GENETIC ANALYSIS

OF

MALE INFERTILITY

A thesis submitted

in fulfillment of the

requirements for the

Degree of

Doctor of Philosophy

in

Cellular and Molecular Biology

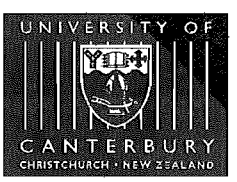
in the

School of Biological Sciences

University of Canterbury

by

Tanya Penina Harris



University of Canterbury

2004

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Approximately one in twenty men has impaired spermatogenesis due to mutation of genes involved in the establishment or maintenance of fertility. Our understanding of male infertility is complicated by the variable phenotypes produced by similar genetic changes, largely due to the practise of screening a single fertility gene in isolation.

This thesis aimed to increase our understanding of the role of synergistic mutations in relation to differences in semen quality. Each sample was analysed for mutation in: CAG trinucleotide repeat variation in the X-linked androgen receptor *(AR)* gene, micro-deletion within the three Y chromosome azoospermic factor (AZF) regions, and CAG trinucleotide repeat variation and exonuclease domain mutation in the nuclear polymerase gamma (*POLγ*) gene. These genes have been associated with reduced semen quality in past research.

Each gene region was amplified by polymerase chain reaction (PCR), followed by sequencing. Suspected AZF micro-deletions were confirmed by Southern blot hybridisation. Associations with semen quality were evaluated using either a t-test or G-test for independence at $\alpha=0.05$.

Yq AZF micro-deletions were observed in 6.6% (14/211) of men with poor semen quality but not in normozoospermic samples (0/104); $P<0.001$). Micro-deletion frequency was greatest in azoospermic and severely oligoasthenozoospermic individuals (15% and 11.5%, respectively).

AR CAG repeat length ranged from 9-38 CAG repeats in the normozoospermic population (n=98) and 13–31 CAG repeats in men with poor semen quality (n=119). Variation in *AR* CAG trinucleotide repeat number was not significantly related to poor semen quality ($P>0.05$).

Variation in *POLγ* CAG repeat number was not significantly different between normozoospermic men (n=93) and men with poor semen quality (n=182); $P>0.05$. No nucleotide changes were observed in any of the three *POLγ* exonuclease motifs (n=83 normozoospermic and 191 non-normozoospermic motif I, 61 and 65 motif II, and 60 and 64 motif III).

Although most gene regions did not show an association with poor semen quality on their own, there was a general trend towards greater severity of impaired spermatogenesis with the presence of both Yq micro-deletion and mitochondrial DNA substitutions or moderately expanded *AR* CAG repeats. These results support the idea that male infertility is a complex process, due to many factors, some of which act dominantly and others act in concert.

CHAPTER ONE INTRODUCTION

1.1 Introduction

The majority of idiopathic male infertility cases are thought to be due to genetic changes in one or more genes involved in the establishment and maintenance of male sexual characteristics and sperm production (De Kretser and Baker 1999; Yong et al. 2000a). Frequently, studies of male infertility are hampered by the considerable variation observed in fertility status resulting from seemingly similar gene mutations (De Kretser and Baker 1999). In addition, many associations between gene mutation and reduced semen quality are considered controversial, showing population specific correlations (Erasmuson et al. 2003). Although recent research indicates there are an enormous number of genes with testes specific expression that are predicted to be involved in the establishment and maintenance of sperm production, most research focuses on associations between a particular gene mutation and reduced semen quality in isolation; very few studies have analysed a set of candidate fertility genes at once. My thesis research aims to analyse several gene regions known or suspected to be involved in normal fertility in a single population of men: the X chromosome androgen receptor gene (*AR*), the Y chromosome azoospermic factor (AZF) region, and the nuclear polymerase gamma (*POLγ*) gene whose gene product has an important role in mitochondrial DNA replication.

These three genes were chosen to cover important areas of fertility research. The Y chromosome has been well studied, and recently DNA sequenced, it is important not only for sex determination but semen quality as well (Skaletsky et al. 2003). The X chromosome is the homologous partner to Y, although they share very little identical sequence and so cannot recombine. Recently nine novel testis-specific genes been found on the X chromosome demonstrating it plays an important role in influencing male semen quality (Wang et al. 2001). The *AR* gene is X-linked but involved in male sexual development (Loy and Yong 2001). Similarly, the *POLγ* gene is autosomal but is involved in replicating the mitochondrial genome which plays an essential role in cellular energy production (Longley et al. 1998). High frequencies of mitochondrial DNA mutation have been implicated in defective sperm function due to a reduction in electron transport

chain enzyme activity (Ruiz-Pesini et al. 1998). It is hypothesised that errors in replication might contribute to the elevated number of single nucleotide polymorphism of the mitochondrial genome (Holyoake et al. 2001). Both the Y chromosome and mitochondrial genome generate enormous research interest; and the discovery of fertility genes on X has renewed research interest in it.

1.1.1 Male Fertility

Male sexual development is a complex process involving the interaction of numerous gene products and pathways, many of which have not been characterised, others whose role is ambiguous, and more that we have not yet identified. Recently more than a thousand novel testes-specific genes were discovered by expression analysis in mouse spermatogonia (Wang et al. 2001; Schultz et al. 2003). A surprising number of these were encoded on the X chromosome and expressed in premeiotic male germ cells suggesting a role in spermatogonial cell establishment and proliferation (Wang et al. 2001). Importantly, this indicates that females play a larger role in shaping the fertility status of men than ever before suspected, because akin to mitochondrial DNA, these factors are inherited through the female lineage since X carrying sperm create female embryos (Wang et al. 2001). Hence, although the Y chromosome encodes more than just the male determining gene *SRY*, it does not hold the monopoly on spermatogenesis genes, and like male pattern baldness the mother is an important determinant of male fertility, yet indicators of abnormality may not be apparent in her.

Male infertility affects approximately one in twenty men (De Kretser and Baker 1999; Yong et al. 2000b). Rarely is the cause attributable to an obvious clinical condition, such as cryptorchidism, testicular atrophy, epididymal obstructions, or infections; more often, the reason for poor semen quality is unexplained; otherwise known as idiopathic male infertility. For these men the underlying basis is likely to be due to genetic mutation of a gene whose normal function is essential for normal sperm production, however diagnosis is hindered by the deficits in our understanding and knowledge of the genes involved in

fertility (Yong et al. 2000b; McLachlan and de Kretser 2001). Diagnosis is further complicated by the fact that semen quality does not always explain fertility status, since men with normozoospermia can be infertile and men with reduced semen quality can father offspring without medical intervention (Matzuk and Lamb 2002; Thangaraj et al. 2003; Jensen et al. 2004).

The use of animal models such as *Drosophila* and mouse has aided our understanding of the roles of specific genes in germ cell establishment and maturation. Saturation mutagenesis screening in *Drosophila* gave weight to the idea that infertility can result from the combined effect of mutation in many genes (Hackstein et al. 2000). The ability to manipulate the mouse genome has provided a means of deciphering the order of events in meiosis, allowed the creation of mouse models of diseases that show associated infertility, identified important signalling pathways and proteins involved in reproduction and faciliated targeted deletion of candidate spermatogenesis genes in order to determine function (Cooke and Saunders 2002; Matzuk and Lamb 2002). More than 200 mouse models of infertility have been created contributing to our understanding of male gametogenesis and reproduction using targeted gene disruption and chemical mutagenesis, and even spontaneous mutants, such as hypogonadal (*hpg*) and testicular feminisation (*tfm*) (Matzuk and Lamb 2002; Barchi and Jasin 2003; Libby et al. 2003).

In addition mouse models can be used to trial potential infertility treatments; currently germ cell transplantation is showing promise as both a means of expressing gene mutation and preserving the reproductive capacity of young men undergoing cancer treatment or infertility (Cooke and Saunders 2002; Shinohara et al. 2003).

1.1.2 The Testes

In addition to germ cells the testes contain two other important cell types; Leydig cells, which produce male sex androgen (testosterone), and Sertoli cells. Sertoli cells are a fixed population of non dividing cells that are linked by tight junctions to form a physical

barrier between the body's circulatory system and the testicular environment so that male germ cells are maintained in an immunologically privileged location (Browder et al. 1991; Shinohara et al. 2003). This separation is necessary because spermatogenesis begins at puberty long after the body's immune system has developed; if no barrier existed, sperm cells would be recognised as foreign bodies and eliminated by the immune system. Sertoli cells produce androgen-binding protein which transports testosterone intracellularly and acts as a testosterone reservoir within the seminiferous tubule, maintaining a high level of testosterone around the developing spermatocytes (Browder et al. 1991).

Germ cells are arranged specifically according to the maturation stage within seminiferous tubules. Spermatogonia (stem cells) and young spermatocytes are near the basal layer close to Sertoli cells; as they mature they progress towards the centre of the tubule for transport into the lumen and the epididymus (Browder et al. 1991).

1.1.3 Spermatogenesis

Spermatogenesis involves three specific phases: proliferation of spermatogonia, meiosis, and spermiogenesis; at any stage abnormal function can result in reduced sperm production (Matzuk and Lamb 2002). In the initial stages of spermatogenesis, spermatogonia divide by mitosis generating two daughter cells, one that remains to divide again (self-renewal) and a primary spermatocyte that enters the first meiotic division to produce secondary spermatocytes, and then the reductive meiotic division to produce round, haploid spermatids (Figure 1.1) (Foresta et al. 2001).

Spermatogonia have the highest proliferation rate in the body. A balance between apoptosis stimulating and inhibiting proteins is essential for normal germ cell establishment. Female primordial follicles, which proliferate only in the developing embryo, produce a finite quantity of oocytes. In contrast, male germ cells proliferate throughout the lifetime of a male. The spermatogonia of a 40 year old man have

replicated 26 times that of a female ovum (Figure 1.1). It is this difference that is thought to result in the higher rate of base substitution in males than females (Crow 2000).

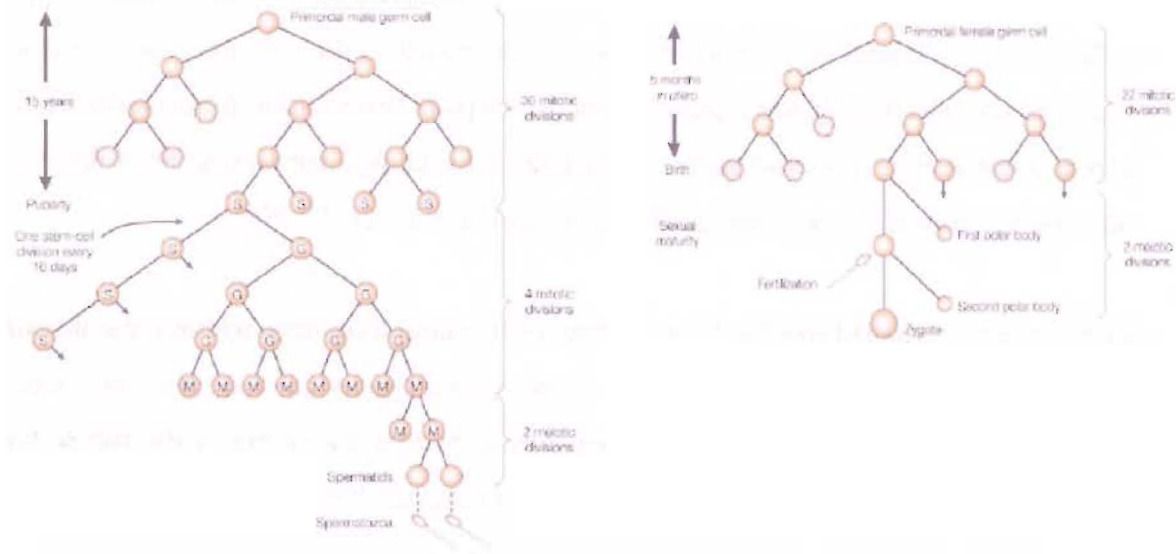


Figure 1.1. Cell divisions during oogenesis and spermatogenesis. The total replicative cell divisions in the life of an ovum is 23. In males the total number is dependant on the age of the man. S, stem cell; G, gonial cells; M, meiotic cells. Reproduced from Crow (2000).

Spermatozoa are specialized cells, dedicated to fertilisation of the oocyte (Figure 1.2) (Vernet et al. 2004). The sperm cells that are released from the germinal epithelium of the mammalian testis are unable to fertilise an oocyte (without the aid of artificial reproductive techniques) (Aitken et al. 2004). Immature spermatids acquire functional competence during two post-testicular maturation stages (Aitken et al. 2004). Spermatids undergo the first stage of maturation as they travel along the epididymus which results in nuclear condensation and a reduction in cytoplasm content, replacement of histone proteins with protamines, formation of the acrosome, arrangement of mitochondria in the midpiece, and motility (Foresta et al. 2001; Jimenez et al. 2004; Vernet et al. 2004). In humans it takes approximately 10 days for spermatozoa to reach the storage area at the end of the epididymidis (Vernet et al. 2004).

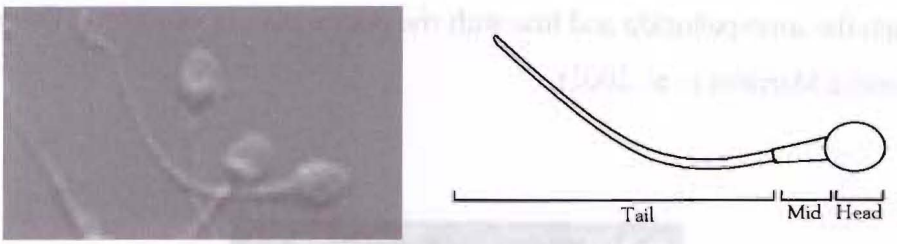


Figure 1.2. A human sperm cell. Microscope image on left (Jimenez et al. 2004), and schematic representation on right. Mature spermatids have three parts, the head, midpiece (Mid), and tail. Within the head is the acrosome vesicle. Mitochondria are arranged in a helix surrounding a central axial filament in the midpiece.

The second stage; capacitation, occurs only once sperm cells have entered the female reproductive tract (Taylor 2001; Aitken et al. 2004). The molecular basis of sperm capacitation is believed to involve the establishment of a unique, cell-specific signal transduction pathway involving a redox-regulated, cAMP-mediated, tyrosine phosphorylation cascade (Aitken 1997; Aitken et al. 2004). The initial stages of capacitation involve a rise in intracellular calcium and cAMP levels, the production of ROS, and the development of highly vigorous movement (hyperactivation) (Aitken 1997). The production of low levels of ROS seems to be important in the initiation and regulation of tyrosine phosphorylation since scavenging of H_2O_2 results in a loss of spermatocyte sensitivity to progesterone and zona-pellucida glycoproteins (Aitken 1997). The final stage of capacitation involves the influx of calcium into the sperm head in two waves (Aitken 1997; Tomes et al. 2004).

Thus once spermatocytes reach the oocyte they are primed for action, but in order to penetrate the oocyte (and achieve fertilisation), they must undergo the acrosome reaction (Aitken 1997). The acrosome is a membrane-bound vesicle that overlies the nucleus of the mature spermatozoon (Figure 1.3) (Tomes et al. 2004). Sperm cells undergo the acrosome reaction upon contact with progesterone, in the surrounding follicular fluid, or glycoproteins (ZP3) of the oocyte zona-pellucida (Gonzalez-Martinez et al. 2002). This triggers fusion of the acrosome vesicle with the sperm plasma membrane followed by exocytosis of the acrosomal contents (hydrolytic enzymes) allowing the spermatocyte to

pass through the zona-pellucida and fuse with the oocyte plasma membrane (Jaiswal et al. 1998; Gonzalez-Martinez et al. 2002).

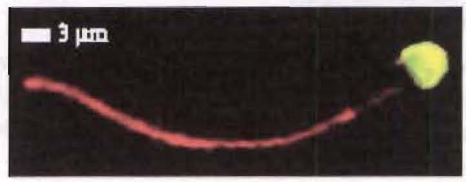


Figure 1.3. Staining of the acrosome vesicle. A spermatid showing staining of the tail (red) and acrosome vesicle (yellow) in the sperm head. A scale line of 3μm is given. Reproduced from Tomes et al. (2004).

The ability of the spermatocyte to fuse with the oocyte membrane is enhanced by a sperm plasma membrane rich in polyunsaturated fatty acids (Vernet et al. 2004). However, this leaves spermatozoa susceptible to lipid peroxidation as a result of reactive oxygen species (ROS) attack. Paradoxically, evidence suggests sperm cells actively produce low concentrations of ROS (O_{2}-, and H_2O_2) presumably because they play an important positive role in the regulation of capacitation and actually enhance the ability of spermatocytes of fuse with the oocyte membrane (Aitken et al. 1989; Vernet et al. 2004). This necessitates a balance between ROS production and scavenging, since high concentrations of ROS damage sperm cells but removal inhibits capacitation and fertilisation (Taylor 2001; Aitken et al. 2004).

1.1.4 Artificial Reproductive Techniques

Low numbers of mature sperm can be found even in azoospermic men. These sperm are capable of achieving fertilisation when injected directly into an egg, in a process called intracytoplasmic sperm injection (ICSI) (Mulhall et al. 1997). It was presumed that epididymal and especially testicular sperm lacked the ability to fertilise an oocyte since testicular sperm had not undergone the necessary maturation changes (formation of the acrosome, and chromatin and protein changes) and neither type of sperm cell has undergone capacitation within the female reproductive tract. However, the relative

success of artificial reproductive techniques (ART) suggests that the changes that take place during maturation and capacitation are not absolutely necessary for normal embryonic development (Okabe et al. 1998). Hence the advent of ART has removed the selection pressure on fertility. However, it seems unlikely that this practise will increase the quality of sperm in future generations; we may only be prolonging the existence of deleterious traits in our gene pool. Widespread concern surrounds the possibility that infertility-causing genetic mutations could be passed from father to sons or grandsons who then have a greater likelihood of experiencing infertility as adults (Aitken et al. 1998; De Kretser and Baker 1999; Foresta et al. 2001).

1.2 Thesis Research Aims

The objective of my thesis was to elucidate the role of mutation in a number of genes known or suspected to perform a vital role in male fertility, in a single population of males, in an effort to better understand why such variation in semen quality exists. The research aimed to establish how the frequency of synergistic mutations (multiple mutations in several fertility genes) relates to the occurrence and severity of impaired spermatogenesis by analysing the following three gene regions for the presence or absence of mutation: (1) Micro-deletion within the three Y chromosome azoospermic factor (AZF) regions; (2) CAG trinucleotide repeat variation in the X-linked androgen receptor *(AR)* gene; and (3) CAG trinucleotide repeat variation and the three exonuclease motifs of the nuclear polymerase gamma (*POLγ*) gene.

The Y chromosome has a well established role in both sex-determination and sperm production. Deletions within one particular region, the azoospermic factor (AZF) region are found in approximately 10% of men with idiopathic infertility (Foresta et al. 2001). Although there is abundant evidence linking candidate spermatogenesis genes within each of the three AZF regions and cases of reduced semen quality, there is no absolute relationship between particular Y chromosome micro-deletion and characteristics of semen quality (Foresta et al. 2001). In several such studies it is thought that the particular

AZF deletion interval is likely to extend beyond the set of STS (sequence tagged site) markers into adjoining AZF regions (Foresta et al. 2000). To avoid obscuring relationships between particular combinations of Yq11 micro-deletions and semen characteristics a selection of STS markers representing all three AZF regions were screened for micro-deletion in this study.

The X-linked *AR* gene has a long established role in male sexual development and has also been implicated in causing male infertility in otherwise healthy men through moderate expansion of a trinucleotide repeat region. Many population studies have reported a correlation between low sperm count and *AR* CAG repeat length and suggest the association is due to a reduction in male-vigour gene transcription initiation by the AR with increasing numbers of CAG trinucleotide repeats. However, the association between moderately expanded repeat regions and reduced sperm production is controversial, with numerous population studies reporting no such association. The hypothesis is supported by several *in vitro* studies demonstrating a subtle reduction in AR activity with increases in the *AR* CAG repeat length. In this study variation in trinucleotide repeat number was evaluated in relation to semen quality.

The accumulation of mitochondrial DNA (mtDNA) single nucleotide polymorphisms (SNPs) and deletions is linked with an increased risk of reduced semen quality (Kao et al. 1995; Kao et al. 1998; Holyoake et al. 2001). The suggestion that aberrant polymerase gamma (*POLγ*) replication might result in the increased frequency of mtDNA polymorphisms and deletions in these men has been somewhat supported by cell culture experiments and DNA screening in PEO (progressive external ophthalmoplegia) patient families. Substitution of several highly conserved amino acid residues within the exonuclease and polymerase domains (D198A, E200A, Y955C) results in error-prone DNA replication and an increased mitochondrial DNA mutational load during cell culture (Spelbrink et al. 2000). These same substitutions have been observed in PEO patients whose polymerases show low dNTP binding-affinity leading to error-prone mtDNA synthesis (Lamantea et al. 2002; Copeland et al. 2003). Each sample was screened for mutation within the three exonuclease motifs that make up the exonuclease

domain. Importantly, 91 samples (44 normozoospermic and 47 men with poor semen quality) from the study population have been previously analysed for the presence or absence of single nucleotide polymorphisms (SNPs) of the mitochondrial genome enabling analysis of the potential relationship between *POLγ* mutation and mitochondrial polymorphisms.

There is already some evidence linking the *POLγ* gene and reduced semen quality; the amino-terminal CAG trinucleotide repeat region of *POLγ* has recently been associated with oligozoospermia and normozoospermic-infertility (Rovio et al. 2001; Jensen et al. 2004). The CAG trinucleotide repeat is unique to the human *POLγ* and is commonly found to encode 10 CAG units in the population (Rovio et al. 1999). The contribution of *POLγ* CAG repeat variation to fertility status was analysed in this population.

Each sample was analysed for mutation in each of the three genes and mutation frequencies compared against semen characteristics of the sample population. It is predicted that men with reduced semen quality will display a greater frequency of nucleotide changes in these genes, and that the highest mutational load will be observed in men with the most severely reduced semen quality (azoospermia and severely oligoasthenozoospermia).

References:

Aitken, R.J. 1997. Molecular mechanisms regulating human sperm function. *Mol Hum Reprod* 3: 169-73.

Aitken, R.J., J.S. Clarkson, and S. Fishel. 1989. Generation of reactive oxygen species, lipid peroxidation, and human sperm function. *Biol Reprod* 41: 183-97.

Aitken, R.J., E. Gordon, D. Harkiss, J.P. Twigg, P. Milne, Z. Jennings, and D.S. Irvine. 1998. Relative impact of oxidative stress on the functional competence and genomic integrity of human spermatozoa. *Biol Reprod* 59: 1037-46.

Aitken, R.J., A.L. Ryan, M.A. Baker, and E.A. McLaughlin. 2004. Redox activity associated with the maturation and capacitation of mammalian spermatozoa. *Free Radic Biol Med* 36: 994-1010.

Barchi, M. and M. Jasin. 2003. Seeking new meiotic genes. *Proc Natl Acad Sci U S A* 100: 15287-9.

Browder, L.W., C.A. Erickson, and W.R. Jeffery. 1991. *Developmental Biology.* Saunders College Publishers, Philedelphia.

Cooke, H.J. and P.T. Saunders. 2002. Mouse models of male infertility. *Nat Rev Genet* 3: 790-801.

Copeland, W.C., M.V. Ponamarev, D. Nguyen, T.A. Kunkel, and M.J. Longley. 2003. Mutations in DNA polymerase gamma cause error prone DNA synthesis in human mitochondrial disorders. *Acta Biochim Pol* 50: 155-67.

Crow, J.F. 2000. The origins, patterns and implications of human spontaneous mutation. *Nat Rev Genet* 1: 40-7.

De Kretser, D.M. and H.W. Baker. 1999. Infertility in men: recent advances and continuing controversies. *J Clin Endocrinol Metab* 84: 3443-50.

Erasmuson, T., I.L. Sin, and F.Y. Sin. 2003. Absence of association of androgen receptor trinucleotide expansion and poor semen quality. *Int J Androl* 26: 46-51.

Foresta, C., A. Ferlin, and E. Moro. 2000. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum Mol Genet* 9: 1161-9.

Foresta, C., E. Moro, and A. Ferlin. 2001. Y chromosome microdeletions and alterations of spermatogenesis. *Endocr Rev* 22: 226-39.

