr;50(4):774-85.

Rey R, Mebarki F, Forest MG, Mowszowicz I, Cate RL, Morel Y, Chaussain JL, Josso N.: Anti-mullerian hormone in children with androgen insensitivity. J Clin Endocrinol Metab. 1994 Oct;79(4):960-4.

Rice NA, Nadeau OW, Yang Q, Carlson GM: The calmodulin-binding domain of the catalytic gamma subunit of phosphorylase kinase interacts with its inhibitory alpha subunit: evidence for a Ca2+ sensitive network of quaternary interactions. J Biol Chem. 2002 Apr 26;277(17):14681-7. Epub 2002 Feb 14.

Riento K, Ridley AJ: Rocks: multifunctional kinases in cell behaviour. Nat Rev Mol Cell Biol. 2003 Jun;4(6):446-56.

Roelink, H., Porter, J.A., Chiang, C., Tanabe, Y., Chang, D.T., Beachy, P.A., Jessell, T.M. (1995). Floor plate and motor neuron induction by different concentrations of the amino-terminal cleavage product of sonic hedgehog autoproteolysis. Cell 81, 445-455.

Roy F, Therrien M: MAP kinase module: the Ksr connection. Curr Biol. 2002 Apr 30;12(9):R325-7.

Van Roijen JH, Van Assen S, Van Der Kwast TH, De Rooij DG, Boersma WJ, Vreeburg JT, Weber RF: Androgen receptor immunoexpression in the testes of subfertile men.

J Androl. 1995 Nov-Dec;16(6):510-6.

Rossi P, Dolci S, Albanesi C, Grimaldi P, Geremia R: Direct evidence that the mouse sex-determining gene Sry is expressed in the somatic cells of male fetal gonads and in the germ cell line in the adult testis. Mol Reprod Dev. 1993 Apr;34(4):369-73.

Rossi P, Dolci S, Albanesi C, Grimaldi P, Ricca R, Geremia R: 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. 1993 Jan;155(1):68-74.

Ryu JH, Essner R, Ohta T, Kuriyama R: Filamentous polymers induced by overexpression of a novel centrosomal protein, Cep135. Microsc Res Tech. 2000 Jun 1;49(5):478-86.

Saarma M, Sariola H: Other neurotrophic factors: glial cell line-derived neurotrophic factor (GDNF). Microsc Res Tech. 1999 May 15-Jun 1;45(4-5):292-302.

Sakai Y, Koyama Y, Fujimoto H, Nakamoto T, Yamashina S: Immunocytochemical study on fibrous sheath formation in mouse spermiogenesis using a monoclonal antibody.
Anat Rec. 1986 Jun;215(2):119-26.

Sariola H, Meng X: GDNF-induced seminomatous tumours in mouse--an experimental model for human seminomas? APMIS. 2003 Jan;111(1):192-6.

Sassone-Corsi P.: Transcriptional checkpoints determining the fate of male germ cells. Cell. 1997 Jan 24;88(2):163-6. Review.

Sassone-Corsi P: Molecular clocks. PERpetuating the PASt. Nature. 1997 Oct 2;389(6650):443-4.

Schalles U, Shao X, van der Hoorn FA, Oko R: Developmental expression of the 84-kDa ODF sperm protein: localization to both the cortex and medulla of outer dense fibers and to the connecting piece. Dev Biol. 1998 Jul 15;199(2):250-60.

Schlecht U, Demougin P, Koch R, Hermida L, Wiederkehr C, Descombes P, Pineau C, Jegou B, Primig M (2004) Expression profiling of Mammalian male meiosis and gametogenesis identifies novel candidate genes for roles in the regulation of fertility. Mol Biol Cell, *15*, 1031-43.

Schmidt EE, Schibler U: Developmental testis-specific regulation of mRNA levels and mRNA translational efficiencies for TATA-binding protein mRNA isoforms. Dev Biol. 1997 Apr 1;184(1):138-49.

Schneider H, Brueckner M: Of mice and men: dissecting the genetic pathway that controls left-right asymmetry in mice and humans. Am J Med Genet. 2000 Winter;97(4):258-70. Review.

Scmidt M.C., Zhou Q, Berk AJ: Spl Activates Transcription without Enhancing DNA-Binding Activity of the TATA Box Factor. Mol. Cell. Biol. 9:3299-3307 1989.

Schmitz AA, Govek EE, Bottner B, Van Aelst L: Rho GTPases: signaling, migration, and invasion. Exp Cell Res. 2000 Nov 25;261(1):1-12.

Schroer TA, Bingham JB, Gill SR: Actin-related protein 1 and cytoplasmic dynein-based motility - what's the connection? Trends Cell Biol. 1996 Jun;6(6):212-5.

Schweizer A, et al.: Neonatal lethality in mice deficient in XCE, a novel member of the endothelin-converting enzyme and neutral endopeptidase family. J Biol Chem 1999 Jul 16;274(29):20450-6

Seabra MC, Mules EH, Hume AN: Rab GTPases, intracellular traffic and disease. Trends Mol Med. 2002 Jan;8(1):23-30.

Seo HR, Kwan YW, Cho CK, Bae S, Lee SJ, Soh JW, Chung HY, Lee YS: PKCalpha induces differentiation through ERK1/2 phosphorylation in mouse keratinocytes. Exp Mol Med. 2004 Aug 31;36(4):292-9.

Shao X, Tarnasky HA, Lee JP, Oko R, van der Hoorn FA: Spag4, a novel sperm protein, binds outer dense-fiber protein Odf1 and localizes to microtubules of manchette and axoneme. Dev Biol. 1999 Jul 1;211(1):109-23.

Shen WH, Moore CC, Ikeda Y, Parker KL, Ingraham HA.: Nuclear receptor steroidogenic factor 1 regulates the mullerian inhibiting substance gene: a link to the sex determination cascade. Cell. 1994 Jun 3;77(5):651-61.

Shimamura, K., Hartigan, D.J., Martinez, S., Puelles, L., Rubenstein, J.L.R. (1995). Longitudinal organization of the anterior neural plate and neural tube. Development 121, 3923-3933.

Shimizu K, Nagamachi Y, Tani M, Kimura K, Shiroishi T, Wakana S, Yokota J. Molecular cloning of a novel NF2/ERM/4.1 superfamily gene, ehm2, that is expressed in high-metastatic K1735 murine melanoma cells. Genomics 2000; 65: 113-120.

Si Y and Olds-Clarke: Evidence for the involvement of calmodulin in mouse sperm capacitation. Biol Reprod. 2000 May;62(5):1231-9., 2000

Si Y, Okuno M: Extrusion of microtubule doublet outer dense fibers 5-6 associating with fibrous sheath sliding in mouse sperm flagella. J Exp Zool. 1995 Nov 1;273(4):355-62.

Si Y, Okuno M: Activation of mammalian sperm motility by regulation of microtubule sliding via cyclic adenosine 5'-monophosphate-dependent phosphorylation. Biol Reprod. 1995 Nov;53(5):1081-7.

Si Y, Okuno M: Multiple activation of mouse sperm motility. Mol Reprod Dev. 1993 Sep;36(1):89-95.

Si Y, Okuno M: The sliding of the fibrous sheath through the axoneme proximally together with microtubule extrusion. Exp Cell Res. 1993 Sep;208(1):170-4.

Sibony M, Segretain D, Gasc JM: Angiotensin-converting enzyme in murine testis: step-specific expression of the germinal isoform during spermiogenesis. Biol Reprod. 1994 May;50(5):1015-26. Erratum in: Biol Reprod 1994 Aug;51(2):345.

Sibony M, Segretain D, Gasc JM.: Angiotensin-converting enzyme in murine testis: step-specific expression of the germinal isoform during spermiogenesis. Biol Reprod. 1994 May;50(5):1015-26.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN: A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature. 1990 Jul 19;346(6281):240-4.

Sinisi AA, Pasquali D, Notaro A, Bellastella A: Sexual differentiation. J Endocrinol Invest. 2003;26(3 Suppl):23-8.

Smith KM, Gaultier A., Cousin H., Alfandari D., White JM., and DeSimone DW. 2002. Disintegrin and cystine-rich domains regulate ADAM protease activity. J Cell Biol. 159(5): 893-902.

Smith CA, McClive PJ, Western PS, Reed KJ, Sinclair AH: Conservation of a sex-determining gene. Nature. 1999 Dec 9;402(6762):601-2.

Sorrentino V, Giorgi M, Geremia R, Besmer P, Rossi P: Expression of the c-kit proto-oncogene in the murine male germ cells. Oncogene. 1991 Jan;6(1):149-51.

Stenmark H, Vitale G, Ullrich O, Zerial M.: Rabaptin-5 is a direct effector of the small GTPase Rab5 in endocytic membrane fusion. Cell. 1995 Nov 3;83(3):423-32.

Stetler-Stevenson WG: The role of matrix metalloproteinases in tumor invasion, metastasis, and angiogenesis. Surg Oncol Clin N Am. 2001 Apr;10(2):383-92, x.

Stetler-Stevenson WG, Yu AE: Proteases in invasion: matrix metalloproteinases. Semin Cancer Biol. 2001 Apr;11(2):143-52.

Stockinger W, Brandes C, Fasching D, Hermann M, Gotthardt M, Herz J, Schneider WJ, Nimpf J: The reelin receptor ApoER2 recruits JNK-interacting proteins-1 and -2. J Biol Chem. 2000 Aug 18;275(33):25625-32.

Sun,W.D. and Chantler,P.D.: A unique cellular myosin II exhibiting differential expression in the cerebral cortex. J Biochem. Biophys. Res. Commun. 175 (1), 244-249 (1991)

Sun,W. and Chantler,P.D.: Cloning of the cDNA encoding a neuronal myosin heavy chain from mammalian brain and its differential expression within the central nervous system. J. Mol. Biol. 224 (4), 1185-1193 (1992).

Suh KS, Tatunchak TT, Crutchley JM, Edwards LE, Marin KG, Yuspa SH: Genomic structure and promoter analysis of PKC-delta. Genomics. 2003 Jul;82(1):57-67.

Surace EI, Haipek CA, Gutmann DH: Effect of merlin phosphorylation on neurofibromatosis 2 (NF2) gene function. Oncogene. 2004 Jan 15;23(2):580-7.

Sutton KA, Jungnickel MK, Wang Y, Cullen K, Lambert S, Florman HM: Enkurin is a novel calmodulin and TRPC channel binding protein in sperm. Dev Biol. 2004 Oct 15;274(2):426-35.

Swain A, Lovell-Badge R: A molecular approach to sex determination in mammals. Acta Paediatr Suppl. 1997 Nov;423:46-9.

Swain A, Narvaez V, Burgoyne P, Camerino G, Lovell-Badge R: Dax1 antagonizes Sry action in mammalian sex determination. Nature. 1998 Feb 19;391(6669):761-7.

Symolon H, Schmelz EM, Dillehay DL, Merrill AH Jr: Dietary soy sphingolipids suppress tumorigenesis and gene expression in 1,2-dimethylhydrazine-treated CF1 mice and ApcMin/+ mice. J Nutr. 2004 May;134(5):1157-61.

Tadokoro Y, Yomogida K, Ohta H, Tohda A, Nishimune Y: Homeostatic regulation of germinal stem cell proliferation by the GDNF/FSH pathway. Mech Dev. 2002 Apr;113(1):29-39.

Takeuchi K, Sato N, Kasahara H, Funayama N, Nagafuchi A, Yonemura S, Tsukita S, Tsukita S. Perturbation of cell adhesion and microvilli formation by antisense oligonucleotides to ERM family members. J. Cell Biol. 1994; 125: 1371-1384

Tamai KT, Monaco L, Nantel F, Zazopoulos E, Sassone-Corsi P.: Coupling signalling pathways to transcriptional control: nuclear factors responsive to cAMP. Recent Prog Horm Res. 1997;52:121-39; discussion 139-40.

Tafuri SR, Familari M, Wolffe AP: A mouse Y box protein, MSY1, is associated with paternal mRNA in spermatocytes. J Biol Chem. 1993 Jun 5;268(16):12213-20.

Tanaka R, Ono T, Sato S, Nakada T, Koizumi F, Hasegawa K, Nakagawa K, Okumura H, Yamashita T, Ohtsuka M, Asagoe K, Yamasaki O, Noguchi Y, Iwatsuki K, Nakayama E: Over-expression of the testis-specific gene TSGA10 in cancers and its immunogenicity. Microbiol Immunol. 2004;48(4):339-45.

Tanoue T, Nishida E: Molecular recognitions in the MAP kinase cascades. Cell Signal. 2003 May;15(5):455-62.

Tanoue T, Nishida E: Docking interactions in the mitogen-activated protein kinase cascades. Pharmacol Ther. 2002 Feb-Mar;93(2-3):193-202. Review

Tash JS, Bracho GE: Identification of phosphoproteins coupled to initiation of motility in live epididymal mouse sperm. Biochem Biophys Res Commun. 1998 Oct 20;251(2):557-63.

Tash JS: Protein phosphorylation: the second messenger signal transducer of flagellar motility. Cell Motil Cytoskeleton. 1989;14(3):332-9.

Tatsura H, Nagao H, Tamada A, Sasaki S, Kohri K, Mori K: Developing germ cells in mouse testis express pheromone receptors. FEBS Lett. 2001 Jan 19;488(3):139-44.

Tokunou M, Niki T, Saitoh Y, Imamura H, Sakamoto M, Hirohashi S. Altered expression of the ERM proteins in lung adenocarcinoma. Lab Invest 2000; 80: 1643-1650.

Tres LL, Kierszenbaum AL: Sak57, an acidic keratin initially present in the spermatid manchette before becoming a component of paraaxonemal structures of the developing tail. Mol Reprod Dev. 1996 Jul;44(3):395-407.

Turner RM.: Tales from the tail: what do we really know about sperm motility? J Androl. 2003 Nov-Dec;24(6):790-803.

Ueda S, Mizuki M, Ikeda H, Tsujimura T, Matsumura I, Nakano K, Daino H, Honda Zi Z, Sonoyama J, Shibayama H, Sugahara H, Machii T, Kanakura Y.: Critical roles of c-Kit tyrosine residues 567 and 719 in stem cell factor-induced chemotaxis: contribution of src family kinase and PI3-kinase on calcium mobilization and cell migration. Blood. 2002 May 1;99(9):3342-9.

Uetake Y, Terada Y, Matuliene J, Kuriyama R: Interaction of Cep135 with a p50 dynactin subunit in mammalian centrosomes. Cell Motil Cytoskeleton. 2004 May;58(1):53-66.

Um JY, Choi BM, Kim JS, Rim JS, Kim HM, Chung HT: Expression of protein kinase C delta gene in germ cells. J Urol. 1995 Sep;154(3):1237-40.

Vanhaesebroeck B, Alessi DR.: The PI3K-PDK1 connection: more than just a road to PKB. Biochem J. 2000 Mar 15;346 Pt 3:561-76.

Van Roijen JH, Van Assen S, Van Der Kwast TH, De Rooij DG, Boersma WJ, Vreeburg JT, Weber RF.: Androgen receptor immunoexpression in the testes of subfertile men. J Androl. 1995 Nov-Dec;16(6):510-6.

Veldhuisen B. , Saris J.J. , De haij S. , Hayashi T. , Reynolds D.M. , Mochizuki T. , Elles R. , Fossdal R. , Bogdanova N. , Van dijk M.A. , Coto E. , Ravine D. , Nørby S. , Verellen-dumoulin C. , Breuning M.H. , Somlo S. , Peters D.J.: A spectrum of mutations in the second gene for autosomal dominant polycystic kidney disease (PKD2). Am. J. Hum. Genet. 61: 547- 555 (1997)

Ventela S, Mulari M, Okabe M, Tanaka H, Nishimune Y, Toppari J, Parvinen M: Regulation of acrosome formation in mice expressing green fluorescent protein as a marker. Tissue Cell. 2000 Dec;32(6):501-7.

Vereb M, Agulnik AI, Houston JT, Lipschultz LI, Lamb DJ, Bishop CE.: Absence of DAZ gene mutations in cases of non-obstructed azoospermia. Mol Hum Reprod. 1997 Jan;3(1):55-9.

Verhey KJ, Meyer D, Deehan R, Blenis J, Schnapp BJ, Rapoport TA, Margolis B: Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. J Cell Biol. 2001 Mar 5;152(5):959-70.

Viard I, Wehrli P, Jornot L, Bullani R, Vechietti JL, Schifferli JA, Tschopp J, French LE: Clusterin gene expression mediates resistance to apoptotic cell death induced by heat shock and oxidative stress. J Invest Dermatol. 1999 Mar;112(3):290-6.

Vojtek AB, Der CJ: Increasing complexity of the Ras signaling pathway. J Biol Chem. 1998 Aug 7;273(32):19925-8.

Vogt PH: Molecular genetics of human male infertility: from genes to new therapeutic perspectives. Curr Pharm Des. 2004;10(5):471-500. Review.

Vyssokikh M, Zorova L, Zorov D, Heimlich G, Jurgensmeier J, Schreiner D, Brdiczka D: The intra-mitochondrial cytochrome c distribution varies correlated to the formation of a complex between VDAC and the adenine nucleotide translocase: this affects Bax-dependent cytochrome c release. Biochim Biophys Acta. 2004 Feb 2;1644(1):27-36.

Wehling M: Specific, nongenomic actions of steroid hormones. Annu Rev Physiol. 1997;59:365-93.

Welch JE, Brown PR, O'Brien DA, Eddy EM.: Genomic organization of a mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells. Dev Genet. 1995;16(2):179-89.

Weil D, Blanchard S, Kaplan J, Guilford P, Gibson F, Walsh J, Mburu P, Varela A, Levilliers J, Weston MD, et al: Defective myosin VIIA gene responsible for Usher syndrome type 1B.

Whitmarsh AJ, Kuan CY, Kennedy NJ, Kelkar N, Haydar TF, Mordes JP, Appel M, Rossini AA, Jones SN, Flavell RA, Rakic P, Davis RJ: Requirement of the JIP1 scaffold protein for stress-induced JNK activation. Genes Dev. 2001 Sep 15;15(18):2421-32.

Williams, J. B.; Rexer, B.; Sirripurapu, S.; John, S.; Goldstein, R.; Phillips, J. A., III; Haley, L. L.; Sait, S. N. J.; Shows, T. B.; Smith, C. M.; Gerhard, D. S.: The human HNP36 gene is localized to chromosome 11q13 and produces alternative transcripts that are not mutated in multiple endocrine neoplasia, type 1 (MEN I) syndrome. *Genomics* 42: 325-330, 1997.

Woo JH, Lim JH, Kim YH, Suh SI, Min do S, Chang JS, Lee YH, Park JW, Kwon TK: Resveratrol inhibits phorbol myristate acetate-induced matrix metalloproteinase-9 expression by inhibiting JNK and PKC delta signal transduction. Oncogene. 2004 Mar 11;23(10):1845-53.

Worby CA, Dixon JE: Sorting out the cellular functions of sorting nexins. Nature reviews, Molecular Cell Biology. 2002 Dec. 3: 919-931

Wu X, Wang F, Rao K, Sellers JR, Hammer JA: Rab27a is an essential component of melanosome receptor for myosin Va. Mol Biol Cell. 2002 May;13(5):1735-49.

Wu XS, Rao K, Zhang H, Wang F, Sellers JR, Matesic LE, Copeland NG, Jenkins NA, Hammer JA: Identification of an organelle receptor for myosin-Va. Nat Cell Biol. 2002 Apr;4(4):271-8.
Xu W, Cooper GM.: Identification of a candidate c-mos repressor that restricts transcription of germ cell-specific genes. Mol Cell Biol. 1995 Oct;15(10):5369-75.

Yamada, T., Pfaff, S.L., Edlund, T., and Jessell, T.M. (1993). Control of cell pattern in the neural tube: motor neuron induction by diffusible factors from notochord and floor plate. Cell 73, 673-686.

Yang J, Chennathukuzhi V, Miki K, O'Brien DA, Hecht NB: Mouse testis brain RNA-binding protein/translin selectively binds to the messenger RNA of the fibrous sheath protein glyceraldehyde 3-phosphate dehydrogenase-S and suppresses its translation in vitro. Biol Reprod. 2003 Mar;68(3):853-9.

Yomogida K, Yagura Y, Tadokoro Y, Nishimune Y: Dramatic expansion of germinal stem cells by ectopically expressed human glial cell line-derived neurotrophic factor in mouse Sertoli cells. Biol Reprod. 2003 Oct;69(4):1303-7. Epub 2003 Jun 11.

Yoshida K, Oshima H: The regulation of testicular function by paracrine mechanism. Nippon Naibunpi Gakkai Zasshi. 1994 Dec 20;70(10):1047-54.

Yoshida A, Minowa MT, Takamatsu S, Hara T, Oguri S, Ikenaga H, Takeuchi M. Tissue specific expression and chromosomal mapping of a human UDP-N-acetylglucosamine: alpha1,3-d-mannoside beta1, 4-N-acetylglucosaminyltransferase. Glycobiology 1999; 9: 303-310.

Zhang et al., 2004 SPAM1

Zhang L, Fu Z, Binkley C, Giordano T, Burant CF, Logsdon CD, Simeone DM: Raf kinase inhibitory protein inhibits beta-cell proliferation. Surgery. 2004 Sep;136(3):708-15.

Zhang SS, Carrillo AJ, Darling DS.: Expression of multiple thyroid hormone receptor mRNAs in human oocytes, cumulus cells, and granulosa cells. Mol Hum Reprod. 1997 Jul;3(7):555-62.