Gomas, K. P. The Association of Single Nucleotide Polymorphisms in Mitochondrial ND1 and ND2 Genes with Human Male Fertilising Potential. Christchurch, University of Canterbury. 2002. p133. (Thesis: MSc: Cellular and Molecular Biology).

Gonzalez-Martinez, M.T., M.A. Bonilla-Hernandez, and A.M. Guzman-Grenfell. 2002. Stimulation of voltage-dependent calcium channels during capacitation and by progesterone in human sperm. *Arch Biochem Biophys* 408: 205-10.

Hackstein, J.H., R. Hochstenbach, and P.L. Pearson. 2000. Towards an understanding of the genetics of human male infertility: lessons from flies. *Trends Genet* 16: 565-72.

Holyoake, A.J. Aspects of mitochondrial DNA mutations in relation to human male fertilising potential. Christchurch, University of Canterbury. 1998. p154. (Thesis: PhD: Cellular and Molecular Biology).

Holyoake, A.J., P. McHugh, M. Wu, S. O'Carroll, P. Benny, I.L. Sin, and F.Y. Sin. 2001. High incidence of single nucleotide substitutions in the mitochondrial genome is associated with poor semen parameters in men. *Int J Androl* 24: 175-82.

Jaiswal, B.S., A. Cohen-Dayag, I. Tur-Kaspa, and M. Eisenbach. 1998. Sperm capacitation is, after all, a prerequisite for both partial and complete acrosome reaction. *FEBS Lett* 427: 309-13.

Jensen, M., H. Leffers, J.H. Petersen, A. Nyboe Andersen, N. Jorgensen, E. Carlsen, T.K. Jensen, N.E. Skakkebaek, and E. Rajpert-De Meyts. 2004. Frequent polymorphism of the mitochondrial DNA polymerase gamma gene (POLG) in patients with normal spermiograms and unexplained subfertility. *Hum Reprod* 19: 65-70.

Jimenez, A., W. Zu, V.Y. Rawe, M. Pelto-Huikko, C.J. Flickinger, P. Sutovsky, J.A. Gustafsson, R. Oko, and A. Miranda-Vizuete. 2004. Spermatocyte/spermatid-specific thioredoxin-3, a novel golgi apparatus-associated thioredoxin, is a specific marker of aberrant spermatogenesis. *J Biol Chem.*

Kao, S., H.T. Chao, and Y.H. Wei. 1995. Mitochondrial deoxyribonucleic acid 4977-bp deletion is associated with diminished fertility and motility of human sperm. *Biol Reprod* 52: 729-36.

Kao, S.H., H.T. Chao, and Y.H. Wei. 1998. Multiple deletions of mitochondrial DNA are associated with the decline of motility and fertility of human spermatozoa. *Mol Hum Reprod* 4: 657-66.

Lamantea, E., V. Tiranti, A. Bordoni, A. Toscano, F. Bono, S. Servidei, A. Papadimitriou, H. Spelbrink, L. Silvestri, G. Casari, G.P. Comi, and M. Zeviani. 2002. Mutations of mitochondrial DNA polymerase gammaA are a frequent cause of autosomal dominant or recessive progressive external ophthalmoplegia. *Ann Neurol* 52: 211-9.

Libby, B.J., L.G. Reinholdt, and J.C. Schimenti. 2003. Positional cloning and characterization of Mei1, a vertebrate-specific gene required for normal meiotic chromosome synapsis in mice. *Proc Natl Acad Sci U S A* 100: 15706-11.

Longley, M.J., P.A. Ropp, S.E. Lim, and W.C. Copeland. 1998. Characterization of the native and recombinant catalytic subunit of human DNA polymerase gamma: identification of residues critical for exonuclease activity and dideoxynucleotide sensitivity. *Biochemistry* 37: 10529-39.

Loy, C.J. and E.L. Yong. 2001. Sex, infertility and the molecular biology of the androgen receptor. *Curr Opin Obstet Gynecol* 13: 315-21.

Matzuk, M.M. and D.J. Lamb. 2002. Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4 Suppl: s41-9.

McLachlan, R.I. and D.M. de Kretser. 2001. Male infertility: the case for continued research. *Med J Aust* 174: 116-7.

Mulhall, J.P., R. Reijo, R. Alagappan, L. Brown, D. Page, R. Carson, and R.D. Oates. 1997. Azoospermic men with deletion of the DAZ gene cluster are capable of completing spermatogenesis: fertilization, normal embryonic development and pregnancy occur when retrieved testicular spermatozoa are used for intracytoplasmic sperm injection. *Hum Reprod* 12: 503-8.

Okabe, M., M. Ikawa, and J. Ashkenas. 1998. Male infertility and the genetics of spermatogenesis. *Am J Hum Genet* 62: 1274-81.

Rovio, A., V. Tiranti, A.L. Bednarz, A. Suomalainen, J.N. Spelbrink, N. Lecrenier, A. Melberg, M. Zeviani, J. Poulton, F. Foury, and H.T. Jacobs. 1999. Analysis of the

trinucleotide CAG repeat from the human mitochondrial DNA polymerase gene in healthy and diseased individuals. *Eur J Hum Genet* 7: 140-6.

Rovio, A.T., D.R. Marchington, S. Donat, H.C. Schuppe, J. Abel, E. Fritsche, D.J. Elliott, P. Laippala, A.L. Ahola, D. McNay, R.F. Harrison, B. Hughes, T. Barrett, D.M. Bailey, D. Mehmet, A.M. Jequier, T.B. Hargreave, S.H. Kao, J.M. Cummins, D.E. Barton, H.J. Cooke, Y.H. Wei, L. Wichmann, J. Poulton, and H.T. Jacobs. 2001. Mutations at the mitochondrial DNA polymerase (POLG) locus associated with male infertility. *Nat Genet* 29: 261-2.

Ruiz-Pesini, E., C. Diez, A.C. Lapena, A. Perez-Martos, J. Montoya, E. Alvarez, J. Arenas, and M.J. Lopez-Perez. 1998. Correlation of sperm motility with mitochondrial enzymatic activities. *Clin Chem* 44: 1616-20.

Schultz, N., F.K. Hamra, and D.L. Garbers. 2003. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A* 100: 12201-6.

Shinohara, T., K.E. Orwig, M.R. Avarbock, and R.L. Brinster. 2003. Restoration of spermatogenesis in infertile mice by Sertoli cell transplantation. *Biol Reprod* 68: 1064-71.

Skaletsky, H., T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.P. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, and D.C. Page. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825-37.

Spelbrink, J.N., J.M. Toivonen, G.A. Hakkaart, J.M. Kurkela, H.M. Cooper, S.K. Lehtinen, N. Lecrenier, J.W. Back, D. Speijer, F. Foury, and H.T. Jacobs. 2000. In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *J Biol Chem* 275: 24818-28.

Taylor, C.T. 2001. Antioxidants and reactive oxygen species in human fertility. *Environmental Toxicology and Pharmacology* 10: 189-98.

Thangaraj, K., M.B. Joshi, A.G. Reddy, A.A. Rasalkar, and L. Singh. 2003. Sperm mitochondrial mutations as a cause of low sperm motility. *J Androl* 24: 388-92.

Tomes, C.N., C.M. Roggero, G. De Blas, P.M. Saling, and L.S. Mayorga. 2004. Requirement of protein tyrosine kinase and phosphatase activities for human sperm exocytosis. *Dev Biol* 265: 399-415.

Vernet, P., R.J. Aitken, and J.R. Drevet. 2004. Antioxidant strategies in the epididymis. *Mol Cell Endocrinol* 216: 31-9.

Wang, P.J., J.R. McCarrey, F. Yang, and D.C. Page. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 27: 422-6.

Weir, F. Association of polymorphisms of the mitochondrial COI gene with human male infertility. Christchurch, University of Canterbury. 2003. p124. (Thesis: MSc: Cellular and Molecular Biology).

Wu, M. Human sperm mitochondrial DNA point mutations and spermatogenesis dysfunction. Christchurch, University of Canterbury. 2000. p114. (Thesis: MSc: Cellular and Molecular Biology).

CHAPTER TWO MATERIALS AND GENERAL METHODOLOGY

Yong, E.L., J. Lim, Q. Wang, A. Mifsud, Y.C. Ong, and C.K. Sim. 2000a. Genetics of male infertility: role of androgen receptor mutations and Y-microdeletions. *Ann Acad Med Singapore* 29: 396-400.

Yong, E.L., L.S. Lim, Q. Wang, A. Mifsud, J. Lim, Y.C. Ong, and K.S. Sim. 2000b. Androgen receptor polymorphisms and mutations in male infertility. *J Endocrinol Invest* 23: 573-7.

2.1 Materials

All chemicals used in this research were molecular biology grade. Abbreviations are listed in Appendix A. Chemicals and suppliers are listed in Appendix B. Water was deionised and ultra-filtrated (NANOpure ultra pure water system, Barnstead, IO, USA). All water was sterilized by autoclaving before use.

2.2 Sample Population

All individuals were recruited from the New Zealand Centre for Reproductive Medicine in Christchurch, New Zealand. Semen was initially given by couples for analysis of a suspected fertility problem. Because a couple's infertility may be due to one or either partner, semen samples show large variation in semen characteristics from severely reduced semen quality to high sperm count and motility. Individuals with semen parameters within the normal range ($\geq 20x10^6$ sperm/ml and $\geq$ 50% motility) were classified as normozoospermic and included in the control population, although fertility was not proven. Each individual gave semen samples by masturbation after three days abstinence, their consent was obtained for the sample to be used further in molecular research. Semen parameters such as sperm count and motility were recorded by staff at the New Zealand Centre for Reproductive Medicine at the time of donation, and in addition, morphology and testicular histology noted for some individuals. This data was entered into our own laboratory database. The entire study population along with semen parameters and any other noted factors are recorded in Appendix C.

Ethical approval for the research project was gained from the Canterbury Ethics Committee, research progress was reported annually.

Semen samples were stored in 200 μl aliquots at -80°C to avoid DNA damage by repeated freeze/thawing. Extracted seminal DNA was stored at -20°C or 4°C while in daily use. A total of 321 individuals were used in this research after being categorised into one of

seven semen classifications based on World Health Organisation guidelines (Table 2.1) (WHO 1999). However, DNA extracted from 29 samples was not in sufficient quantity for use in analysis of all genes. Thus, 292 individuals were used for the majority of the research (Appendix C). This group includes 50 samples where research has been previously conducted to evaluate the frequency of single nucleotide polymorphisms within their mitochondrial DNA (mtDNA), including the *ND* genes *1*, *2*, *5*, and *6*, *ATPase 6*, and *8*, and *COI*, *II* and *III* (Holyoake 1998; Wu 2000; Gomas 2002; Weir 2003).

Of the 292 individuals in the study population 195 were men with poor semen quality (19 Azoospermic, 48 Severe Oligoasthenozoospermic, 17 Severe Oligozoospermic, 28 Oligoasthenozoospermic, 27 Oligozoospermic, and 56 Asthenozoospermic) and 97 were men with normal semen quality. Sample population details and the number of samples analysed in each gene region are given in Table 2.1 below.

Table 2.1. Classification of semen samples.

Semen Classification	Sperm Count (10^6/ml)	Sperm Motility (%)	n	*AR* CAG	Yq M.del	*POLγ* CAG	Exo I	Exo II & III
Azoospermia	0	0	19	12	19	11	19	5
Severe Oligoasthenozoospermia	0-5	<50	48	31	48	47	46	16
Severe Oligozoospermia	0-5	>50	17	11	17	14	16	6
Oligoasthenozoospermia	5-20	<50	28	26	28	28	28	11
Oligozoospermia	5-20	>50	27	24	27	27	27	9
Asthenozoospermia	≥20	<50	56	53	56	53	55	18
Normozoospermia	≥20	>50	97	89	97	92	83	61
Total			292	246	292	272	274	126

Detail on samples and gene analysis can be found in Appendices C and D. N refers to the total sample number in the sample population subgroup. Each column gives the number of samples analysed in each subgroup for each gene region. Semen classifications are shortened in some tables to: Az, azoospermia; SOAs, severely oligoasthenozoospermia; SO, severely oligozoospermia; OAs, oligoasthenozoospermia; O, oligozoospermia; As, asthenozoospermia; N, normozoospermia; and 'zoospermia' to ZS. The last five columns indicate numbers of individuals analysed in each gene region.

2.3 General Methodology

There are several aspects of the research method that were common for all gene regions investigated. These are outlined below so that the methods section of each chapter is condensed, and only particular characteristics noted.

2.3.1 Human Seminal DNA Extraction

Total genomic DNA was extracted from semen samples using a standard Phenol/Chloroform extraction method as outlined in Kao *et al.* (1995). The extraction was carried out in 1.5 ml Eppendorf tubes; if required, semen samples were transferred to these tubes from 0.5 ml Eppendorf tubes after thawing. Approximately 200 μl of semen sample (200-300 μl if possible from men with poor semen quality) was thawed on ice after which sperm cells were pelleted by centrifugation at 6800 x g, at 4°C for 10 min in an Eppendorf centrifuge (model 5417R). The supernatant was poured off leaving the pellet of sperm cells behind. This pellet was then resuspended in 100 μl of extraction buffer/lysis mix (0.1 M Tris-HCl (pH 8.0), 0.1 M NaCl, 20 mM EDTA (pH 8.0) / 1% (w/v) SDS, 10 mM dithiothreitol (DTT), and 100 μg/ml proteinase K) by gently flicking the tube. Tubes were wrapped in parafilm and incubated at 50°C in a waterbath overnight (16-20 hours). After incubation, samples were centrifuged briefly to collect condensation. An equal volume (100 μl) of phenol saturated with TE (pH 8.0) buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)) was added and tubes gently inverted in order to extract proteins from the lysed sperm cell mix. Tubes were centrifuged at 15300 x g for 5 min at room temp to separate the aqueous and organic phases. The aqueous phase was then transferred to a fresh tube using a wide bore pipette to avoid mechanical shearing of DNA. The residual interphase and organic layer were back extracted with an equal volume (~100 μl) of TE (pH 8.0) to maximally recover remaining DNA. The combined aqueous phases were extracted with an equal volume (~200 μl) of phenol:chloroform:isoamyl alcohol (25:24:1) and the aqueous layer transferred to a fresh tube. To remove residual phenol from the sample the aqueous layer was then extracted with an equal volume (~200 μl) of chloroform:isoamyl alcohol (24:1) and transferred to a

fresh tube. This extraction was repeated if a white interphase was observed after centrifugation. Finally, DNA was precipitated out of solution overnight (16-20 hrs) by adding 0.1 volumes of 3 M sodium acetate (NaOAc, pH 7.0) and two volumes of cold absolute ethanol (stored at -20°C). The DNA pellet was collected by centrifugation at 13800 x g at 4°C for 20 min, and then washed in 300 μl of cold 70% ethanol (stored at -20°C). The DNA was repelleted and washed for a total of three times before air drying for seven to ten min. The DNA pellet was resuspended in 50 μl TE (pH 8.0) (30 μl if the sample was from a non-normozoospermic man). DNA was stored at -20°C until required for use at which time it was stored at 4°C.

2.3.1.1 Visualisation of Extracted Genomic DNA

Successful extraction of genomic DNA was confirmed by electrophoresis of 3 μl of extracted DNA with 1 x orange G loading buffer (30% glycerol, 0.35% orange G) through a 1.5% (w/v) agarose gel made up with 1 x TAE (0.04 M Tris-acetate/0.001 M EDTA (pH 8.0)) buffer. Electrophoresis was carried out at 5 V/cm for 20 min. Gels were stained with ethidium bromide (0.5 μg/ml) on a shaker for 20 min and the DNA visualised using an ultra-violet (UV) transilluminator (300 nm).

2.3.1.2 Quantification of Genomic DNA

The purity and yield of extracted DNA was calculated by reading the absorbance of 2 μl of DNA in 498 μl of TE (pH 8.0), vortexed to mix, at 260 nm (A_{260}) and 280 nm (A_{280}) on a spectrophotometer (model LKB Biochrom Ultraspec II 4050). DNA concentration was calculated based on the absorbance reading at 260 nm (1 optical density (OD) unit at 260 nm equals 50 μg/ml double-stranded DNA) and adjusted for volume (500 μl to 1 ml) and dilution factor (2 μl to 1 μl):

$$\text{DNA concentration (ds DNA) } (\mu g/ml) = \frac{A_{260} \times 50\ \mu g/ml}{2_{(\text{volume factor})} \times 2_{(\text{dilution factor})}}$$

(Sambrook et al. 1989).

DNA purity was estimated using the ratio of light absorbance at 260 nm to that at 280 nm (A_{260}:A_{280}). An optimally pure DNA sample will have a ratio equal to 1.8, less than 1.8 indicates contamination due to the presence of proteins in the sample (high A_{280}), while greater than that suggests RNA contamination (Sambrook et al. 1989).

2.3.2 Polymerase Chain Reaction (PCR) Analysis

2.3.2.1 Primer Design

All primers used to conduct this research were based on DNA sequences submitted to the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). Primers were designed so that they were located in the regions of DNA flanking the gene of interest so that amplification proceeded through the entire sequence. Primer pairs were designed using the Primer3 programme (http://www.path.com.ac.uk/cgi.bin/primer3.cgi) and checked for specificity, secondary DNA structure, primer dimer and multiple priming sites using the PCR simulation programme Amplify 1.2 (Engels 1993). All primer pairs were designed based on an optimal primer length of 20 base pairs (bp), similar melting temperature (Tm), and GC content of around 50%. Primers were chemically synthesised by Invitrogen, Life Technologies Ltd (Auckland, New Zealand). Specific primer pairs for each gene region are given in the corresponding chapter.

2.3.2.2 PCR Reactions

PCR reaction conditions: annealing temperature and magnesium (Mg^{2+}) were optimised for each primer pair using a temperature gradient PCR machine (Eppendorf Mastercycler® gradient). Each new PCR primer pair was optimised by amplification in two normozoospermic samples with a median annealing temperature of 58°C, varying 5°C either side to produce a gradient temperature range between 53°C and 63°C. Magnesium concentration was set at the standard 10 x buffer plus Mg^{2+} concentration of 1.5 mM for temperature optimisation. None of the PCR reactions in this research required a higher magnesium concentration than 1.5 mM.

Each 25 μl PCR reaction mix contained approximately 200 ng of genomic DNA plus 0.5 units of *Taq* DNA polymerase (Roche), 1 x PCR buffer (10 mM Tris-HCl, 1.5 mM $MgCl_2$, 50 mM KCl, pH 8.3), 200 μM of each deoxyribonucleoside triphosphate (dNTP; dATP, dGTP, dCTP, and dTTP), and 0.4 μM of each primer. The multiplex PCR employed for the Y chromosome sequence tagged site (STS) deletion analysis (Chapter Three) required some additional optimising for the multiplex PCR reaction conditions where five sets of primer pairs were added, each with its own optimum annealing temperature. Therefore, some primer concentrations in the multiplex were less than 0.4 μM, and the annealing temperature of the PCR programme higher than for other gene regions to reduce non-specific priming. All reactions had a drop of mineral oil added to the tube before cycling to prevent evaporation.

PCR reactions were cycled in either a MJ Research Inc. (PTC-100) or Eppendorf (Mastercycler® gradient) thermal cycler machine. PCR programmes followed standard programme cycling times, although CAG trinucleotide repeat sequences required longer amplification times and a greater number of cycles than other DNA sequences amplified (Chapters Four, and Five).

PCR products were stored at 4°C after completion of thermal cycling until required.

2.3.2.3 Visualisation of PCR Products

To ascertain the quantity of gene product generated by the PCR programme, 3 μl was electrophoresed through a 1.5% (w/v) agarose gel (2% for Yq multiplex products) with 1 x TAE buffer at between 7 and 10 V/cm for 20 min (a longer running time and lower voltage was used for Yq multiplex products). Agarose gels were stained with ethidium bromide (0.5 μg/ml) for 15-20 min and illuminated under UV (as per DNA extraction gels).

2.3.2.4 Purification of PCR Products

PCR products were precipitated and washed to remove contaminants, such as unincorporated primers or dNTPs, before sequencing. The cycled PCR reaction was transferred from the reaction tube onto parafilm then into a fresh tube to remove excess mineral oil. One volume of 4 M ammonium acetate and two volumes of isopropanol were added and tubes inverted several times to mix. The reaction was incubated at room temp for 20 min (25 min if the PCR product concentration was low) and then centrifuged at 15300 x g at room temp for 20 min. The supernatant was drawn off carefully and discarded, and the pellet washed with 100 μl of 70% ethanol (stored at room temp). The pellet was recentrifuged for 10 min and washed again, before air-drying for 5 min. The precipitated PCR product was resuspended in 15 μl of TE (pH 8.0) buffer and stored at 4°C.

2.3.2.5 Visualisation of the Purified PCR Product

To ensure retention, and calculate approximate concentration of the amplified product, 3 μl of the purified PCR product was electrophoresed through a 1.5% (w/v) agarose gel as per PCR reaction visualisation.

2.3.3 DNA Sequencing

Various methods of DNA sequencing were employed throughout this research; ^{32}P and ^{33}P labelled sequencing, and fluorescent T3-tagged sequencing using a LICOR DNA sequencer (model 4000L). Each will be discussed in the relevant chapter.

2.4 Results

2.4.1 DNA Extraction

Genomic DNA was extracted from the semen of 215 men with poor semen quality and 106 men with normozoospermia as defined by WHO (1999) (Figure 2.1) (refer Appendix C for details of sample population characteristics).

DNA yield and purity varied. In general, DNA yield from both normozoospermic and non-normozoospermic samples was similar (approximately 2.5 μg from 200 μl of semen sample) where original semen samples were of similar viscosity (high) and volume (2-4 ml). For some azoospermic and severe oligozoospermic samples, one or both of the volume and viscosity was low. In these cases the DNA yield and purity tended to be lower (around 0.6 μg of genomic DNA from 100-200 μl of semen sample; purity of greater than 1.8 (ranging between 1.7 and 2.1)) in an effort to maximise yield. DNA purity was generally closer to 1.8 (1.7-2.0) in all other samples.

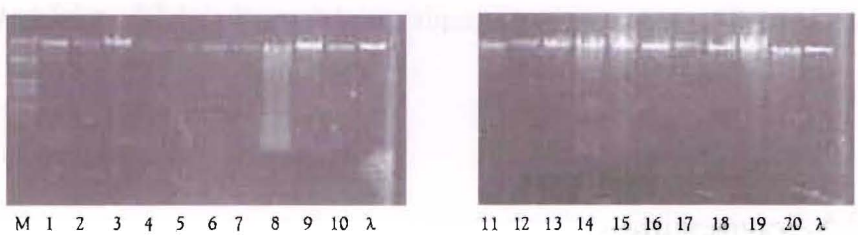


Figure 2.1. Electrophoresis images of genomic DNA after extraction. The marker (M) is λH/E (lamba DNA digested with *HindIII* and *EcoRI* restriction enzymes); the top band is 21226 bp. Samples numbered 1 to 10 have semen classifications of O, Az, O, SOAs, SOAs, SO, SOAs, SOAs, SOAs, N, and 11 to 20 are SO, SOAs, OAs, OAs, OAs, O, O, OAs, O, and OAs. A 20 ng λ DNA size reference was loaded in the last lane of each gel.

References:

Engels, W.R. 1993. Contributing software to the internet: the Amplify program. *Trends Biochem Sci* 18: 448-50.

Gomas, K.P. The Association of Single Nucleotide Polymorphisms in Mitochondrial ND1 and ND2 Genes with Human Male Fertilising Potential. Christchurch, University of Canterbury. 2002. p133 (Thesis: MSc: Cellular and Molecular Biology).

Holyoake, A.J. Aspects of mitochondrial DNA mutations in relation to human male fertilising potential. Christchurch, University of Canterbury. 1998. p 154 (Thesis: PhD: Cellular and Molecular Biology).

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Weir, F. Association of polymorphisms of the mitochondrial COI gene with human male infertility. Christchurch, University of Canterbury. 2003. p124 (Thesis: MSc: Cellular and Molecular Biology).

WHO. 1999. *WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interactions.* Cambridge University Press., Cambridge.

Wu, M. Human sperm mitochondrial DNA point mutations and spermatogenesis dysfunction. Christchurch, University of Canterbury. 2000. p114 (Thesis: MSc: Cellular and Molecular Biology).

CHAPTER THREE Y CHROMOSOME MICRO-DELETION

3.1 Introduction

DNA sequence analysis is forcing scientists to rethink assumptions about the Y chromosome. Thought to be a decaying mess of partial sequences, pseudogenes, and repetitive elements, it now seems to encode a surprisingly complex system of near perfect amplicons arranged into longer palindromes, separated by hundreds and even thousands of base pairs (Skaletsky et al. 2003). It no longer appears that the Y chromosome is passively losing genes to mutation, but actively correcting them to preserve the integrity of its spermatogenesis genes.

3.1.1 The Human Sex Chromosomes

The human sex chromosomes X and Y evolved from a pair of homologous autosomes some 300 million years ago (Figure 3.1). The Y chromosome has a special status as the sex-determining chromosome. It encodes the *SRY* gene (sex-determining region of the Y), which stimulates a cascade of gene activity resulting in the development of male sexual characteristics (Graves 2002).

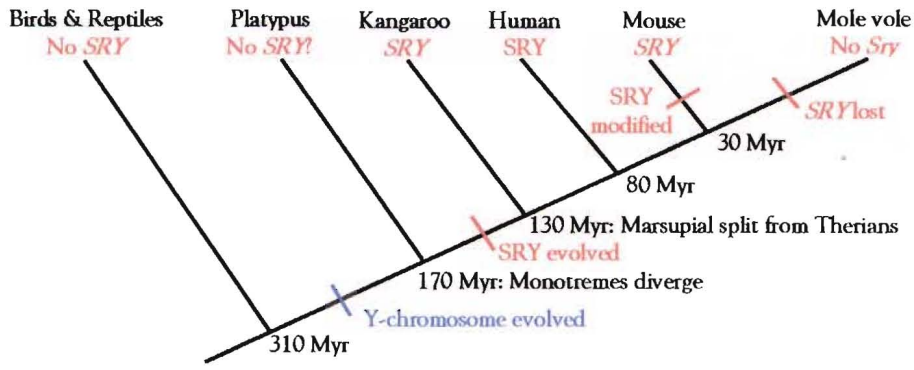


Figure 3.1. Evolutionary timescale of mammalian Y chromosome and *SRY*. Sequence comparison between mammalian groups indicates an approximate timescale for Y chromosome and *SRY* evolution. Detection of sex-specific *SRY* in both Platypus and Echidna (Monotremes) by PCR and Southern Blot have failed, however, two *SOX* genes are autosomal. Adapted from Graves (2002).

Four separate inversion events beginning in distal Yq and ending with proximal Yp triggered the end of sexual recombination between X and Y (Lahn and Page 1999a; Skaletsky et al. 2003). While the loss of recombination allowed Y to specialise in male vigour genes it also made it vulnerable to mutation and consequent loss of functionally important genes by Müller's ratchet (Charlesworth 1978; Rice 1998). A comparison of X and Y sequences suggests that in 300 million years the Y chromosome has lost 1500 genes, 5 genes per million years; it currently encodes between 50 and 78 genes, a mere 3% of what it once did (Graves 2002; Skaletsky et al. 2003).

However, should the Y chromosome lose all its genes the fate of humankind is unlikely to result in a race of women, instead the future of human sex-determination is likely to lie with another *SRY* analogue. This has already happened in two species of the mole vole *Ellobius* which achieve dimorphic sexual development with an identical sex chromosome complement in both sexes; one active X chromosome, while a close relative retains the *SRY* sex-determination mechanism. The alternative means of sex-determination in these voles remains unclear but once discovered it will give insight into how one system of sex-determination evolves from the last (Graves 2002). It is hypothesised that a new sex-determining gene could stimulate another burst of human evolution; the accumulation of new sex-benefit genes at this new site will begin sex-chromosome evolution all over again and, if more than one sex-determining gene arises, may eventuate in human populations that cannot interbreed (Graves 2002). Moreover, if this new successor is not on X or Y would we lose X as well, as new versions of its genes accumulate on the new sex-determining chromosome pair?

In addition to directing testis development the Y chromosome is vital for sperm production, a characteristic that was not at first apparent until identification of several azoospermic men with deletions of distal portions of the Y chromosome long arm suggested the presence of 'fertility genes' on Y (Tiepolo and Zuffardi 1976).

3.1.2 The Y Chromosome

The human Y chromosome encodes roughly 63 million base pairs (Mb) of DNA, 95% of which is unlike sequences on its partner X, termed the MSY (male-specific region of Y) (Figure 3.2). Only the short (<3 Mb) pseudo-autosomal regions at either end of Y enable it to synapse with X during meiosis and undergo random assortment into gametic daughter cells. Two-thirds (40 Mb) of the MSY is made up of highly repetitive heterochromatic DNA sequences (Figure 3.2) (Skaletsky et al. 2003).

The euchromatic portion of MSY encompasses approximately 23 Mb, divided into three sequence classes dispersed amongst each other along MSY. It encodes 156 transcription units, half of which encode proteins, divided among nine MSY-specific gene families (Hawley 2003; Skaletsky et al. 2003). In all, the MSY encodes 27 proteins or protein families; twelve expressed ubiquitously throughout the body and eleven expressed predominantly in the testes (Skaletsky et al. 2003).

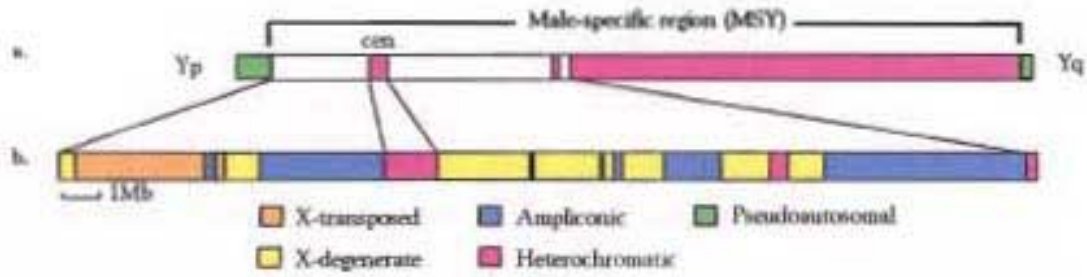


Figure 3.2. The male-specific region of the Y chromosome. (a) Schematic representation of the Y chromosome. The MSY extends over the majority of the Y chromosome, including an extensive (40 Mb) region of heterochromatin in Yq. (b) Enlargement of the euchromatic region of MSY (23Mb), illustrating the dispersed nature of the three euchromatic sequence types, and relative proportions of each. An approximate 1 Mb scale bar is shown. From Skaletsky et al. (2003).

'X-transposed' euchromatic sequences are 99% similar to present Xq21 sequences and result from a single transposition event from the X chromosome to Yp after the human

and chimpanzee lineages diverged. It makes up 15% of MSY and encodes only two genes, *TGIF2LY* (TGF (beta)-induced transcription factor 2-like) expressed in the testes, and *PCDH11Y* (Protocadherin 11 Y) expressed in the brain (Skaletsky et al. 2003).

'X-degenerate' sequences are remnants of the autosomal ancestor of X and Y. These sequences encode single-copy or pseudogene homologs of 27 X chromosome genes. Y chromosome pseudogenes show sequence similarity to the exons and introns of their functional X-linked gene counterpart. X-degenerate sequences are older and have lower sequence homology to X-linked genes than X-transposed sequences. *SRY* is the only testis-specific X-degenerate sequence; however, all 12 ubiquitously expressed protein-coding transcripts are made up of X-degenerate euchromatin, suggesting that X-degenerate transcripts predominantly encode house-keeping genes (Skaletsky et al. 2003).

The majority (70%) of MSY euchromatin is 'ampliconic', so called because each sequence is duplicated, or amplified to multiple-copy along the Y chromosome. These sequences are unique to the Y chromosome and encode the highest density of genes; 13.3 transcription units per Mb, compared with 2.2 and 0.6 for X-degenerate and X-transposed sequences. In total, ampliconic euchromatin encodes 60 protein-coding genes divided into nine gene families all expressed solely, or primarily in the testes, suggesting these genes are important for spermatogenesis and male sexual development. Ampliconic sequences were derived by three processes: amplification of ancestral autosomal genes, translocation and amplification of autosomal genes, and retroposition and amplification of autosomal genes (Figure 3.3) (Lahn and Page 1999b; Skaletsky et al. 2003).

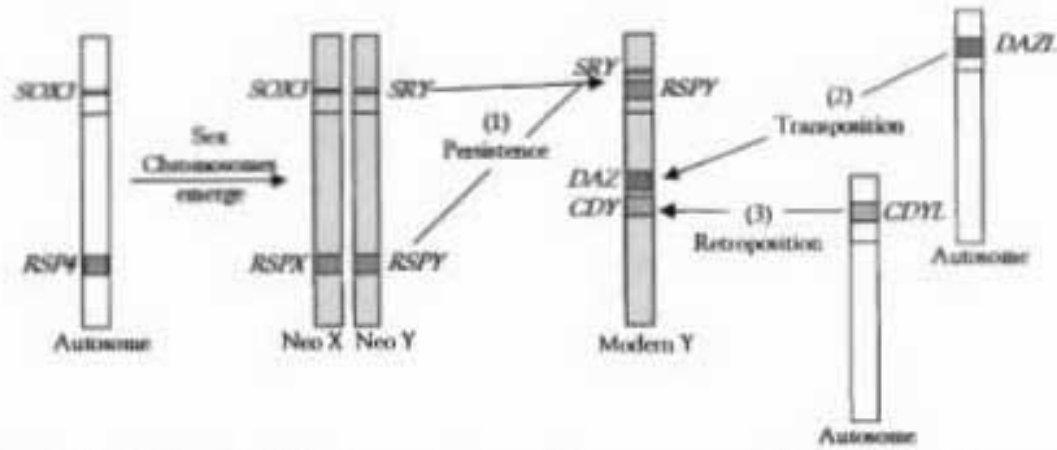


Figure 3.3. Origins of Y-linked sequence types. The emergence of SRY set apart the new sex chromosomes from other autosomes, and their ancestor autosome. (1) Persistence: remnant genes remain from the ancestral autosome. (2) Transposition: The transfer of autosomal genes (*DAZL*) to the Y, homologs of these sequences are found on modern autosomes. (3) Retroposition: reverse transcription and integration of mRNA transcripts gave rise to gene sequences such as *CDY*. Adapted from Lahn and Page (1999b).

The most notable feature of ampliconic sequences are the palindromes made up of DNA sequences that read the same in the opposite direction on both strands of the double helix. The eight palindromes range in size from a tiny 9 kilobases (kb) to 1.45 Mb and are highly identical; each arm of the palindrome pair is 99.94% - 99.997% identical to its partner. Of the nine testes-specific MSY gene families, eight are found within the palindromes, six of which are located exclusively within them; this includes the four copies of the *DAZ* (deleted in azoospermia) gene. Interestingly, six of the eight palindromes existed before the human lineage split from the chimpanzee (Skaletsky et al. 2003).

These features suggest the MSY sequence structure has functional importance. In fact, an alternate explanation for the multiple copies of important spermatogenesis genes is that frequent non-reciprocal gene conversion between arms of the palindromes maintains Y chromosome gene integrity in the absence of recombination, by swapping the sequence on one arm of the palindrome with the sequence on the other; in essence recombination with itself (Skaletsky et al. 2003). As yet is it unclear exactly how this self-repair mechanism works, however, the intricate process required to generate the amplicons and

palindromes, which in some instances encodes the only copies of vital spermatogenesis genes, suggests it was favoured by selection (Kuroda-Kawaguchi et al. 2001).

3.1.2.1 Micro-deletions of AZF

Defective spermatogenesis is the most common form of male infertility. In 1976, Tiepolo and Zuffardi observed that a region of the Y chromosome long arm (Yq) adjacent to the heterochromatic region was absent in six patients with severe spermatogenic failure; azoospermia. This region, termed AZF (azoospermic factor) was further characterised into three sub-regions, AZFa, b and c in proximal, middle and distal Yq11, based on the clustering of micro-deletions (Figure 3.4) (Vogt et al. 1996).

A fourth region lying between AZFb and c, AZFd was later described as a result of a single study (Kent-First et al. 1999). This small region beginning distal to *RBMY2* spans only eight STS', however, commonly only the three original AZF regions are referred to.

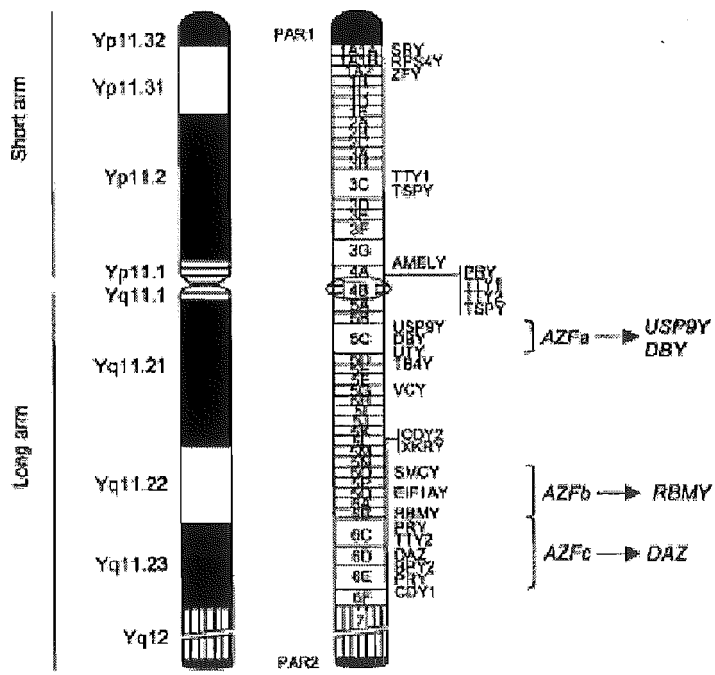


Figure 3.4. A schematic illustration of the human Y chromosome showing cytological banding and deletion intervals. On the right is a deletion map of the Y chromosome as defined by Vollrath et al. (1992) and the genes that map within in them. AZF regions and candidate spermatogenesis genes are indicated. Adapted from Foresta et al. (2001).

Approximately 10-15% of severely oligozoospermic and azoospermic men have micro-deletions within Yq11. Nearly all are *de novo* and each sub-region is associated with a different phenotype and deletion frequency (Foresta et al. 2001). Deletions within AZFc are most common ($3x10^{-4}$) (Yen 2001), but produce variable phenotypes from severe to mild spermatogenic arrest. Deletion of AZFa is associated with severe infertility; very early spermatogenic arrest characterised by an absence of germ cells, Sertoli-cell only (SCO) syndrome, and small testis volume (Foresta et al. 2001). AZFb deletions can also produce variable phenotypes but generally result in premeiotic disruption to spermatogenesis where spermatogonial cells are present (often in reduced numbers) but do not undergo meiosis to produce haploid spermatocytes. (Foresta et al. 2001; Krausz et al. 2003).

Many AZF deletions share the same proximal and distal breakpoints, suggesting they occur by a common mechanism. The majority of both AZFa and c deletion intervals have similar breakpoints bounded by repeat sequences, non-homologous recombination between these repeats eliminates the spermatogenesis genes between them (Kamp et al. 2000; Sun et al. 2000; Kuroda-Kawaguchi et al. 2001). The 800 kb AZFa region encodes only three genes, *USP9Y, DBY* and *UTY*; intrachromosomal recombination between two nearly identical (94%) HERV15 endogenous retroviruses completely eliminates both *USP9Y* and *DBY* (Kamp et al. 2000).

The AZFc region consists almost entirely of long direct- and inverted repeat units divided into six amplicon families which make up three larger palindromes (Figure 3.5). Each amplicon is 99.8% identical to other amplicons in its family and at least one amplicon lies in the opposite orientation. Sequencing revealed eleven gene families: five known (*RBMY1, PRY, DAZ, BPY2, and CDY*), and six new families (*CSPG4CY, GOLGA2LY* and *TTY3-6*), comprising a total of 27 potentially functional genes all expressed solely or predominantly in the testes. Remarkably, there are no single-copy genes (Kuroda-Kawaguchi et al. 2001).

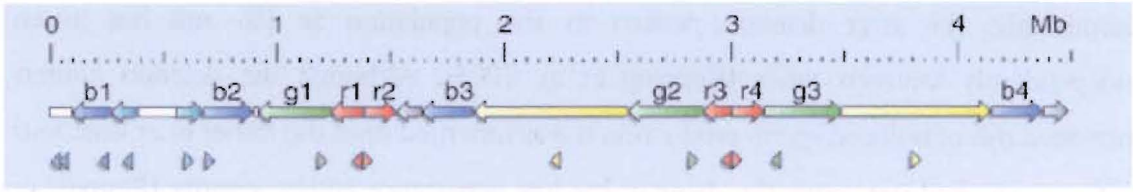


Figure 3.5. AZFc. The entire 4.5 Mb DNA segment sequenced by Kuroda-Kawaguchi et al. (2001). The AZFc region (defined by common deletions) extends from amplicon b2 to b4. Amplicons and their orientation are indicated by the numbered coloured arrows, those belonging to the same family are the same colour (blue, turquoise, green, red, grey, or yellow). Protein-coding genes are indicated by triangles. The largest palindrome, P1 is centred between r3/r4 and extends from b3 to b4; within its arms lie two more palindromes P1.1 and P1.2. P2 encompasses r1 and r2, and P3 extends from b1 to b2. Adapted from Repping et al. (2003).

Most AZFc deletions (98%) involve the removal of a 3.5 Mb segment, flanked by two 229 kb direct repeats (amplicons b2 and b4), which completely eliminates seven gene families altogether, including all four copies of the *DAZ* gene (Figure 3.6) (Kuroda-Kawaguchi et al. 2001). Homologous recombination between other amplicons is predicted to account for the variation in AZFc-deletion phenotypes. For example, the gr/gr deletion spans 1.6 Mb and is significantly associated with reduced spermatogenesis (Figure 3.6). This deletion does not eliminate any of the eleven AZFc testes-specific gene families but does reduce the copy number of eight, adding support for the idea that dosage of some genes is important for normal sperm production.

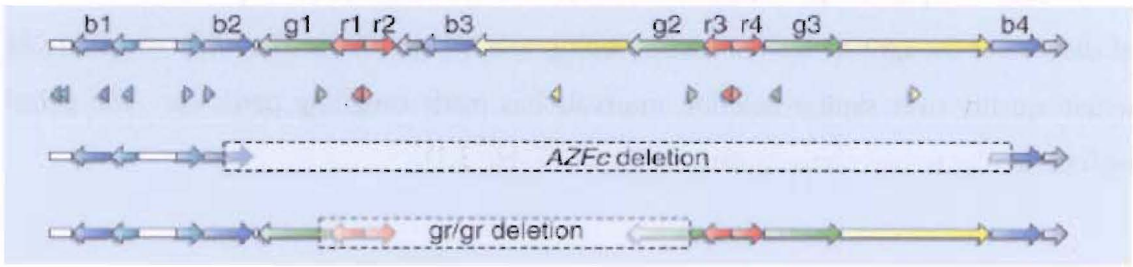


Figure 3.6. Common deletion intervals of the AZFc region. The reference sequence and protein-coding genes are indicated at top. The deletion breakpoints have not been located precisely but lie somewhere in the amplicon indicated. The AZFc (b2/b4) deletion removes the entire region containing all the members of seven testes-specific gene families, and almost always results in reduced spermatogenesis. The gr/gr (g1-r2/g2-r4) deletion eliminates half of AZFc and is associated with an increased risk of spermatogenic failure. Adapted from Tyler-Smith and McVean (2003).

Surprisingly, the gr/gr deletion persists in the population at 2% and has arisen independently fourteen times (Repping et al. 2003). Although the deletion confers increased risk of reduced sperm production it was inherited from the father in at least four of those screened suggesting the deletion has low penetrance and/or severity (Repping et al. 2003).

Breakpoint analysis in four AZFb patients indicates that a common deletion mechanism is operating with AZFb deletions also; however, lack of flanking repeat sequences suggests the mechanism is different to that of AZFa and c. Other near identical repeat sequences to one found in AZFb have been identified within AZFc, suggesting that some AZFb and AZFc deletions extend over the boundary between them (Ferlin et al. 2003).

3.1.2.2 Spermatogenesis Genes within AZF

It is assumed that there is at least one spermatogenesis gene encoded within each AZF region (Vogt 1998). DNA sequencing and expression analysis suggests there are a variety of candidate spermatogenesis genes encoded within the three AZF regions, many of which have unknown function and importance in fertility (Lahn and Page 1997; Vogt 1998). These seem to fall into two categories: Y-specific genes in multiple copies, or Y-linked genes with X-linked homologs. The Y-specific genes have testis-specific expression and are predicted to play in spermatogenesis, while X-Y linked genes are expressed in a variety of tissues are thought to have a housekeeping role (Vogt 1998). The wide variability in semen quality over similar deletion intervals has made assigning particular AZF genes with particular semen characteristics difficult (Table 3.1).

Table 3.1. Candidate spermatogenesis genes within Yq11.

Gene Symbol	Gene Name	Protein Homolog	Expression	Yq11 location	Homolog[a]
BPY1	Basic protein Y, pI 9	Novel	Testis only	D8*	?
BPY2	Basic protein Y, pI 10	Novel	Testis only	AZFc	No
CDY	Chromo domain Y	Chromatin package protein	Testis only	D10-11, 23-24*	(6)
DAZ	Deleted in AZoospermia	RNA binding RRM proteins	Testis only	AZFc	*DAZL*(3p32)
DBY	DEAD Box Y	RNA helicases	Multiple	AZFa	*DDX3* (Xp11.3-p11.23)
EIF1AY	Eukaryotic translation initiation-translation factor 1A Y	Translation initiation factor	Multiple plus testis specific	AZFb	*EIF1AX*(X)
HSFY	Heat shock transcription factor2, Y chromosome	Heat shock element DNA binding domain	unknown	AZFb	*HSF2* (6)
PRY	PTPN-13 like on Y	Protein tyrosin phosphatase	Testis only	AZFb	No
RBMY	RNA binding motif	RNA binding RRM proteins	Testis only	AZFb, c	*RBMX* (X+ autosomal)
SMCY	Selected mouse cDNA Y	H-Y antigen HLA B7	Multiple	AZFb	*SMCX*(X)
TSPY	Testis-specific protein Y encoded	SET/NAP-1 regulated cell proliferation	Testis only	AZFb	No
TTY1	Testis transcript Y1	No-protein encoded RNA	Testis only	AZFc	No
TTY2	Testis transcript Y2	No-protein encoded RNA	Testis only	AZFc	No
USP9Y (DFFRY)	*Drosophila* fat facets related Y	C-terminal ubiquitin hydrolase	Multiple	AZFa	*USP9X (DFFRX)* (Xp11.4)
UTY	Ubiquitous transcribed Y	H-Y antigen HYD	Multiple	AZFa	*UTX*(X)
XKRY	XK related Y	Putative membrane transport protein	Testis only	D10-11*	No

Table adapted from Vogt (1998) based on information given in Brown et al. (1998), Foresta et al. (2000), Ferlin et al. (2003), and Stouffs et al. (2004). [a] Parentheses indicate the chromosomal location of spermatogenesis gene homologs. *Yq11 locations with an asterisk indicate the reference is to the Vogt et al. (1996) deletion map; D8 corresponds to 5G (sY94), D10-11 to 5H-K (sy97-sY109), and D23-24 to just distal to *RBM1* (Vollrath et al. (1992); Reijo et al. (1995); Vogt et al. (1996)).

3.1.2.2.1 Candidate Spermatogenesis Genes for AZFa

AZFa deletions are associated with the most severe infertility pathologies suggesting that there may be several genes within AZFa that are crucial for normal germ cell development (Foresta et al. 2000). The proximal half of AZFa contains pseudogenes with

homology to Xp22, the remainder consists of three genes designated *USP9Y*, *DBY* and *UTY*. All three genes have some elements consistent with that of spermatogenesis genes (Foresta et al. 2000).