Zhong, H., DeMarzo, A. M., Laughner, E., Lim, M., Hilton, D. A., Zagzag, D., Buechler, P., Isaacs, W. B., Semenza, G. L. & Simons, J. W.: Overexpression of hypoxia-inducible factor 1-alpha in common human cancers and their metastases. Cancer Res 59:5830-5835 1999.

Zhou W, Goldberg E.: A dual-function palindromic sequence regulates testis-specific transcription of the mouse lactate dehydrogenase c gene in vitro. Biol Reprod. 1996 Jan;54(1):84-90.

Zhu D, Dix DJ, Eddy EM.: HSP70-2 is required for CDC2 kinase activity in meiosis I of mouse spermatocytes. Development. 1997 Aug;124(15):3007-14.

Zinkham WH, Kupchyk L, Blanco A, Isensee H: Polymorphism of lactate dehydrogenase isozymes in pigeons. Nature. 1965 Oct 16;208(7):284-6.

Appendices

Appendix A: Sequences, alignments and sequences analysis

The sequences of the inserts that encodes activating protein via AD with Tsga10 bait construct in yeast two hybrid assay

A1

CGTGGGCGGGGGGAATTGTGGACACAAAGAGAGTACGCATTAACCGCGGGGATAGCGCTTG
CTAGTCCGCCAGAGCATACATGAATACGTAGATAACTGCAAAAACACACACGCCAAGTAAA
AAAGCCTACTTCTTTATTAAGCAAAGTGTGTAAAGGTTGCCATAAAGGACAAACAAGTGAA
GTATTTTATCCATTACAGTGGCTGGAATAAAAATTGGGATGAATGGGTGCCAGAAAGCAGA
GTGCTCAAATACGTGGACGCCAATTTGCAGAAACAGCGGGAACTTCAAAAGGCCAACCAGG
AACAATATGCAGAGGGCAAGATGAGAGGGGCTGCTCCAGGGAAGAAGACATCCGGTCTGC
AACAGAAAAATGTTGAAGTGAAAACAAAAAAGAACAAACAGAAAACACCTGGAAACGGAG
ATGGTGGCAGTACCAGTGAGACCCCTCAGCCTCCTAGGAAGAAGCGAGCCCGGGTAGATCC
AACCGTTGAAAATGAAGAAACATTCATGAACAGAGTTGAAGTTAAAGTGAAGATTCCTGAG
GAGCTGAAACCATGGCTTGTAGATGACTGGGACTTGATCACCAGACAAAAGCAGCTTTTTTA
TCTACCTGCCAAGAAGAATGTGGATTCCATATTGGAGGATTATGCAAACTATAAGAAGTCTC
GAGGAAATACAGATAATAAGGAGTATGCTGTTAATGAGGTTGTGGCAGGGATAAAGGAGTA
CTTCAATGTAATGTCGGGGCACTCAGCTACTCTACAAAATATGAGAGACCACAGNATGCTGA
AATTTCTGGCCGATCACCCGGGATGCACCCCATGTTCCCAGGGTGTATTGGAAGCGCCCACA
CTTAACCTGAGAAATTGTTTTG

A'6

GGGAGGGGTTGACAAAAGAGATCGAATTACCCGGGGATCCGTCGATCCTGCAGAGATCTAT
GAATCGTAGATACTGAAAAACCCCGCCAAGTAACACAAGCGTGGAGGTCCAGGATGGCCCT
AATGAGGATGGGAATATGTTCATGCGGCCAGGGAAGCTGTCTGACTATTTTCCAAAACCATA
CCCCAACCCTGAGGCTGCAAGAGCTGCTAACAATGGAGCTTTACCCCCTGACCTCAGCTACA
TCGTTCGAGCAAGGCACGGTGGTGAGGACTATGTGTTTTCCTTGCTCACTGGCTACTGTGAA
CCTCCCACTGGGGTGTCATTGCGAGAAGGCCTCTATTTCAACCCTTACTTTCCTGGCCAGGCC
ATTGGCATGGACTCCTCCCATCTACACAGAAGTCTTGGAGTATGACGATGGTACCCCAGCTA
CCATGTCACAAGTAGCCAAGGATGTTGCCACCTTCCTTCGCTGGGCATCAGAGCCAGAGCAT
GACCATCGAAAACGCATGGGACTCAAGATGTAGTTGATGATGGGCTTGCTGCTGCCCCTGAC
CTATGCCATGAAGCGGCATAAGTGGTCAGTCCTGAAGAGTCGAAAGCTGGCTTATCGGCCA
CCCAAGTGACCCTGTTCAGTATCTGCTTGTCATCTTGCCTGAACAAGCTCTCAGCAGACCAG
GAACAGTCCGAGGGCCTTTCAGGCCTGGCCCCCTTTTCATCAGATACCAGGGGCCAGGGAG
ACCAGACTCTGGCCTTCTGATCCTCCCTTAGCCCTTCATTGTGGAATTAANTAGGTTCTCAAA
CAAAAAAAAAAAAGTCACCGCGAATTTTGATTGTTATCGGAACCCAAGTAAAAAAAAA

F2

AGGNTTACAAAAAGAGATCGAATACCAGTGTGACGAATTCACCGGTGAATTCGTCACCTGG
GCTGGAGAAAAACCTGGCGACGGCGGCACGGAACCTGGAGGCCCTGGAGGACGACCTGGA
CTTCCTGCGGGACCAGTTCACCACCACGGAAGTCAACATGGCGCGCGTTTACAACTGGGACG
TGAAGCGCAGGAACCGGGACGATGGCGGCGCCGCCAAGAGCAAAGCGTGAGGGCGCCGAC
CACTGCGCTCGCCCGCCCGATGTCCGTCTCACGCCTGAACATTAATGTATAGCTCACCGGTTT
AATGTATATCACACCCGATTTGTTTTTCCCTCCCTCCCCTCCCAACCCCACCTTTTTTCTTTAC
AACTGGAGAAATGATTGTTTTCTTTCAAGAAAAACCACAGTGGATTCATCTACATTAAAAAG
TACTCTGAATAAATCACCTATACCTAGAGTGACAATTGCCAACCAAATGCCTAATAAAATAT
ATNCTTGAGCCC

F3

GTTNAACAAAAGAGATCGAATTACCAGTGTGACGAATTCACCAAGACAGCGTCTGAGCTCT
CCAAGTCCATGGAGTCCATGCGGGGGCATTTGCAGGCACAGCTTCGCTGCAAAGAGGCTGA
GAACAGCCGCCTGTGCATGCAGATCAAGAATTTAGAGCGCAGTGGGAACCAGCACAAGGCG
GAAGTAGAGGCCATCATGGAGCAGCTAAAGGAACTGAAGCAAAAGGGAGACCGAGACAAA
GAGACCCTGAAGAAGGCCATCCGAGCCCAGAAGGAGCGAGCTGAGAAGAGTGAGGAGTAT
GCCGAGCAGCTACATGTGCAGCTGGCCGACAAGGACCTTTATGTTGCTGAAGCTTTATCTAC
TCTGGAGTCATGGAGGAGCCGTTACAACCAAGTTGTGAAAGACAAAGGAGACCTTGAGTTG
GAGATCATTGTCCTAAATGACCGGGTGACAGATCTTGTAAACCAACAACAGAGCTTGGAGG
AGAAGATGCGGGAAGACCGGGACAGCTTGGTAGAGAGACTGCACCGGCAAACTGCTGAGT
ATTCTGCCTTCAAGCTAGAGAACGAGAGGCTCAAGGCTAGCTTTGCCCCTATGGAGGACAA
GCTCATTCAAGCTCCCTTGGAGGTCCAGCACTGAAAGCTTCATGANAACTACAAGGGATATC
CACACCTCAACCAGTATAACAATTGAGCGACAATTTGTTTTTCATAACCCCACNAAACAATC

F5

AAATTGCCAGTGTGACGAATTGCACCCCCAGAGGTGAAATGCAGGAGAAGAAGCAGTAGAG
GACTGTACTCCATCGAAGTGTCTGGAACTTTTCTGACAGCCAGAGACAGTCTTACTGGCTGT
CCTAGAACTGGATATATAGACCAAGCTGGCCTCAAACTCACAGAGATCCAACCCAGAGCAC
TGCGATTAAAGGTGTGGCCACCACACCTGGCTTAATGGATTTTTTTATATGAATGTACCTGTT
CATCCAGGGGAGAAAGTATTGTTTGTATCCATCTCAGTGTCCAGAGAGGAGTGTGAGTACTC
TGCACCATCACATACCACCATTCCTCAGTGCTACGGGTTCCACAGACATCACTTCTGGGTTG
CAGACATCACTGGAAGAATGTTACAAACCTTTTAAACATGGAATTGAGTTGATTTTAAGTAA
ACTTATTTGTGTACTGAAAAAAACCTGCAGAGATCTATGAATCGTAGATACTGAAAAACCCC
GCAAGTAAACAAAAA

L1

AANAAATGCAAAAAGAGATCGAATGCCAGATGTGACGAATTCTCCGGTGAATCGTCACACT
GGAATTCCCGGGGATCCGTCGACCTGCAGAGATCTATGAATCGTAGATACTGAAAAACCCC
GCAAGTACACAACA

L3

GGGGTTACAAAAAGAGATCGAATACCAGGACGAATTCACTCGAGCTGGCAGGTAGGGGTGG
GAGAGGACAACATGCCAATATGCTGAACACTGACCAACTGTTGGTTAGGGCCATATCGGAG
AGACCAAGGAAGAGCACAGAGCCCTTGGTTCCAGGGACACCTACAGGACTGTTCAGCCTTC
TATCAAACATCTCTCCGGAGGAGCAGGGCCGCCTTGGGAGCGGAGACAGTCTCCAGAGTCA
GAGCTGCCAGCAGCAAAGGTCTTACAGCGCAGGGCAGACAACAAAGAAAGAACGAAAACC
CCGGAGGCGCAACAAGAAAGGACGTGGCTCAGCTGAGGCAGAGGATCTCTTCAGTTCACCT
CGGAAGCCTTCCTTCCCTTTCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGCCAGGCTCT
GCTTCAGTCAGGCTCCTCTGTGGCCGTGGGCCATCGATCCTTGCTCTTCCCTTCAGCAGCCCC
CCAGTGCAAGACCAGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTTCGAGCCTGCCAC
AAGTCGGAAGTGCAAAACCTGGGAAGGAGATATCAGCCTGGGGCCTGGGCCAACCTTTCCC
TGCCTCTGGGCAAGGCTGAGAGCTATTGGCTATATCGACCCACCTTCTGGGTCAACAGGGAA
TTGGTCATTGGTCTTAGTGTGATGATGTCTCAGAGGTAGAAGGCCAGAATGCTGACCAGGCT
AATATATGTCTCAGCACTGGGGAACTGACCAGCTCCGAGGGCTAGGGTTGATCTTGGATGAT
GATTTGATCTCCAAGGTTATTGAGGAGGATGAACCCTTCCCGACAAAGGGCCATTTGTGTNC
TTGGTATGATTTTGTAAAAACGCCCAAATATTGTGTACCTTCCACTCCTTTGTGGAGGGGCTC
CCCCCCTTCTCNAATTTAAGG

S3

ACTNACAAAAAGAGATCGAAtTACCAGGACGAATTCACCTAGCTGGCAGGTAGGGGTGGGA
GAGGACAACATGCCAATATGCTGAACACTGACCAACTGTTGGTTAGGGCCATATCGGAGAG
ACCAAGGAAGAGCACAGAGCCCTTGGTTCCAGGGACACCTACAGGACTGTTCAGCCTTCTAT
CAAACATCTCTCCGGAGGAGCAGGGCCGCCTTGGGAGCGGAGACAGTCTCCAGAGTCAGAG
CTGCCAGCAGCAAAGGTCTTACAGCGCAGGGCAGACAACAAAGAAAGAACGAAAACCCCG
GAGGCGCAACAAGAAAGGACGTGGCTCAGCTGAGGCAGAGGATCTCTTCAGTTCACCTCGG
AAGCCTTCCTTCCCTTTCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGCCAGGCTCTGCTT
CAGTCAGGCTCCTCTGTGGCCGTGGGCCATCGATCCTTGCTCTTCCCTTCAGCAGCCCCCCAG
TGCAAGACCAGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTTCGAGCCTGCCACAAGT
CGGAAGTGCAAAACCTGGGAAGGAGATATCAGCCTGGGGCCTGGGCCAACCTTTCCCTGCC
TCTGGGCAAGGCTGAGAGCCAAGGGCTAGAGCGACCCACCTTCTGGTCAACAGGGAAAGGG

TCAGGGTCAGAGTGTGAGGATGTCTCAGAGGTAGAAGGCCAGAATGCTGACGAGGCTgAGG
AAAAGTCTCACCACTGGGGAACTGcCCCAGCTCCCAGGGCAGGGTTGACCTGGAGGAGAGT
GATCTCGAGGTATGAGAGAGACCAACGGAAAAGCACTTTTCCCAGGAGGATTTGAAAAGGC
AAAAA

U14

GGGGGTGTGAAACAAAAGGAGGATGCAATACCGCGGGGATCGATTGTACCGCCAGAGGAT
ACTGATGAATCGTCAGATACTGAAAAACACCCGGCGACGCTAAAGAAGCGGGAAGATGGG
GAACCACAGACCAAGCAGCTCAGAGAAGAGGAAGAGGAAGTTGGGAAGCACAGGGGCCTC
AGGCTTCGGAACTATGTCCCAGAGGATGAAGACCTGAAGAGGAGGCGGGTTCCCCAGGCCA
AACCAGTTGCAGTGGAAGAGAAGGTGAAGGAGCAGCTGGAGGCTGCCAAGCCAGAGCCTG
TCATAGAGGAAGTGGACCTGGCCAACCTGGCACCCCGGAAACCCGATTGGGATCTCAAGAG
AGATGTGGCTAAGATTTCTAGAGAAGTTAGAAAAGCGGACCCAGAGAGCCATTGCAGAGCT
GATCCGTGAACGGCTAAAAGGGCAGGAAGACAACCTGGCCTCTGCAGTGGACGCCACCACC
GGACAAGAGGCTTGTGACTCTGACTGAGATGTCTTCTGAGTCCCACCTGTTCAGAACACTCA
TCCTGGGGGAGGATAGTCTCAGCAGGGATGGGGTTAGACTGGCTAGCACTTCTGGGTCTTGA
GAACCGAGCCAACCCACTGAGTCTCAGCTTCTGTGGGCTGTGACCAGTTCATCTCCCAAGGG
TGCCCACCCCCTGAGTGGGTCGAAGCTGAGATCTGTACATATGTATGTTAACGGAGCCTTTT
AACCTTTAGAAATAACCTCCATTTAGTAGAGAAAAAAAAAAAAAAAAAGGTGGACCTGGCG
GAGATCTTTGGANTCTGTAATTCCTGGAAAACCCCCGCAGTACCCAAAGAAA

Y1

TGACAAAAAGAGATCGAATTGCCAGTGTGACGAATTCTACCCCAGGGCTGGCGCGCGGATG
AAAAATGGCGAAATGGTGGGTAGCCGGTGGGAAACCTGACGGGATGGGGCGCGGAGGTGG
CACTGTCGTGGGAGCCTGCTGAGGTGAGCGCAGCTGACCTGGCGACCTCGGGCCGGCGCCT
GACAGCAGACACGGACTCCGAGGGTCGCCGAGGGCCCCGCCGAGAGCCGGAGGCAATGGA
TGAGCAGAGTGTGGAGAGCATTGCTGAGGTTTTCCGATGCTTCATTTGTATGGAGAAATTGC
GGGATGCCCGACTGTGTCCTCACTGCTCCAAGCTCTGTTGTTTCAGCTGTATTAGGCGCTGGC
TGACAGAGCAGAGAGCTCAGTGTCCTCATTGTCGTGCTCCACTCCAGCTACGGGAACTTGTA
AATTGTCGCTGGGCGGAAGAGGTGACACAGCAGCTGGATACTCTCCAGCTGTGTAGTCTCAC
CAAACATGAAGAAAATGAGACAGACAAATGTGAGAATCACCATGAAGAACTTAGTGTGTGT
TGCTGGACTTGTAAGACGTGTATCTGCCATCAGTGTGCACTTTGGGGAGGAATGCATGGTGG
ACACACGTTTAAACCTTTGGCAGAAATCTATGAACAGCATGTCACTAAAGTGAACGAAGAG
GTAGCTAAACTTCGTCGACCTGCAGAGATCTATGATTCGTAGATACTGAATACCCGATTACC
ACAAAGAAAAAA

Tsga10 amino acid sequence showing exons/introns boundaries, the exon 16 (underlined) and two halves of Tsga10 (N,C-termini) which were used in Yeast two hybrid. An 18-aa is common in both halves (labeled in green).

MMRNRSKSPRRPSPTSRAANCDVELLKSTARDREELKCMLEKYERHLAEIQGNVKVLTSE
RDKTFLLYEQAQEEIARLRREMMKSCKSPKSTTAHAILRRVETERDVAFTDLRRMTTERD
SLRERLKIAQETAFNEKAHLEQRIEELECTVHNLDDERMEQMANMTLMKETITTVEKEMK
SLARKAMDTESELGRQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDE
KIDNFTRQNIAQREEISILGATLNDLAKEKECLQACLDKKSENIASLGESLAMKEKTISG
MKNIIAEMEQASRQSTEALIMCEQDISRMR

RQLDETNDELGQIARERD

ILAHENDNLQEQFAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEVNKLKNILKSEESE
NRQIMEQLRKANEDAENWENKARQTEAENNTLKLELITAEAEGNRLKEKVDALNREVEQHL
NAERSYKSQIATLHKSLVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVA
KDQEIEMMENELDSARSEIELLRSQMTNERISMQNLEALLVANRDKEYQSQIALQEKESEI
QLLKEHLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQELRRQNY
SSNAYNLGPMKPNTKCHSPERAHHRSPDRGLDRSLEENLCYRDF

Nucleotide and deduced amino acid sequences of mouse *Tsga10*. Number of nucleotide and amino acid are shown on each line. The complete (and alternative splicing isoforms) sequences of Tsga10 has been deposited in GenBank under Accession No. **AF530050.** Exon 16 is completely spliced out and underlined (Acc. No. **AY618264).** *Tsga10* contains EZR/myosin, ATPase dependent seggeregation and metallopeptidase domains *(Shaded)*. **Myosin Class II/ERM (Ezrin/Radixin/Moesin) domain (aa # 34-648)** 'ESEIQLLKE' is a MapKKK motif **ATPase dependent chromosome segregation domain (aa # 131-493)** **Membrane bound metallopeptidase domain (aa # 318-578)**

gcagacagagattattataagactgaagctcagaatttaagaaaaatgatgagaaataga
A D R D Y Y K T E A Q N L R K M M R N R
tctaagagtccgagacgcccatcgccaacttcccgggctgcaaactgtgatgtagagctt
S K S P R R P S P T S R A A N C D V E L
ttgaagtcaactgcaagagaccgtgaagagctaaagtgtatgctggaaaaatatgaacgt
L K S T A R D R E E L K C M L E K Y E R
catttggcagaaatccagggcaacgtcaaggttctcacatctgagagagacaagaccttc
H L A E I Q G N V K V L T S E R D K T F
ctcctttatgagcaggcacaggaagaaattgctcgacttcgacgagaaatgatgaaaagc
L L Y E Q A Q E E I A R L R R E M M K S
tgtaagtctcctaagtcaaccacagcacatgctattcttcgtcgggtagagacggagaga
C K S P K S T T A H A I L R R V E T E R
gatgtagccttcactgacctacgaagaatgaccacagagcgagacagtctgagagaaagg
D V A F T D L R R M T T E R D S L R E R
ctcaagattgctcaagagacagcgttcaacgagaaggctcacttggaacagcggatagag
L K I A Q E T A F N E K A H L E Q R I E

```
gagctggagtgcacagttcacaaccttgatgatgagcgcatggaacagatggcaaacatg
 E  L  E  C  T  V  H  N  L  D  D  E  R  M  E  Q  M  A  N  M
actttgatgaaggaaaccataaccactgtggaaaaagaaatgaaatcattagcaagaaag
 T  L  M  K  E  T  I  T  T  V  E  K  E  M  K  S  L  A  R  K
gcaatggacaccgagagtgagcttggcagacaaaaagcagagaataattctttgagactt
 A  M  D  T  E  S  E  L  G  R  Q  K  A  E  N  N  S  L  R  L
ttatatgaaaacacagaaaaagatctctctgatactcagcgacatcttgctaagaaaaaa
 L  Y  E  N  T  E  K  D  L  S  D  T  Q  R  H  L  A  K  K  K
tatgagctacagcttactcaggagaaaattatgtgcttggatgaaaaaattgataatttc
 Y  E  L  Q  L  T  Q  E  K  I  M  C  L  D  E  K  I  D  N  F
acgaggcaaaatattgcacagcgagaagaaatcagcattcttggtgcaaccctcaatgac
 T  R  Q  N  I  A  Q  R  E  E  I  S  I  L  G  A  T  L  N  D
ctggctaaagaaaaggaatgcctgcaagcgtgtctggataaaaagtctgagaacattgca
 L  A  K  E  K  E  C  L  Q  A  C  L  D  K  K  S  E  N  I  A
tcccttggcgagagcttggcaatgaaagaaaaaaccatttcaggcatgaagaatatcatt
 S  L  G  E  S  L  A  M  K  E  K  T  I  S  G  M  K  N  I  I
gctgagatggaacaggcatcaagacagtctactgaagccctcattatgtgcgaacaagat
 A  E  M  E  Q  A  S  R  Q  S  T  E  A  L  I  M  C  E  Q  D
atttccaggatgcgccggcagttggacgagacaaatgatgagctgggtcagattgccagg
 I  S  R  M  R  R  Q  L  D  E  T  N  D  E  L  G  Q  I  A  R
gagagagatatcttggctcatgagaatgacaatcttcaagaacagtttgccaaagtcaaa
 E  R  D  I  L  A  H  E  N  D  N  L  Q  E  Q  F  A  K  V  K
caagaaaaccaggcactgtccaaaaaactgaatgatactcataacgaactcagtgacata
 Q  E  N  Q  A  L  S  K  K  L  N  D  T  H  N  E  L  S  D  I
aagcagaaggtccaggacacgaatctggaggttaacaagttgaagaacatattaaagtct
 K  Q  K  V  Q  D  T  N  L  E  V  N  K  L  K  N  I  L  K  S
gaggaatctgagaaccggcaaataatggaacaactccgaaaagccaatgaagatgctgaa
 E  E  S  E  N  R  Q  I  M  E  Q  L  R  K  A  N  E  D  A  E
aactgggaaaataaggcccgccagacagaggcagaaaacaacaccctcaaactggaactt
 N  W  E  N  K  A  R  Q  T  E  A  E  N  N  T  L  K  L  E  L
atcactgctgaagcagagggcaacagattaaaggaaaaagttgatgcccttaacagagag
 I  T  A  E  A  E  G  N  R  L  K  E  K  V  D  A  L  N  R  E
gttgagcaacacctaaacgcagagcgctcttacaaatcccagattgcaactttacacaag
 V  E  Q  H  L  N  A  E  R  S  Y  K  S  Q  I  A  T  L  H  K
tctcttgtgaagatggaggaggagcttcagaaggttcagtttgagaaggtgtctgctctc
 S  L  V  K  M  E  E  E  L  Q  K  V  Q  F  E  K  V  S  A  L
gcagatttgtcttccacaagggaactctgcataaaactcgactcaagcaaagaacttctt
 A  D  L  S  S  T  R  E  L  C  I  K  L  D  S  S  K  E  L  L
aatcgacagctggttgccaaagatcaggaaatagaaatgatggagaatgagctggactca
 N  R  Q  L  V  A  K  D  Q  E  I  E  M  M  E  N  E  L  D  S
gcgcgctctgaaattgaactgctccggagtcagatgacaaacgagaggatctccatgcag
 A  R  S  E  I  E  L  L  R  S  Q  M  T  N  E  R  I  S  M  Q
aatctcgaggctctgctggtggccaaccgggacaaagagtaccagtctcagatagcactg
 N  L  E  A  L  L  V  A  N  R  D  K  E  Y  Q  S  Q  I  A  L
caggaaaaggagtctgagatccagttgctgaaggagcacctctgcctggctgagaacaaa
 Q  E  K  E  S  E  I  Q  L  L  K  E  H  L  C  L  A  E  N  K
atggccatccagagcagagacgtggcacagttcagaaatgttgtaacgcagttagaagca
 M  A  I  Q  S  R  D  V  A  Q  F  R  N  V  V  T  Q  L  E  A
gatttggacattaccaaaaggcaactcgggacagaacgttttgaaagggaaagggctgtc
 D  L  D  I  T  K  R  Q  L  G  T  E  R  F  E  R  E  R  A  V
caagaacttcgccgccagaattactcaagcaatgcctataatttgggtccaatgaagcca
 Q  E  L  R  R  Q  N  Y  S  S  N  A  Y  N  L  G  P  M  K  P
aatacaaaatgccactcaccagagcgtgctcaccatcgatctcctgaccgaggcctcgac
 N  T  K  C  H  S  P  E  R  A  H  H  R  S  P  D  R  G  L  D
cgatcattggaagagaatctttgctatagagatttctgacatgtgaaaagattctccaca
 R  S  L  E  E  N  L  C  Y  R  D  F  -
```

Alignment of mouse *Tsga10* nucleotide sequence

```
                                 .................................................
NM_207228   (     1) ----------GCGCCGGACGGAAGTGGAGCGTAGGCGCCGGGGAGGTGGAAGAAGAGCCC
BC066782    (     1) ----------GCGCCGGACGGAAGTGGAGCGTAGGCGCCGGGGAGGTGGAAGAAGAGCCC
AF530050    (     1) --------------------------------------------------------------
AK030254    (   128) tcgctacgtgctcgctcgggtgcgcgcgcacgatccgtagtctccgcgagagggcgtgtc

                     ......+++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (    51) CGCTCAGGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTTTCTGAAGCTGG
BC066782    (    51) CGCTCAGGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTTTCTGAAGCTGG
AF530050    (     1) ------GGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTTTCTGAAGCTGG
AK030254    (   188) gagaCAGGTTAACACCAAAGAACATCAGTTTGGTCTTGGGGCTCCAGTTTCTGAAGCTGG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   111) AAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGATAGATTCATCTC
BC066782    (   111) AAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGATAGATTCATCTC
AF530050    (    55) AAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGATAGATTCATCTC
AK030254    (   248) AAGAAACCAGAATACTTTCCAACTAGAACAAGAAGTGAGAACCCAAGATAGATTCATCTC

                     ++++++++++ ++++++++++++++++.++.....++++++++++++++++++++++++
NM_207228   (   171) TACACTGAAA-TTACAGATTGAAGATCTCAAACAGACAAATCATGACTTGGAAGAATATG
BC066782    (   171) TACACTGAAA-TTACAGATTGAAGATCTCAAACAGACAAATCATGACTTGGAAGAATATG
AF530050    (   115) TACACTGAAAtTTACAGATTGAAGATCTCAAACAGACAAATCATGACTTGGAAGAATATG
AK030254    (   308) TACACTGAAA-TTACAGATTGAAGATCTC-AAACAGAAAATCATGACTTGGAAGAATATG
                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   230) TTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGACTTAGCCAACC
BC066782    (   230) TTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGACTTAGCCAACC
AF530050    (   175) TTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGACTTAGCCAACC
AK030254    (   366) TTAGGAAACTCTTGGATAGTAAAGAGGCGGTAAGCACTCAAGTAGATGACTTAGCCAACC

                     ++++++++++++++++++++++++++++++++++++++++++..++++++++++++++++
NM_207228   (   290) ACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGAGAAATTACAAA
BC066782    (   290) ACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGAGAAATTACAAA
AF530050    (   235) ACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACATGCAGAGAAATTACAAA
AK030254    (   426) ACAATGAGCACCTTTGTAAAGAGTTGATTAAACTTGACCAACTAGCAGAGAAATTACAAA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   350) AAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGCAAAGATTGAAC
BC066782    (   350) AAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGCAAAGATTGAAC
AF530050    (   295) AAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGCAAAGATTGAAC
AK030254    (   486) AAGAAAAAAATTTTGTGGTGGACACTGCCGACAAGGAACTTGAAGAAGCAAAGATTGAAC

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++.+
NM_207228   (   410) TCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAGGCTTAAGTCCA
BC066782    (   410) TCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAGGCTTAAGTCCA
AF530050    (   355) TCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAGGCTTAAGTCTA
AK030254    (   546) TCATTTGCCAGCAAAATAATATAACAGTATTAGAAGATACAATCCAAAGGCTTAAGTCCA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   470) TAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGATTCCTTTGTCA
BC066782    (   470) TAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGATTCCTTTGTCA
AF530050    (   415) TAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGATTCCTTTGTCA
AK030254    (   606) TAATTTTAGAGACTGAAAAGGCACAAAATACATCTCCTTCTAGACTGGATTCCTTTGTCA

                     ++++++++++++++++++++++++.+...+..........+..   +.++.+..+++++
NM_207228   (   530) AGACTTTAGAAGCAGACAGAGATTATTATAAGACTGAAGCTCAG---AATTTAAGAAAAA
BC066782    (   530) AGACTTTAGAAGCAGACAGAGATTATTATAAGACTGAAGCTCAG---AATTTAAGAAAAA
AF530050    (   475) AGACTTTAGAAGCAGACAGAGATTAgttataagaactgaagctcagaATTTAAGAAAAAA
AK030254    (   666) AGACTTTAGAAGCAGACAGAGATTATTATAAGACTGAAGCTCAG---AATTTAAGAAAAA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   587) TGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCGGGCTGCAAACT
BC066782    (   587) TGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCGGGCTGCAAACT
```

```
AF530050    (   535) TGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCGGGCTGCAAACT
AK030254    (   723) TGATGAGAAATAGATCTAAGAGTCCGAGACGCCCATCGCCAACTTCCCGGGCTGCAAACT

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   647) GTGATGTAGAGCTTTTGAAGTCAACTGCAAGAGACCGTGAAGAGCTAAAGTGTATGCTGG
BC066782    (   647) GTGATGTAGAGCTTTTGAAGTCAACTGCAAGAGACCGTGAAGAGCTAAAGTGTATGCTGG
AF530050    (   595) GTGATGTAGAGCTTTTGAAGTCAACTGCAAGAGACCGTGAAGAGCTAAAGTGTATGCTGG
AK030254    (   783) GTGATGTAGAGCTTTTGAAGTCAACTGCAAGAGACCGTGAAGAGCTAAAGTGTATGCTGG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   707) AAAAATATGAACGTCATTTGGCAGAAATCCAGGGCAACGTCAAGGTTCTCACATCTGAGA
BC066782    (   707) AAAAATATGAACGTCATTTGGCAGAAATCCAGGGCAACGTCAAGGTTCTCACATCTGAGA
AF530050    (   655) AAAAATATGAACGTCATTTGGCAGAAATCCAGGGCAACGTCAAGGTTCTCACATCTGAGA
AK030254    (   843) AAAAATATGAACGTCATTTGGCAGAAATCCAGGGCAACGTCAAGGTTCTCACATCTGAGA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   767) GAGACAAGACCTTCCTCCTTTATGAGCAGGCACAGGAAGAAATTGCTCGACTTCGACGAG
BC066782    (   767) GAGACAAGACCTTCCTCCTTTATGAGCAGGCACAGGAAGAAATTGCTCGACTTCGACGAG
AF530050    (   715) GAGACAAGACCTTCCTCCTTTATGAGCAGGCACAGGAAGAAATTGCTCGACTTCGACGAG
AK030254    (   903) GAGACAAGACCTTCCTCCTTTATGAGCAGGCACAGGAAGAAATTGCTCGACTTCGACGAG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   827) AAATGATGAAAAGCTGTAAGTCTCCTAAGTCAACCACAGCACATGCTATTCTTCGTCGGG
BC066782    (   827) AAATGATGAAAAGCTGTAAGTCTCCTAAGTCAACCACAGCACATGCTATTCTTCGTCGGG
AF530050    (   775) AAATGATGAAAAGCTGTAAGTCTCCTAAGTCAACCACAGCACATGCTATTCTTCGTCGGG
AK030254    (   963) AAATGATGAAAAGCTGTAAGTCTCCTAAGTCAACCACAGCACATGCTATTCTTCGTCGGG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   887) TAGAGACGGAGAGAGATGTAGCCTTCACTGACCTACGAAGAATGACCACAGAGCGAGACA
BC066782    (   887) TAGAGACGGAGAGAGATGTAGCCTTCACTGACCTACGAAGAATGACCACAGAGCGAGACA
AF530050    (   835) TAGAGACGGAGAGAGATGTAGCCTTCACTGACCTACGAAGAATGACCACAGAGCGAGACA
AK030254    (  1023) TAGAGACGGAGAGAGATGTAGCCTTCACTGACCTACGAAGAATGACCACAGAGCGAGACA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (   947) GTCTGAGAGAAAGGCTCAAGATTGCTCAAGAGACAGCGTTCAACGAGAAGGCTCACTTGG
BC066782    (   947) GTCTGAGAGAAAGGCTCAAGATTGCTCAAGAGACAGCGTTCAACGAGAAGGCTCACTTGG
AF530050    (   895) GTCTGAGAGAAAGGCTCAAGATTGCTCAAGAGACAGCGTTCAACGAGAAGGCTCACTTGG
AK030254    (  1083) GTCTGAGAGAAAGGCTCAAGATTGCTCAAGAGACAGCGTTCAACGAGAAGGCTCACTTGG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1007) AACAGCGGATAGAGGAGCTGGAGTGCACAGTTCACAACCTTGATGATGAGCGCATGGAAC
BC066782    (  1007) AACAGCGGATAGAGGAGCTGGAGTGCACAGTTCACAACCTTGATGATGAGCGCATGGAAC
AF530050    (   955) AACAGCGGATAGAGGAGCTGGAGTGCACAGTTCACAACCTTGATGATGAGCGCATGGAAC
AK030254    (  1143) AACAGCGGATAGAGGAGCTGGAGTGCACAGTTCACAACCTTGATGATGAGCGCATGGAAC

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1067) AGATGGCAAACATGACTTTGATGAAGGAAACCATAACCACTGTGGAAAAAGAAATGAAAT
BC066782    (  1067) AGATGGCAAACATGACTTTGATGAAGGAAACCATAACCACTGTGGAAAAAGAAATGAAAT
AF530050    (  1015) AGATGGCAAACATGACTTTGATGAAGGAAACCATAACCACTGTGGAAAAAGAAATGAAAT
AK030254    (  1203) AGATGGCAAACATGACTTTGATGAAGGAAACCATAACCACTGTGGAAAAAGAAATGAAAT

                     ++++++++++++++++++++++++++++++++++++++++++++++.++++++++++++++
NM_207228   (  1127) CATTAGCAAGAAAGGCAATGGACACCGAGAGTGAGCTTGGCAGACAAAAAGCAGAGAATA
BC066782    (  1127) CATTAGCAAGAAAGGCAATGGACACCGAGAGTGAGCTTGGCAGACAAAAAGCAGAGAATA
AF530050    (  1075) CATTAGCAAGAAAGGCAATGGACACCGAGAGTGAGCTTGGCAGACAAAAAGCAGAGAATA
AK030254    (  1263) CATTAGCAAGAAAGGCAATGGACACCGAGAGTGAGCTTGGCAGACGAAAAGCAGAGAATA

                     ++++++++++++++++++++++++++++++++++++++.++++++++++++++++++
NM_207228   (  1187) ATTCTTTGAGACTTTTATATGAAAACACAGAAAAAGATCTCTCTGATACTCAGCGACATC
BC066782    (  1187) ATTCTTTGAGACTTTTATATGAAAACACAGAAAAAGATCTCTCTGATACTCAGCGACATC
AF530050    (  1135) ATTCTTTGAGACTTTTATATGAAAACACAGAAAAAGATCTTTCTGATACTCAGCGACATC
AK030254    (  1323) ATTCTTTGAGACTTTTATATGAAAACACAGAAAAAGATCTCTCTGATACTCAGCGACATC

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1247) TTGCTAAGAAAAAATATGAGCTACAGCTTACTCAGGAGAAAATTATGTGCTTGGATGAAA
BC066782    (  1247) TTGCTAAGAAAAAATATGAGCTACAGCTTACTCAGGAGAAAATTATGTGCTTGGATGAAA
```

```
AF530050    (  1195) TTGCTAAGAAAAAATATGAGCTACAGCTTACTCAGGAGAAAATTATGTGCTTGGATGAAA
AK030254    (  1383) TTGCTAAGAAAAAATATGAGCTACAGCTTACTCAGGAGAAAATTATGTGCTTGGATGAAA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1307) AAATTGATAATTTCACGAGGCAAAATATTGCACAGCGAGAAGAAATCAGCATTCTTGGTG
BC066782    (  1307) AAATTGATAATTTCACGAGGCAAAATATTGCACAGCGAGAAGAAATCAGCATTCTTGGTG
AF530050    (  1255) AAATTGATAATTTCACGAGGCAAAATATTGCACAGCGAGAAGAAATCAGCATTCTTGGTG
AK030254    (  1443) AAATTGATAATTTCACGAGGCAAAATATTGCACAGCGAGAAGAAATCAGCATTCTTGGTG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1367) CAACCCTCAATGACCTGGCTAAAGAAAAGGAATGCCTGCAAGCGTGTCTGGATAAAAAGT
BC066782    (  1367) CAACCCTCAATGACCTGGCTAAAGAAAAGGAATGCCTGCAAGCGTGTCTGGATAAAAAGT
AF530050    (  1315) CAACCCTCAATGACCTGGCTAAAGAAAAGGAATGCCTGCAAGCGTGTCTGGATAAAAAGT
AK030254    (  1503) CAACCCTCAATGACCTGGCTAAAGAAAAGGAATGCCTGCAAGCGTGTCTGGATAAAAAGT

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1427) CTGAGAACATTGCATCCCTTGGCGAGAGCTTGGCAATGAAAGAAAAAACCATTTCAGGCA
BC066782    (  1427) CTGAGAACATTGCATCCCTTGGCGAGAGCTTGGCAATGAAAGAAAAAACCATTTCAGGCA
AF530050    (  1375) CTGAGAACATTGCATCCCTTGGCGAGAGCTTGGCAATGAAAGAAAAAACCATTTCAGGCA
AK030254    (  1563) CTGAGAACATTGCATCCCTTGGCGAGAGCTTGGCAATGAAAGAAAAAACCATTTCAGGCA

                     ++++++++++++++++++++++++++++++++++++++++++++++++.++++++++++
NM_207228   (  1487) TGAAGAATATCATTGCTGAGATGGAACAGGCATCAAGACAGTCTACTGAAGCCCTCATTA
BC066782    (  1487) TGAAGAATATCATTGCTGAGATGGAACAGGCATCAAGACAGTCTACTGAAGCCCTCATTA
AF530050    (  1435) TGAAGAATATCATTGCTGAGATGGAACAGGCATCAAGACAGTCTACTGAAGCCCTCATTA
AK030254    (  1623) TGAAGAATATCATTGCTGAGATGGAACAGGCATCAAGACAGTCTACTGAGGCCCTCATTA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1547) TGTGCGAACAAGATATTTCCAGGATGCGCCGGCAGTTGGACGAGACAAATGATGAGCTGG
BC066782    (  1547) TGTGCGAACAAGATATTTCCAGGATGCGCCGGCAGTTGGACGAGACAAATGATGAGCTGG
AF530050    (  1495) TGTGCGAACAAGATATTTCCAGGATGCGCCGGCAGTTGGACGAGACAAATGATGAGCTGG
AK030254    (  1683) TGTGCGAACAAGATATTTCCAGGATGCGCCGGCAGTTGGACGAGACAAATGATGAGCTGG

                     ++++++++++++++++++++++++++++++++++.+++++..++++++++++++++++
NM_207228   (  1607) GTCAGATTGCCAGGGAGAGAGATATCTTGGCTCATGAGAATGACAATCTTCAAGAACAGT
BC066782    (  1607) GTCAGATTGCCAGGGAGAGAGATATCTTGGCTCATGAGAATGACAATCTTCAAGAACAGT
AF530050    (  1555) GTCAGATTGCCAGGGAGAGAGATATCTTGGCTCATCAGAATACCAATCTTCAAGAACAGT
AK030254    (  1743) GTCAGATTGCCAGGGAGAGAGATATCTTGGCTCATGAGAATGACAATCTTCAAGAACAGT

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1667) TTGCCAAAGTCAAACAAGAAAACCAGGCACTGTCCAAAAAACTGAATGATACTCATAACG
BC066782    (  1667) TTGCCAAAGTCAAACAAGAAAACCAGGCACTGTCCAAAAAACTGAATGATACTCATAACG
AF530050    (  1615) TTGCCAAAGTCAAACAAGAAAACCAGGCACTGTCCAAAAAACTGAATGATACTCATAACG
AK030254    (  1803) TTGCCAAAGTCAAACAAGAAAACCAGGCACTGTCCAAAAAACTGAATGATACTCATAACG

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1727) AACTCAGTGACATAAAGCAGAAGGTCCAGGACACGAATCTGGAGGTTAACAAGTTGAAGA
BC066782    (  1727) AACTCAGTGACATAAAGCAGAAGGTCCAGGACACGAATCTGGAGGTTAACAAGTTGAAGA
AF530050    (  1675) AACTCAGTGACATAAAGCAGAAGGTCCAGGACACGAATCTGGAGGTTAACAAGTTGAAGA
AK030254    (  1863) AACTCAGTGACATAAAGCAGAAGGTCCAGGACACGAATCTGGAGGTTAACAAGTTGAAGA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1787) ACATATTAAAGTCTGAGGAATCTGAGAACCGGCAAATAATGGAACAACTCCGAAAAGCCA
BC066782    (  1787) ACATATTAAAGTCTGAGGAATCTGAGAACCGGCAAATAATGGAACAACTCCGAAAAGCCA
AF530050    (  1735) ACATATTAAAGTCTGAGGAATCTGAGAACCGGCAAATAATGGAACAACTCCGAAAAGCCA
AK030254    (  1923) ACATATTAAAGTCTGAGGAATCTGAGAACCGGCAAATAATGGAACAACTCCGAAAAGCCA

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1847) ATGAAGATGCTGAAAACTGGGAAAATAAGGCCCGCCAGACAGAGGCAGAAAACAACACCC
BC066782    (  1847) ATGAAGATGCTGAAAACTGGGAAAATAAGGCCCGCCAGACAGAGGCAGAAAACAACACCC
AF530050    (  1795) ATGAAGATGCTGAAAACTGGGAAAATAAGGCCCGCCAGACAGAGGCAGAAAACAACACCC
AK030254    (  1983) ATGAAGATGCTGAAAACTGGGAAAATAAGGCCCGCCAGACAGAGGCAGAAAACAACACCC

                     ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228   (  1907) TCAAACTGGAACTTATCACTGCTGAAGCAGAGGGCAACAGATTAAAGGAAAAAGTTGATG
BC066782    (  1907) TCAAACTGGAACTTATCACTGCTGAAGCAGAGGGCAACAGATTAAAGGAAAAAGTTGATG
AF530050    (  1855) TCAAACTGGAACTTATCACTGCTGAAGCAGAGGGCAACAGATTAAAGGAAAAAGTTGATG
AK030254    (  2043) TCAAACTGGAACTTATCACTGCTGAAGCAGAGGGCAACAGATTAAAGGAAAAAGTTGATG
```

```
                    ++++++++++++++++++++++..................................
NM_207228    (  1967) CCCTTAACAGAGAGGTTGAGCAACACCTAAACGCAGAGCGCTCTTACAAATCCCAGATTG
BC066782     (  1967) CCCTTAACAGAGAGGTTGAGCAACACCTAAACGCAGAGCGCTCTTACAAATCCCAGATTG
AF530050     (  1915) CCCTTAACAGAGAGGTTGAGCAACACCTAAACGCAGAGCGCTCTTACAAATCCCAGATTG
AK030254     (  2103) CCCTTAACAGAGAGGTTGAGCAA-----------------------------------

                    ..........................................................
NM_207228    (  2027) CAACTTTACACAAGTCTCTTGTGAAGATGGAGGAGGAGCTTCAGAAGGTTCAGTTTGAGA
BC066782     (  2027) CAACTTTACACAAGTCTCTTGTGAAGATGGAGGAGGAGCTTCAGAAGGTTCAGTTTGAGA
AF530050     (  1975) CAACTTTACACAAGTCTCTTGTGAAGATGGAGGAGGAGCTTCAGAAGGTTCAGTTTGAGA
AK030254     (  2126) ----------------------------------------------------------

                    ..........................................................
NM_207228    (  2087) AGGTGTCTGCTCTCGCAGATTTGTCTTCCACAAGGGAACTCTGCATAAAACTCGACTCAA
BC066782     (  2087) AGGTGTCTGCTCTCGCAGATTTGTCTTCCACAAGGGAACTCTGCATAAAACTCGACTCAA
AF530050     (  2035) AGGTGTCTGCTCTCGCAGATTTGTCTTCCACAAGGGAACTCTGCATAAAACTCGACTCAA
AK030254     (  2126) ----------------------------------------------------------

                    ..........................................................+++++++
NM_207228    (  2147) GCAAAGAACTTCTTAATCGACAGCTGGTTGCCAAAGATCAGGAAATAGAAATGATGGAGA
BC066782     (  2147) GCAAAGAACTTCTTAATCGACAGCTGGTTGCCAAAGATCAGGAAATAGAAATGATGGAGA
AF530050     (  2095) GCAAAGAACTTCTTAATCGACAGCTGGTTGCCAAAGATCAGGAAATAGAAATGATGGAGA
AK030254     (  2126) ----------------------------------------------------aTGGAGA

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2207) ATGAGCTGGACTCAGCGCGCTCTGAAATTGAACTGCTCCGGAGTCAGATGACAAACGAGA
BC066782     (  2207) ATGAGCTGGACTCAGCGCGCTCTGAAATTGAACTGCTCCGGAGTCAGATGACAAACGAGA
AF530050     (  2155) ATGAGCTGGACTCAGCGCGCTCTGAAATTGAACTGCTCCGGAGTCAGATGACAAACGAGA
AK030254     (  2133) ATGAGCTGGACTCAGCGCGCTCTGAAATTGAACTGCTCCGGAGTCAGATGACAAACGAGA

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2267) GGATCTCCATGCAGAATCTCGAGGCTCTGCTGGTGGCCAACCGGGACAAAGAGTACCAGT
BC066782     (  2267) GGATCTCCATGCAGAATCTCGAGGCTCTGCTGGTGGCCAACCGGGACAAAGAGTACCAGT
AF530050     (  2215) GGATCTCCATGCAGAATCTCGAGGCTCTGCTGGTGGCCAACCGGGACAAAGAGTACCAGT
AK030254     (  2193) GGATCTCCATGCAGAATCTCGAGGCTCTGCTGGTGGCCAACCGGGACAAAGAGTACCAGT

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2327) CTCAGATAGCACTGCAGGAAAAGGAGTCTGAGATCCAGTTGCTGAAGGAGCACCTCTGCC
BC066782     (  2327) CTCAGATAGCACTGCAGGAAAAGGAGTCTGAGATCCAGTTGCTGAAGGAGCACCTCTGCC
AF530050     (  2275) CTCAGATAGCACTGCAGGAAAAGGAGTCTGAGATCCAGTTGCTGAAGGAGCACCTCTGCC
AK030254     (  2253) CTCAGATAGCACTGCAGGAAAAGGAGTCTGAGATCCAGTTGCTGAAGGAGCACCTCTGCC

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2387) TGGCTGAGAACAAAATGGCCATCCAGAGCAGAGACGTGGCACAGTTCAGAAATGTTGTAA
BC066782     (  2387) TGGCTGAGAACAAAATGGCCATCCAGAGCAGAGACGTGGCACAGTTCAGAAATGTTGTAA
AF530050     (  2335) TGGCTGAGAACAAAATGGCCATCCAGAGCAGAGACGTGGCACAGTTCAGAAATGTTGTAA
AK030254     (  2313) TGGCTGAGAACAAAATGGCCATCCAGAGCAGAGACGTGGCACAGTTCAGAAATGTTGTAA

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2447) CGCAGTTAGAAGCAGATTTGGACATTACCAAAAGGCAACTCGGGACAGAACGTTTTGAAA
BC066782     (  2447) CGCAGTTAGAAGCAGATTTGGACATTACCAAAAGGCAACTCGGGACAGAACGTTTTGAAA
AF530050     (  2395) CGCAGTTAGAAGCAGATTTGGACATTACCAAAAGGCAACTCGGGACAGAACGTTTTGAAA
AK030254     (  2373) CGCAGTTAGAAGCAGATTTGGACATTACCAAAAGGCAACTCGGGACAGAACGTTTTGAAA

                    ++++++++++++++++...++++++++++++++.++++++++++.++++++++.+
NM_207228    (  2507) GGGAAAGGGCTGTCCAAGAACTTCGCCGCCAGAATTACTCAAGCAATGCCTATAATTTGG
BC066782     (  2507) GGGAAAGGGCTGTCCAAGAACTTCGCCGCCAGAATTACTCAAGCAATGCCTATAATTTGG
AF530050     (  2455) GGGAAAGGGCTGTCCAA---CTTCGCCGCCAGAATTATTCAAGCAATGCTTATAATTTCG
AK030254     (  2433) GGGAAAGGGCTGTCCAAGAACTTCGCCGCCAGAATTACTCAAGCAATGCCTATAATTTGG

                    ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
NM_207228    (  2567) GTCCAATGAAGCCAAATACAAAATGCCACTCACCAGAGCGTGCTCACCATCGATCTCCTG
BC066782     (  2567) GTCCAATGAAGCCAAATACAAAATGCCACTCACCAGAGCGTGCTCACCATCGATCTCCTG
AF530050     (  2512) GTCCAATGAAGCCAAATACAAAATGCCACTCACCAGAGCGTGCTCACCATCGATCTCCTG
AK030254     (  2493) GTCCAATGAAGCCAAATACAAAATGCCACTCACCAGAGCGTGCTCACCATCGATCTCCTG

                    ++++++++++++++++++++++++++++.+..+.+++....++..+.......+.....
NM_207228    (  2627) ACCGAGGCCTCGACCGATCATTGGAAGAGAATCTTTGCTATAGAGATTTCTGACATGTGA
BC066782     (  2627) ACCGAGGCCTCGACCGATCATTGGAAGAGAATCTTTGCTATAGAGATTTCTGACATGTGA
AF530050     (  2572) ACCGAGGCCTCGACCGATCATTGGAAGAGTAAGTGTGCACAGGACCTGGAGTGTAGCCAT
AK030254     (  2553) ACCGAGGCCTCGACCGATCATTGGAAGAGTAAGTGTGCACAGGACCTGGAGTGTAGCCAT
```

```
                    ....+...+.........+.+..+...+++...+.+.....................
NM_207228   (  2687) AAAGATTCTCCACAGCCTTGGAAAGGTCAAAGATACAAATGGATTTTTTTTTTTTTTTT
BC066782    (  2687) AAAGATTCTCCACAGCCTTGGAAAGGTCAAAGATACAAATGGATTTTTTTTTTTTTTTT
AF530050    (  2632) GTTCACAGTGAGATTTGAGGAAGCGAGGAAACTGATA----------------------
AK030254    (  2613) GTTCACAGTGAGATTTGAGGAAGCGAGGAAACTGATA----------------------

                    ............................................................
NM_207228   (  2747) TTTTGCTATGTGATTGCATTTATCTTTTGAATGCTTGACAGGGTTAAATGTATTTATTAA
BC066782    (  2747) TTTTGCTATGTGATTGCATTTATCTTTTGAATGCTTGACAGGGTTAAATGTATTTATTAA
AF530050    (  2669) ------------------------------------------------------------
AK030254    (  2650) ------------------------------------------------------------

                    ..........................................+++.....+.+.+.+.+...
NM_207228   (  2807) CTTTGTGCCTCTGAGTCTGTTCTCGAGTGCCCTAACACAGCTATCTGGGATGATTGATGT
BC066782    (  2807) CTTTGTGCCTCTGAGTCTGTTCTCGAGTGCCCTAACACAGCTATCTGGGATGATTGATGT
AF530050    (  2669) ----------------------------------------CTAAACCTGGTTAATAAGAC
AK030254    (  2650) ----------------------------------------CTAAACCTGGTTAATAAGAC

                    +++...+++..+...+++..+...++++..+.............................
NM_207228   (  2867) ATGCAAACAAAATGCATATGGCTCTTTCCACACAGTAATAGTGAGTGTGAAATCTGCTTA
BC066782    (  2867) ATGCAAACAAAATGCATATGGCTCTTTCCACACAGTAATAGTGAGTGTGAAATCTGCTTA
AF530050    (  2689) ATGATGACACCAAAAATAAAGGGTTTTCATCTataaaaaaaaaa----------------
AK030254    (  2670) ATGATGACACCAAAAATAAAGGGTTTTCATCT----------------------------

                    ............................................................
NM_207228   (  2927) CATCACTATTGAGCACTGGTTTGTTAACTACCCGGAGTAAATAAAGAAGCTTGAAAAAGA
BC066782    (  2927) CATCACTATTGAGCACTGGTTTGTTAACTACCCGGAGTAAATAAAGAAGCTTGAAAAAGA
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  2987) GGGAAAAGTGTTTTTACAAAACTGATTTCCTTTCCTTTCTTGCTATCTCAAATGACTGGT
BC066782    (  2987) GGGAAAAGTGTTTTTACAAAACTGATTTCCTTTCCTTTCTTGCTATCTCAAATGACTGGT
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  3047) TGGTTAGATTTTTTTTTCTTTGTGATTTATATTCTCATGGCTGGAGAAGCTGGGCCAAGA
BC066782    (  3047) TGGTTAGATTTTTTTTTCTTTGTGATTTATATTCTCATGGCTGGAGAAGCTGGGCCAAGA
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  3107) TGGGCACAGCCTAGCAGCACTGGCCTGGGAGCCAGTATCACATTTTGGAACCAACCTGAC
BC066782    (  3107) TGGGCACAGCCTAGCAGCACTGGCCTGGGAGCCAGTATCACATTTTGGAACCAACCTGAC
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  3167) ATAACTTCCATGCCACAGTCTACCTTCAAAGTAGGAGTAGGCCAGGGCTCTCGGGGCACA
BC066782    (  3167) ATAACTTCCATGCCACAGTCTACCTTCAAAGTAGGAGTAGGCCAGGGCTCTCGGGGCACA
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  3227) CTGTTGCAGTGTTCAGCCCAGGCACCATGCGTTTTTACAGTTGTGTTTGCTCTTACTGTT
BC066782    (  3227) CTGTTGCAGTGTTCAGCCCAGGCACCATGCGTTTTTACAGTTGTGTTTGCTCTTACTGTT
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------
                    ............................................................
NM_207228   (  3287) GTGAGAAAGGTAGGTGATGATCTATCTATCCACATGTAAGATTTTTTTAAAATATGTACT
BC066782    (  3287) GTGAGAAAGGTAGGTGATGATCTATCTATCCACATGTAAGATTTTTTTAAAATATGTACT
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------

                    ............................................................
NM_207228   (  3347) CTTAATAAGTTTTCTTCAAAACATTAAAGTACCGCAATGAGCTTACTTGCCATTTAAAAA
BC066782    (  3347) CTTAATAAGTTTTCTTCAAAACATTAAAGTACCGCAATGAGCTTACTTGCCATTTAAAAA
AF530050    (  2733) ------------------------------------------------------------
AK030254    (  2702) ------------------------------------------------------------
```

```
                       ............................................................
NM_207228     (  3407) AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
BC066782      (  3407) AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AF530050      (  2733) ------------------------------------------------------------
AK030254      (  2702) ------------------------------------------------------------

                       .................................
NM_207228     (  3467) AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
BC066782      (  3467) AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AF530050      (  2733) ---------------------------------
AK030254      (  2702) ---------------------------------
```