USP9Y (also known as *DFFRY)* is a homolog of the X-linked gene *USP9X* (*DFFRX*), and both are related to the *Drosophila* developmental gene fat facets *(faf)*; a member of a protein family that prevents protein degradation by removing ubiquitin markers. While deletion of *faf*, *USP9X*, and *USP9Y* commonly results in severe disruption to fertility, the ubiquitous expression of *USP9Y* is inconsistent with a specialised role in spermatogenesis. It is possible that *USP9Y* functions indirectly, preventing protein degradation of important testes-specific gene products expressed in spermatogonia or early spermatocytes (Brown et al. 1998).

DBY encodes two transcripts, one large ubiquitously expressed transcript and a smaller testes-specific transcript. The protein encoded by *DBY*, DDXY belongs to the DEAD box family of proteins that are predicted to function as ATP-dependent RNA helicases. *UTY* is largely uncharacterised and rarely included in the deletion interval (Foresta et al. 2000).

Typically, AZFa deletion genotypes do not match well with phenotype; a study of nine individuals indicated that elimination of *DBY* did not always produce a SCO phenotype, nor did deletion of *USP9Y* (Figure 3.7) (Sargent et al. 1999; Foresta et al. 2000).

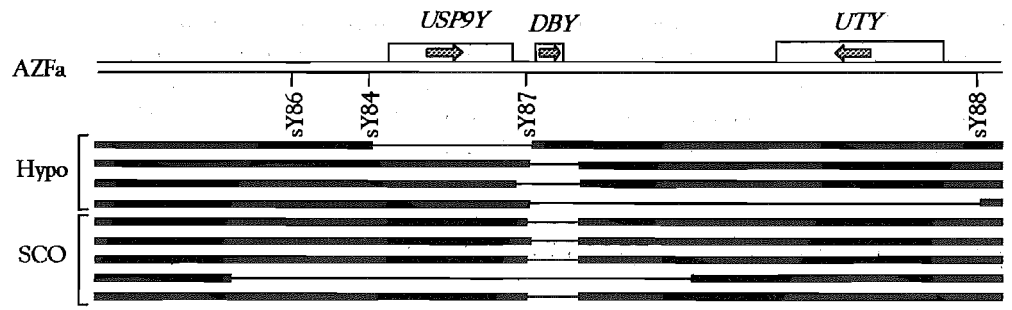


Figure 3.7. Schematic representation of AZFa spermatogenesis genes and deletion intervals. The three AZFa genes are indicated at the top, and Yq STS markers below them. The "Hypo" deletion intervals resulted in hypospermatogenesis (severely reduced sperm production); "SCO" deletion intervals resulted in Sertoli-cell only syndrome. Figure reproduced from Foresta et al. (2000).

3.1.2.2.2 Candidate Spermatogenesis Genes for AZFb

The *RBMY* (RNA-binding motif) family of genes are widely thought to be the most likely candidate spermatogenesis genes for AZFb. They are expressed only in the testis, throughout all developmental stages except mature spermatozoa, and are homologous to the mouse gene *rbm* which causes infertility when deleted. Yet, AZFb deletion breakpoint analysis indicates that some deletions of AZFb that result in reduced sperm production do not remove either *RBMY1* or *RBMY2*.

Of other AZFb genes, *SMCY* and *PRY* are thought to be unlikely candidates. *SMCY* is ubiquitously expressed, encodes a histocompatibility antigen, and transgene expression does not restore sperm production in XSxr^{b}O male mice. The two active *PRY* genes (*PRY1* and *2*) are more likely to be involved in the apoptosis of sperm cells than spermatogenesis (Mazeyrat et al. 2001; Stouffs et al. 2004).

Three other AZFb genes show more promising features consistent with spermatogenesis genes. *EIF1AY* is expressed exclusively in the testis and encodes the Y-linked isoform of eIF-1A, an essential translation initiation factor (Ferlin et al., 2003). The role of *EIF1AY* in spermatogenesis is not known, nor have any individuals with a specific elimination of this gene been reported, yet the abundance of testis-specific transcripts suggest some role in spermatogenesis (Foresta et al. 2001).

The two *CDY2* genes that have testis-specific expression, lie within the AZFb deletion interval, and function in the replacement of histone proteins with protamines during spermiogenesis (Ferlin et al. 2003)

Although the function of the novel candidate spermatogenesis gene *HSFY in vivo* has not been elucidated, expressed sequence tags (ESTs) associated with it are expressed in the testis. *HSFY* is a homolog of the *HSF2* gene on chromosome 6 and contains a heat shock factor DNA binding domain, for this reason it is predicted to function as a transcriptional activator specifically binding to heat shock promoter elements (Ferlin et al. 2003).

3.1.2.2.3 Candidate Spermatogenesis Gene for AZFc

The best candidate spermatogenesis factor for AZFc is the *DAZ* gene family, located in two clusters each of two genes, one in each orientation (3'-5', 5'-3') (Figure 3.8). All four genes are exclusively expressed in the testis and encode varying numbers of RNA recognition motifs; it is not known whether this confers different functions (Foresta et al. 2001).

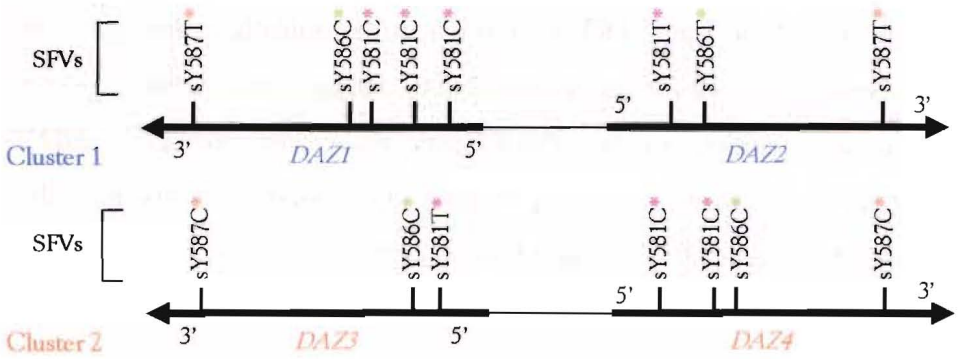


Figure 3.8. Schematic illustration of the two *DAZ* gene clusters. Two STS markers targeted to sequence family variants (SFVs) distinguish between the two gene clusters (sY587T* or C*), between genes within each cluster (sY581C* and T*), and between *DAZ2* and the other three *DAZ* genes (sY586C* and T*). From de Vries et al. (2002).

Initially, seven *DAZ* genes were proposed but more recent research suggests there are only four, virtually identical (99.9%) gene copies (Szmulewicz et al. 2002). Research suggests that at least three of the four *DAZ* genes are transcribed in the testis during spermatogenesis and that the extent to which sperm production is impaired might depend on the number of *DAZ* genes deleted, and which cluster, in a gene dosage effect (de Vries et al. 2002; Fernandes et al. 2002).

DAZ family proteins are found in human germ cells at many developmental stages. In gonocytes, DAZL and DAZ proteins are distributed throughout the nucleus and cytoplasm, but predominate in the nuclei of spermatogonia. In early (round) spermatocytes the proteins are found exclusively in the cytoplasm, but only DAZL is

observed in the cytoplasm of elongating spermatids (Reijo et al. 2000). This suggests that DAZ proteins have multiple roles during gametogenesis, most likely during meiosis and in the establishment of spermatogonial stem cell populations (Reijo et al. 2000). The similar but not identical expression patterns of *DAZ* and *DAZL* suggests they have different functions (Reijo et al. 1995; Reijo et al. 1996; Reijo et al. 2000).

3.1.3 Y Chromosome Micro-deletion Analysis

Commonly, analysis of the AZF regions in population studies, and even fertility testing at reproductive health centres, has taken the form of sequence tagged site (STS) micro-deletion detection by PCR. Micro-deletions do not necessarily represent deletion of a entire gene, some STS markers tag DNA sequences whose function is unknown but assumed to be important to male fertilising potential (Foresta et al. 2001).

However, the presence of other identical gene copies or sequences (pseudogenes) suggests that PCR results are in fact more complicated than first realised. In the case of the *DAZ* gene family, if the STS marker sequence is present in all four gene copies, a positive PCR result indicates between one and four gene copies is intact, yet if *DAZ* dosage is important the presence of only one copy may actually result in reduced sperm production. A negative PCR result indicates that all four copies are deleted (Foresta et al. 2001). Recently, more sensitive STS markers have been developed that distinguish between individual DAZ clusters, and genes (Figure 3.8) (de Vries et al. 2002; Ferlin et al. 2002).

Thus the methods used to study Yq micro-deletion impact on the research outcome and conclusions that can be drawn from it. The association between deletions of particular spermatogenesis genes and identifiable semen characteristics is strengthened when study populations are selected based on testicular phenotype. However, depending on the screening method employed, seemingly similar deletion intervals can be observed among varied testicular phenotypes because deletions in some men might extend beyond the particular interval analysed in the study (Foresta et al. 2001).

Additionally, it is not fully understood whether X-linked or autosomal homolog expression alleviates or changes the pathology resulting from deletion of the Y-linked copy. This may be particularly true for the *DAZ* gene family, since *DAZL* is the only *DAZ* protein expressed in mature spermatids. Furthermore, without DNA analysis of male relatives it is difficult to distinguish natural polymorphisms from those that cause impaired spermatogenesis (Foresta et al. 2001).

3.1.4 Aim of Research

The Yq micro-deletion analysis aims to establish whether deletions within the three AZF regions contribute towards defective spermatogenesis in this study population. It is probable that mutations within more than one region, or multiple mutations in one region create more severely impaired spermatogenesis; this will be determined and then examined alongside the androgen receptor and *POLγ* data to further ascertain what mutations are present together in relation to semen quality (Chapter Six).

3.2 Method

3.2.1 Yq Multiplex PCR

3.2.1.1 Yq Multiplex PCR Primer Pairs

Fourteen Y chromosome long arm STS' were selected to span AZFa (sY84 and sY86), AZFb (sY127 and sY134) and AZFc (sY146, sY149, sY156, sY157, sY158, sY239, sY254 (*DAZ*), sY269, sY277) and sY14 (*SRY*) was used as a control marker (Figure 3.9).

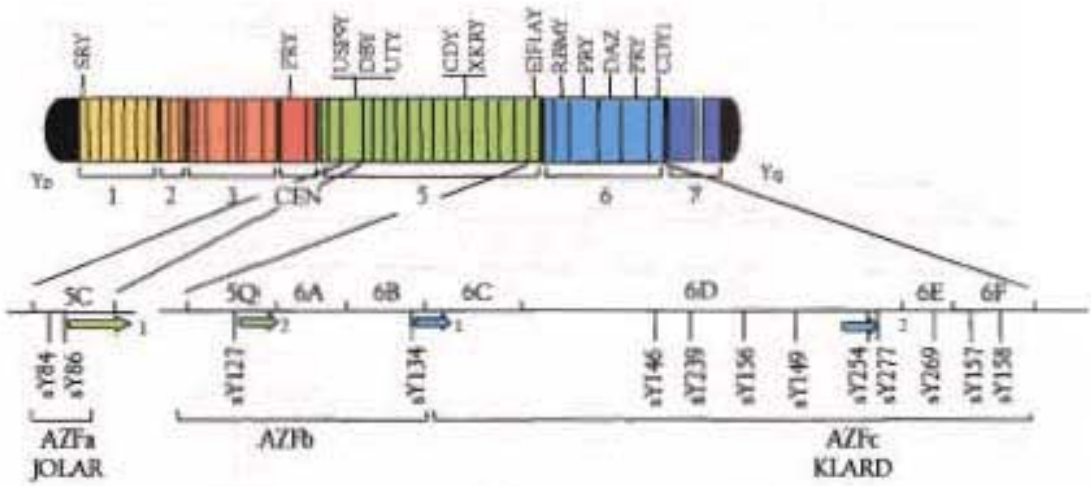


Figure 3.9. Y chromosome deletion interval map. The interval map was adapted from Vollrath (1992), using information published in Kent-first et al. (1999) and Vogt (1998). Each interval is numbered 1-7 and is represented by a different colour. Bands within each interval represent subintervals which are lettered (not indicated except in enlargement). Interval 4 (red) includes the centromere (CEN), and interval 7 (dark blue) encodes the 40Mb block of heterochromatin (interrupted). The pseudoautosomal regions are indicated at either end of the chromosome in black. AZFa is located within deletion interval 5C; AZFb spans intervals 5Q-6B; and AZFc spans 6C-6E (Vogt et al. 1998). The location of some spermatogenesis genes whose location is known with respect to STS' used in the study are indicated with coloured arrows; green arrow 1 represents *USP9Y*, *DBY* and *UTY*; green arrow 2 *EIF1AY*; blue arrow 1 *RBMY1*; blue arrow 2 *DAZ*.

Primer sequences were selected from those published in Vollrath (1992), and Reijo et al. (1995 and 1996) (Table 3.2). The fourteen PCR primer pairs that amplify these STS markers were divided into four groups based on product size then tested in a multiplex PCR to ensure positive amplification of each STS (Table 3.3).

Table 3.2. Yq STS primer pairs.

STS	Locus	Left Primer	Right Primer	Product Size (bp)
sY14	SRY	gaatattcccgctctccgga	gctggtgctccattcttgag	472
sY84	DYS273	agaagggtctgaaagcaggt	gcctactacctggaggcttc	300
sY86	DYS148	gtgacacacagactatgcttc	acacacagagggacaaccct	290
sY127	DYS218	ggctcacaaacgaaaacaaa	ctgcaggcagtaataaggga	280
sY134	DYS224	gtctgcctcaccataaaacg	accactgccaaaactttcaa	290
SY146	DYF52S1	acaaaaatgtggctcaggga	aaatagtgtgcccacccaaa	173
SY149	DYS1	tgtcacactgccctaatcct	tggtcatgacaaaagacgaa	132
SY156	DYS239	aggaactggcaggattagcc	atgtcagggtttcctttgcc	950
SY157	DYS240	cttaggaaaaagtgaagccg	cctgctgtcagcaagataca	285
SY158	DYS241	ctcagaagtcctcctaatagttcc	acagtggtttgtagcgggta	231
SY239		cattcatcttccctttgaagg	atgcaagtcgcaggaaatct	201
SY254	*DAZ*	gggtgttaccagaaggcaaa	gaaccgtatctaccaaagcagc	380
SY269		ctctgggacaagtgttccttg	cattggcatgaatgtgtattca	94
SY277		gggttttgcctgcatacgtaatta	cctaaaagcaattctaaacctccag	300

All primer sequences are given in the 5' to 3' direction.

Table 3.3. Yq STS multiplex groups.

Group	STS	Product Size (bp)[a]	AZF Region
1	sY269	94	AZFc
	sY146	173	AZFc
	sY158	231	AZFc
	sY277	300	AZFc (*DAZ*)
	sY156	950	AZFc
2	sY149	132	AZFc
	sY239	201	AZFc
	sY157	285	AZFc
	sY254	380	AZFc (*DAZ*)
	sY14	472	SRY
3	sY84	300	AZFa
	sY134	290	AZFb
4	sY86	290	AZFa
	sY127	280	AZFb

[a]Multiplex product sizes are given in base pairs (bp) and in ascending order of size.

3.2.1.2 Yq Multiplex PCR Conditions

PCR conditions were based on multiplex conditions published in Kent-First et al. (1999). The four multiplex PCR's followed the same cycling programme which consisted of an initial denaturation at 96°C for 2 min, followed by 34 cycles of denaturation at 94°C,

primer annealing at 61°C, and primer extension at 72°C, all at 1 min cycles, and a final extension at 72°C for 10 min. This programme was later trimmed back to 30 s denaturation and annealing, and a 40 s extension for 30 cycles, and a 5 min final extension. Since the PCR product was not required for sequencing, and in order to conserve DNA, the PCR reactions were conducted in 6.25 μl reaction volumes, and 6 μl electrophoresed through an agarose or a polyacrylamide gel.

3.2.1.3 Yq Multiplex PCR Visualisation

Size differences within groups 1 and 2 (AZFc) (Table 3.3) were not less than 20 bp and could be visualised as distinct bands after electrophoresis through a 2% (w/v) agarose gel with 1 x TAE buffer for 75 min at 8 V/cm.

Groups 3 and 4 (AZFa and b) (Table 3.3) were more similar in size and therefore, to avoid conducting four individual PCR's, the products were visualised by electrophoresis through a 5% non-denaturing polyacrylamide gel (5% acrylamide, 1 x TBE, 0.07% APS; 10 cm x 7 cm x 0.75 mm; Mini PROTEAN 3 (Biorad)) with 1 x TBE running buffer (0.045 M Tris-borate, 0.001M EDTA (pH 8.0))(Appendix F) and 1 x loading dye III (0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol in water). Gels were stained and visualised as described in Chapter Two Section 2.3.1.1.

3.2.1.4 Yq Multiplex PCR Reactions

Each multiplex PCR was run with a sY14 (*SRY*) positive control. Individual products were scored for a positive or negative result. The presence of a band indicated the STS marker was intact, while the absence of a band indicated a possible deletion. STS primers that gave a negative PCR result were repeated three times with controls for DNA concentration and the primer pair. If the band failed to be detected on all three occasions the product was probed with a labelled primer in a Southern blot hybridisation.

3.2.2 Southern Blot Hybridisation of Negative PCR Results

A total of thirty samples were analysed in four Southern blot hybridisations (Table 3.4).

Table 3.4. Samples analysed by Southen blot hybridisation.

	Sample[a]	Probed STS	Control STS
Southern blot 1[b]	1469	157	SRY
	2275	157	SRY
	2307	156 & 254	SRY & 149
	4646	156	149 & 157
	8055	156	149 & 157
	95-22	156 & 254	SRY
	RS	156 & 254	SRY, 149 & 157
	SM	156 & 254	SRY & 149
Probe 1	2050		SRY, 149, 156, 157, 254
Southern blot 2	2307	254	158
	3002	156	254, 158
	4646	156	158
	8055	156	no other STS
	95-22	156 & 254	158
	RS	156	146
	SM	254	158
Probe 2	4020		146, 156, 158, 254
Southern blot 3	2696	156	158
	3047	149	158
	4071	149	158
	4290	149	158
	4449	158	149
	5544	149	158
	BT	146 & 158	156
	CP	146 & 158	149
	SF96-283	156	149
	WT	156	254
	ZC	146	254
Probe 3a[c]	SF95-590		149, 157, 239, 254
Probe 3b[c]	TP		146, 158, 156, 277
Southern blot 4[d]	6290	146, 149 & 157	239 & 254
	AM302J	146 & 158	277
	CP	146 & 156	149 & 158
	SF94-380	149	158

[a] Some samples were repeated by probing several PCR reactions.
[b]Membrane 1 was reprobed with probe 2.
[c]Southen Blot 3 had two probes added to the hybridization.
[d]Southern Blot 4 was probed with probes 3a and 3b.

To set up the Southern blot experiments PCR samples were electrophoresed through a 1.5% (w/v) agarose gel (10 cm x 7 cm x 0.5 cm) with 1 x TAE buffer at 7 V/cm for 75 min. To assist transfer of DNA to the nitrocellulose membrane, the gel was illuminated under UV for 2 min after visualisation to shear the DNA into smaller fragments.

3.2.2.1 DNA Transfer

After UV irradiation the gel was depurinated by treatment with 0.125 M HCl. The gel was placed in a container, covered in 0.25 M HCl (~200 ml) and gently shaken for 15 min to break up the two-ringed purine nucleotides of the DNA. This solution was then decanted off and replaced with a denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 30 min (shaking). This solution hydrolyses the phosphodiester backbone at sites of depurination, denaturing the double-stranded DNA into single-stranded. The gel is then placed onto the transfer apparatus, lying wells-down onto a wick of three long strips of 3MM paper (Whatman chromatography paper) (pre-wetted in the transfer solution (1.5 M NaCl, 0.25M NaOH)). After eliminating air bubbles from underneath the gel, the nitrocellulose membrane (Hybond XL (Amersham Biosciences)) was placed with forceps on top of the gel. Three pieces of 3MM were placed on top of the membrane (both the membrane and 3MM are cut to the size of the gel) followed by a stack of tissue paper and a weight. Plastic (Gladwrap) was placed around the gel, touching its sides to prevent 'short circuiting' where solution is pulled past the gel and membrane by the paper above touching the wet wick. The system was left to transfer DNA onto the membrane overnight.

3.2.2.2 Pre-hybridisation of the Nitrocellulose Membrane

At all times the nitrocellulose membrane was handled with forceps to avoid contamination and dirty smears. The transfer system is first disassembled keeping the gel and membrane together. The locations of the well spaces are marked through the gel onto the membrane with a soft pencil, as is the location of the marker lane. The membrane is

briefly illuminated under UV (300 nm) to fix the DNA, then left to air dry between sheets of clean 3MM for 60 min while the pre-hybridisation solution (6 x SSC, 5 x Denhardts (2% Ficoll Type 400, 2% polyvinyl pyrrolidine, 2% bovine serum albumin), 0.5% SDS, 0.5% blocking agent) is prepared. Once dry the membrane was washed in 2 x SSC (0.3 M NaCl, 0.3 M (tri) sodium citrate) rolled up lengthwise and placed into a preheated (65°C) bottle (Hybaid, SciTech). Pre hybridisation solution was added (0.1 ml/cm^2) and the membrane was pre-hybridised in an oven (Hybaid, SciTech) at 65°C, turning for 1 hr.

3.2.2.3 Preparation of ^{32}P-labelled Hybridisation Probe

The PCR products to be used as hybridisation probes were first precipitated to remove unincorporated dNTP's and non-annealed primer pairs. The purified PCR products were then labelled with ^{32}P-dCTP using the Random Primed DNA labelling Kit (Boehringer Mannheim). The 9 μl of precipitated PCR reaction was denatured for 10 min at 97°C then added to the labelling mixture which consisted of 3 μl of dNTPs (1:1:1 mix of dATP, dGTP and dTTP; all 0.5 mM), and 2 μl of hexanucleotides (random 6 bp primers in 10 x reaction buffer). The ^{32}P-dCTP (3000 Ci/mmol; 5 μl) and 1 μl of klenow enzyme (2 u/μl in 50% glycerol (w/v)) were added last. The tube was then covered with a boiling hat, and heated at 37°C in a heating block for 10 min. The reaction is stopped by boiling for 10 min and then stored on watery ice until required.

3.2.2.4 Probe-Membrane Hybridisation

The labelled probe was added to 15 ml of the pre-hybridisation solution (Appendix F) and poured gently into the hybaid bottle avoiding the membrane. The probe is allowed to hybridise to DNA on the membrane overnight at 65°C.

3.2.2.5 Washing of the Radioactively Labelled Membrane

The radioactive hybridisation mix was poured off into a 50 ml falcon tube and stored at -20°C for future use. Excess, unbound probe was washed off the membrane in 2 x SSC for 5 min at 65°C, with agitation. This solution was discarded and washing was repeated with 2 x SSC for a further 15 min. A final wash in 2 x SSC, 0.1% SDS was conducted for 30 min. The membrane radioactivity was then measured on a Geiger counter (Technical Associates, CA, USA), to establish whether a further, more stringent wash (1 x SSC) was required. If the background count was low and count over the DNA was between 10,000 and 20,000 counts/minute/cm^2 (cpm) the third wash was omitted. The membrane was then wrapped in gladwrap for autoradiography.

3.2.2.6 Southern Blot Visualisation by Autoradiography

X-ray film (Kodak Biomax MR) was exposed to the membrane in an autoradiography cassette, with an amplifying screen. A membrane that read 10,000 to 20,000 cpm was exposed for 1 hr before developing, readings less than that required longer exposure. After developing the presence or absence of a band was scored.

3.2.3 Statistical Analysis of Micro-deletion Prevalence

A G-test of independence was performed to test the null hypothesis that semen quality is independent of the presence of Yq AZF micro-deletions, against the alternate hypothesis that semen quality is related to the presence of Yq AZF micro-deletions (Sokal and Rohlf 1995; Zar 1996).

3.3 Results

Y chromosome AZF micro-deletions were only observed in men with poor semen quality ($P<0.001$) (Table 3.5) and only within the AZFc locus; no micro-deletions of men with normal semen parameters, or of AZFa or b were identified. Y chromosome micro-deletions predominated in men with the most severe classes of spermatogenic failure (azoospermia or severe oligoasthenozoospermia; 10/15 observed micro-deletions).