Alignment of Human and mouse TSGA10 amino acid sequences (Exon sixteen which spliced out completely is marked in blue. (aa # 176-557: mitotic checkpoint domain, aa # 469-538: Exon 16)

```
Human       MMRSRSKSPRRPSPTARGANCDVELLKTTTRDREELKCMLEKYERHLAEIQGNVKVLKSE 60
Mice        MMRNRSKSPRRPSPTSRAANCDVELLKSTARDREELKCMLEKYERHLAEIQGNVKVLTSE 60
            ***.***********:*.*********:*:***************************.**

Human       RDKIFLLYEQAQEEITRLRREMMKSCKSPKSTTAHAILRRVETERDVAFTDLRRMTTERD 120
Mice        RDKTFLLYEQAQEEIARLRREMMKSCKSPKSTTAHAILRRVETERDVAFTDLRRMTTERD 120
            *** ***********:********************************************

Human       SLRERLKIAQETAFNEKAHLEQRIEELECTVHNLDDERMEQMSNMTLMKETISTVEKEMK 180
Mice        SLRERLKIAQETAFNEKAHLEQRIEELECTVHNLDDERMEQMANMTLMKETITTVEKEMK 180
            ******************************************:*********:*******

Human       SLARKAMDTESELGRQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDE 240
Mice        SLARKAMDTESELGRQKAENNSLRLLYENTEKDLSDTQRHLAKKKYELQLTQEKIMCLDE 240
            ************************************************************

Human       KIDNFTRQNIAQREEISILGGTLNDLAKEKECMQACLDKKSENIASLGESLAMKEKTISG 300
Mice        KIDNFTRQNIAQREEISILGATLNDLAKEKECLQACLDKKSENIASLGESLAMKEKTISG 300
            ********************.***********:***************************

Human       MKNIIAEMEQASRQCTEALIVCEQDVSRMRRQLDETNDELAQIARERDILAHDNDNLQEQ 360
Mice        MKNIIAEMEQASRQSTEALIMCEQDISRMRRQLDETNDELGQIARERDILAHQNTNLQEQ 360
            **************.*****:****:**************.***********:* *****

Human       FAKAKQENQALSKKLNDTHNELNDIKQKVQDTNLEVNKLKNILKSEESENRQMMEQLRKA 420
Mice        FAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEVNKLKNILKSEESENRQIMEQLRKA 420
            ***.******************.*****************************:*******

Human       NEDAENWENKARQSEADNNTLKLELITAEAEGNRLKEKVDSLNREVEQHLNAERSYKSQI 480
Mice        NEDAENWENKARQTEAENNTLKLELITAEAEGNRLKEKVDALNREVEQHLNAERSYKSQI 480
            *************:**:***********************:*******************

Human       STLHKSVVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIEMRE 540
Mice        ATLHKSLVKMEEELQKVQFEKVSALADLSSTRELCIKLDSSKELLNRQLVAKDQEIEMME 540
            :*****:*************************************************** *

Human       NELDSAHSEIELLRSQMANERISMQNLEALLVANRDKEYQSQIALQEKESEIQLLKEHLC 600
Mice        NELDSARSEIELLRSQMTNERISMQNLEALLVANRDKEYQSQIALQEKESEIQLLKEHLC 600
            ******:**********:******************************************

Human       LAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQELRRQNYSSNAYHM 660
Mice        LAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGTERFERERAVQ-LRRQNYSSNAYNF 659
            ********************************************** ***********::

Human       SSTMKPNTKCHSPERAHHRSPDRGLDRSLEENLCYRDF 698
Mice        G-PMKPNTKCHSPERAHHRSPDRGLDRSLEE------- 689
            . .****************************
```

Alignment of Tsga10 (AF530050) and cp431 (XM_136734) aminoacids sequences

```
                    +                 *+.  *     .   ** *     ..    * *         + +
530050      (     1) L----------------TPKNISL----VLGLQF-LKLEETRILSNZN-------KKZEP
136734      (     1) MgpekassrsperraggGPRRYVLarvnTKEHQFgLGAPVSEAGRNQNtfqleqeVRTQD

                    +. *.* .************************************************** *
530050      (    33) KIDSSLhNLQIEDLKQTNHDLEEYVRKLLDSKEAVSTQVDDLANHNEHLCKELIKLDQHA
136734      (    61) RFISTL-KLQIEDLKQTNHDLEEYVRKLLDSKEAVSTQVDDLANHNEHLCKELIKLDQLA

                    ************************************************************
530050      (    93) EKLQKEKNFVVDTADKELEEAKIELICQQNNITVLEDTIQRLKSIILETEKAQNTSPSRL
136734      (   120) EKLQKEKNFVVDTADKELEEAKIELICQQNNITVLEDTIQRLKSIILETEKAQNTSPSRL

                    ************   . . .  .************************************
530050      (   153) DSFVKTLEADRDZ1ZELKLRIZEKMMRNRSKSPRRPSPTSRAANCDVELLKSTARDREEL
136734      (   180) DSFVKTLEADRDY-YKTEAQNLRKMMRNRSKSPRRPSPTSRAANCDVELLKSTARDREEL

                    ************************************************************
530050      (   213) KCMLEKYERHLAEIQGNVKVLTSERDKTFLLYEQAQEEIARLRREMMKSCKSPKSTTAHA
136734      (   239) KCMLEKYERHLAEIQGNVKVLTSERDKTFLLYEQAQEEIARLRREMMKSCKSPKSTTAHA


                    ************************************************************
530050      (   273) ILRRVETERDVAFTDLRRMTTERDSLRERLKIAQETAFNEKAHLEQRIEELECTVHNLDD
136734      (   299) ILRRVETERDVAFTDLRRMTTERDSLRERLKIAQETAFNEKAHLEQRIEELECTVHNLDD

                    ************************************************************
530050      (   333) ERMEQMANMTLMKETITTVEKEMKSLARKAMDTESELGRQKAENNSLRLLYENTEKDLSD
136734      (   359) ERMEQMANMTLMKETITTVEKEMKSLARKAMDTESELGRQKAENNSLRLLYENTEKDLSD

                    ************************************************************
530050      (   393) TQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQREEISILGATLNDLAKEKECLQAC
136734      (   419) TQRHLAKKKYELQLTQEKIMCLDEKIDNFTRQNIAQREEISILGATLNDLAKEKECLQAC

                    ************************************************************
530050      (   453) LDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQSTEALIMCEQDISRMRRQLDET
136734      (   479) LDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQSTEALIMCEQDISRMRRQLDET

                    ****************+* ****************************************
530050      (   513) NDELGQIARERDILAHQNTNLQEQFAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEV
136734      (   539) NDELGQIARERDILAHENDNLQEQFAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEV

                    ************************************************************
530050      (   573) NKLKNILKSEESENRQIMEQLRKANEDAENWENKARQTEAENNTLKLELITAEAEGNRLK
136734      (   599) NKLKNILKSEESENRQIMEQLRKANEDAENWENKARQTEAENNTLKLELITAEAEGNRLK

                    ************************************************************
530050      (   633) EKVDALNREVEQHLNAERSYKSQIATLHKSLVKMEEELQKVQFEKVSALADLSSTRELCI
136734      (   659) EKVDALNREVEQHLNAERSYKSQIATLHKSLVKMEEELQKVQFEKVSALADLSSTRELCI

                    ************************************************************
530050      (   693) KLDSSKELLNRQLVAKDQEIEMMENELDSARSEIELLRSQMTNERISMQNLEALLVANRD
136734      (   719) KLDSSKELLNRQLVAKDQEIEMMENELDSARSEIELLRSQMTNERISMQNLEALLVANRD

                    ************************************************************
530050      (   753) KEYQSQIALQEKESEIQLLKEHLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGT
136734      (   779) KEYQSQIALQEKESEIQLLKEHLCLAENKMAIQSRDVAQFRNVVTQLEADLDITKRQLGT

                    ********** ************.******************************
530050      (   813) ERFERERAVQ-LRRQNYSSNAYNFGPMKPNTKCHSPERAHHRSPDRGLDRSLEE
136734      (   839) ERFERERAVQeLRRQNYSSNAYNLGPMKPNTKCHSPERAHHRSPDRGLDRSLEE
```

Alignment of S3 and L3 inserts with candidate activating domain –FLJ32880 hypothetical protein-

```
                 + ++++++++++++++++ +++++++++++++++++ + ++++++++++++++++
S3        (1) -ACTNACAAAAAGAGATCGAAtTACCAGGACGAATTCAC-CTAGCTGGCAGGTAGGGGT
L3        (1)ggGGTTACAAAAAGAGATCGAA-TACCAGGACGAATTCACtCGAGCTGGCAGGTAGGGGT

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3       (58) GGGAGAGGACAACATGCCAATATGCTGAACACTGACCAACTGTTGGTTAGGGCCATATCG
L3       (60) GGGAGAGGACAACATGCCAATATGCTGAACACTGACCAACTGTTGGTTAGGGCCATATCG

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (118) GAGAGACCAAGGAAGAGCACAGAGCCCTTGGTTCCAGGGACACCTACAGGACTGTTCAGC
L3      (120) GAGAGACCAAGGAAGAGCACAGAGCCCTTGGTTCCAGGGACACCTACAGGACTGTTCAGC

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (178) CTTCTATCAAACATCTCTCCGGAGGAGCAGGGCCGCCTTGGGAGCGGAGACAGTCTCCAG
L3      (180) CTTCTATCAAACATCTCTCCGGAGGAGCAGGGCCGCCTTGGGAGCGGAGACAGTCTCCAG

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (238) AGTCAGAGCTGCCAGCAGCAAAGGTCTTACAGCGCAGGGCAGACAACAAAGAAAGAACGA
L3      (240) AGTCAGAGCTGCCAGCAGCAAAGGTCTTACAGCGCAGGGCAGACAACAAAGAAAGAACGA

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (298) AAACCCCGGAGGCGCAACAAGAAAGGACGTGGCTCAGCTGAGGCAGAGGATCTCTTCAGT
L3      (300) AAACCCCGGAGGCGCAACAAGAAAGGACGTGGCTCAGCTGAGGCAGAGGATCTCTTCAGT

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (358) TCACCTCGGAAGCCTTCCTTCCCTTTCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGC
L3      (360) TCACCTCGGAAGCCTTCCTTCCCTTTCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGC

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (418) CAGGCTCTGCTTCAGTCAGGCTCCTCTGTGGCCGTGGGCCATCGATCCTTGCTCTTCCCT
L3      (420) CAGGCTCTGCTTCAGTCAGGCTCCTCTGTGGCCGTGGGCCATCGATCCTTGCTCTTCCCT

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (478) TCAGCAGCCCCCCAGTGCAAGACCAGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTT
L3      (480) TCAGCAGCCCCCCAGTGCAAGACCAGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTT

                ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++
S3      (538) CGAGCCTGCCACAAGTCGGAAGTGCAAAACCTGGGAAGGAGATATCAGCCTGGGGCCTGG
L3      (540) CGAGCCTGCCACAAGTCGGAAGTGCAAAACCTGGGAAGGAGATATCAGCCTGGGGCCTGG

                ++++++++++++++++++++++++++++++++++ +  +++++ + ++++++++++++
S3      (598) GCCAACCTTTCCCTGCCTCTGGGCAAGGCTGAGAGCCAAGGGCTAGAGCGACCCACCTTC
L3      (600) GCCAACCTTTCCCTGCCTCTGGGCAAGGCTGAGAGCTATTGGCTATATCGACCCACCTTC

                + ++++++++++++  +++++  ++++  +++++++ ++++++++++++++++++++++
S3      (658) T-GGTCAACAGGGAAAGGGTCA-GGGTCAGAGTGTGAGGATGTCTCAGAGGTAGAAGGCC
L3      (660) TgGGTCAACAGGGAATTGGTCAtTGGTCTTAGTGTGATGATGTCTCAGAGGTAGAAGGCC

                ++++++++++ +++++ +  + + ++++++ ++++++++++++  ++++++++ ++++
S3      (716) AGAATGCTGACGAGGCTgAGGAAAAGTCTCACCACTGGGGAACTGcCCCAGCTCCCAGGG
L3      (720) AGAATGCTGACCAGGCT-AATATATGTCTCAGCACTGGGGAACTG-ACCAGCTCCGAGGG

                + ++++++++ + ++++ ++ ++  ++++++ + +++ ++ ++++   ++  ++
S3      (776) C-AGGGTTGA-CCTGGAGGA-GA-GTGATCT-CGAGG-TA-TGAGAGAGACCAA------
L3      (778) CtAGGGTTGAtCTTGGATGAtGAtTTGATCTcCAAGGtTAtTGAGGAGGATGAAcccttc

                 ++   ++  + ++ ++   + +   + + ++ ++++  ++++ ++  +++
S3      (823) -CG--GAAAAG-CACTT---TTCCCAGGAGGA-TTTG--AAAAGGCAAAAA---------
L3      (838) cCGacAAAGGGcCATTTgtgTNCTTGGTATGAtTTTGtaAAAACGCCCAAAtattgtgta

S3      (864) ----------------------------------------------
L3      (898) ccttccactcctttgtggaggggctccccccttctcnaatttaagg
```

Mice FLJ32880m cDNA Sequence

```
ATGGAGATGG CAGAGAGGGC AGTTCCAGAG CTGGCAGGAA CTTGTCACTT   50
CCTCAGCACT AAGTCCAGAG CCCTGGGTGA CATCCAGAAA AATAGGTGCC  100
TTGGGAGTGG TGACAGTCTC CAGAGTCAGA GCTGCCAGCT ACAAAGATCT  150
TACAGCGCAG GGCAGACAAC AAAGAAAGAG CGAAAGGCCC GGAGGCGGAA  200
CAAGAAAGGC CGTGGCTCAG CTGAGGCAGA GGATCTCTTC AGTCCACCTT  250
CTCGAAAGCC TTCCTTCCCT TTCCAGTGGG CCTGGGAGAG CTTCATCATA  300
GATGGCCAGG CTCTGTTTCA GTCAAGTTCC TCTATGGCCA TGGGCCATCG  350
ACCCCTGCTC CTCCCTCCAG CAGCCCCCCA GTGCAAGTCC AGGCACAAGT  400
CAGTAGCCAA CCTCTCAGAA GACCTTCGAG CCTGCCACAA GATGGAAGAG  450
CAAAGCGTGG GGAGGAGACA TCACCTAGGG GGCTGGGCCA ACTTTCCCCT  500
GCCTCTGAGC AAAGTGGAGA GTCAAGGGCT AGATCAACCT AGCTTCTGGT  550
TATCAGGGAA AGGGTCAGGA TCAGAGTCTG AGGATGTCTT AGAGATAGAA  600
GGCCAGAATT CTGAAGAGGC TGAGAAAAGT CTCAGCTCTG GGGAACTGCC  650
TCAGCTCCCA GGGCAGAGCT TGATCTTGGA GGAGGAATTG ATATCAGAGG  700
CGATGGAGGA GGAACACAGC GACCCCCACA AGGGAAAGGG CAGTTCTGTC  750
AATAAGGGGG GGAATTCTGG AGAGAAGGGC TCAGAAGAGG GGGAGCTGCA  800
GAGCCACAAC CAGGGAAGCA GCTCTAGCTT CAACAGTCTC CGAAAGCCAC  850
AGAAGGGGAT CTCAAGGGCC AAGGACCTGA AGGGGCCCTG GGACCTGGAG  900
CGGCTACACA GGCAGTTACA GGAAGAACTG GAATGTGGTC CCCAGAAGCA  950
GACCTGGAAG GCGCTGCGGG CAGCTGTCCA GGCCTCTGCC AGGAACAAGA 1000
AGACCCCCAT CTTGGGAGAT GATGAAAGTT TCTTGTCTGC CAACTTCCCT 1050
AATCGTACTT TCCATAAACG ACAGGAAGCC ACCAGGAACC TGCTCCGGGC 1100
CTGGGAACAG CAGCAGGTGA AGGAGAGGCA GCAGGCCGAG GTGCGCAGAG 1150
CCCGGGAGCA GCAGGTGCAG CAGCAGGTGG CTCGCTGCCT GGCAGCCTAC 1200
ACATCCGGAG GGAACAGAGG GACCCTGGCA CCTCAGCGCA AACTAGAGGA 1250
GCTAAGGCGC AAGGAGCGAC AGCGTTTTGC TGAGTACCAA GCAGAGCTCC 1300
AAGGTATCCA GCATAGGGTG CAGGCCCGGC CCTTCCTGTT CCAGCAGGCC 1350
ATCAGACCAA CGCCAGACTC ACTGCAAATC GGCGCTTCTC CCAGGTGCTG 1400
TCAGCACTGG GGGTAGACGA GGACCAGCTG CTGGCCGAGG CAGGCAATGC 1450
TGAGAGCACC CCCCGGAAGC ACAGGAGCAA CCGATCACTT AGAGCAGAAA 1500
TGGAGCCCTC TTCTCAGAGC CCCCCCAAAA CAGAGCCAAC CAGCAGCCAG 1550
CCTGGAAGAC GCCCCTCCCC AACCCTGGAC CCAGACTACA GTCCCCGAGA 1600
GAAAAATTAA ACGCTTTATG GCC
```

Length: **556 aa**, molecular weight: **62391 Da**, CRC64 checksum: **9F52411EBC1488D4**

```
MGQDTDMLNT YQQLVRTPSV RPGQDVRLQA PGTRTGLLKL LSTVSQDKQG CLGSGDGVPN   60
QDLQQRPQSS RQTAKKDRKP RGQSKKGQGS EESEDHFPLL PRKPSFPFQW AWESIATDVR  120
AVLQPSSPTP GHQALPMPSS FSQRQSRRKS TANLPEAHGC CWKTEAQNLK ARQQLGAWGG  180
VSIPTGKGEL GSEPPSGLQL PGRRPGSGSA SDKQVQLQSL GAEEAERGLS SGVLPQRPRR  240
GSISEEEQFS EATEEAEEGE HRTPCRRRAG CQRKGQISGE EASDEGEVQG QSQGSSPSFN  300
NLRRRQWRKT RAKELQGPWD LEKLHRQLQR DLDCGPQKLP WKTLRAAFQA SKRNGKAYAS  360
GYDETFVSAN LPNRTFHKRQ EATRSLLQAW ERQRQEERQQ AELRRARTQH VQRQVAHCLA  420
AYAPRGSRGP GAAQRKLEEL RRQERQRFAE YQAELQGIQH RVQARPFLFQ QAMQANARLT  480
VTRRFSQVLS ALGLDEEQLL SEAGKVDREG TPRKPRSHRS MGVRMEHSPQ RPPRTEPTGS  540
QPDRHYNPSL DPECSP
```

(82.6 %: nuclear 13.0 %: cytoskeletal 4.3 %: vesicles of secretory system) Similar to hypothetical protein FLJ32880 [Homo sapi] → P05143 Proline-rich protein MP-3 (Mus musculus Chromosome 19); 27%

```
Human (    1) agttctacggtgcggatgggggatggggcaggatggggcaggacaccgatATGCTAAATA
Mice  (    1) ------------------------------------------------------------
Rat   (    1) ------------------------------------------------ATGCTGAACA

                .  ..... .......+.+.... .+.+..+.+...+.++.++  + ..+.+.. ...+
Human (   61) CCTACCAACAGTTGGTTAGGACCCCGTCGGTGCGACCAGGGCAGGAC-GTGCGGCTCCAG
Mice  (    1) -----------------ATGG--AGATGGCAGAG---AGGGCAGTTC-CAGAGCTGGCAG
Rat   (   11) CTGACCAACTGTTGGTTAGGGCCATATCGGAGAGACCAAGGAAGAGCaCAGAGCC-CTTG

                + .+..+. ++  . .+.+..++  +..++. .. ++... ..+.....++..++ .+..
Human (  120) GCTCCAGGAACCAGAACGGGGCT--GCTCAAGCTGCTGTCGACCGTCTCTCAGGACAAGC
Mice  (   38) GAACTTGTCACTTCCTCAGCACTaaGTCCAGAGCCCTGGGTGACA----TCCAGAAAAAT
Rat   (   70) GTTCCAGGGACACCTACAGGACT--GTTCAGCCTTCTATCAAACATCTCTCCGGAGGAGC


              +..+..++++.+++++.++.++..++.+.+..+.+++.++...++.++..++.++..++.
Human (  178) AGGGCTGCCTGGGGAGTGGTGATGGTGTGCCAAATCA-AG--ACCTGCAGCAGAGGCCTC
Mice  (   94) A--GGTGCCTTGGGAGTGGTGACAGTCTCCAGAGTCAGAGCTGCCAGCTACAAAGATCTT
Rat   (  128) AGGGCCGCCTTGGGAGCGGAGACAGTCTCCAGAGTCAGAGCTGCCAGCAGCAAAGGTCTT
                +.+++.++.++++++++.+++++++.++ ++.++..++.++.+.+..+.+++++++++.+
Human (  235) AGAGCTCAAGGCAGACAGCAAAGAAGGACCGCAAGCCCAGGGGCCAGAGCAAGAAAGGAC
Mice  (  152) ACAGCGCAGGGCAGACAACAAAGAAAGAGCGAAAGGCCCGGAGGCGGAACAAGAAAGGCC
Rat   (  188) ACAGCGCAGGGCAGACAACAAAGAAAGAACGAAAACCCCGGAGGCGCAACAAGAAAGGAC

                ..+++++.+..+++.+.++.++++.++++......+ . ++++.++.++.++++++++.+
Human (  295) AGGGCTCTGAAGAGTCTGAAGATCACTTCCCGCTTCTTCCTCGGAAACCCTCCTTCCCCT
Mice  (  212) GTGGCTCAGCTGAGGCAGAGGATCTCTTCAGTCCACCTTCTCGAAAGCCTTCCTTCCCTT
Rat   (  248) GTGGCTCAGCTGAGGCAGAGGATCTCTTCAGTTCAC---CTCGGAAGCCTTCCTTCCCTT


              ++++++++++++++++++++.++..++.+++++.++.++++.++.+++++.++.+.+++.
Human (  355) TCCAGTGGGCCTGGGAGAGCATCGCCACAGATGTCCGGGCTGTGCTTCAGCCAAGCTCC-
Mice  (  272) TCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGCCAGGCTCTGTTTCAGTCAAGTTCCT
Rat   (  305) TCCAGTGGGCCTGGGAGAGCTTCATCATAGATGGCCAGGCTCTGCTTCAGTCAGGCTCCT

                +.....++ ..+++++.+.+ ++.+++.+ + ++++++..+...+.+++++.+..++.++
Human (  414) CCA--ACCCCTGGCCACCAAGCCCTGCCCATgCCCTCC-TCGTTCTCCCAGCGTCAGTCC
Mice  (  332) CTATGGCCATGGGCCATCGACCCCTGCTCCT-CCCTCCAGCAGCCCCCCAGTGCAAGTCC
Rat   (  365) CTGTGGCCGTGGGCCATCGATCCTTGCTCTT-CCCTCCAGCAGCCCCCCAGTGCAAGACC

                 ++++.++++++....+++++++++.++++.+.++.+.+...++++...+++...++.+
              +
Human (  471) AGGCGCAAGTCCACGGCCAACCTCCCAGAGGCCCATGGCTGCTGCTGGAAGACAGAGGCG
Mice  (  391) AGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTTCGAGCCTGCCACAAGATGGAAGAG
Rat   (  424) AGGCACAAGTCAGTAGCCAACCTCTCAGAAGACCTTCGAGCCTGCCACAAGTCGGAAGTG
                ++++.+.++.....++++.+.++.+. ++.+.+++++.+... +..++.+.+++ +++ +
Human (  531) CAAAACCTGAAGGCGAGACAGCAGCTGGGAGCCTGGGGCGGTGTCTCCATCCCTaCTG-G
Mice  (  451) CAAAGCGTGGGGAGGAGACATCACCTAGGGGGCTGGGCCAACTTTCCCCTGCCT-CTGAG
Rat   (  484) CAAAACCTGGGAAGGAGATATCAGCCTGGGGCCTGGGCCAACCTTTCCCTGCCT-CTGGG

                +++.+  +.+++ +.+++...+++.+ .++.+       .+..+++.... .+++  +..
Human (  590) CAAAG--GGGAG-CTAGGATCAGAGCCCCCCAgtggtctCCAGCTGCCTGGGAGGagGCC
Mice  (  510) CAAAGTGGAGAGTCAAGGGCTAGATCAACCTA-------GCTTCTGGTTATCAGG--GAA
Rat   (  543) CAAGGCTGAGAGCCAAGGGCTAGAGCGACCCA-------CCTTCTGGTCAACAGG--GAA

                +++.+++++.++.+.++..+   ++.+.+++. ..++ +..++.+++.+..+.++++ ++
Human (  647) AGGATCAGGATCGGCGTCCGacaAGCAGGTC---CAGCTGCAAAGCCTGGGTGCTGAGGA
Mice  (  561) AGGGTCAGGATCAGAGTCTG---AGGATGTCTTAGAGATAGAAGGCCAGAATTCTGAAGA
Rat   (  594) AGGGTCAGGGTCAGAGTGTG---AGGATGTCTCAGAGGTAGAAGGCCAGAATGCTGACGA
```

```
                +++.++++...++++.++..++++++..+++++.++++.+++..++....+ +.++.++.
Human (  704) GGCCGAGAGGGGTCTGAGTTCTGGGGTGCTGCCCCAGCGCCCCAGGAGGGGGTCGATCTC
Mice  (  618) GGCTGAGAAAAGTCTCAGCTCTGGGGAACTGCCTCAGCTCCCAGGGCAGAGCTTGATCTT
Rat   (  651) GGCTGAGAAAAGTCTCAGCACTGGGGAACTGCCCCAGCTCCCAGGGCAAGGTTTGACCTT

                   .++++++++.....+ +++++++..+..+++++++   .  ++ +......+++.+.
Human (  764) AGAGGAGGAGCAATTTTCAGAGGCCACAGAGGAGGCTGAgGAG-GGAGAGCACAGGActc
Mice  (  678) GGAGGAGGAATTGATATCAGAGGCGATGGAGGAGGAACA-CAGcGACCCCCACAAG----
Rat   (  711) GGAGGAGGAGTTGATCTCAGAGGTGATGGAGGAGG------AG-GAACACAACAGGA---

                      +++  +++++.+.+.++++  +. ++++++ .++.++++++++++.+.+.+++.
Human (  823) cctgcAGAAggAGGGCTGGT-TGTCagAGAAAGGGGCAGATTTCTGGAGAGGAAGCCTCC
Mice  (  733) -----GGAA--AGGGCAGTTCTGTC--AATAAGGGGGGGAATTCTGGAGAGAAGGGCTCA
Rat   (  761) ------GAA--AGGGCAGTTCTGTC--AACAAGGGGAGGAATTCTGGAGAGAAGGGCTCA


                ++.++.++.++..+++++.++++.+.++++++.++++++.+.+.++.+++++..++++++
Human (  882) GATGAGGGAGAAGTGCAGGGCCAGAGCCAGGGGAGCAGCCCCAGCTTCAACAACCTCCGA
Mice  (  784) GAAGAGGGGGAGCTGCAGAGCCACAACCAGGGAAGCAGCTCTAGCTTCAACAGTCTCCGA
Rat   (  811) GAAGAAGGGGAGCTGCAGAGCCACAACCAGGGAAGCAGCTCTAACTCCAACAGCCTCCGA

                +.+..++....+.+++ ..++.++++++++++ .+..+++++++.++++++++++..+
Human (  942) AGGCGACAATGGAGGAAGACAAGGGCCAAGGAGCTGCAGGGGCCATGGGACCTGGAGAAG
Mice  (  844) AAGCCACAGAAGGGGATCTCAAGGGCCAAGGACCTGAAGGGGCCCTGGGACCTGGAGCGG
Rat   (  871) AAGTCACCAAAGGGGACCTCAGGGGCCAAGGAATTCAAGGGGCCCTGGGACCTGGAGCGG

                ++ +++++++++.+++++...++..++++..++++.+++++.++++.+.++++++++.+.
Human ( 1002) CTGCACAGGCAGCTACAGAGAGACCTGGACTGTGGCCCCCAAAAGCTGCCCTGGAAGACT
Mice  (  904) CTACACAGGCAGTTACAGGAAGAACTGGAATGTGGTCCCCAGAAGCAGACCTGGAAGGCG
Rat   (  931) CTCCACAGGCAGTTACAGGAGGAATTGGAAAGTGGTCCCCAGAAGCAGACCTGGAAGGCG

                .++.++++.++..++++++++++.....+.++....+++.++.....+..++++.+ ++.
Human ( 1062) TTGAGGGCTGCCTTCCAGGCCTCCAAGCGGAATGGAAAGGCCTATGCCTCGGGATACGAT
Mice  (  964) CTGCGGGCAGCTGTCCAGGCCTCTGCCAGGAACAAGAAGACCCCCATCTTGGGAGATGAT
Rat   (  991) TTGCGGGCAGCTGTCCAGGCCTCTGCCAGAAACAGGAAGACCCCCGTCACGGGAGAAGAG

                ++++.++++.++.++++++++.++++++++++.++.+++++.++++++++++.++++++
Human ( 1122) GAAACTTTCGTGTCTGCCAACCTCCCTAATCGCACCTTCCACAAACGACAGGAAGCCACC
Mice  ( 1024) GAAAGTTTCTTGTCTGCCAACTTCCCTAATCGTACTTTCCATAAACGACAGGAAGCCACC
Rat   ( 1051) GAAAGTTTCTTGACTGCCAACTTCCCTAATCGTACCTTCCATAAACGACAGGAGGCCACC

                +++
Human ( 1182) AGG------------------------------------------------------------
Mice  ( 1084) AGG------------------------------------------------------------
Rat   ( 1111) AGGacttgggctcttagagacacgtacttcttcccacagtccctcacctctggaccacat


Human ( 1185) ------------------------------------------------------------
Mice  ( 1087) ------------------------------------------------------------
Rat   ( 1171) gctcctccatgctaccttggccttcctcgagctcctgatccccatggacatccccccttt

                        +.++++++.+.+++++++++.+.+++++.+..+.+++++..+++++++++.
Human ( 1185) ---------AGCCTGCTGCAGGCCTGGGAGCGGCAGCGGCAGGAGGAGCGGCAGCAGGCC
Mice  ( 1087) ---------AACCTGCTCCGGGCCTGGGAACAGCAGCAGGTGAAGGAGAGGCAGCAGGCC
Rat   ( 1231) ctcaccagaAACCTGCTGCAGGCCTGGGAACAGCAGCAGCTGAAGGAGAAGCAGCAGGCG

                +++ ++++..+.++.+++...+++++ ++.++++.+++++++++.+..++.+++++.+++
Human ( 1236) GAGCTGCGGCGGGCCCGGACACAGCATGTACAGCGGCAGGTGGCCCACTGCCTGGCAGCC
Mice  ( 1138) GAGGTGCGCAGAGCCCGGGAGCAGCAGGTGCAGCAGCAGGTGGCTCGCTGCCTGGCAGCC
```

```
Rat   ( 1291) GAGATGCGCAGGGCTCGGGAGCAGCAAGTGCAGCAGCAGGTGGCCCGGTGTCTGGCGGCC
              +++.+..+..++++++.+...+++.++++++. +  +.++++++++.++.++++++++.+
Human ( 1296) TACGCACCCAGAGGGAGC--CGGGGCCCTGGGGCggCCCAGCGCAAGCTGGAGGAGCTGA
Mice  ( 1198) TACACATCCGGAGGGAACAGAGGGACCCTGGCAC--CTCAGCGCAAACTAGAGGAGCTAA
Rat   ( 1351) TACACGCCAGGAGGGAACAGAGGGACCCTGGGTC--CTCAGCGCAAACTAGAGGAGCTGA

              +++++.++++++++++++.+++++.++++++++.++.+++++.++.++.++++++++.+
Human ( 1354) GGCGCCAGGAGCGACAGCGCTTTGCTGAGTACCAGGCGGAGCTGCAAGGCATCCAGCACA
Mice  ( 1256) GGCGCAAGGAGCGACAGCGTTTTGCTGAGTACCAAGCAGAGCTCCAAGGTATCCAGCATA
Rat   ( 1409) GGCGCAAGGAGCGACAGCGTTTTGCCGAGTACCAAGCAGAGCTCCAGGGCATCCAGCATA

              +.++++++++++++++++++++++++++++++.++ +++.++++.+++.+.+++++ +
Human ( 1414) GGGTGCAGGCCCGGCCCTTCCTGTTCCAGCAGGCTATgCAGGCCAATGCCCGGCTCACCG
Mice  ( 1316) GGGTGCAGGCCCGGCCCTTCCTGTTCCAGCAGGCCAT-CAGACCAACGCCAGACTCACTG
Rat   ( 1469) GAGTGCAGGCCCGGCCCTTCCTGTTCCAGCAGGCCAT-CAGACCAATGCCAGGCTCACAG

              ..+..++++++++++++++++++++++++++.++++..+.++.+++++ +++++++++.
Human ( 1474) TCACTCGGCGCTTCTCCCAGGTGCTGTCAGCACTGGGGCTGGATGAGGAGCAGCTGCTGT
Mice  ( 1375) CAAATCGGCGCTTCTCCCAGGTGCTGTCAGCACTGGGGGTAGACGAGGACCAGCTGCTGG
Rat   ( 1528) CAAACCGGCGCTTCTCCCAGGTGCTGTCAGCATTGGGAGTGGATGAGGAACAGCTGCTGG

               +.++++++++.++       ....+++.+++.++++.++++.+.+++++++ ++.+.+++
Human ( 1534) CTGAGGCAGGAAAggtggaCAGAGAGGGCACCCCCAGGAAACCCAGGAGCCACAGGTCAA
Mice  ( 1435) CCGAGGCAGGCAA------TGCTGAGAGCACCCCCCGGAAGCACAGGAGCAACCGATCAC
Rat   ( 1588) CTGAGGCAGGCAA------TGCAGAGGGCATCCCCCGGAAACACAGGAGCTACCGGTCAT

              +..+.+....+++++++..++++.+++++++ +++++.+..+++++.++.+++.++++++
Human ( 1594) TGGGGGTGAGAATGGAGCACTCTCCTCAGAGGCCCCCAAGGACAGAACCCACCGGCAGCC
Mice  ( 1489) TTAGAGCAGAAATGGAGCCCTCTTCTCAGAGCCCCCCCAAAACAGAGCCAACCAGCAGCC
Rat   ( 1642) TTGGAGTAGAAATGGAGTCCTCTCCTCAGAGTCCCCCAAAGACAGAGCCAACCAGCAGCC

              ++++++..++.+.+.....+++.+.+++++.... .. . ......+.++++..++++++
Human ( 1654) AGCCTGACAGGCACTACAACCCCAGCCTGGACCCGGAGTGCAGTCCCTGAGA-TAAAATT
Mice  ( 1549) AGCCTGGAAGACGCCCCTCCCCAACCCTGGACCCAGACTACAGTCCCCGAGAGAAAAATT
Rat   ( 1702) AGCCTGGAAGACACCCTTCCCCAACCCTGG---------------ACTGAGAGAAAAATT
              +++ ++++++++++.
Human ( 1713) AAAGGCTTTATGGC-
Mice  ( 1609) AAACGCTTTATGGCC
Rat   ( 1747) AAATGCTTTATGGCC
```