3.3.1 Yq Multiplex PCR Analysis

A total of 292 individuals were screened for deletions of 14 Y chromosome STS markers, spread throughout the three AZF regions, two each in AZFa and b, and nine in AZFc, plus *SRY* (sY14) as a positive control. An additional 23 samples were screened for deletion in the AZFc region only, because DNA quantity was insufficient for analysis of AZFa and b also. Mulitplex PCR reactions of groups 1 and 2 (AZFc STS markers) were visualised on 2% (w/v) agarose gels, any indistinct or absent bands were amplified by PCR again (Figure 3.10).

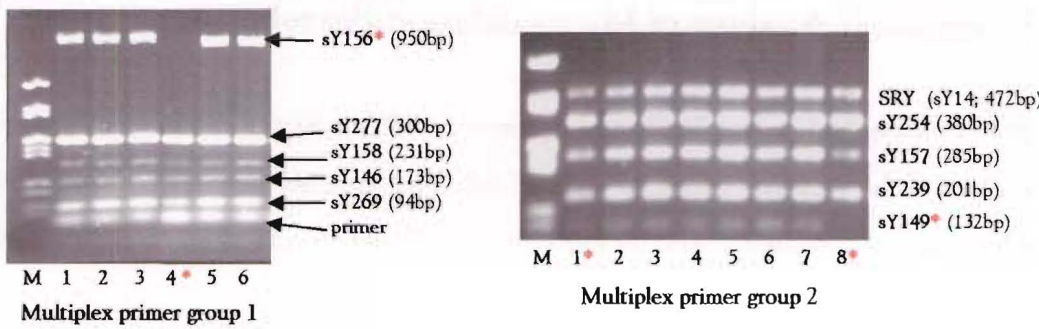


Figure 3.10. Electrophoresis images of Yq multiplex PCR groups 1 and 2. Multiplex PCR group 1 STS markers are indicated to the right of the left-hand agarose gel image with band size in basepairs. M signifies the molecular weight marker lane; pBS *(HaeIII)*. Individuals 1-6 have semen classifications of N, As, O, SOAs, SO, and SOAs. The red asterisks by individual 4 and sY156 highlight the absence of the STS marker band in that sample. Group 2 STS markers and band size are indicated in the right-hand image. The size marker is as per group 1. Individuals 1-8 have N, As, N, As, N, N, N, N, semen classifications. Red asterisks beside individuals 1 and 8, and sY149 indicate two samples with indistinct bands in the sY149 STS marker position.

Separation of multiplex PCR bands of groups 3 and 4 (AZFa and b) was achieved by electrophoresis through 5% PAGE (Figure 3.11).

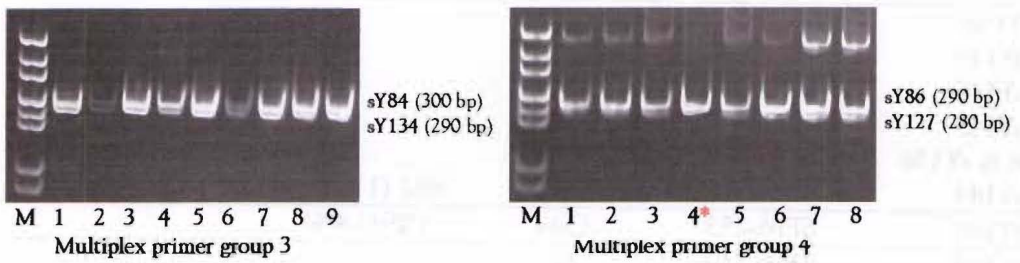


Figure 3.11. Electrophoresis images of Yq multiplex PCR groups 3 and 4. Both groups have one STS marker from AZFa (sY84 or sY86) and one from AZFb (sY127 or sY134). The marker lane is pBS *(HaeIII)* as per Figure 3.14. Individuals depicted in the gel image of group 3 have the following semen classifications, N, Az, N, As, SOAs, SO, SO, As, and N (1-9). Individuals in the image of group 4 are SOAs, As, OAs, N, OAs, F, OAs, and N (1-8). The asterisk beside individual 4 is indicative of a sample whose sY127 band is indistinct and would be repeated. Group 4 primers amplified additional non-specific bands approximately 580bp in size, these are likely to be the result either of a combination of the two primer pairs sY86 and sY127, or a second priming site in the genome.

Of the total 315 samples analysed for Yq11 micro-deletion, 23 displayed three consistently negative PCR results, including one individual who had three possible AZFc micro-deletions. These samples were further analysed by Southern blot hybridisation.

3.3.2 Southern Blot Analysis

Southern blot hybridisation analysis confirmed that 14 of the 23 individuals did have deletions of AZFc STS' (Table 3.5) (Figure 3.12), including one individual who had two non-adjacent STS deletions within AZFc, and a third marker confirmed as intact (CP, Table 3.5 and 3.6); this total includes three individuals who were only screened for AZFc deletions (asterisked Table 3.5).

Table 3.5. Samples identified with Yq micro-deletions.

STS Deleted	Sample	Semen Classification	Frequency[a]	Comments
sY149	4071	Az		Vasectomy Reversal[b]
sY158	4449	Az		
sY149*	5544	Az	3/20(15%)*	
sY156*	2696	SOAs		
sY149	4290	SOAs		
sY149	6290	SOAs		
sY146	BT	SOAs		
sY146 & sY156	CP	SOAs		
sY146	Zc	SOAs	6/52 (11.5%)*	
sY156	SF96-283	OAs	1/29(3.4%)	
sY149	SF96-380	O		
sY158*	AM302J	O	2/28(7%)*	
sY149	3047	As		
sY156	WT	As	2/56(3.6%)	

[a]Frequencies are displayed as the proportion of the total number of individuals in the semen classification. The asterisked* micro-deletions indicate samples for which the extracted DNA volume was not sufficient for screening in all three AZF regions; they have only been screened for deletion of the nine AZFc STS'. [b]The semen quality of this individual was analysed after his vasectomy reversal; his semen quality prior to that is unknown. Neither the SO nor N semen classification groups had any micro-deletions identified (0/21 and 0/104, respectively). For more detail refer Appendix D.

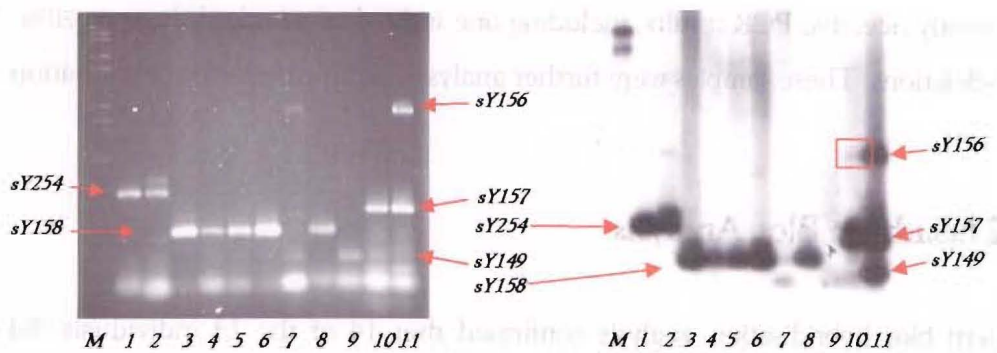


Figure 3.12 Southern blot hybridisation 3. On the left is the agarose gel image before transfer to the nitrocellulose membrane. Various STS marker bands of note are indicated. The right-hand image is a scan of the autoradiography film after developing. The red box highlights a very faint possible band in lane 10 that was being probed for, although the sY157 positive is much darker. Southern blot 4 indicates this is not a real band (Figure 3.13). M is λH/E (lamba DNA digested with *HindIII* and *EcoRI* restriction enzymes). All bands apparent on the Southern blot represent positive control bands; however the sample with the red box was repeated in Southern 4 below.

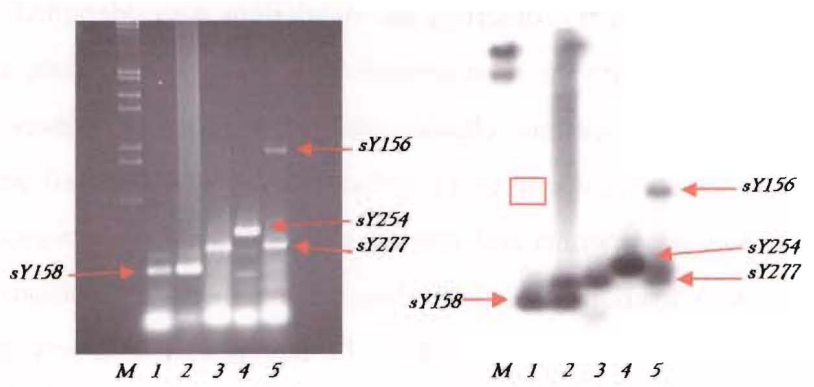


Figure 3.13. Southern blot hybridisation 4. The agarose gel image on the left shows additional non-specific bands that are not detected by the probe in the autoradiography image on the right. The red box indicates where a sY156 band was probed for; its absence suggests that the band-like smudge in Southern blot 3 was not a real band. There is a band in sample 2 that looks like a sY277 band although it does not appear with ethidium bromide staining (left), it may be a well placed spot like the one near the top of the gel since no sY277 primers were added to that sample. M is λH/E.

The remaining nine samples had the presence of DNA corresponding to the STS marker band confirmed by Southern blot hybridisation. This indicated that DNA, albeit at a very small concentration was present due to successful primer binding and amplification through the STS during the multiplex PCR reaction (Table 3.6).

Table 3.6. Samples with STS band presence confirmed by Southern blot.

Sample	Southern blot Number	STS confirmed	Semen Classification
8055	1	156	Az
4646	1	156	SOAs
2275	1	157	OAs
2307	1	156 & 254	O
SM	1	156 & 254	O
RS	1	156 & 254	As
1469	1	157	N
8055	2	156	Az
4646	2	156	SOAs
2307	2	254	O
95-22	2	156 & 254	O
SM	2	254	O
RS	2	156	As
3002	2	156	N
CP	3	158	SOAs

Sample CP appears in both Tables 3.5 and 3.6 because it had some deletions confirmed (Table 3.5) and one STS band verified as present (Table 3.6). Semen classification is abbreviated as per Table 3.5.

Exactly two-thirds (10/15) of the observed micro-deletions were identified in the two semen classification groups with the most severely affected spermatogenesis; azoospermia and severely oligoasthenozoospermia, characterised by the complete absence of sperm cells, or extremely low sperm count (<5.0 million sperm/ml) combined with reduced motility (<50%). Both azoospermia and severely oligoasthenozoospermia groups showed markedly higher micro-deletion frequencies than other semen classifications (15% and 11.5% respectively; Table 3.5). In total, 14 of the 211 (6.6%) men with poor semen quality were identified with micro-deletions; none of the 104 normozoospermic individuals were identified with any AZF micro-deletions.

3.3.3 Statistical Analysis

The difference in Yq micro-deletion frequency between normozoospermic and non-normozoospermic groups was statistically tested using a G-test of independence. A 2x2 contingency table (Table 3.7) was created in order to evaluate the relationship.

Table 3.7. Yq micro-deletion contingency table.

		NormoZS	Non normoZS	Total
Yq m-d*	obs	0	14	14
	exp	4.62	9.38	
No Yq m-d*	obs	104	197	301
	exp	99.38	201.62	
Total		104	211	315

*Yq m-d refers to samples identified with AZF STS micro-deletions; the "No Yq m-d" row indicates numbers of samples that did not have any micro-deletions. "obs" is observed numbers of individuals and "exp" is expected frequencies.

The null hypothesis is that semen quality is unrelated to micro-deletion presence; against the alternate hypothesis that they are related. A William's correction was used as recommended by Sokal and Rolhf (1995).

At an α level of 0.05, $G_{adj}=11.09$, and therefore, because $\chi^2_{0.05,1}=3.841$ is less than 11.09, the null hypothesis is rejected; semen quality is related to the presence of Yq AZF micro-deletions ($P<0.001$).

3.4 Discussion

3.4.1 Comparison with Published Yq Micro-deletion Frequencies

Screening of the Yq AZF region identified 14 individuals with Yq micro-deletions. These were restricted to men with poor semen quality (6.6%) and were most commonly observed in men with severe spermatogenic defects. Not surprisingly this difference in deletion frequency between normozoospermic and non-normozoospermic individuals was statistically significant ($P<0.001$). A comparison of other Y chromosome micro-deletion studies suggests that the deletion frequency commonly ranges from three to seven percent among men with poor semen quality (Vogt et al. 1996; Girardi et al. 1997; Pryor et al. 1997; Kent-First et al. 1999), but is significantly higher if the study population is intentionally selected based on semen characteristics (Kent-First et al. 1999). This suggests that more meaningful conclusions regarding particular male semen characteristics and the genetic mutations that cause them would come from analysis of highly selected study populations. While this seems rather obvious, it is harder to put into practise since gathering enough individuals of a particular phenotype so that results have significance can be difficult.

Nevertheless, while studies such as this one may not necessarily highlight which Yq chromosomal regions are associated with particular phenotypes, the fact that Yq deletions are consistently found in 5-10% of men with impaired spermatogenesis suggests they play an important role in the predisposition to reduced sperm production (Qureshi et al. 1996; Pryor et al. 1997; Grimaldi et al. 1998; Liow et al. 1998; Kent-First et al. 1999; Krausz et al. 1999; Kerr et al. 2000; Dohle et al. 2002; Madgar et al. 2002; Erasmuson et al. 2003).

Yq micro-deletions are extremely rare in men with normal semen parameters. Of the published data analysed only three studies have reported finding any deletions within normozoospermic controls; 4/200 (2%) (Pryor et al. 1997)), 8/920 (0.87%) (Kent-First et al. 1999) and 1/141 (0.71%) (Kerr et al. 2000). While several deletion intervals are associated with azoospermia, JOLAR in AZFa, and KLARD in AZFc, the presence of

small deletions of the Y chromosome's AZF region do not preclude the possibility of fathering offspring. However, many of the fathers found to harbour micro-deletions reported having trouble conceiving and several choose to utilise alternative means of having children (Vogt et al. 1996; Pryor et al. 1997). Sperm from men with Y chromosome micro-deletions are capable of achieving fertilisation through ICSI (intracytoplasmic sperm injection) although as expected sons inherited the AZF deletions present in their father (Jiang et al. 1999; Kamischke et al. 1999; Page et al. 1999).

3.4.2 Micro-deletion within AZFc.

Of the 15 micro-deletions identified, the majority of deletions were of the sY149 marker (6/15). This STS falls close to the *DAZ* marker (sY254), but has no sequence homology with the *DAZ* sequence (*Entrez* accession numbers NM_004081.3 and NM_020420.1, National Centre for Biotechnology Information (NCBI) www.ncbi.nlm.nih.gov) according to BLAST (Basic local alignment search tool; NCBI, (Altschul et al. 1990)). Interestingly, although the sY277 marker has been reported to be an additional marker for the *DAZ* gene family (Reijo et al. 1996) it did not align with either of the *DAZ* gene sequences submitted to the *Entrez* database (given above), while sY254 aligned with both.

The four deleted AZFc markers (sY146, sY149, sY156, and sY158) do not have sequence similarity to any of the favoured AZFc spermatogenesis genes, and are therefore presumed to bind to intervening regions, pseudogenes and introns within AZFc (Skaletsky et al. 2003). Given that most deletions of AZFc are larger than 1Mb and remove all five gene families (*BPY2*, *CDY1*, *DAZ*, *PRY*, and *RBMZ*) this suggests that the majority of the individuals in the study population did not have large deletions of that nature, since STS' dispersed between the AZFc spermatogenesis genes were amplified successfully. However, it is possible that the deletions observed may extend into spermatogenesis genes on their boundaries and result in reduced spermatogenesis.

The locations of the AZFc STS markers used in this study given on the NCBI website (www.ncbi.nlm.nih.gov) suggest that the individual identified with two AZFc micro-deletions (CP, Table 3.5) appears to have two adjacent deletions (Figure 3.14). However, these are separated by thousands of base pairs and may not necessarily represent a large deletion. Analysis of the father's and brother's AZF region would indicate whether the deletions were inherited or are the result of a single or two separate *de novo* deletion events. Notably the gene positions of the *DAZ* and *RBMY* gene families, *BPY2*, and *CDY* are different between the NCBI website, de Vries et al. (2002), and Skaletsky et al. (2003) making definitive conclusions difficult.

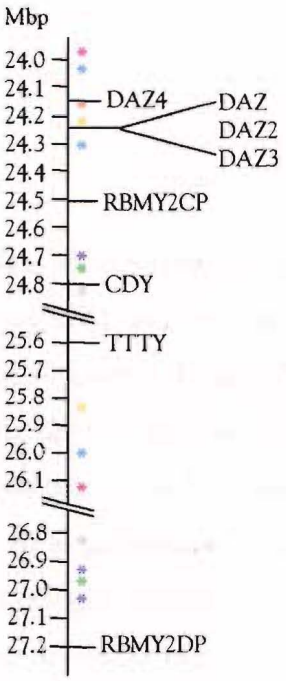


Figure 3.14. Schematic map of AZFc showing location of STS markers. STS marker locations were identified using the Reverse e-PCR web-based programme ((Schuler 1997) NCBI website www.ncbi.nlm.nih.gov). All potential binding sites are given within the *Homo sapiens* genome sequence (taxon reference 9606). Other Y chromosome genes or pseudogenes located on the Y chromosome STS map viewer are noted. The approximate nucleotide location along the Y chromosome is given on the left hand side and corresponds to Yq11.22-23. RBMY2CP and DP are RBMY family 2 C and D pseudogenes.

- sY156
- sY146
- sY149, sY254, and sY277
- sY254
- sY157
- sY269
- sY158

3.4.3 Absence of Micro-deletions within AZFa and b.

The failure to identify micro-deletions of AZFa is not necessarily surprising since they are more often observed associated with very severe reductions in semen quality (10% of SCO patients (Foresta et al., 2000)) and only a single individual in the entire study population was noted as having this type of testicular pathology. Furthermore, the two STS' selected to screen for AZFa deletions fall proximal to all three candidate spermatogenesis genes so that only large deletions that eliminate all of *USP9Y* and

proximal to that would be detected (Figure 3.9). Perhaps use of sY87 which is located between *USP9Y* and *DBY* might detect deletion intervals corresponding to elimination of one or other, or both genes (Figure 3.7).

As with AZFa, STS markers selected for AZFb did not actually bind to any candidate spermatogenesis genes, but to the DNA sequences surrounding them. However, deletions involving the AZFb STS' used in this study do occur in both AZFa and b deletion intervals. STS sY127 was part of a common deletion observed in four men with severe spermatogenic failure that encompassed the *EIF1AY* and *SMCY* genes (Ferlin et al. 2003). A further eight men in the same study had deletions of the entire AZFa interval encompassing both sY127 and sY134.

3.4.4 Yq STS Deletion Analysis by PCR

It is likely that deletion frequency of multicopy genes has been grossly underestimated since PCR will only detect a deletion if all copies are missing. The *DAZ* gene family for example, is present in two clusters each with two *DAZ* genes. It is thought that only one of the four *DAZ* copies needs be intact for the STS primer pair to bind and amplify a product by PCR, yet deletion of three DAZ copies is likely to have a profound impact on an individual's semen quality. Therefore, in the time since sequencing of the MSY more sensitive approaches have been developed. Alternatively, restriction digestion of two new STS', sY581 and sY587 which target subtle sequence differences located in the introns of some *DAZ* gene family members, is able to distinguish between the four copies. In this way it is possible by means of elimination to determine which gene copies of the four are present, and which are deleted (Saxena et al. 2000; Ferlin et al. 2002). Notably this technique identified deletions of *DAZ* copies 1 and 2 in two men with severe spermatogenic failure that had previously been cleared of *DAZ* deletions by standard sY254 STS deletion analysis (Ferlin et al. 2002).

It is possible therefore, that deletions of *DAZ* genes and other multicopy families such as *RBMY* were missed in the AZF deletion screening conducted in this study, which might explain why no deletions of *DAZ* (sY254) were observed.

3.4.5 Yq Deletion Analysis by Southern Blot

Analysis of samples using Southern blot hybridisation verified both the presence and absence of DNA representing a PCR-amplified STS band (Tables 3.5 and 3.6). The significance of confirming the presence of an indistinct PCR band using Southern hybridisation, may suggest one of several possibilities. Firstly, the PCR reaction mix was inhibited in the same way, on at least three different occasions. Generally, problems with PCR reactions occurred in samples with low sperm count, where DNA concentration was low and there is the possibility that reaction inhibitors were retained in the sample during DNA extraction in an effort to maximise DNA yield. Alternatively, low concentration PCR products might suggest that the copy number for the Y chromosome gene is reduced, with the result that amplification is lower because there are less primer binding sites. This is less likely to be the answer in this case, since a BLAST alignment suggested that none of the STS' used corresponded to any of the candidate spermatogenesis gene families identified to date.

REFERENCES

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215: 403-10.

Brown, G.M., R.A. Furlong, C.A. Sargent, R.P. Erickson, G. Longepied, M. Mitchell, M.H. Jones, T.B. Hargreave, H.J. Cooke, and N.A. Affara. 1998. Characterisation of the coding sequence and fine mapping of the human DFFRY gene and comparative expression analysis and mapping to the Sxrb interval of the mouse Y chromosome of the Dffry gene. *Hum Mol Genet* 7: 97-107.

Charlesworth, B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A* 75: 5618-22.

de Vries, J.W., M.J. Hoffer, S. Repping, J.M. Hoovers, N.J. Leschot, and F. van der Veen. 2002. Reduced copy number of DAZ genes in subfertile and infertile men. *Fertil Steril* 77: 68-75.

Dohle, G.R., D.J. Halley, J.O. Van Hemel, A.M. van den Ouwel, M.H. Pieters, R.F. Weber, and L.C. Govaerts. 2002. Genetic risk factors in infertile men with severe oligozoospermia and azoospermia. *Hum Reprod* 17: 13-6.

Erasmuson, T., I.L. Sin, and F.Y. Sin. 2003. Absence of association of androgen receptor trinucleotide expansion and poor semen quality. *Int J Androl* 26: 46-51.

Ferlin, A., E. Moro, A. Rossi, B. Dallapiccola, and C. Foresta. 2003. The human Y chromosome's azoospermia factor b (AZFb) region: sequence, structure, and deletion analysis in infertile men. *J Med Genet* 40: 18-24.

Ferlin, A., E. Moro, A. Rossi, and C. Foresta. 2002. A novel approach for the analysis of DAZ gene copy number in severely idiopathic infertile men. *J Endocrinol Invest* 25: RC1-3.

Fernandes, S., K. Huellen, J. Goncalves, H. Dukal, J. Zeisler, E. Rajpert De Meyts, N.E. Skakkebaek, B. Habermann, W. Krause, M. Sousa, A. Barros, and P.H. Vogt. 2002. High frequency of DAZ1/DAZ2 gene deletions in patients with severe oligozoospermia. *Mol Hum Reprod* 8: 286-98.

Foresta, C., A. Ferlin, and E. Moro. 2000. Deletion and expression analysis of AZFa genes on the human Y chromosome revealed a major role for DBY in male infertility. *Hum Mol Genet* 9: 1161-9.

Foresta, C., E. Moro, and A. Ferlin. 2001. Y chromosome microdeletions and alterations of spermatogenesis. *Endocr Rev* 22: 226-39.

Girardi, S.K., A. Mielnik, and P.N. Schlegel. 1997. Submicroscopic deletions in the Y chromosome of infertile men. *Hum Reprod* 12: 1635-41.

Graves, J.A.M. 2002. The rise and fall of SRY. *Trends in Genetics* 18: 259-264.

Grimaldi, P., C. Scarponi, P. Rossi, M.R. March, A. Fabbri, A. Isidori, G. Spera, C. Krausz, and R. Geremia. 1998. Analysis of Yq microdeletions in infertile males by PCR and DNA hybridization techniques. *Mol Hum Reprod* 4: 1116-21.

Hawley, R.S. 2003. The human Y chromosome: rumors of its death have been greatly exaggerated. *Cell* 113: 825-8.