BAB71488.1 ***(Human TSGA10IP)***

MGQDTDMLNTYQQLVRTPSVRPGQDVRLQAPGTRTGLLKLLSTVSQDKQGCLGSGDGVPNQD
LQQRPQSSRQTAKKDRKPRGQSKKGQGSEESEDHFPLLPRKPSFPFQWAWESIATDVRAVLQPSS
PTPGHQALPMPSSFSQRQSRRKSTANLPEAHGCCWKTEAQNLKARQQLGAWGGVSIPTGKGEL
GSEPPSGLQLPGRRPGSGSASDKQVQLQSLGAEEAERGLSSGVLPQRPRRGSISEEEQFSEATEEA
EEGEHRTPCRRRAGCQRKGQISGEEASDEGEVQGQSQGSSPSFNNLRRRQWRKTRAKELQGPW
DLEKLHRQLQRDLDCGPQKLPWKTLRAAFQASKRNGKAYASGYDETFVSANLPNRTFHKRQEA
TRSLLQAWERQRQEERQQAELRRARTQHVQRQVAHCLAAYAPRGSRGPGAAQRKLEELRRQE
RQRFAEYQAELQGIQHRVQARPFLFQQAMQANARLTVTRRFSQVLSALGLDEEQLLSEAGKVD
REGTPRKPRSHRSMGVRMEHSPQRPPRTEPTGSQPDRHYNPSLDPECSP

XM_355133. ***(Mus musculus Tsga10ip)***

MEMAERAVPELAGTCHFLSTKSRALGDIQKNRCLGSGDSLQSQSCQLQRSYSAGQTTKKERKAR
RRNKKGRGSAEAEDLFSPPSRKPSFPFQWAWESFIIDGQALFQSSSSMAMGHRPLLLPPAAPQCK
SRHKSVANLSEDLRACHKMEEQSVGRRHHLGGWANFPLPLSKVESQGLDQPSFWLSGKGSGSE
SEDVLEIEGQNSEEAEKSLSSGELPQLPGQSLILEEELISEAMEEEHSDPHKGKGSSVNKGGNSGE
KGSEEGELQSHNQGSSSSFNSLRKPQKGISRAKDLKGPWDLERLHRQLQEELECGPQKQTWKAL
RAAVQASARNKKTPILGDDESFLSANFPNRTFHKRQEATRNLLRAWEQQQVKERQQAEVRRAR
EQQVQQQVARCLAAYTSGGNRGTLAPQRKLEELRRKERQRFAEYQAELQGIQHRVQARPFLFQ
QAIRPTPDSLQIGASPRCCQHWG

XM_341991. ***(Rattus norvegicus Tsga10ip)***

MLNTDQLLVRAISERPRKSTEPLVPGTPTGLFSLLSNISPEEQGRLGSGDSLQSQSCQQQRSYSAG
QTTKKERKPRRRNKKGRGSAEAEDLFSSPRKPSFPFQWAWESFIIDGQALLQSGSSVAVGHRSLL
FPPAAPQCKTRHKSVANLSEDLRACHKSEVQNLGRRYQPGAWANLSLPLGKAESQGLERPTFW
STGKGSGSECEDVSEVEGQNADEAEKSLSTGELPQLPGQGLTLEEELISEVMEEEEHNRRKGSSV
NKGRNSGEKGSEEGELQSHNQGSSSNSNSLRKSPKGTSGAKEFKGPWDLERLHRQLQEELESGP
QKQTWKALRAAVQASARNRKTPVTGEEESFLTANFPNRTFHKRQEATRTWALRDTYFFPQSLTS
GPHAPPCYLGLPRAPDPHGHPPFLTRNLLQAWEQQQLKEKQQAEMRRAREQQVQQQVARCLA
AYTPGGNRGTLGPQRKLEELRRKERQRFAEYQAELQGIQHRVQARPFLFQQAIRPMPGSQQTGA
SPRCCQHWEWMRNSCWLRQAMQRASPGNTGATGHLE

Appendix B: Tsga10 protein analysis

SMART Tsga10 study

Defensins, also known as alpha-defensins, are a family of structurally related cysteine-rich peptides active against many Gram-negative and Gram-positive bacteria, fungi, and enveloped viruses. Defensins kill cells by forming voltage-regulated multimeric channels in the susceptible cell's membrane. They play a significant role in innate immunity to infection and neoplasia (nucleotide # 10-39)

Methyl-accepting chemotaxis-like domains- chemotaxis sensory transducer (nucleotide # 244-449).

Anaphylatoxin homologous domain: C3a, C4a and C5a anaphylatoxins are protein fragments generated enzymatically in serum during activation of complement molecules C3, C4, and C5. They induce smooth muscle contraction (nucleotide #332-363).

Saposin (B) Domains: Present in multiple copies in prosaposin and in pulmonary surfactant-associated protein B (nucleotide #361-436).

Description and alignments of Tsga10 residues with potential matches

Myosin tail- described in result.

Alignment of Myosin_tail_1 vs Tsga10/25-765

```
                *->dlerqkreleeqlkrkeselsqlslklEdEqalvaqLqkkikeleaR
                     +++ +e +  l    se+     ++  Eq+++ q    i+ l  +
TSGA10     25     -ARVNTKEHQFGLGAPVSEAGRNQNTFQLEQEVRTQDR-FISTLKLQ 69

                IeELeEeLEaERaARaKaEkqRaDLsrELEeLsERLeEagGaTaaQiEln
                Ie+L                  + +  LEe+  +L ++  a + Q++
TSGA10     70 IEDL------------------KQTNHDLEEYVRKLLDSKEAVSTQVD-- 99

                kKREaELaKLRrdLEEanlqhEealatLRKKHqdainElsdQieqLqKqK
                               +    +E    +L K  q a        e LqK K
TSGA10    100 ---------------DLANHNEHLCKELIKLDQLA--------EKLQKEK 126
                akaEKeKsqlqaEvddllaqldqiaKaKlnaEKkakqlEsQlsElqvKld
                         +  + +d   +l+ ++ + +++   +  lE+++++l++ +
TSGA10    127 --------NFVVDTADK--ELEEAKIELICQQNNITVLEDTIQRLKSIIL 166

                ElqRqlnDltsqKsRLqsENsdLtrqleEaEaqvsqlsklKsqlesQLEe
                E++++   ++++ sRL s +  L  +   +++ +++l k   + +
TSGA10    167 ETEKA---QNTSPSRLDSFVKTLEADRDYYKTEAQNLRKMMRNRSKS--- 210

                AkRslEEEsReRanLqaqlrnlehDlDslrEqlEEEsEAKaeleRqLsKa
                  R+    sR+ +     l++ ++D + l+  lE       ++eR L++
TSGA10    211 -PRRPSPTSRAANCDVELLKSTARDREELKCMLE-------KYERHLAEI 252

                naeiqqw.....rsKfEsEgalraEElEElKkKlnqkisElEeaaEaana
                +  +  ++++++++ +  E a+  EE   l++ +++     + +  +a+
TSGA10    253 QGNVKVLtserdKTFLLYEQAQ--EEIARLRREMMKSCKSPKSTTAHAIL 300

                KcssLEKtKsRLqsElEDlqievEranaaaseLEKKqknFDKilaEwKkk
                +++  E+         +++  ++ r  ++  +L  + k +  ++  +K+
TSGA10    301 RRVETER---------DVAFTDLRRMTTERDSLRERLKIAQETAFNEKAH 341

                vdelqaEletAqreaRnlstElfrlKneleElkDqvEaLrRENKnLqdEi
                +++ + Ele       nl  E +   +     k+++ ++++E K+L+
TSGA10    342 LEQRIEELEC---TVHNLDDERMEQMANMTLMKETITTVEKEMKSLA--- 385

                kDLtdqLgEgGRnvHELEKarRrLEaEkdELqaALeEAEaALeqeEsKvl
                + + d+ +E GR             aE + L   +e  E +L ++
TSGA10    386 RKAMDTESELGRQ----------KAENNSLRLLYENTEKDLSDTQRHLA 424

                RaqvElsqiRsEiERRLaEKEEEfEn.tRKnhqr..aiesLqasLaEaEa
                    El q+ +E   ++   +E ++n tR n  ++++i  L a+L    a
TSGA10    425 KKKYEL-QLTQE---KIMCLDEKIDNfTRQNIAQreEISILGATL-NDLA 469

                KgKaEalR.lKKKLEgdInELEiaLDhaNkanaeaqKnvKkyqqqvkeLQ
                K K  + + l KK E  I+ L ++L +  k+ +  +Kn+
TSGA10    470 KEKECLQAcLDKKSE-NIASLGESLAMKEKT-ISGMKNI---------- 506
                tqvEeeQRaredareqlavaERRataLqaElEELrvaLeqaeRaRKqAEt
                  +E+eQ+ r+ +++ + + E   + +   l+E    L q  R R
TSGA10    507 -IAEMEQASRQSTEALI-MCEQDISRMRRQLDETNDELGQIARERDILAH 554
```

```
                ElaEaservneLtaqnssLiaqKRKLEgelaalqsDLDEavnElkaAeER
                E + ++e+ + +   n +L+   +   +el+++   + +++ E
TSGA10    555 ENDNLQEQFAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEVNKLKNI 604

                akkaqaDaarLaeELrqEQehsqklErlRKqLEsqvKeLqvRL..dEaEa
                 k   ++  +  e+Lr+  e  +  E+   q E ++  L+  L ++EaE
TSGA10    605 LKSEESENRQIMEQLRKANEDAENWENKARQTEAENNTLKLELitAEAEG 654

                aAlkgGKkvIqKLEaRVReLEaELdgEqRRhaetqKnlRKaeRrvKELqf
                + lk+      K  a  Re+E+ L +E R ++ ++++l K+  ++ E  +
TSGA10    655 NRLKE------KVDALNREVEQHLNAE-RSYKSQIATLHKSLVKMEEELQ 697

                QvEEDkKnle....rlQDLvDKLqaKiKtyKRQlEEaEEiaqinlsKyRk
                 v  +k   + + +++++L  KL +   ++ RQl       ++
TSGA10    698 KVQFEKVSALadlsSTRELCIKLDSSKELLNRQL------VAKD------ 735
                aQreLEdAEERADqAEsslnklRqreaKsRrs<-*
                  +e+E +E   D A s++ +lR   +  R s
TSGA10    736 --QEIEMMENELDSARSEIELLRSQMTNERIS    765
```

EB1 (microtubule binding) - The human EB1 protein was originally discovered as a protein interacting with the C-terminus of the APC protein. This interaction is often disrupted in colon cancer, due to deletions affecting the APC C-terminus. Several EB1 orthologues are also included in this family. The interaction between EB1 and APC has been shown to have a potent synergistic effect on microtubule polymerization. Neither of EB1 or APC alone has this effect. It is thought that EB1 targets APC to the + ends of microtubules, where APC promotes microtubule polymerization. This process is regulated by APC phosphorylation by Cdc2, which disrupts APC-EB1 binding. Human EB1 protein can functionally substitute for the yeast EB1 homologue Mal3. In addition, Mal3 can substitute for human EB1 in promoting microtubule polymerization with APC.

Alignment of EB1 vs Tsga10/136-175

```
                *->gLEkERDFYFsKLRdIEiLCQeThetensslvdvkLeLvkriqaILY
                   +LE+++          IE++CQ+     n+ +v +    ++r+++I+
Tsga10    136    ELEEAK---------IELICQQ-----NNITVLE--DTIQRLKSIIL 166

                ATeegfehp<-*
                +Te
Tsga10    167   ETEKAQNTS    175
```

MAP (E-MAP-115): The organisation of microtubules is controlled by tissue-specific microtubule-associated proteins (MAPs). This protein is a microtubule-stabilising protein predominantly expressed in cell lines of epithelial origin. It plays a role in epithelial polarization and differentiation during embryogenesis.

Alignment of E-MAP-115 vs Tsga10/215-388

```
                *->AP.gkaens.aAlsKptAGTTDaeEAtrLLAEK..RRqAReQrEkEE
                   +P++ a+n ++ l K tA   D+eE   +L EK +R +A  Q
Tsga10   215       SPtSRAANCdVELLKSTA--RDREELKCML-EKyeRHLAEIQGNV-- 256

                qERreqeErdrrkREElkrraaEeRtlrrEeEARrrEeeraRekeekKKq
                   ++  Erd+     l   a Ee  +r+  E  +     +   + +
Tsga10   257 --KVLTSERDKTFL--LYEQAQEEI-ARLRREMMKSCKSPKSTTAHAILR 301
                EGEEKRK.AGEElkRkAEEeLLLKekrEqEeqErAmiqkQKEeAEaraRE
                  E  R  A  +l+R   E+     +rE+ +             E+   E
Tsga10   302 RVETERDvAFTDLRRMTTER---DSLRERLKIAQ----------ETAFNE 338

                eA..ErvRlEREkhfqqeEQERlERkKRLeEIMKRTRksevsekVKkeDs
                 A+ E+   E E        ER+E    +   MK T  +   e   k+
Tsga10   339 KAhlEQRIEELECTVHNLDDERMEQMANMT-LMKETITTVEKEM--KSLA 385

                keG<-*
                +
Tsga10   386 RKA    388
```

MAD- described in result.