Jiang, M.C., Y.R. Lien, S.U. Chen, T.M. Ko, H.N. Ho, and Y.S. Yang. 1999. Transmission of de novo mutations of the deleted in azoospermia genes from a severely oligozoospermic male to a son via intracytoplasmic sperm injection. *Fertil Steril* 71: 1029-32.

Kamischke, A., J. Gromoll, M. Simoni, H.M. Behre, and E. Nieschlag. 1999. Transmission of a Y chromosomal deletion involving the deleted in azoospermia (DAZ) and chromodomain (CDY1) genes from father to son through intracytoplasmic sperm injection: case report. *Hum Reprod* 14: 2320-2.

Kamp, C., P. Hirschmann, H. Voss, K. Huellen, and P.H. Vogt. 2000. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum Mol Genet* 9: 2563-72.

Kent-First, M., A. Muallem, J. Shultz, J. Pryor, K. Roberts, W. Nolten, L. Meisner, A. Chandley, G. Gouchy, L. Jorgensen, T. Havighurst, and J. Grosch. 1999. Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. *Mol Reprod Dev* 53: 27-41.

Kerr, N.J., J. Zhang, F.Y. Sin, P. Benny, and I.L. Sin. 2000. Frequency of microdeletions in the azoospermia factor region of the Y-chromosome of New Zealand men. *N Z Med J* 113: 468-70.

Krausz, C., C. Bussani-Mastellone, S. Granchi, K. McElreavey, G. Scarselli, and G. Forti. 1999. Screening for microdeletions of Y chromosome genes in patients undergoing intracytoplasmic sperm injection. *Hum Reprod* 14: 1717-21.

Krausz, C., G. Forti, and K. McElreavey. 2003. The Y chromosome and male fertility and infertility. *Int J Androl* 26: 70-5.

Kuroda-Kawaguchi, T., H. Skaletsky, L.G. Brown, P.J. Minx, H.S. Cordum, R.H. Waterston, R.K. Wilson, S. Silber, R. Oates, S. Rozen, and D.C. Page. 2001. The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat Genet* 29: 279-86.

Lahn, B.T. and D.C. Page. 1997. Functional Coherence of the Human Y Chromosome. *Science* 278: 675-80.

-. 1999a. Four evolutionary strata on the human X chromosome. *Science* 286: 964-7.

-. 1999b. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nat Genet* 21: 429-33.

Liow, S.L., F.J. Ghadessy, S.C. Ng, and E.L. Yong. 1998. Y chromosome microdeletions, in azoospermic or near-azoospermic subjects, are located in the AZFc (DAZ) subregion. *Mol Hum Reprod* 4: 763-8.

Madgar, I., L. Green, M. Kent-First, R. Weissenberg, R. Gershoni-Baruch, B. Goldman, and E. Friedman. 2002. Genotyping of Israeli infertile men with idiopathic oligozoospermia. *Clin Genet* 62: 203-7.

Mazeyrat, S., N. Saut, V. Grigoriev, S.K. Mahadevaiah, O.A. Ojarikre, A. Rattigan, C. Bishop, E.M. Eicher, M.J. Mitchell, and P.S. Burgoyne. 2001. A Y-encoded subunit of the translation initiation factor Eif2 is essential for mouse spermatogenesis. *Nat Genet* 29: 49-53.

Page, D.C., S. Silber, and L.G. Brown. 1999. Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection, but are likely to transmit the deletion and infertility. *Hum Reprod* 14: 1722-6.

Pryor, J.L., M. Kent-First, A. Muallem, A.H. Van Bergen, W.E. Nolten, L. Meisner, and K.P. Roberts. 1997. Microdeletions in the Y chromosome of infertile men. *N Engl J Med* 336: 534-9.

Qureshi, S.J., A.R. Ross, K. Ma, H.J. Cooke, M.A. Intyre, A.C. Chandley, and T.B. Hargreave. 1996. Polymerase chain reaction screening for Y chromosome microdeletions: a first step towards the diagnosis of genetically-determined spermatogenic failure in men. *Mol Hum Reprod* 2: 775-9.

Reijo, R., R.K. Alagappan, P. Patrizio, and D.C. Page. 1996. Severe oligozoospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet* 347: 1290-3.

Reijo, R., T.Y. Lee, P. Salo, R. Alagappan, L.G. Brown, M. Rosenberg, S. Rozen, T. Jaffe, D. Straus, O. Hovatta, and et al. 1995. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* 10: 383-93.

Reijo, R.A., D.M. Dorfman, R. Slee, A.A. Renshaw, K.R. Loughlin, H. Cooke, and D.C. Page. 2000. DAZ family proteins exist throughout male germ cell development and transit from nucleus to cytoplasm at meiosis in humans and mice. *Biol Reprod* 63: 1490-6.

Repping, S., H. Skaletsky, L. Brown, S.K. van Daalen, C.M. Korver, T. Pyntikova, T. Kuroda-Kawaguchi, J.W. de Vries, R.D. Oates, S. Silber, F. van der Veen, D.C. Page, and S. Rozen. 2003. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* 35: 247-51.

Rice, W.R. 1998. Male fitness increases when females are eliminated from gene pool: implications for the Y chromosome. *Proc Natl Acad Sci U S A* 95: 6217-21.

Sargent, C.A., C.A. Boucher, S. Kirsch, G. Brown, B. Weiss, A. Trundley, P. Burgoyne, N. Saut, C. Durand, N. Levy, P. Terriou, T. Hargreave, H. Cooke, M. Mitchell, G.A. Rappold, and N.A. Affara. 1999. The critical region of overlap defining the AZFa male infertility interval of proximal Yq contains three transcribed sequences. *J Med Genet* 36: 670-7.

Saxena, R., J.W. de Vries, S. Repping, R.K. Alagappan, H. Skaletsky, L.G. Brown, P. Ma, E. Chen, J.M. Hoovers, and D.C. Page. 2000. Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* 67: 256-67.

Schuler, G.D. 1997. Sequence mapping by electronic PCR. *Genome Res* 7: 541-50.

Skaletsky, H., T. Kuroda-Kawaguchi, P.J. Minx, H.S. Cordum, L. Hillier, L.G. Brown, S. Repping, T. Pyntikova, J. Ali, T. Bieri, A. Chinwalla, A. Delehaunty, K. Delehaunty, H. Du, G. Fewell, L. Fulton, R. Fulton, T. Graves, S.F. Hou, P. Latrielle, S. Leonard, E. Mardis, R. Maupin, J. McPherson, T. Miner, W. Nash, C. Nguyen, P. Ozersky, K. Pepin, S. Rock, T. Rohlfing, K. Scott, B. Schultz, C. Strong, A. Tin-Wollam, S.P. Yang, R.H. Waterston, R.K. Wilson, S. Rozen, and D.C. Page. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423: 825-37.

Sokal, R.R. and F.J. Rohlf. 1995. *Biometry: The Principles and Practise of Statistics in Biological Research.* W. H. Freeman and Company, New York.

Stouffs, K., W. Lissens, G. Verheyen, L. Van Landuyt, A. Goossens, H. Tournaye, A. Van Steirteghem, and I. Liebaers. 2004. Expression pattern of the Y-linked PRY gene suggests a function in apoptosis but not in spermatogenesis. *Mol Hum Reprod* 10: 15-21.

Sun, C., H. Skaletsky, S. Rozen, J. Gromoll, E. Nieschlag, R. Oates, and D.C. Page. 2000. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum Mol Genet* 9: 2291-6.

Szmulewicz, M.N., L.M. Ruiz, E.P. Reategui, S. Hussini, and R.J. Herrera. 2002. Single-nucleotide variant in multiple copies of a deleted in azoospermia (DAZ) sequence - a human Y chromosome quantitative polymorphism. *Human Heredity* 53: 8-17.

Tiepolo, L. and O. Zuffardi. 1976. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum Genet* 34: 119-24.

Vergnaud, G., D.C Page, M.C. Simmler, L. Brown, F. Rouyer, B. Noel, D. Botstein, A. de la Chapelle, J. Weissenbach. 1986. A deletion map of the human Y chromosome based on DNA hybridization. *Am J Hum Genet* 38:109-24.

Vogt, P.H. 1998. Human chromosome deletions in Yq11, AZF candidate genes and male infertility: history and update. *Mol Hum Reprod* 4: 739-44.

Vogt, P.H., A. Edelmann, S. Kirsch, O. Henegariu, P. Hirschmann, F. Kiesewetter, F.M. Kohn, W.B. Schill, S. Farah, C. Ramos, M. Hartmann, W. Hartschuh, D. Meschede, H.M. Behre, A. Castel, E. Nieschlag, W. Weidner, H.J. Grone, A. Jung, W. Engel, and G. Haidl. 1996. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* 5: 933-43.

Vollrath, D., S. Foote, A. Hilton, L.G. Brown, P. Beer-Romero, J.S. Bogan, D.C. Page. 1992. The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* 258:52-9.

Yen, P. 2001. The fragility of fertility. *Nat Genet* 29: 243-4.

Zar, J. 1996. *Biostatistical Analysis Third Edition.* Prentice-Hall International, Inc, Upper Saddle River, NJ.

CHAPTER FOUR ANDROGEN RECEPTOR TRINUCLEOTIDE EXPANSION

4.1 Introduction

The androgen receptor (AR) is not essential for life, just for maleness. As a consequence, it is the most frequently mutated human steroid receptor, with over 300 different mutations documented (Yong et al. 2000). These changes fall into two classes: increases in length of a polymorphic CAG trinucleotide repeat in the transactivation domain, or missense mutations of the ligand-binding domain; both reduce the transcription-activating function of the AR. It is predicted that mutations of the androgen receptor gene cause up to 10% of idiopathic male infertility (Yong et al. 2000).

4.1.1 The Androgen Receptor and Male Sexual Development.

Human embryos follow the same developmental path until the sixth week of gestation when male embryos are prompted to differentiate down the male developmental pathway by members of the SOX (SRY-like HMG box) gene family on the Y chromosome, which stimulate a cascade of gene activities resulting in male sexual development (Yong et al. 2000).

At puberty a surge of male androgen (the steroid sex hormones; testosterone and dihydrotestosterone (DHT)) triggers the differentiation of male secondary sexual organs and initiates spermatogenesis (Mifsud et al. 2001). This response of the male body to androgen is mediated by the AR. Binding of androgen to the ligand-binding domain (LBD) in the cytoplasm of the cell stimulates a conformational change in the AR in a region of the LBD that is made up of twelve α-helices. Folding of this region prevents further androgen binding and results in the dissociation of other proteins bound to the AR to reveal functional domains where transcriptional co-activators can bind. The androgen-bound AR translocates into the nucleus where it forms a dimer complex with another androgen-bound AR (Wong et al. 1993; Roy et al. 2001). The DNA-binding domain (DBD) facilitates binding of the androgen-AR homodimer complex to androgen-response elements in the regulatory regions (promoter and enhancer sequences) of target

genes. Co-activator proteins (transcription factors) bind to the transactivation domain (TAD) of the AR and fine-tune the amount of target-gene transcription activated by the AR (Gottlieb et al. 1999).

Male sexual development is not an androgen-driven process, but androgens are required for normal development; differentiation of Wolffian Ducts, inhibition of female sexual organ development (Quigley et al. 1995). Sperm production also requires high concentrations of androgen and a functional AR (Quigley et al. 1995).

4.1.2 The *AR* Gene

The *AR* gene is a single copy gene located on the long arm of the X-chromosome (Xq11). The coding region comprises nearly 2760 base pairs of DNA spread over eight exons. The AR protein is derived from a single polypeptide that varies in length depending on the number of trinucleotide repeats present in two polymorphic repeat regions within exon one (Figure 4.1) (Lubahn et al. 1988; Quigley et al. 1995).

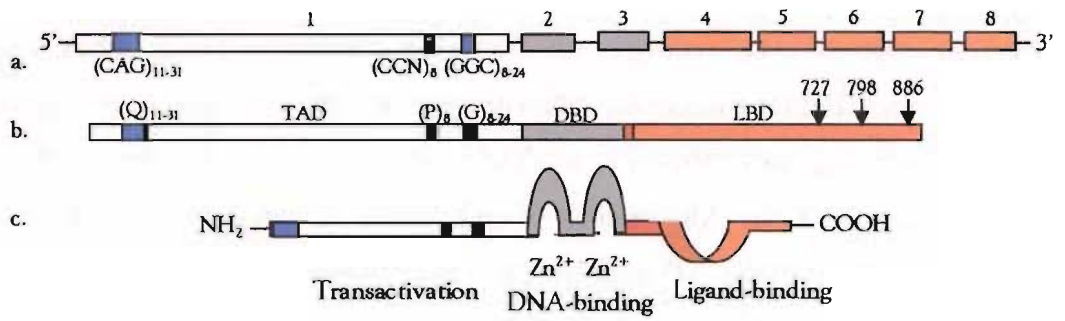


Figure 4.1. Schematic representation of the *AR* gene and protein. (a) The *AR* gene. Exons are colour-coded to protein domains. (b) The protein has three functional domains: the transactivation domain (TAD) from exon 1, the DNA-binding domain (DBD) from exons 2 and 3, and the ligand-binding domain (LBD) from exons 4 to 8. The protein is approximately 919 amino acids in length but varies depending on repeat number within two polymorphic repeat tracts; a polyglutamine (Q)(CAG)$_{11\text{-}31}$ and polyglycine (G)(GGC)$_{8\text{-}24}$ (blue boxes). The third trinucleotide repeat is an eight-unit proline repeat tract (P)(CCN)$_8$. Arrows indicate mutation hot-spots (Wang et al (1998b); Yong et al. (2000)). The hinge region containing the nuclear-targeting sequence is indicated by the red box. (c) AR protein domains. Adapted from Quigley et al. (1995).

The AR belongs to the steroid-receptor superfamily which also includes receptors for oestrogen, adrenal and thyroid hormones, retinal acid and vitamin D. All of which encode the three functional domains plus the hinge region (Figure 4.1) (Quigley et al. 1995).

The TAD is the largest AR domain comprising more than half the protein (~537 amino-acids (aa)). It has the least conserved DNA sequence and is the most variable in size between members of the steroid receptor family. Within the TAD are three trinucleotide repeat DNA regions; two polymorphic homopolymeric repeats; the CAG-encoded glutamine repeat (11-31 repeats; approximately Q^{58}-Q^{79}), and the GGC-encoded glycine repeat (8 to 24 repeats; roughly G^{449}-G^{472}), and a single eight-unit heteropolymeric (CCN) proline repeat (P^{372}-P^{379}) (Simental et al. 1991; Jenster et al. 1992).

The central, cysteine-rich DBD is 90 aa long and has the most highly conserved amino-acid sequence among steroid receptor family members. This region folds to create two loop structures that together form a single unit comprising two zinc finger motifs. Each motif is created by the binding of a zinc ion to four cysteine residues in each of the two exons. Three residues at the base of the first zinc finger (exon 2), glycine^{557}, serine^{578}, and valine^{581} determine the DNA specificity controlled by the DBD of the AR by interacting with enhancer sequences in 5' regulatory regions of target genes. The second zinc finger is basic and functions to stabilise the DNA/AR association by interacting with the phosphate backbone and mediating receptor dimerisation (Lubahn et al. 1988; Marcelli et al. 1990; Jenster et al. 1992; Quigley et al. 1995).

The hinge region is encoded by the 5' end of exon four, and contains all but two of the amino-acids that make up the nuclear-targeting sequence (arg^{629}-lys-leu-lys-lys^{633}) which directs the transportation of the AR from the cytoplasm to the nucleus upon androgen binding. The first two amino-acids of the targeting sequence are encoded at the 3' end of the TAD (arg-lys) (Jenster et al. 1992; Quigley et al. 1995).

The ligand-binding domain (LBD; 250 aa) facilitates the specific, high-affinity binding of male sex steroids to the receptor. This domain is also predicted to be the binding site of inhibitory proteins that prevent AR transcription activation in the absence of ligand and is involved in dimerisation of receptors (Housley et al. 1990; Quigley et al. 1995).

4.1.3 Co-activator Proteins Enhance AR Transactivation

A number of proteins form complexes with the AR and enhance its transactivation activity. Androgen receptor-associated protein 24 (ARA24, also known as GTP-binding nuclear protein RAN) is a member of the Ran (Ras-related nuclear protein) family of proteins. It functions as a GTP-binding protein and is predicted to be involved in the transport of proteins out of the nucleus, cell-cycle progression, and chromatin condensation (Hsiao et al. 1999). ARA24 enhances AR-mediated transcription activation by binding to the polyglutamine region. The binding affinity of ARA24 for the AR is inversely related to the number of CAG repeat units encoded in the polyglutamine region; interactions become stronger as glutamine number decreases resulting in enhanced AR transactivation capability, and are weakened with increasing numbers of CAG repeat units reducing transactivational activity (Hsiao et al. 1999).

Research indicates that co-activators bind at several different places on the AR (Irvine et al. 2000). The TAD contains a ligand-independent transcription-activation functional domain (AF1) that is repressed by the LBD when no androgen is bound. The AF1 region comprises two overlapping, but distinct transcriptional activation units: TAU1 and TAU5 at amino-acid residues 101-370 and 360-528, respectively (Figure 4.2). Both AF1 and AF2 are required for optimal transactivation activity of the AR (Irvine et al. 2000).

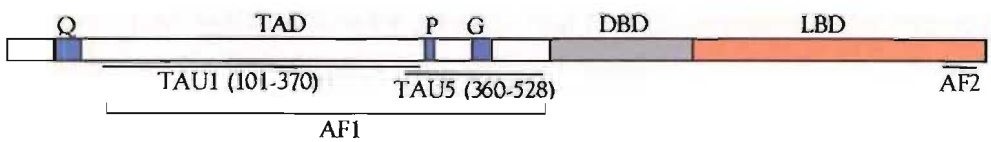


Figure 4.2. The location of transcription-activating functional domains within the TAD.

The LBD also contains an activation function region, AF2, where ligand binding induces a conformational change that repositions AF2 so that it can interact with AR transcription co-activators (Figure 4.2). Transcription activation of androgen-target genes requires both AF1 and AF2; however elimination of virtually any stretch of the TAD results in a dramatic reduction (80%) in ligand-dependent transactivation activity by the AR (Irvine et al. 2000).

4.1.4 Androgen Insensitivity Syndromes

Normal function of the AR is essential in the developing embryo to ensure proper differentiation of the male genitals both internally and externally. Mutations that alter the ligand- or DNA-binding activity of the AR are associated with a subtle reduction in AR target gene activation and cause "androgen-insensitivity" syndromes (AIS). At its most severe, AR dysfunction results in complete androgen insensitivity syndrome (CAIS), characterized by sex reversal in otherwise healthy 46XY individuals; patients with CAIS have normal testosterone levels but entirely feminine reproductive phenotype. Mutations of the *AR* gene that result in a less severe impairment to AR functioning result in partial androgen insensitivity syndrome (PAIS) where the physiology of male genitalia can appear anywhere from the normal male phenotype to almost completely feminine genitalia. PAIS patients commonly experience defective spermatogenesis and male infertility (Marcelli et al. 1990; McPhaul 1999; Lim et al. 2000).

Malfunction of the AR is associated with many androgen-related conditions, such as abnormal genital development, spinal bulbar muscular atrophy, breast, ovarian and postate cancer, and even baldness and acne (Yong et al. 2000).

Most studies of *AR* mutation have identified defects in androgen- and DNA-binding domains because these changes produce characteristic AIS phenotypes in affected males (Puscheck et al. 1994). Mutations range from single nucleotide changes that cause missense amino-acid substitutions to large deletions of the gene (Loy and Yong 2001). Substitutions within the DNA-binding domain impair the capacity of the receptor to bind

to target DNA sequences within promoter and enhancer regions. Substitutions within the LBD have more varied effects on receptor function but often result in disruption of associations between the transcription co-activators and the TAD (Loy and Yong 2001). Both types of mutation result in reduced DNA binding by the AR, leading to diminished transcriptional activation and reducing AR regulation of target gene activity during sexual development (Lim et al. 2000; Loy and Yong 2001). Two regions within the LBD are 'hot-spots' for amino-acid substitutions, these regions; spanning residues 726-776 and 828-886 (Figure 4.1) are often found associated with AIS phenotypes. A single substitution within this region is sufficient to impair receptor functioning and drastically alter the sexual appearance of individuals (Puscheck et al. 1994; Wang et al. 1998a; Wang et al. 1998b; Loy and Yong 2001).

Although almost the entire TAD region is important for normal receptor transactivation, few deleterious mutations have been characterised; only 18 mutations have been observed in AIS patients. Most of these mutations produce a premature stop codon resulting in a severely truncated and non-functional AR protein. The remainder results in a reduction in transactivation; deletion of amino-acid residues 101-370 reduces transactivation activity to 50% of wild-type. While these mutations are predicted to be the cause of the reduced spermatogenesis in these men, they do not always result in infertility. At least one of the substitutions has also been observed in normozoospermic controls, and many more produce varying degrees of AIS phenotypes between different individuals (Wang et al. 1998b; Mifsud et al. 2001). This suggests that genetic background is important in the determination of fertility. The lack of TAD mutations identified in AIS individuals suggests this domain does not contribute significantly to this pathology. It does not necessarily follow however, that the domain is infrequently mutated since genetic studies of the AR have largely comprised screening for mutation in individuals with some degree of androgen insensitivity symptoms (Loy and Yong 2001). Alternatively, it is hypothesised that mutation of the TAD might alter expression patterns of genes essential for aspects of male sexual development not necessarily associated with outward appearance, for example, sperm production (Puscheck et al. 1994).

4.1.5 The *AR* Gene and Trinucleotide Repeat Expansion

The polymorphic polyglutamine region encoded by the CAG trinucleotide repeat varies normally between members of the population (11 to 31 repeat units), and between ethnicities (Edwards et al. 1992). Men of African descent tend to have shorter CAG repeat alleles, with an average of 18 repeats, Asian men have on average, 23-24 repeats (Mifsud et al. 2001), and Caucasian men fall in between with an average of 21 repeats (Quigley et al. 1995; Hardy et al. 1996). Glutamine-rich regions enhance transcription factor binding, therefore, the AR polyglutamine region is predicted to function by regulating the amount of gene transcription initiated by the AR, through its interaction with co-activators such as, TIF2 (transcriptional intermediary factor 2), ARA24 (androgen-receptor associated protein 24), and P160-coactivators (Gerber et al. 1994; Hsiao et al. 1999; Irvine et al. 2000; Yong et al. 2000).

The AR polyglutamine region, like other trinucleotide repeat tracts, can undergo either expansion or contraction in repeat number during DNA replication. Gross expansion (>36 CAG repeats) of the repeat tract results in spinal bulbar muscular atrophy (SBMA; commonly known as Kennedy's Disease) (La Spada et al. 1991). Kennedy's disease is a severe, progressive, adult-onset neuromuscular disease that causes weakness of the muscles in the extremities, and those controlling speech and swallowing. Secondary symptoms commonly include AIS gynecomastia (the development of breast tissues in men, often during adolescence), testicular atrophy and severely reduced spermatogenesis (oligozoospermia or azoospermia) (Yong et al. 2000). Due to the X-linked nature of the disease it is maternally inherited and only affects men; female carriers are usually phenotypically normal (Yong et al. 2000).

In contrast, a reduction in glutamine repeat number increases the risk of androgen-driven cancers, specifically prostate, breast, and ovarian; less than 20 *AR* CAG repeats doubles the risk of developing prostate cancer (Sartor et al. 1999). Contraction of the *AR* repeat disrupts receptor specificity so that it becomes activated by ligands other than male androgen, resulting in hyperstimulation and prolific growth of target tissues. The X-linked

and maternally inherited nature of *AR* mutation correlates with genetic features of cancer heredity, where history in a brother indicates greater risk than in the father (Giovannucci et al. 1997). Interestingly, breast cancer risk is heightened in women who carry the BRCA1 gene mutation and have long AR polyglutamine regions (>28 CAG repeats) (Rebbeck et al. 1999).

4.1.6 Mechanism of Trinucleotide Repeat Expansion

The *AR* CAG trinucleotide repeat is a member of a class of DNA called microsatellites; a stretch of DNA made up of units, 1-6 bp in length, repeated multiple times in a head to tail orientation; in this case CAGCAG...$(CAG)_{11\text{-}31}$. Most microsatellites are polymorphic in unit copy number between individuals, differing by small increases and decreases in repeat number within the normal range, and are not associated with disease (Kovtun et al. 2001). The dramatic increase in trinucleotide repeat number associated with disease often occurs during transmission from parent to offspring, and repeat number correlates with both disease severity, and age of onset; the longer the allele the earlier the age of onset and more severe the disease symptoms (also known as anticipation) (Kovtun et al. 2001).

Although the mechanism of microsatellite expansion is not fully understood it is thought to be due to misaligned pairing of repeat stretches and formation of secondary structures (Kovtun and McMurray 2001). Of the ten possible trinucleotide repeat sequences, three are especially prone to expansion: CAG, CGG, and GAA. These three sequences have an enhanced ability to form stable secondary structures through nucleotide base-pairing (McMurray 1999). For example, the CAG-CTG repeat forms a hairpin structure through G-C base-pairing with one mismatch base, A-A or T-T (Figure 4.3). In contrast, repeat stretches that rarely undergo expansion in repeat number contain base interruptions to the triplet repeat sequence that inhibit the formation of hairpin structures by increasing the number of mismatched bases (Chung et al. 1993; Kovtun et al. 2001). In yeast, CAG-CTG repeat stretches with two interruptions were repaired better, and were more stable than repeats with only one interruption (Kovtun et al. 2001).

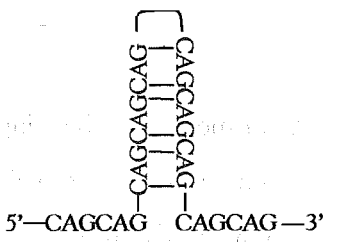


Figure 4.3. Hairpin structure formed by CAG trinucleotide repeat DNA sequences. The hairpin structure is stabilised by additional hydrogen bonding between C-G base pairs, the interruption of a trinucleotide repeat DNA stretch by a different base increases the proportion of mismatched bases and reduces the stability of the structure. Adapted from Kovtun et al. (2001).

The longer the trinucleotide repeat region becomes the more likely it is to undergo further expansion (Leeflang et al. 1995). Analysis of the Huntington's disease (HD) CAG repeat observed that repeat lengths in the normal range had a mutational frequency of 0.6%, while moderately expanded HD alleles (30 and 36) had a mutation rate of 11% and 53% respectively, and disease-causing repeat lengths (38-51 CAG repeats) mutated 92%-99% of the time (Leeflang et al. 1995; Kovtun et al. 2001). It is thought the propensity for higher mutation frequencies with increasing repeat length is due to the greater stability conferred by the higher level of hydrogen bonding; thus while long repeat regions (>25 units) form hairpin structures as quickly as shorter repeats they unpair more slowly, and remain looped for longer to be trapped into DNA or inhibit replication and repair enzymes (Kovtun et al. 2001).

Research in bacteria and yeast suggests that trinucleotide repeat expansion occurs in these organisms via the incorporation of DNA loops after replication slippage or DNA strand break repair (Kovtun et al. 2001; Mirkin and Smirnova 2002). Replication slippage occurs when the polymerase enzyme has difficulty replicating the trinucleotide repeat region, the enzyme briefly dissociates itself from the replication fork and the template strand, and re-attaches at an earlier repeat dragging the nascent strand with it, forming a loop of extra DNA. Resumption of DNA synthesis results in an increase in repeat number equal to the number of repeats the enzyme travelled back (Figure 4.4a) (Kovtun et al. 2001; Mirkin and Smirnova 2002).

Breaks in DNA can be created by the formation of DNA loops at the single- or double-strand breakpoint. Repair then occurs either by gene conversion; non-reciprocal

homologous recombination where DNA is copied but not exchanged, or complementary "fill-in" by repair enzymes (Figure 4.4b) (Kovtun et al. 2001).

Alternatively it is hypothesised that incorporation of RNA primers by the flap-endonuclease enzyme could result in microsatellite expansion. The removal of RNA primers from Okasaki fragments is inhibited when the RNA primer is composed of trinucleotide repeat DNA sequences; instead of being removed the primers are displaced and form loops that are then ligated to the 3' end of the adjacent Okasaki fragment and are replicated into DNA during the following replication round (Figure 4.4c) (Kovtun et al. 2001; Mirkin and Smirnova 2002).

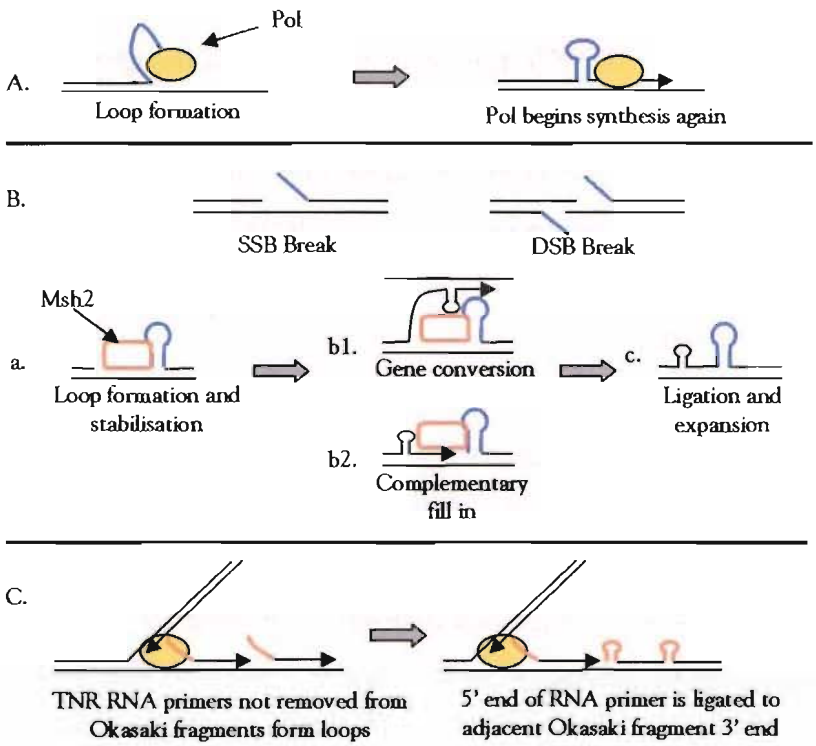


Figure 4.4. Three models of trinucleotide repeat expansion. A. Replication Slippage, the Polymerase enzyme (Pol; yellow) disassociates itself from the template strand when it encounters difficulty replicating the trinucleotide repeat (TNR); the nascent strand forms a loop (blue) that expands the TNR by the number of repeats pol moved back. B. Break repair at single- (SSB) or double-strand breaks (DSB) by either gene conversion, using a non-homologous template, or complementary fill allows expansion of the TNR of the disassociated strand (blue). Other TNR loops (black) that form as synthesis proceeds may be incorporated at this time too. C. Okasaki fragment RNA primers (orange) can form loops when they consist of TNR sequences. These are not removed by the flap-endonuclease and are incorporated into the DNA by ligation. Adapted from Kovtun et al. (2001), Kovtun and McMurray (2001), Mirkin and Smirnova (2002).

Research aiming to establish the mechanism and timing of expansion in mammalian cells has been more difficult. Transgenic mouse models suggest that expansion takes place in both somatic and germ cells, and that both occur in a cell developmental-stage dependent manner (Kovtun et al. 2001). Expansion in germ cells was observed in post-meiotic cells through the incorporation of repeat hairpin structures during DNA break repair (Kovtun and McMurray 2001). The expansion mechanism in both somatic and germ cells appears reliant on the presence of a *mut*S homolog-2 (Msh2) protein suggesting microsatellite expansion in somatic cells occurs via a mechanism similar to germ cells (Kovtun and McMurray 2001).

4.1.7 *AR* CAG Repeat Tract Expansion and Male Infertility

It is predicted that mutations of the androgen receptor gene cause up to 10% of idiopathic male infertility (Yong et al. 2000). Normal spermatogenesis requires high levels of male androgen, yet many men with reduced sperm production have normal hormone levels and do not display AIS symptoms suggesting the cause lies elsewhere in the pathway. The AR mediates the transcription of genes involved in male sexual development and sperm production through its interaction with co-activators which bind to the polyglutamine region of the TAD in a CAG repeat number dependent manner. High-normal CAG repeat numbers weaken the interaction and reduce transactivation while very long CAG repeat alleles cause Kennedy's disease where onset is accompanied by dramatically reduced sperm production. Therefore, it is predicted that moderate expansion might be sufficient to reduce AR transcription activation of male genes such that sperm production is significantly decreased but remains beneath the threshold for pathological protein activity and androgen insensitivity phenotypes (Figure 4.5).

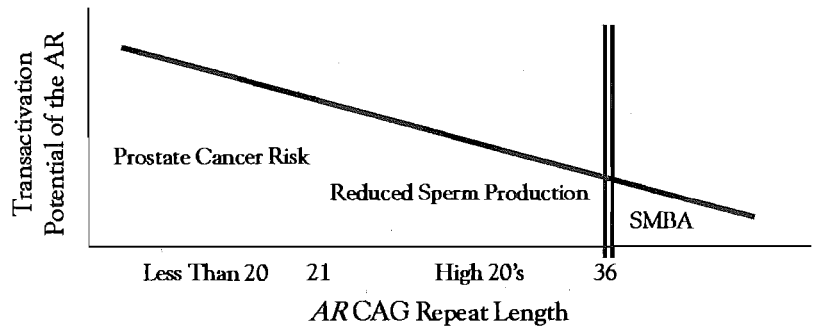


Figure 4.5. Schematic graph illustrating the inverse relationship between *AR* CAG repeat length and the transactivation capability of the AR. Long alleles (>26 CAG units) results in low transactivation and increased risk of decreased sperm production. Very long *AR* CAG repeat lengths above the double line cause SBMA and reduced sperm production. Short repeat lengths increase the transactivation potential of the AR resulting in hyperstimulation of tissues increasing the risk of cancer. Adapted from Yong et al. (2000).

Research in a number of populations in the wake of this hypothesis suggests that indeed moderate increases in repeat size (≥26 CAG repeats) are significantly associated with reduced sperm production. Moreover, the greatest difference in repeat number from normozoospermic controls is observed in men with the most severely impaired spermatogenesis (azoospermia) (Tut et al. 1997; Dowsing et al. 1999; Yoshida et al. 1999; Lim et al. 2000; Mifsud et al. 2001). However, not all studies of male populations have observed a statistically significant association between *AR* polyglutamine number and semen quality (Giwercman et al. 1998; Dadze et al. 2000), and criticism of the findings has centred on the fact that significant risk of decreased sperm production might result from an increase in CAG repeat number of only one to two repeats above the mean. The results of my own research indicated there was no association between AR CAG repeat length and reduced sperm production in the population studied (Erasmuson et al. 2003)(with Appendices). Furthermore, experimental and control populations often display significant overlap in CAG repeat number distribution. This means there are normozoospermic men with high numbers of glutamine repeats for whom diminished AR activity is not causing a decrease in sperm production, and men with reduced sperm production with low numbers of repeats for whom a different mechanism is the cause of their impaired spermatogenesis, suggesting that *AR* mutation causes only rare cases of

reduced sperm production. The question remains whether the correlation is merely statistical or a true physiological effect built up by a small change in AR activity over the course of the male's lifetime?

Support for the hypothesis has come from *in vitro* studies of the transactivation activity of AR's with low-normal (<20), normal (20-21) and high-normal (>26) numbers of glutamine repeats (Puscheck et al. 1994; Kazemi-Esfarjani et al. 1995; Tut et al. 1997; Dowsing et al. 1999; Irvine et al. 2000; Mifsud et al. 2001). A 6% reduction in AR transactivation activity was noted with an increase in repeat number from 21-29 repeats (Irvine et al. 2000). Surprisingly, these expression assays indicated that while repeat numbers in the higher end of the normal range did indeed have significantly lower levels of transcription activation, deletion of the entire polyglutamine repeat did not eliminate transactivational activity (Kazemi-Esfarjani et al. 1995). In fact, AR's with no CAG repeats had higher levels of transcription activation than those with 'normal' numbers of CAG units (20-21) (Kazemi-Esfarjani et al. 1995). On the basis of this research a second hypothesis was proposed which suggests that instead of decreasing transactivation, *AR* trinucleotide expansion instead increases an inhibitory effect of glutamine units on transactivation activity. Hence, the loss of transactivation function in the expanded AR's of Kennedy's patients and infertile men might be the result of an exaggeration of the AR's ability to bind protein, leading to the reduction in gene transcription activation of male target genes (Kazemi-Esfarjani et al. 1995). It is likely that X chromosomes with very low numbers of *AR* CAG repeats would be lost from the gene pool by the reproductive death of individuals from aggressive androgen-driven cancers.

One final controversy is that while Kennedy's disease patients commonly experience some fertility problems during puberty and adolescence most do not develop oligo- or azoospermia until disease manifestation at 40 or 50 years of age, and so, for most of the patient's life their reproductive capacity is normal, but they have massively expanded AR polyglutamine tracts. Moreover, foetal sexual differentiation and development of Kennedy's patients is normal, androgen-insensitivity appears rather suddenly at disease onset (La Spada et al. 1991).

Most likely the reduction in transcriptional activation in expanded *AR* CAG repeat tracts acts to compound effects of mutations in other fertility genes, so that men can have increased numbers of CAG repeats but not experience reduced sperm production if all other fertility genes are intact. It is widely appreciated that men with poor semen quality are likely to harbour more than one mutation of genes involved in sperm production and sexual development. Once a threshold of mutations is surpassed severe defects result in severely impaired spermatogenesis: azoospermia (a form of truncation selection by reproductive death) (Vogt 1995). However, few if any studies have analysed the same population of men for mutations in a variety of recognised fertility genes, to examine if synergistic effects exist. In any case, since the AR is vital to the proper establishment of fertility it is likely to be the causal factor in at least some cases of male infertility. The concern is then the ability of men with poor semen quality to conceive children via intracytoplasmic sperm injection (ICSI) and pass on genetic changes that cause impaired spermatogenesis to their children (Vogt 1995). In the case of expanded AR CAG alleles, ICSI daughters will be 'carriers' passing the unstable allele to sons in whom further expansion might occur, leading to a higher frequency of Kennedy's disease in succeeding generations (Hawkins et al. 1999; Vogt 1999; Cram et al. 2000).

Transmission of extremely expanded alleles with 46 CAG repeats was examined over four generations in one family where six males had Kennedy's disease. All normal alleles segregated stably, in contrast, the expanded allele was markedly unstable, increasing seven times and decreasing once in a total of 17 transmissions. The increase was usually by one CAG unit; however, expansion by five repeats was noted once (Takiyama et al. 1995).

Further evidence for the largely stable nature of CAG allele inheritance was observed in analysis of CAG number in 92 fathers and their 99 ICSI conceived daughters. Only four of the 99 daughters showed expansion or contraction of the original paternal *AR* CAG allele (20→24, 22→23, 26→18, and 26→22). This mutation frequency is thought to be consistent with the low level of mosaicism in CAG number between gametes of a normal healthy individual. Therefore, a child conceived by ICSI is not more likely to inherit an

AR CAG allele with a greater number of repeats than one from a healthy father who falls within the 'normal' range, 11-35(Vogt 1999). Nonetheless, the actual risk of transmitting impaired spermatogenesis is dependent on how significant slight differences in CAG number are to receptor function.

4.1.8 Aim of Research

By analysing *AR* CAG repeat variation I aim to establish whether moderate expansion of the AR polyglutamine region is a significant risk factor of poor semen quality in this study population. There is evidence for reduced transcription activation by the AR *in vitro* and in several other populations but the relationship remains controversial. Trinucleotide repeat expansion is an acknowledged mechanism leading to pathological protein activity, but it is less widely accepted that the small increases in *AR* CAG repeat length postulated to cause reduced transcription of fertility genes will produce a large enough effect to result in impaired spermatogenesis even over the lifetime of an individual. These results will be evaluated alongside Y chromosome micro-deletion and *POLγ* data to assess the contribution of these three regions to semen quality (Chapter Six).

4.2 Method

A total of 272 individuals were analysed for variation in the *AR* CAG repeat region, 251 of these formed the basis for a peer review paper (with Appendices).

4.2.1 PCR Analysis

Specific PCR primers were designed to border the *AR* CAG repeat region and amplify through it. Primer sequences were derived from previously published research (Sartor et al. 1999) and confirmed by PCR simulation assessing the specificity of priming within the androgen receptor gene sequence published on the NCBI web site (version number NM_000044.2) using the computer programme Amplify 1.2 (Engels 1993). Primer sequences and parameters are specified in Table 4.1 below.

Table 4.1. *AR* CAG trinucleotide repeat PCR primer characteristics.

Primer Name	Nucleotide location	Sequence	GC content	Tm (1M Na^+)	Product Size (bp)
AndroRF1	1183-1206	tccagaatctgttccagagcgtgc	54%	72.4°C	288
AndroRR2	1447-1470	gctgtgaaggttgctgttcctcat	50%	69.4°C	

AndroRF1 is the forward primer and RR2 the reverse. Primer sequences are given in the 5'-3' direction. Refer Appendix G for nucleotide sequence.

The PCR reaction mix used standard concentrations of each additive (as per Chapter Two). PCR analysis was conducted in a MJ Research Inc. PTC-100 thermo-cycler. The PCR programme consisted of an initial denaturation at 94°C for four minutes, followed by 30 cycles of 94°C, 58°C, and 72°C for one min each, and a final ten min extension at 72°C.

Initially 5 μl of the PCR reaction was electrophoresed through a 2% (w/v) agarose gel with 1 x orange G, at 8 V/cm for 20 min, then stained and visualised under UV illumination; on later gels this volume was reduced to 3 μl to maximise product yield for DNA sequencing.

4.2.2 DNA Sequencing

Precise quantification of CAG repeat number was established by DNA sequencing of purified PCR products for 272 individuals. Direct DNA sequencing using dideoxy termination dNTPs was conducted using either γ ^{32}P-dCTP 5'-labelled primer sequencing or internally labelled α ^{33}P-dCTP with the Thermosequenase (Amersham Pharmacia Biotech) or Amplicycle (Applied Biosystems) sequencing kit systems respectively. Both isotopes had specific activities of 3000 Ci/mM. Because one sequencing run was sufficient to read the entire CAG repeat region, sequencing reaction volumes were reduced to one quarter of the suggested volume given in the DNA sequencing kits.

4.2.2.1 5' γ ^{32}P-dCTP Labelled Primer Sequencing Reactions

Both primers were tested to ascertain which produced the clearest sequencing reads with the result that the AndroRR2 (reverse) primer was used for the majority of sequencing. The AndroRR2 primer was end-labelled with γ ^{32}P-dCTP (NEN, 3000 Ci/mmol) by mixing 0.625 μl (approximately equal to 20 pMol) of AndroRR2 primer with 1 x T4 polynucleotide kinase (PNK) buffer (0.5 M Tris-HCl, pH 7.6, 100 mM $MgCl_2$, 100 mM 2-mercaptoethanol), 1.125 μl (16 pMoles) of γ ^{32}P-dCTP isotope, and 2.5 units of T4 PNK enzyme (2.75 μl of 0.9 u/μl) the final labelled primer volume was 5 μl (quarter reactions). The primer mixture was heated in a thermal cycler (MJ Research Inc. PTC-100) at 37°C for 30 mins, and then stopped by heating at 95°C for 5 min to inactivate the kinase enzyme. The primer was then stored in the radiation room at -20°C until required.

Cycle sequencing was carried out by combining 200 fmoles of purified PCR product with 1 μl of sequencing buffer (260 mM Tris-HCl (pH 9.5), 65 mM $MgCl_2$) plus 2 pMoles of γ ^{32}P-dCTP labelled AndroRR2 primer (above), and 4 u of Thermo Sequenase DNA polymerase (4 u/μl; 0.0006 U/μl *Thermoplasma acidophilum* inorganic pyrophosphatase; 20 mM Tris-HCl (pH 8.5), 1 mM dithiothreitol (DTT), 100 mM KCl, 0.1 mM EDTA, 0.5% Tween®20, 0.5% Nonidet®P-40, 50% glycerol); these are half reactions with a total reaction volume of 9.25 μl. The reaction was thoroughly mixed before dispensing 2 μl aliquots into each of the individual ddNTP termination tubes (2 μl; each consists of 150 μm of the three other dNTPs, 150 μM deaza-termination nucleotide plus 1.5 μM ddnon-termination nucleotide), mixed again, and a drop of oil added.

Sequencing reactions were cycled in a thermal cycler (Eppendorf Mastercycler® gradient) in the following programme: 96°C for 2 min, 30 cycles of 96°c 30 s, 58°C 30 s, 72°C 45 s, and a final 10 cycles of 95°C 30 s, and 72°C 45 s. Sequencing stop solution (2 μl; 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF) was added as soon as the cycling was complete and reactions stored at 4°C in the radiation room until required.

4.2.2.2 α ^{33}P-dCTP Internally Labelled DNA Sequencing

Internally labelled cycle sequencing reactions were prepared in half or quarter volumes. A bulk reaction mix was made up consisting of n+1 (n=the number of samples) volumes of 0.25 μl of AndroRR2 (8 μM), 0.25 μl of α ^{33}P-dCTP and finally 2 μl of 10 x cycling mix (0.25 u/μl AmpliTaq DNA Polymerase, CS, in 500 mM Tris-HCl (pH 8.9), 100 mM KCl, 25 mM $MgCl^2$, 0.25% (v/v) Tween®20). After thoroughly mixing, 2.5 μl was aliquoted to individual tubes containing 100-200 fmoles of DNA (made up with ddH_20 to 12.5 μl) to make a final DNA sequencing reaction volume of 15 μl. Finally, 3 μl of each DNA reaction mix was aliquoted to one of four tubes each containing 1 μl of a different ddNTP termination solutions (each contains 22.5 μM c7dGTP (addition of 7-deaza-2'-deoxy-guanosine-5'-triphosphate alters the base pairing of guanine residues enhancing migration of GC rich sequences through the gel), 10 μM dATP, dCTP and dTTP, plus one of: 80

μM of ddGTP, if G termination mix tube; 600 μM ddATP if A termination mix; 300 μM ddCTP if C termination mix; or 900 μM ddTTP if T termination mix in 10 mM Tris-HCl, 0.1 mM EDTA (pH7.5)).

The reaction mixes were overlaid with a drop of oil then cycled in a preheated (95°C) thermal cycler for an initial 2 min at 95°C followed by 35 cycles of 95°C, 68°C, 72°C each for one min. Once amplification was complete 2 μl of stop solution was added, and the reactions stored at -20°C until required.

4.2.2.3 Electrophoresis of Sequencing Reactions.

All sequencing reaction products were electrophoresed through a 6% denaturing polyacrylamide gel (6% acrylamide, 1 x TBE, 8.3 M UREA; Appendix F) (40 cm x 34 cm x 0.4 mm; model S2 electrophoresis tank; BRL Life Technologies Ltd). Once polymerised, the gel was pre-run for 30 min to ensure an even heat throughout. Samples were thawed and denatured at 95°C for 4 min just before pre-running was complete, then placed on watery ice to prevent re-annealing. Before loading, well spaces were flushed of urea using the 1 x TBE buffer in the upper buffer tank, and then 3 μl of each termination mix was added in the A, G, C, T nucleotide order. Gels were run for a total of 90 min with 500 ml 1 x TBE (0.04 M Tris-borate, 0.001 M EDTA) electrophoresis buffer at a constant 70 W power (approximately 1500 V and 45 mA), before being transferred to a piece of 3MM (cut-to-size), placed on a gel-drier (model 583; Biorad) and dried for 60 min at 80°C under vacuum.

4.2.2.4 Autoradiography

Gels were checked for radioactivity with a Geiger counter then placed in close contact with an X-ray film (Kodak MR Biomax), secured inside an autoradiography cassette, black plastic, and Perspex. Films were left to expose for increasingly longer periods as the isotope lost activity, for example, fresh sequences (> 1000 cpm) requires approximately

12-24 hr exposure but as activity decayed films required greater exposure times (^{32}P films were left to expose at -80°C).