Alignment of MAD vs Tsga10/308-898

```
                *->DdlesnttvmadssndrsilrSalnkflsarLegsfseLgvstscty
                   D+   ++t++    ++  + r+     l+ rL+
TSGA10   308       DV---AFTDL----RRMTTERD----SLRERLK-------------- 329
                rcRemvkielkdalsslekqvqesmallqdlestnsksraelirlkfkLi
                            a  ++   ++e + l+q           + +l+  +
TSGA10   330 -----------IAQETA---FNEKAHLEQ----------RIEELECTVH 354

                qlenElaqkelehkraqleLekkastlaerYereidknqeLltdlKqLee
                +l  E     + + +a   L k+              +   +   K L
TSGA10   355 NLDDE-----RMEQMANMTLMKE-------------TITTVEKEMKSLAR 386

                seakaeselaeakErlkqleqkldkLseelrkdaesKkL........qkE
                +    esel ++k+++++l++           +  +K+L++++++  +k+
TSGA10   387 KAMDTESELGRQKAENNSLRLLY---------ENTEKDLsdtqrhlaKKK 427

                kedqqedakesisklknqlsemqlrAqnaetelkllekeledlkeqleel
                 e+q+ ++k  i +l  +++ +         +   +++e+  l   l +l
TSGA10   428 YELQLTQEK--IMCLDEKIDNFT-------RQNIAQREEISILGATLNDL 468
```

```
                qkelSGFEqLRaEaeqkLqslkasqaeradnDDeqlikhLeeelkeyeqe
                 ke+   E L a  ++k  +  ++ ++ a    e+ i+     + e eq
TSGA10   469 AKEK---ECLQACLDKKSENIASLGESLAMK--EKTISGMKNIIAEMEQA 513

                vevvKsmkeqllrl.peLEreleqlreEnkkLksskenneLLkEeledlq
                      + e l   + +  r   ql e n  L +  +   +L  e+  lq
TSGA10   514 S---RQSTEALIMCeQDISRMRRQLDETNDELGQIARERDILAHENDNLQ 560

                skLeRaEkmrskladLelenekLeaeLksWeiYNDSDDDDDNNVNNNDNN
                          +++a    en+ L ++L+
TSGA10   561 ----------EQFAKVKQENQALSKKLNDTH------------------- 581

                NNNKNDNNNDNNNDTSNNNnLlqnigLkNnlrNtPedisrkiveLQkeel
                                      n     +l+    di +k+ ++  e+
TSGA10   582 ----------------------N-----ELS----DIKQKVQDTNLEVN 599

                ilteknssltsdlknLktanqqLqlerqqalaeitelkkkleelkalnrR
                +l++   s +s +++       + + +++a+  +++ ++k+   +a+n
TSGA10   600 KLKNILKSEESENRQ-------IMEQLRKANEDAENWENKARQTEAENNT 642

                LqrrlsLvtkErDlLRaildsLYSAqNNALydsenteteaSnqlvrrLee
                L + l+++  E  +L++ +d      N  +    n+e +++ q +
TSGA10   643 LKLELITAEAEGNRLKEKVDA----LNREVEQHLNAERSYKSQIA----- 683

                GmHn..isaEdmvQkvdsykakmEaltvqLdvkeledlggqKdraetle
                 +H + +  E  +Qkv+      E+    +    ++ +l++ ++++  l
TSGA10   684 TLHKslVKMEEELQKVQF-----EK----V---SALADLSSTRELCIKLD 721

                kElkllkeqlssnerslsfvkEalnaedfLrlkiesLErerdrLrseial
                   +ll+ ql+  ++                 +ie  E e+d +rsei+l
TSGA10   722 SSKELLNRQLVAKDQ-----------------EIEMMENELDSARSEIEL 754

                LEmkLehlcLqGDYsasstkVLhmsnnPavkaeqikkntleaLQa...El
                L +++         +  s   L   + + v ++    +   aLQ++++E+
TSGA10   755 LRSQMTN-------ERISMQNL---EALLVANRDKEYQSQIALQEkesEI 794

                ekLkerLkalEegneqaLvdlepivdskiaekElaqLkkqvesaE.....
                + Lke L     ++e +         + i ++ +aq+++ v  +E + +
TSGA10   795 QLLKEHL----CLAENK---------MAIQSRDVAQFRNVVTQLEadldi 831

                .......krnqRLKeVFqtKilEFRkACysLfGYkIdRiddqQrsNGtpe
                ++++ ++ r++R ++V     +E R+  ys   Y +  +         p+
TSGA10   832 tkrqlgtERFERERAV-----QELRRQNYSSNAYNLGPM--------KPN 868


                sryrltS...mYAesdddkLifdlessstskcwlkLletassetfedlId
                +  + +S+++   +s+d  L+ +le++ +
TSGA10   869 T--KCHSperAHHRSPDRGLDRSLEENLCY-------------------- 896

                lwlrkrnsiPAFlAaLTLELFnkrt<-*
                                    r+
TSGA10   897 ----------------------RD    898
```

The ***ERM*** family consists of three closely-related proteins, ezrin, radixin and moesin. Ezrin is known as a constituent of microvilli PUBMED:6885906, radixin as a barbed, end-capping actin-modulating protein PUBMED:2500445, and moesin as a heparin binding protein PUBMED:3046603. A tumour suppressor molecule responsible for neurofibromatosis type 2 (NF2) which is named as merlin (moesin-ezrin-radixin-like protein) is highly similar to ERM proteins. ERM molecules contain 3 domains, a charged C-terminal domain PUBMED:9048483; an N-terminal globular domain; and an extended α-helical domain. Ezrin, radixin and merlin also contain a polyproline region between the helical and C-terminal domains. The N-terminal domain is highly conserved, and is also found in merlin, band 4.1 proteins and members of the band 4.1 superfamily. ERM proteins crosslink actin filaments with plasma membranes. They co-localise with CD44 at actin filament-plasma membrane interaction sites, associating with CD44 via their N-terminal domains and with actin filaments via their C-terminal domains PUBMED:9048483. Interestingly, myosin regulatory light chain interacting protein (MIR) belongs to ERM family (Bornhauser et al., 2003) while is localized specially to neurons in cerebellum and hippocampus. The presence of it in brain neurons during development together with its known effects on neurite outgrowth is suggestive an important function of it in nerve cell motility and cytoskeletal interaction regulation (Olsson et al., 2000).

Alignment of ERM vs Tsga10/425-704

```
                *->NKKGTdLWLGVDALGLNIYekdDrLTPKIGFPWSEIRNISFnDKKFv
                    KK+ +L                 +LT                +K
TSGA10    425     -KKKYEL---------------QLT--------------QEK--- 436

                IKPiDKKApdFv..FYaprlRINKRIL.....aLCmGNHELYMRRRKPDT
                I  +D K+++F+++  a+r  I+  IL+ + ++L
TSGA10    437 IMCLDEKIDNFTrqNIAQREEIS--ILgatlnDL---------------- 468
```

```
                IEVQQMKaQArEEKalKqaERekLerEkkaREiaekekEeaEreqq..El
                               +K+ E  +    kk   ia   +  a++e+  + +
TSGA10    469 ---------------AKEKECLQACLDKKSENIASLGESLAMKEKTisGM 503

                eerlkqmeEdmrraqkeLeeseekaleLeeklkqeqeaaqkLekkqkele
                    ++me + r + ++L   e+ + +   +l ++  +  + + +++ l+
TSGA10    504 KNIIAEMEQASRQSTEALIMCEQDISRMRRQLDETNDELGQIARERDILA 553

                eenrrLeeeaaaseeErerLeaevdeataevakleeerekkeeEteqlqa
                 en+ L+e+ a+ + E + L++++ +   e   +++ ++    E+ +l+
TSGA10    554 HENDNLQEQFAKVKQENQALSKKLNDTHNELSDIKQKVQDTNLEVNKLKN 603

                elreaqekeeeakeellevlaaptappmhhvsapeaELttdngeqleqdd
                 l++     + ++e l                          + + +++
TSGA10    604 ILKSEESENRQIMEQL-----------------------RKANEDAENW 629

                ngeeasadlePkEFiLtdedmkdrsnEiEeeRvteaeknerlqtqLkaLk
                ++ +   + e              +n + +e +t++    rl+++ +aL
TSGA10    630 ENKARQTEAE--------------NNTLKLELITAEAEGNRLKEKVDALN 665

                sELaavrdesK..etanDilHeENsvRaGrDKYkTLRqIRkGNTKqRVDq
                +E +++ + +++ ++++  lH    v+  +     L +           q
TSGA10    666 REVEQHLNAERsyKSQIATLHKSL-VKMEEE----LQKV----------Q 700

                FEsM<-*
                FE
TSGA10    701 FEKV    704
```

Prefoldin: This family comprises of several prefoldin subunits. The biogenesis of the cytoskeletal proteins actin and tubulin involves interaction of nascent chains of each of the two proteins with the oligomeric protein prefoldin (PFD) and their subsequent transfer to the cytosolic chaperonin CCT (chaperonin containing TCP-1).

Alignment of Prefoldin vs Tsga10/440-548

```
                *->lkeeiesLqaelarLqeaieeleksletlktlkkkeedegkevlvpl
                   l+e+i++++++     ++ ee+  + +tl++l k     +ke l
TSGA10    440    LDEKIDNFTRQNI---AQREEISILGATLNDLAK-----EKECLQ-- 476

                gaglfvkaevidtdkvlVdlGagyyvEksleeAieiLdkrieelekqlek
                       a+++++++ + +lG++ ++++     ++ +   +e++++q+ +
TSGA10    477 -------ACLDKKSENIASLGESLAMKEKTISGMKNIIAEMEQASRQSTE 519

                ....leeelekLrdqittleanlqqvqqk<-*
                    +e++++  r+q++++ ++l q++ +
TSGA10    520 alimCEQDISRMRRQLDETNDELGQIARE    548
```

Rabaptin (growth factor activity)- Regeneration of injured axons at neuromuscular junctions has been assumedto be regulated by extra-cellular factors that promote neurite outgrowth. A novel neurite outgrowth factor from chick denervated skeletal muscle has been cloned and characterised. The protein, termed neurocrescin (rabaptin), has been shown to be secreted in an activity-dependent fashion PUBMED:9427343.

Alignment of Rabaptin vs Tsga10/461-588

```
                *->kaiatvsletlqeaadvvkskkrEevakv..qaieqEevaslesslk
                   +   +  l   +e     + kk E +a++++   ++E + s    +
TSGA10    461     LGATLNDLAKEKECLQACLDKKSENIASLgeSLAMKEKTISGMKNII 507

                dsleseeaqiaayreesereiadLrkkvrelGNKLGRRCDmCSNYEKqLq
                  +e++  q +      e++i  +r+++ e    LG
TSGA10    508 AEMEQASRQSTEALIMCEQDISRMRRQLDETNDELG-------------- 543

                GiqiqEAEtRDqvkKllLsraqeedqLEkeMkkaqEdeekLRevvlpmek
                    q A  RD        +a e+d+L  +  k +     L     +
TSGA10    544 ----QIARERDI-------LAHENDNLQEQFAKVKQENQALSKKLNDTHN 582

                eIeaLk<-*
                e +  k
TSGA10    583 ELSDIK    588
```

Intermediate filaments (IF) PUBMED:8771189, PUBMED:3052284, PUBMED:2183847 are proteins which are primordial components of the cytoskeleton and the nuclear envelope. They generally form filamentous structures 8 to 14 nm wide.

Alignment of Filament vs Tsga10/547-830

```
                *->nEKeqmQnLNDRLAsYIdKVRfLEqqNkeLevkieelrqkqsrggpa
                    E+ ++ + ND L     KV    q+N +L  k+++ ++
TSGA10    547     RERDILAHENDNLQEQFAKVK---QENQALSKKLNDTHNEL------ 584

                svsrlyslYet...eieeLRrqidqltnerarlqlEidnlrealedfrkK
                  s ++++ + ++ e + L++ + + ++e+ ++  ++ ++ e++e  ++K
TSGA10    585 --SDIKQKVQDtnlEVNKLKNILKSEESENRQIMEQLRKANEDAENWENK 632

                yedKeDLaaQnqlkdlEialntkeaeLaTaL.eRqeaEndlvgLRaQiAk
                  +             E+++nt+  eL Ta+ e   + + +++L
TSGA10    633 ARQ------------TEAENNTLKLELITAEaEGNRLKEKVDAL------ 664
```

```
                lEslaaRkdlDeaTLarvDLEn....kvEsLqEElaFLKknHeEEvkeLq
                       ++++    +++ +E++ ++++ +L   l  +    eEE+  +q
TSGA10    665 ------NREVE----QHLNAERsyksQIATLHKSLVKM----EEELQKVQ 700

                aqiqdtgqvnVEmDaarqqEwklDLtkaLrEiRaQYEeiAeknrqeaEew
                    ++    ++  + r  E  ++L++                   + e
TSGA10    701 FEKVSA---LADLSSTR--ELCIKLDS------------------SKEL 726

                YksKleeLqtaaarngealrsaKeEitElRRqiQsLeiELqslksqnasL
                 +++l    +++++   +l sa++Ei  lR q+   +i  q+l    ++L
TSGA10    727 LNRQLVAKDQEIEMMENELDSARSEIELLRSQMTNERISMQNL----EAL 772

                ErqlaElEeryeaelaqyqalisqlEeeLqqlreEMarqLrEYQeLLdVK
                  + ++ E++ + +l++ ++ i+ l e L  + + Ma q r+  +  +V
TSGA10    773 LVANRDKEYQSQIALQEKESEIQLLKEHLCLAENKMAIQSRDVAQFRNVV 822

                laLDiEIATYRKLLEGEEsR<-*
                + L++            +
TSGA10    823 TQLEA-----------DLD    830
```

Tropomyosin- is a family of closely related proteins present in muscle and non-muscle cells. In striated muscle, tropomyosin mediates the interactions between the troponin complex and actin so as to regulate muscle contraction. The role of tropomyosin in smooth muscle and non-muscle tissues is not clear. Tropomyosin is an α-helical protein that forms a coiled-coil structure of 2 parallel helices containing 2 sets of 7 alternating actin binding sites PUBMED:6993480. There are multiple cell-specific isoforms, created by differential splicing of the messenger RNA from one gene, but the proportions of the isoforms vary between different cell types.

Alignment of Tropomyosin vs Tsga10/557-587

```
                *->DrLEDeLvaeKekYkaisdeLDqtlneLtgi<-*
                   D L ++ ++ K ++ a+s  L  t neL+ i
Tsga10    557      DNLQEQFAKVKQENQALSKKLNDTHNELSDI    587
```

Appendix C: Potential phosphorylation, glycolyzation and myrtolyzation sites of Tsga10 (689 aa residues)

N-glycosylation site: PA N-{P}-[ST]-{P}

NRSK 4-7 NMTL 164-167 NNSL 200-203 NFTR 244-247
NDTH 376-379 NNTL 438-441 NYSS 652-655

```
1       10        20        30        40
mmrNRSKsprrpsptsraancdvellkstardreelkcmlekyerhlaei 50
qgnvkvltserdktfllyeqaqeeiarlrremmksckspksttahailrr 100
veterdvaftdlrrmtterdslrerlkiaqetafnekahleqrieelect 150
vhnlddermeqmaNMTLmketittvekemkslarkamdteselgrqkaeN 200
NSLrllyentekdlsdtqrhlakkkyelqltqekimcldekidNFTRqni 250
aqreeisilgatlndlakekeclqacldkkseniaslgeslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldetndelgqiarerdil 350
ahendnlqeqfakvkqenqalskklNDTHnelsdikqkvqdtnlevnklk 400
nilkseesenrqimeqlrkanedaenwenkarqteaeNNTLklelitaea 450
egnrlkekvdalnreveqhlnaersyksqiatlhkslvkmeeelqkvqfe 500
kvsaladlsstrelcikldsskellnrqlvakdqeiemmeneldsarsei 550
ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvtqleadlditkrqlgterfereravqelrr 650
qNYSSnaynlgpmkpntkchsperahhrspdrgldrsleenlcyrdf 697
```

cAMP- and cGMP-dependent protein kinase phosphorylation site: PA [RK](2)-x-[ST]

RRPS 10-13 RRMT 113-116

```
1       10        20        30        40
mmrnrskspRRPSptsraancdvellkstardreelkcmlekyerhlaei 50
qgnvkvltserdktfllyeqaqeeiarlrremmksckspksttahailrr 100
veterdvaftdlRRMTterdslrerlkiaqetafnekahleqrieelect 150
vhnlddermeqmanmtlmketittvekemkslarkamdteselgrqkaen 200
nslrllyentekdlsdtqrhlakkkyelqltqekimcldekidnftrqni 250
aqreeisilgatlndlakekeclqacldkkseniaslgeslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldetndelgqiarerdil 350
ahendnlqeqfakvkqenqalskklndthnelsdikqkvqdtnlevnklk 400
nilkseesenrqimeqlrkanedaenwenkarqteaenntlklelitaea 450
egnrlkekvdalnreveqhlnaersyksqiatlhkslvkmeeelqkvqfe 500
kvsaladlsstrelcikldsskellnrqlvakdqeiemmeneldsarsei 550
ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvtqleadlditkrqlgterfereravqelrr 650
qnyssnaynlgpmkpntkchsperahhrspdrgldrsleenlcyrdf 697
```

Protein kinase C phosphorylation site : PA [ST]-x-[RK]

SPR 8-10	TSR 15-17	TAR 29-31	SER 59-61
SCK 85-87	SPK 88-90	TER 103-105	TER 117-119
SLR 121-123	SLR 202-204	TEK 210-212	TQR 217-219
SKK 372-374	TLK 440-442	SYK 475-477	STR 510-512
SSK 520-522	SAR 545-547	TKR 630-632	TER 636-638

```
1       10        20        30        40
mmrnrskSPRrpspTSRaancdvellksTARdreelkcmlekyerhlaei 50
qgnvkvltSERdktfllyeqaqeeiarlrremmkSCKSPKsttahailrr 100
veTERdvaftdlrrmtTERdSLRerlkiaqetafnekahleqrieelect 150
vhnlddermeqmanmtlmketittvekemkslarkamdteselgrqkaen 200
nSLRllyenTEKdlsdTQRhlakkkyelqltqekimcldekidnftrqni 250
aqreeisilgatlndlakekeclqacldkkseniaslgeslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldetndelgqiarerdil 350
ahendnlqeqfakvkqenqalSKKlndthnelsdikqkvqdtnlevnklk 400
nilkseesenrqimeqlrkanedaenwenkarqteaennTLKlelitaea 450
egnrlkekvdalnreveqhlnaerSYKsqiatlhkslvkmeeelqkvqfe 500
kvsaladlsSTRelcikldSSKellnrqlvakdqeiemmeneldSARsei 550
ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvtqleadldiTKRqlgTERfereravqelrr 650
qnyssnaynlgpmkpntkchsperahhrspdrgldrsleenlcyrdf 697
```

Casein kinase II phosphorylation site: PA [ST]-x(2)-[DE]

TARD 29-32	SERD 59-62	TERD 103-106	TERD 117-120
SLRE 121-124	TTVE 173-176	TESE 189-192	TEKD 210-213
TLND 262-265	SLGE 286-289	TNDE 336-339	THNE 378-381
TNLE 392-395	TEAE 434-437	STRE 510-513	SSKE 520-523
SEIE 548-551	TQLE 621-624	SLEE 687-690	

```
1       10        20        30        40
mmrnrsksprrpsptsraancdvellksTARDreelkcmlekyerhlaei 50
qgnvkvltSERDktfllyeqaqeeiarlrremmksckspksttahailrr 100
veTERDvaftdlrrmtTERDSLRErlkiaqetafnekahleqrieelect 150
vhnlddermeqmanmtlmketiTTVEkemkslarkamdTESElgrqkaen 200
nslrllyenTEKDlsdtqrhlakkkyelqltqekimcldekidnftrqni 250
aqreeisilgaTLNDlakekeclqacldkkseniaSLGEslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldeTNDElgqiarerdil 350
ahendnlqeqfakvkqenqalskklndTHNElsdikqkvqdTNLEvnklk 400
nilkseesenrqimeqlrkanedaenwenkarqTEAEnntlklelitaea 450
egnrlkekvdalnreveqhlnaersyksqiatlhkslvkmeeelqkvqfe 500
kvsaladlsSTRElcikldSSKEllnrqlvakdqeiemmeneldsarSEI 550
Ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvTQLEadlditkrqlgterfereravqelrr 650
qnyssnaynlgpmkpntkchsperahhrspdrgldrSLEEnlcyrdf 697
```

Tyrosine kinase phosphorylation site: **PA [RK]-x(2,3)-[DE]-x(2,3)-Y**

RSLEENLCY 686-694

```
1       10        20        30        40
mmrnrsksprrpsptsraancdvellkstardreelkcmlekyerhlaei 50
qgnvkvltserdktfllyeqaqeeiarlrremmksckspksttahailrr 100
veterdvaftdlrrmtterdslrerlkiaqetafnekahleqrieelect 150
vhnlddermeqmanmtlmketittvekemkslarkamdteselgrqkaen 200
nslrllyentekdlsdtqrhlakkkyelqltqekimcldekidnftrqni 250
aqreeisilgatlndlakekeclqacldkkseniaslgeslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldetndelgqiarerdil 350
ahendnlqeqfakvkqenqalskklndthnelsdikqkvqdtnlevnklk 400
nilkseesenrqimeqlrkanedaenwenkarqteaenntlklelitaea 450
egnrlkekvdalnreveqhlnaersyksqiatlhkslvkmeeelqkvqfe 500
kvsaladlsstrelcikldsskellnrqlvakdqeiemmeneldsarsei 550
ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvtqleadlditkrqlgterfereravqelrr 650
qnyssnaynlgpmkpntkchsperahhrspdrgldRSLEENLCYrdf 697
```

N-myristoylation site: **PA G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}**

GATLND 260-265 GLDRSL 683-688

```
1       10        20        30        40
mmrnrsksprrpsptsraancdvellkstardreelkcmlekyerhlaei 50
qgnvkvltserdktfllyeqaqeeiarlrremmksckspksttahailrr 100
veterdvaftdlrrmtterdslrerlkiaqetafnekahleqrieelect 150
vhnlddermeqmanmtlmketittvekemkslarkamdteselgrqkaen 200
nslrllyentekdlsdtqrhlakkkyelqltqekimcldekidnftrqni 250
aqreeisilGATLNDlakekeclqacldkkseniaslgeslamkektisg 300
mkniiaemeqasrqstealimceqdisrmrrqldetndelgqiarerdil 350
ahendnlqeqfakvkqenqalskklndthnelsdikqkvqdtnlevnklk 400
nilkseesenrqimeqlrkanedaenwenkarqteaenntlklelitaea 450
egnrlkekvdalnreveqhlnaersyksqiatlhkslvkmeeelqkvqfe 500
kvsaladlsstrelcikldsskellnrqlvakdqeiemmeneldsarsei 550
ellrsqmtnerismqnleallvanrdkeyqsqialqekeseiqllkehlc 600
laenkmaiqsrdvaqfrnvvtqleadlditkrqlgterfereravqelrr 650
qnyssnaynlgpmkpntkchsperahhrspdrGLDRSLeenlcyrdf 697
```

Potential phosphorylation sites for protein kinase GSK3 (C-terminal +4 S must be prephosphorylated):

- T-117 RMTTERDS
- S-286 NIASLGES
- T-482 QIATLHKS
- T-666 KPNTKCHS

No phosphorylation sites for protein kinase MLCK found.

Potential phosphorylation sites for protein kinase p34cdc2:

- S-8 RSKSPRR
- S-670 KCHSPER
- S-678 HHRSPDR

Potential phosphorylation sites for protein kinase p70s6k:

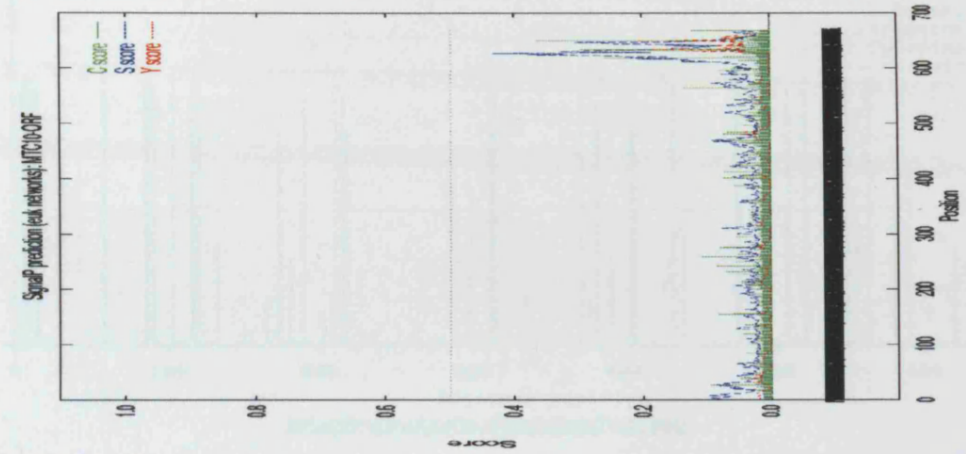
- S-8 RNRSKSPRR

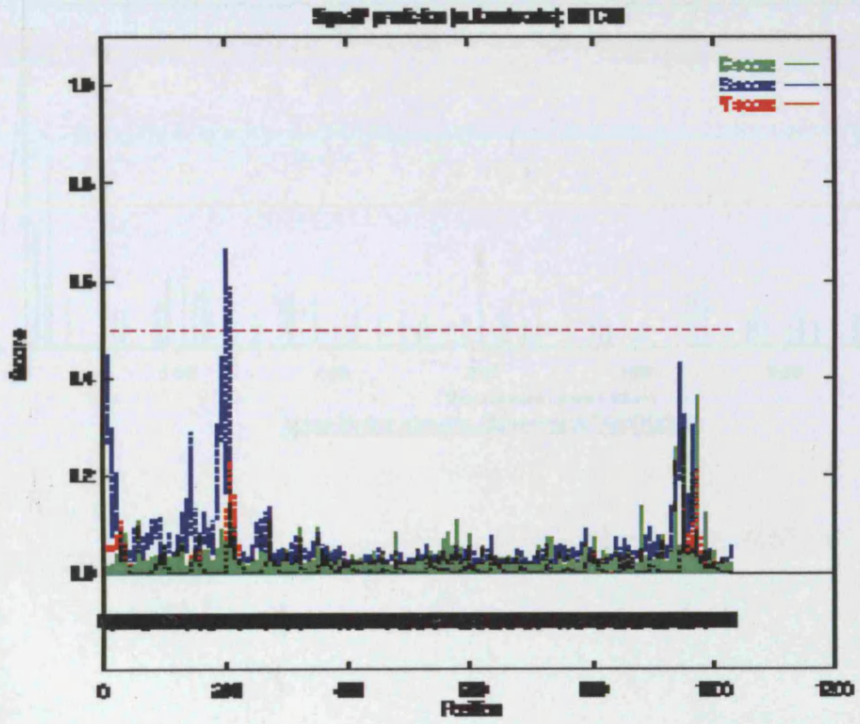
Potential phosphorylation sites for protein kinase PKA:

- S-6 RNRSKSP
- S-13 RRPSPTS
- T-64 RDKTFLL
- T-103 RVETERD
- T-116 RRMTTER
- S-121 RDSLRE
- T-316 RQSTEAL
- T-434 RQTEAE
- S-563 RISMQN

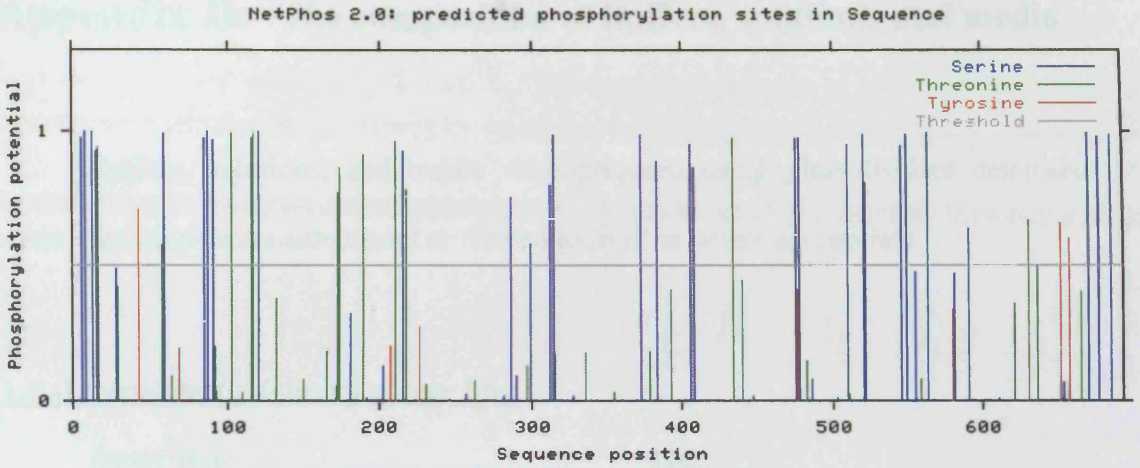
Potential phosphorylation sites for protein kinase PKG:

- S-13 RRPSPTS
- T-116 RRMTTER

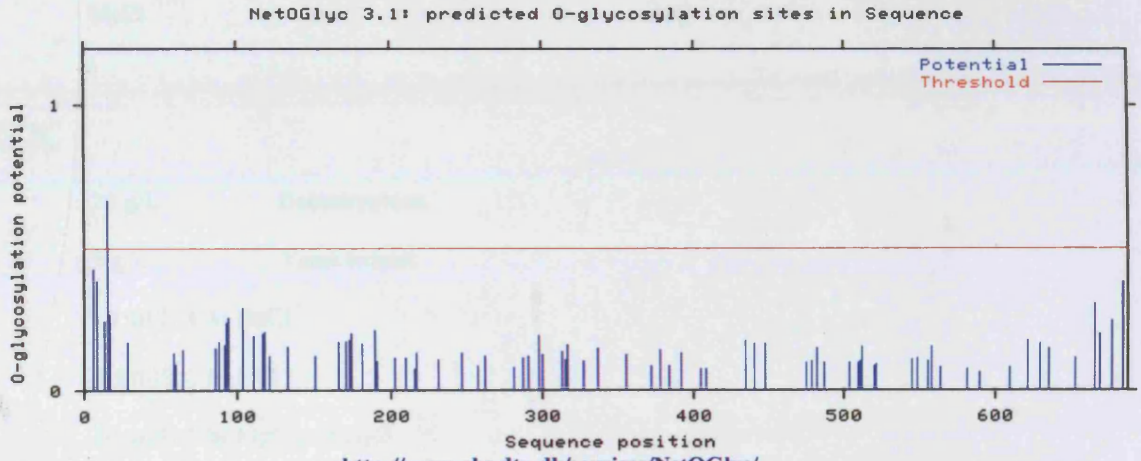




http://www.cbs.dtu.dk/services/TMHMM/



http://www.cbs.dtu.dk/services/NetPhos/



http://www.cbs.dtu.dk/services/NetOGlyc/

Appendix D: The composition of buffers, solutions and media

Buffers, solutions, and media were prepared using glass-distilled deionised water. Solutions were autoclaved or filtered to sterilise where appropriate.

LB (Lauria-Bertani) medium, per litre

Bactotryptone	10g
Yeast extract	5g
NaCl	10g

SOC

20 g/L	Bactotryptone
5 g/L	Yeast extract
10 ml/L	1 M NaCl
2.5 ml/L	1 M KCl
10 ml/L	1 M $MgCl_2 \cdot 6\ H_2O$
10 ml/L	1 M $MgSO_4 \cdot 7\ H_2O$
20 ml/L	1 M glucose
960 ml/L	deionised H_2O

TAE buffer (50X), per litre

Tris base	242g
Glacial acetic acid	57.1 ml
EDTA 0.5M, pH8.0	100 ml

Adjust to pH7.2 and bring the final volume to 1 litre with distilled water

SSC buffer(20X)- 500ml

NaCl	87.7
Na-citrate	44.1g

TE buffer (1X), per litre

Tris base	1.21g
EDTA, sodium salt	0.372g

Solutions for RNA ISH

Hybridisation Buffer

Substance	1 ml	10 mls	Final Conc.
10 x Salts	100 µl	1 ml	1x
formamide	500 µl	5 mls	50%
50% Dextran Sulphate	200µl	2 mls	10%
tRNA (10 mg/ml)	100 µl	1 ml	1 mg/ml
50 x Denhardts	5 µl	50 µl	1 x
ddwater	90 µl	900 µl	

10x Salts

Substance	Amount
NaCl	114 g
Tris HCL pH7.5	14.04 g
Tris Base	1.34 g
$NaH_2PO_4 \cdot 2H_2O$	7.8 g
Na_2HPO_4	7.1 g
0.5 M EDTA	100 mls
water to 1000 mls	

Washing Solution

1x SSC, 50% formamide, 0.1% Tween-20.

MABT

100 mM Maleic acid, 150 mM NaCl, 0.1% Tween-20

For 5x stock: the maleic acid dissolved, titrated with NaOH and the rest was added.

(e.g. for 1 litre: 58 g maleic acid, titrate, 43.8 g NaCl, 5 mls Tween-20)

(a 5X stock of MAB was made and then added the tween as needed as it goes off)

M-Block

MABT + 2% Blocking Reagent[1] + 20% heat-inactivated serum.

([1]Blocking Reagent is Boehringer's own made up in maleic acid at 10% stocks, autoclaved and frozen in aliquots)

500ul stocks of Mblock made and by using 5XMAB, 0.5% tween, 10% blocking reagent; stored at –20C. To make working solution of Mblock, 500ul frozen Mblock was used, then 500ul lamb serum and 1.5ml water were added, and finaly heated at 70C for 30min then cooled.

Staining Buffer

For 25 mls:

Substance	Amount	Final Conc.
5 M NaCl	0.5 mls	100 mM
1 M MgCl	1.25 mls	50 mM
1 M Tris pH9.5	2.5 mls	100 mM
Tween-20	25 µl	0.1%
To 25 mls with water		

Dissolving Buffer

50% Glycerol	2.5ml
dH2O	6.55ml
0.5M Tris-HCl pH6.8	0.8ml
10% SDS	2.5ml
2-Mercaptoethanol	0.6ml
Bromophenol Blue	small end of spatula

2X Treatment Buffer

0.125 M Tris, pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, and 0.05% bromphenol blue

0.5 M Tris, pH 6.8	25%
DTT	0.031 gr/ml

Sample Buffer (for yeast protein extraction)

pH=6.8

Tris-HCl 0.5 M,

NaV (sodium orthovanadate, inhibitor of phosphatase)

SDS-PAGE Running Buffer

pH8.3 - for 4.5litres

0.025M Tris 13.6g

0.192M Glycine 64.8g

0.1% SDS 45ml

TGM Transfer Buffer pH 8.3 - 5 Litres

25mM Tris 15.15g

192mM Glycine 72.0g

20% Methanol 1 litre

Dissolve Tris and Glycine in 4 litres dH_2O then add Methanol. (pH shouldn't need adjusting but check.)

Western Blot Blocking Buffer

5% Fat-free dried milk in PBS

Yeast

SD - synthetic dropout

6.7g	Yeast nitrogen base without amino acids
850ml	H_2O
100ml	appropriate sterile 10xdropout solution *

For plates, 20g agar was added to the above. Nutrients omitted according to the selection required, see below.

10 x dropout solution

200 mg/L	L-Adenine hemisulfate salt
200 mg/L	L-Arganine HCl
200 mg/L	L-Histidine HCl monohydrate
300 mg/L	L-Isoleucine
1000 mg/L	L-Leucine
300 mg/L	L-Lysine HCl
200 mg/L	L-Methionine
500 mg/L	L-Phenylalanine
2000 mg/L	L-Threonine
200 mg/L	L-Tryptophan
300 mg/L	L-Tyrosine
200 mg/L	L-Uracil
1500 mg/L	L-Valine

YPD

20g/L	Difco peptone
10g/L	Yeast extract

For plates, 20g/L agar was added to the above. For Kanamycin-containing medium, add 10-15 mg/L Kanamycin.

PEG/LiAc solution - Polyethylene glycol/Lithium acetate

For 100ml of solution:

80 ml	50% PEG 4000
10 ml	10 x TE
10 ml	10 x LiAc

Herring testes carrier DNA - **Supplied by Clontech** 10µg/µl

DMSO - Dimethyl sulphoxide 100%

PBS - 67mM potassium phosphate pH 7.5

Lyticase - Supplied by Clontech

Phenol:chloroform:isoamyl alcohol - (25:24:1)

Ammonium acetate - 10 M

Solution P1 - *E. coli* mini-prep resuspension buffer (pH = 8)

50 mM	Glu
25mM	Tris HCl
10mM	EDTA

Solution P2 - *E. coli* mini-prep lysis solution

0.2 M	NaOH
1%	SDS

Solution P3 - *E. coli* mini-prep precipitation solution

For 200 µl of solution:

120 µl	5M Potassium acetate
23 µl	37% acetic acid
47 µl	dd H_2O

EtBr - Ethidium bromide

DNA taq pol - *T. aquatieus* thermostable DNA polymerase

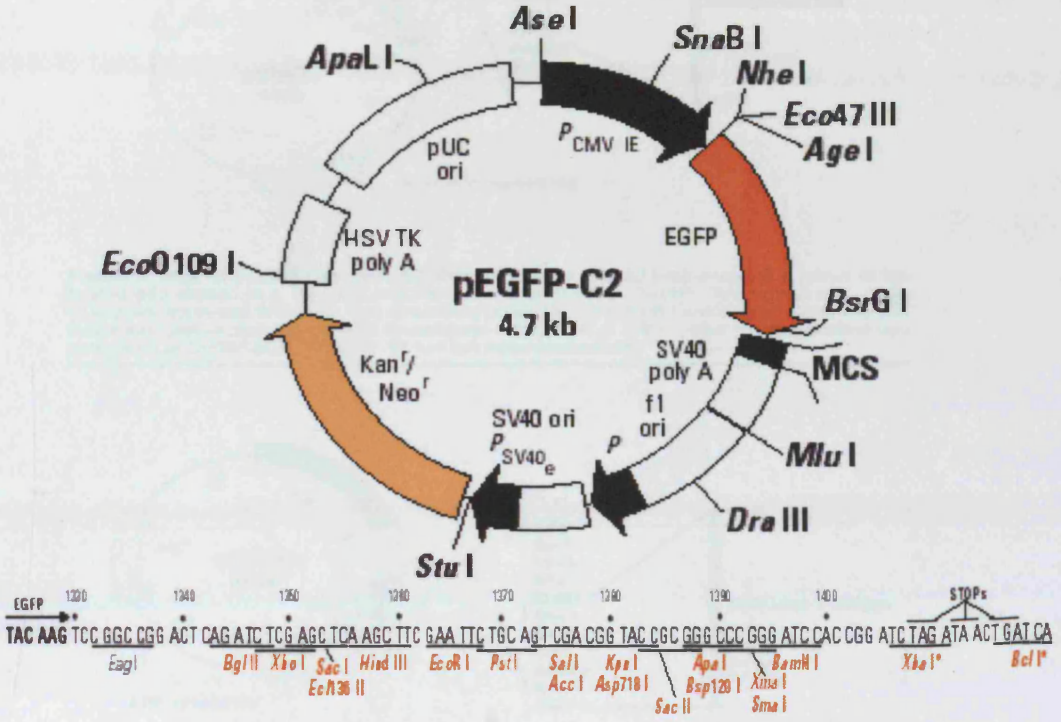
Supplied by Advanced Biotechnologies

10 x buffer IV - Supplied by Advanced Biotechnologies

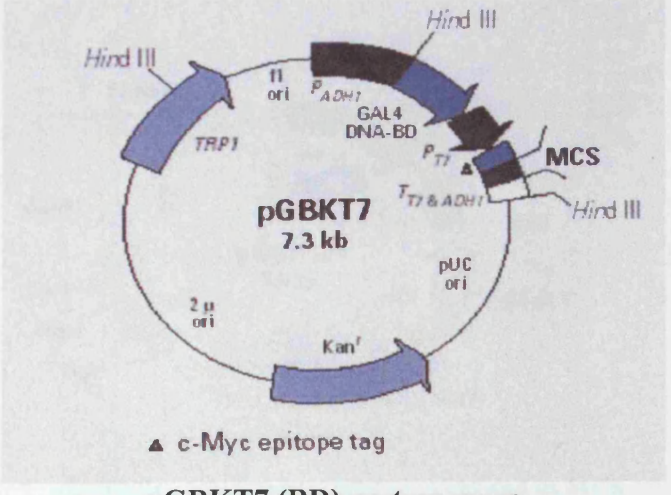
Contents of vial:

1.25 ml of:	200 mM $(NH_4)_2SO_4$	
	750 mM Tris-HCl, pH 8.8 at 25°C	
	0.1% (v/v)	Tween® 20
	15 mM	MgCl

Appendix E : Vectors maps and information



pEGFP-C2 (BD) vectors map and its MCS



pGBKT7 (BD) vectors maps

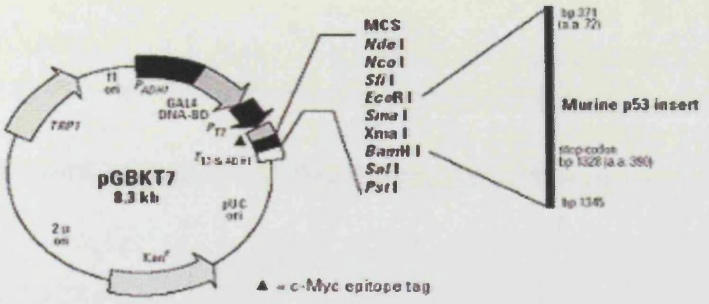


Figure 3. Map of pGBKT7-53 DNA-BD Control Vector. pGBKT7-53 encodes a fusion of the murine p53 protein (a.a. 72–390) and the GAL4 DNA-BD (a.a. 1–147). The murine p53 cDNA (GenBank Accession #K01700) was cloned into pGBKT7 at the *Eco*R I and *Bam*H I sites. The p53 insert was derived from the plasmid described in Iwabuchi *et al.* (1993); plasmid modification was performed at CLONTECH. pGBKT7-53 has not been sequenced.

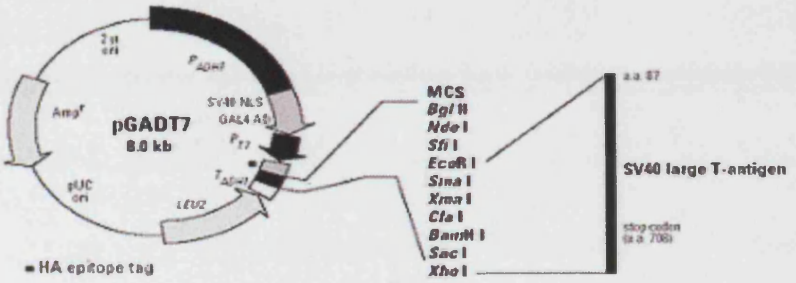
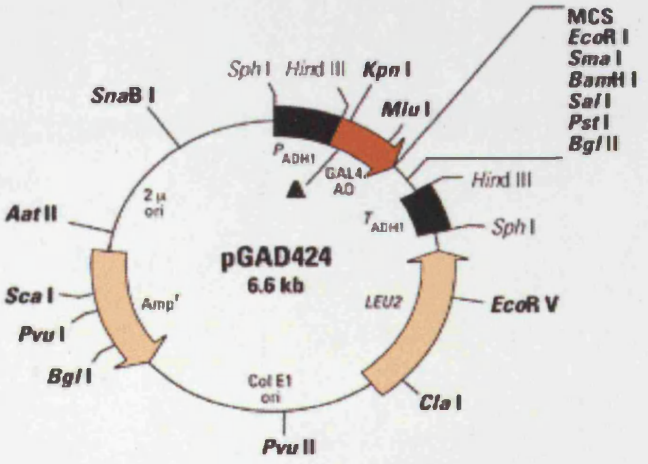


Figure 4. Map of pGADT7-T AD Control Vector. pGADT7-T encodes a fusion of the SV40 large T-antigen (a.a. 86–708) and the GAL4 AD (a.a. 768–881). The SV40 large T DNA (GenBank Locus SV4CG) was derived from a plasmid referenced in Li & Fields (1993) and was cloned into pGADT7 using the *Eco*R I and *Xho* I sites. pGADT7-T has not been sequenced.



pGAD424 plasmid map (BD Clontech) and its MCS