4.2.2.5 Statistical Analysis of CAG Repeat Length Variation

Multiple comparisons were performed to test whether groups of men with poor semen quality (grouped as non-normozoospermia) had significantly greater mean *AR* CAG repeat lengths compared to normozoospermic men, using both one- and two-tailed two-sample t-tests (Zar 1996). The null hypothesis was:

* H_O:$\mu_{\text{non-normozoospermia}} \leq \mu_{\text{normozoospermia}}$ against
* H_A:$\mu_{\text{non-normozoospermia}} > \mu_{\text{normozoospermia}}$ (one-tailed)

and

* H_O:$\mu_{\text{non-normozoospermia}} = \mu_{\text{normozoospermia}}$ against
* H_A:$\mu_{\text{non-normozoospermia}} \neq \mu_{\text{normozoospermia}}$ (two-tailed), $P<0.05$.

There were two groupings of total non-normozoospermia individuals, one including asthenozoospermic men, and one excluding them. The majority of research investigating associations between long AR CAG repeat lengths and semen quality has excluded asthenozoospermic men on the basis that reductions in AR transactivation lower the expression levels of genes involved in spermatogenesis, whereas these men have normal sperm counts but lack normal levels of sperm motility. I have decided to include them in one grouping since presumably sperm motility requires gene transcription of structural elements involved in the organisation of mitochondria and filaments within the sperm tail, and nuclear proteins involved in energy production, some of which may be up-regulated by the AR.

4.3 Results

AR CAG repeat number varied between 8 and 38 repeat units in the total population (n=272), 9 to 38 in normozoospermic controls (n=98), and from 8 to 31 in men with poor semen quality (n=174) (Table 4.2).

Men with normal semen parameters exhibited the lowest mean CAG repeat number (21.62±0.32 repeats), while men with the most severely impaired spermatogenesis; azoospermia had the highest mean CAG repeat number (22.13± 0.7 repeats). This was not a statistically significant difference (Table 4.2).

Although mean *AR* CAG repeat number increased as the severity of reduced sperm production increased, from normozoospermia to oligozoospermia, severe oligozoospermia, and azoospermia, the difference was not statistically significant indicating that CAG repeat distribution was unrelated to semen quality.

4.3.1 PCR Analysis

A single pair of PCR primers was sufficient to amplify the *AR* CAG repeat region. A sole PCR product was obtained of approximately 300 bp, but varied depending on the number of CAG units encoded within the polyglutamine repeat region (Figure 4.6).

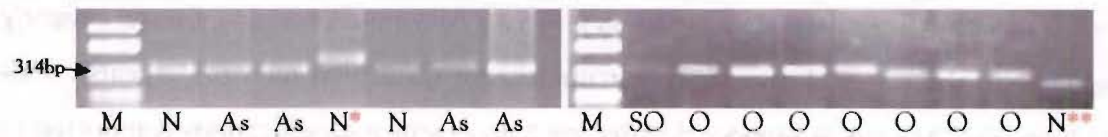


Figure 4.6. Electrophoresis of *AR* CAG PCR products. N signifies NormoZS, As AsthenoZS, SO Severely OligoZS, and O OligoZS. Each PCR product migrates differently through the agarose gel in a size-dependent manner. The asterisked normozs sample has 38 *AR* CAG repeats, and the double-asterisked sample, nine; the remainder have between 19 and 24 CAG repeats.

4.3.2 DNA Sequencing Analysis

All 272 *AR* CAG PCR products were DNA sequenced using either ^{32}P- or ^{33}P-labelled DNA sequencing methods and CAG unit number directly counted from the autoradiography film (Figure 4.7). There was no obvious difference in sequencing clarity between the two methods.

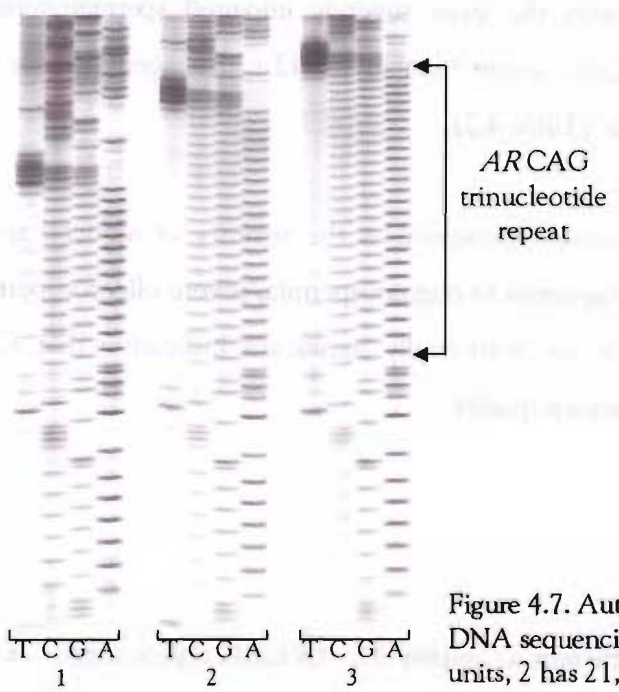


Figure 4.7. Autoradiography image of *AR* CAG DNA sequencing. Individual 1 has 13 CAG repeat units, 2 has 21, and 3 has 26.

Mean CAG repeat number and distribution were remarkably similar among semen classifications; the mean *AR* CAG repeat number for each semen classification differed from the next by less than or equal to half a repeat. All mean repeat lengths were slightly higher than the reported average repeat number for Caucasian men (21 repeats). This similarity in CAG repeat number distribution among semen classifications is reflected in the mean and mode for most groups which centres around 21 repeats. The range of CAG repeat lengths observed in both normozoospermic men and men with poor semen quality (non-normozoospermic) were virtually identical (9-38 and 8-31 respectively) (Figure 4.8). In addition, both groups of men had members with repeat lengths in the very low and

very high range, in fact, one normozoospermic individual has an *AR* CAG allele close to the threshold associated with Kennedy's disease (38 CAG units). The normozoospermic group of men displayed both the highest (38) and lowest (9) *AR* CAG repeat alleles (Figure 4.8).

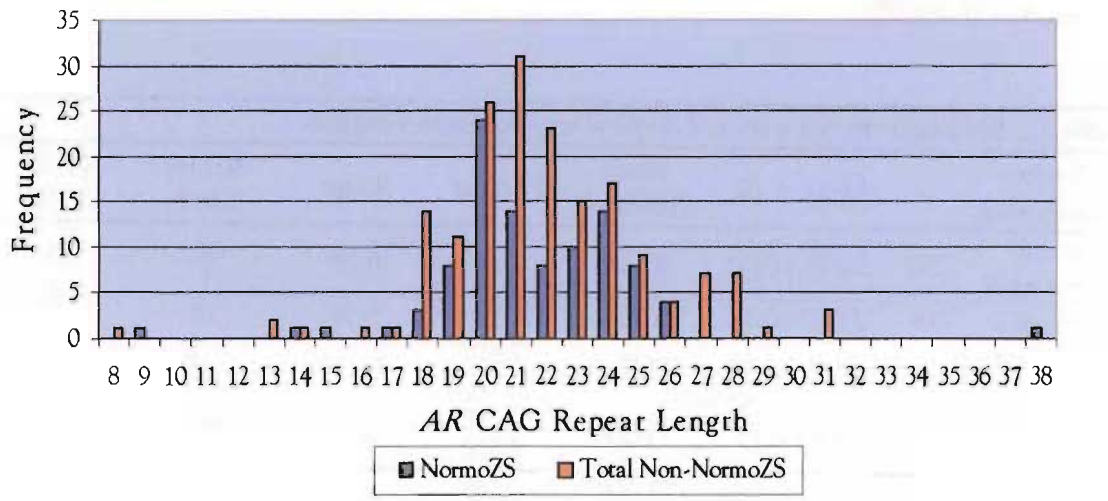


Figure 4.8. *AR* CAG repeat length distribution within the sample population. The graph compares *AR* CAG repeat length distribution between normozoospermic individuals (dark purple bars) and men with poor semen quality (orange bars); CAG repeat number is indicated along the X axis, and the frequency observed in the population up the Y axis.

Men with poor semen quality had a markedly greater proportion of individuals with long *AR* CAG alleles (≥26; 13% compared to 5% for normozoospermic men) (Table 4.2; Figure 4.8). Apart from the azoospermic group, more than 10% of each subgroup of non-normozoospermic men had long *AR* CAG alleles; a marked difference to the proportion of those groups with shorter *AR* CAG alleles. In comparison, normozoospermic men had low frequencies of both long and short repeat alleles (Table 4.2).

4.3.3 Statistical Analysis of Mean *AR* CAG Length.

Mean CAG repeat length of each group of men with poor semen quality, the total group of men with reduced sperm production (excludes asthenozoospermic men), and the non-

normozoospermic group as a whole, was compared to the mean length observed in normozoospermic men using a one-tailed and two-tailed two-sample t-test at an α level of 0.05 (Table 4.2). There was no statistically significant difference in CAG repeat length distribution between semen classifications; men with reduced sperm production were not more likely to have long *AR* CAG repeat lengths than normozoospermic men and no inverse correlation was observed between length of *AR* CAG alleles and sperm production (P>0.05).

Table 4.2. Statistical characteristics of *AR* CAG repeat length variation.

Semen Classification	n	Mean ± SE	Mode (Median)	P Value[a]	Range	%≤16[b] CAG's	%≥26[b] CAG's
Total Population	272	21.79±0.20	20(21)	-	8-38	3%	10%
NormoZS	98	21.62±0.32	20(21)	-	9-38	3%	5%
Total Low Count[c]	119	21.87±0.29	21(21)	0.289 0.577	13-31	2.5%	13%
Total Non-normozs[d]	174	21.9±0.25	21(21)	0.260 0.520	8-31	3%	13%
AZS	15	22.13± 0.7	21, 24[d] (22)	0.258 0.515	18-27	0%	1%
Severely OligoZS	49	21.90±0.51	21(21)	0.323 0.647	13-31	6%	14%
OligoZS	55	21.76±0.41	20(21)	0.394 0.787	18-31	0%	11%
AsZS	55	21.93±0.48	23(22)	0.299 0.597	8-28	4%	13%

(a) The first P value is the one-tailed t-test score, the second is the two-tailed score. (b) These columns reflect the proportion of each semen classification that fall greater than or equal to five repeats either side of the mean (21); the high risk region for reduced spermatogenesis (≥26 CAG repeats (Dowsing et al, 1999)) and greater risk of androgen-driven cancers (≤16 CAG repeats). (c) The combined 'low count' group includes all individuals with reduced sperm count (<20 million sperm/ml). The 'Total Non-normozs' group includes asthenozoospermic individuals. (d) Both 21 and 24 CAG repeats are equally most common among azoospermic men.

The addition of 21 samples did not alter the result of statistical testing comparing *AR* CAG repeat distribution between semen classifications.

4.4 Discussion

Since the suggestion that expanded, yet normal *AR* trinucleotide repeat lengths might be a significant risk factor for reduced sperm production was made by Tut et al. (1997) and Dowsing et al. (1999) many other studies have sought to examine this association in other populations.

There was no significant association between poor semen quality and moderate *AR* CAG repeat expansion in this group of New Zealand Caucasian men. Comparisons of other such studies indicates that there is little variation in mean CAG repeat number (approximately half a repeat) between Caucasian populations of men with reduced sperm count where sample sizes are high (Dadze et al. 2000; Mifsud et al. 2001; Wallerand et al. 2001; Rajpert-De Meyts et al. 2002; Van Golde et al. 2002)(Table 4.3). This might suggest that for those Caucasian populations where the mean CAG repeat number is much greater than 22 repeat units the sample size is not large enough. Because the difference in mean *AR* CAG repeat length between non-normozoospermic and normozoospermic populations is commonly small (usually one to two repeats), large sample sizes are required to generate a normal population distribution in order to evaluate the significance of the association. It is possible that for some of these populations Type II errors are occurring.

Table 4.3. Comparison of AR CAG repeat length variability and association with reduced semen quality in other studies.

Semen Classification	Population Ethnicity	Mean ± SE (n)	Range	P Value[a]	Reference
AZS	Caucasian*	22.1±0.70 (15)	18-27	Not significant	This Study
	Caucasian*	23.0±0.63 (16)	19-26	Not significant	Patrizio et al. 2001
		25.4±0.77 (27)	20-39	p=0.0005	(<1x10^6 spm/ml)
	Caucasian*	23.9±0.50 (37)		Significant	Wallerand et al. 2001
	German	21.4±0.64 (18)	16-26	Not significant	Dadze et al. 2000
		21.9±0.34 (59)	16-29	Not significant	(<1x10^6 spm/ml)
	Caucasian*	24.7±1.0 (10)	20-29	.001<p<.0025	Dowsing et al. 1999
	US (Texan)	22.9±0.73 (23)	17-31	p=0.019	Mifsud et al. 2001
	Chinese	23.8±0.52 (33)	18-33	p=0.043	
	Japanese	23.4±0.52 (30)	19-30	Not significant	Sasagawa et al. 2001

	Japanese	26.5±0.54(41)	20-34	p=0.0013	Yoshida et al. 1999
Severely OliogoZS	Caucasian*	21.9±0.51 (49)	13-31	Not significant	This Study
	Caucasian*	22.2±0.45 (26)	18-26	Not significant	Patrizio et al. 2001
	German	22.3±0.80 (29)	16-34	Not significant	Dadze et al. 2000
	Caucasian*	23.0±1.20 (13)	15-34	.025<p<.05	Dowsing et al. 1999
OligoZS	Caucasian*	21.8±0.41 (55)	18-31	Not significant	This Study
	German	23.0±0.94 (13)	17-28	Not significant	Dadze et al. 2000
	Caucasian*	22.0±1.00 (7)	20-27	Not significant	Dowsing et al. 1999
	US (Texan)	21.7±0.34 (72)	14-29	Not significant	Mifsud et al. 2001
	Chinese	22.9±0.33 (87)	14-32	Not significant	
	Japanese	21.2±0.55 (59)	14-32	Not significant	Komori et al. 1999
Total Subfertile Population[b]	Caucasian*	21.9±0.29 (119)	13-31	Not significant	This Study
	Dane	21.5±0.26 (119)	15-29	Not significant	Rajpert-De Meyts et al. 2002
	Dutch	22.2±0.36 (75)	13-36	Not significant	Van Golde et al. 2002
	Caucasian*	23.5±0.41 (69)	18-39	p=0.03	Patrizio et al. 2001
	German	22.0±0.29 (119)	16-34	Not significant	Dadze et al. 2000
	Swedish	21.9 (33)	16-27	Not significant	Gewercman et al. 1998
	US (Texan)	22.0±0.31 (95)	17-31	p=0.034	Mifsud et al. 2001
	Chinese	23.8 (120)	18-33	p=0.032	
NormoZS	Caucasian*	21.6±0.32 (98)	9-38		This Study
	Dane	21.8±0.32 (110)	14-33		Rajpert-De Meyts et al. 2002
	Dutch	21.7±0.41 (70)	15-31		Van Golde et al. 2002
	Caucasian*	22.0±0.42 (45)	12-30		Patrizio et al. 2001
	Caucasian*	22.2±0.40 (50)			Wallerand et al. 2001
	German	20.8±0.70 (22)	15-26		Dadze et al. 2000
	Caucasian*	20.5±0.30 (32)	17-25		Dowsing et al. 1999
	Swedish	23.2 (294)	8-30		Giwercman et al. 1998
	US (Texan)	20.7±0.52 (55)	8-27		Mifsud et al. 2001
	Chinese	22.4±0.32 (87)	11-29		
	Japanese	23.7±0.44 (51)	17-28		Sasagawa et al. 2001
	Japanese	21.4±0.58 (36)	16-31		Komori et al. 1999
	Japanese	23.9±0.42 (48)	17-30		Yoshida et al. 1999

[a]P Value is significant is $p<0.05$. [b]The total non-normozoospermic population does not include asthenozoospermic individuals. Caucasian* refers to a predominantly Caucasian sample population. (n) indicates the sample size of each semen quality classification in each study.

On the other hand it may be that other factors are operating in populations where no association was found. For example, there is good evidence that *AR* CAG length varies in a race-dependent manner, and while most recent research has attempted to select ethnically homogeneous sample populations it is possible that genealogy has some affect on observed *AR* CAG repeat lengths in individuals who consider themselves Caucasian.

For instance our New Zealand population of men all identify themselves as Caucasian but some may have Polynesian ancestry that may affect *AR* CAG repeat distribution in some semen classification groups or the total population.

In addition, genetic background may influence findings. Many studies seek to rule out other recognised causes of impaired spermatogenesis in the study population, such as Y chromosome micro-deletions, varicocele, hypogonadism and cryptorchidism, but we do not yet fully understand how non-harmful polymorphisms operate to influence our physiology. The same mutations that lead to reduced sperm production in some men might reduce the semen quality of men with a fertility-enhancing polymorphism, but not to the point of oligo- or azoospermia and so confuse associations between gene mutation and infertility (De Kretser and Baker 1999)

The comparison between other studies such as this also indicates that different populations have different CAG repeat length distributions. For example, a German Caucasian population of men with reduced sperm production who have a mean *AR* CAG repeat number of 22 is not statistically different from the normozoospermic group mean of 20.8 repeats (Dadze et al. 2000) but 21.95 repeats is significantly different to 20.72 repeats in a Texan Caucasian population (Table 4.3) (Mifsud et al. 2001).

Furthermore, as was observed in this study, *AR* CAG repeat length ranges are remarkably similar between normozoospermic and non-normozoospermic populations, usually spanning much of the normal polymorphic range (11-31 repeats).

In light of these observations it seems unlikely that a small expansion in the *AR* CAG repeat number within the normal polymorphic range represents a significant risk factor for reduced sperm production in the majority of healthy men. In fact, several studies indicate that the population of men with poor semen quality are not always the group with the greatest proportion of long CAG repeat lengths (≥28), proposed by Tut et al. (1997) as being at especially high risk of reduced sperm production (Giwercman et al. 1998; Komori et al. 1999).

Nevertheless there seems to be sufficient evidence to suggest that moderate expansion and contractions of the AR polyglutamine repeat affect AR functioning such that androgen-related disorders are more commonly found associated with long or short *AR* CAG alleles (Mifsud et al. 2000). For instance, shorter *AR* CAG repeat tracts resulting in higher AR transactivation have been associated with Polycystic Ovary Syndrome (PCOS) characterised by irregular menstruation and anovulatory infertility. PCOS is usually associated with hyperandrogenism due to higher than normal levels of testosterone in women resulting in the inhibition of follicular development and anovuation. However, it seems that increased intrinsic AR activity as a result of short *AR* CAG allele lengths explains how PCOS occurs through hyperandrogenism in women that have low androgen levels (Mifsud et al. 2000). Moreover, ARs with long *AR* CAG repeat regions are associated with undermasculinised male genitalia. Reduced levels of AR intrinsic action due to long *AR* CAG alleles were thought to cause moderate undermasculinisation by compounding the effect of other weak causal factors that on their own do not cause abnormalities (Lim et al. 2001). This lends credibility to the idea that long *AR* CAG alleles might cause a small reduction in AR activity that over the reproductive lifetime of an individual increases the risk of reduced sperm production in men with a particular genetic background.

The rational behind an AR-mediated mechanism resulting in reduced sperm production seems reasonable; it appears to be the causal mechanism in other related disorders (PCOS, abnormal genitalia, AIS) and polyglutamine expansion has a noted *in vitro* effect on the ability of co-activators such as ARA24 and p160 transcription factors to bind and regulate transcription-activation by the AR. Therefore, it seems plausible that at the very least, reduced AR activity as a result of long *AR* CAG repeat regions exacerbates the effect of other weak mutations leading to reduced spermatogenesis in a proportion of men with poor sperm quality. Thus, although it may not represent a significant causal factor on its own, analysis of trinucleotide repeat expansion of the AR polyglutamine region might be of interest if combined with screening of other fertility genes perhaps using DNA micro-array technologies.

References:

Chung, M.Y., L.P. Ranum, L.A. Duvick, A. Servadio, H.Y. Zoghbi, and H.T. Orr. 1993. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* 5: 254-8.

Cram, D.S., B. Song, R.I. McLachlan, and A.O. Trounson. 2000. CAG trinucleotide repeats in the androgen receptor gene of infertile men exhibit stable inheritance in female offspring conceived after ICSI. *Mol Hum Reprod* 6: 861-6.

Dadze, S., C. Wieland, S. Jakubiczka, K. Funke, E. Schroder, B. Royer-Pokora, R. Willers, and P.F. Wieacker. 2000. The size of the CAG repeat in exon 1 of the androgen receptor gene shows no significant relationship to impaired spermatogenesis in an infertile Caucasoid sample of German origin. *Mol Hum Reprod* 6: 207-14.

De Kretser, D.M. and H.W. Baker. 1999. Infertility in men: recent advances and continuing controversies. *J Clin Endocrinol Metab* 84: 3443-50.

Dowsing, A.T., E.L. Yong, M. Clark, R.I. McLachlan, D.M. de Kretser, and A.O. Trounson. 1999. Linkage between male infertility and trinucleotide repeat expansion in the androgen-receptor gene. *Lancet* 354: 640-3.

Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12: 241-53.

Engels, W.R. 1993. Contributing software to the internet: the Amplify program. *Trends Biochem Sci* 18: 448-50.

Erasmuson, T., I.L. Sin, and F.Y. Sin. 2003. Absence of association of androgen receptor trinucleotide expansion and poor semen quality. *Int J Androl* 26: 46-51.

Gerber, H.P., K. Seipel, O. Georgiev, M. Hofferer, M. Hug, S. Rusconi, and W. Schaffner. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* 263: 808-11.

Giovannucci, E., M.J. Stampfer, K. Krithivas, M. Brown, D. Dahl, A. Brufsky, J. Talcott, C.H. Hennekens, and P.W. Kantoff. 1997. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc Natl Acad Sci U S A* 94: 3320-3.

Giwercman, Y.L., C. Xu, S. Arver, A. Pousette, and R. Reneland. 1998. No association between the androgen receptor gene CAG repeat and impaired sperm production in Swedish men. *Clin Genet* 54: 435-6.

Gottlieb, B., L. Pinsky, L.K. Beitel, and M. Trifiro. 1999. Androgen insensitivity. *Am J Med Genet* 89: 210-7.

Hardy, D.O., H.I. Scher, T. Bogenreider, P. Sabbatini, Z.F. Zhang, D.M. Nanus, and J.F. Catterall. 1996. Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. *J Clin Endocrinol Metab* 81: 4400-5.

Hawkins, M.M., C.L. Barratt, A.G. Sutcliffe, and I.D. Cooke. 1999. Male infertility and increased risk of diseases in future generations. *Lancet* 354: 1906-7.

Housley, P.R., E.R. Sanchez, M. Danielsen, G.M. Ringold, and W.B. Pratt. 1990. Evidence that the conserved region in the steroid binding domain of the glucocorticoid receptor is required for both optimal binding of hsp90 and protection from proteolytic cleavage. A two-site model for hsp90 binding to the steroid binding domain. *J Biol Chem* 265: 12778-81.

Hsiao, P.W., D.L. Lin, R. Nakao, and C. Chang. 1999. The linkage of Kennedy's neuron disease to ARA24, the first identified androgen receptor polyglutamine region-associated coactivator. *J Biol Chem* 274: 20229-34.

Irvine, R.A., H. Ma, M.C. Yu, R.K. Ross, M.R. Stallcup, and G.A. Coetzee. 2000. Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Hum Mol Genet* 9: 267-74.

Jenster, G., J.A. van der Korput, J. Trapman, and A.O. Brinkmann. 1992. Functional domains of the human androgen receptor. *J Steroid Biochem Mol Biol* 41: 671-5.

Kazemi-Esfarjani, P., M.A. Trifiro, and L. Pinsky. 1995. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. *Hum Mol Genet* 4: 523-7.

Komori, S., H. Kasumi, R. Kanazawa, K. Sakata, Y. Nakata, H. Kato, and K. Koyama. 1999. CAG repeat length in the androgen receptor gene of infertile Japanese males with oligozoospermia. *Mol Hum Reprod* 5: 14-6.

Kovtun, I.V., G. Goellner, and C.T. McMurray. 2001. Structural features of trinucleotide repeats associated with DNA expansion. *Biochem Cell Biol* 79: 325-36.

Kovtun, I.V. and C.T. McMurray. 2001. Trinucleotide expansion in haploid germ cells by gap repair. *Nat Genet* 27: 407-11.

La Spada, A.R., E.M. Wilson, D.B. Lubahn, A.E. Harding, and K.H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352: 77-9.

Leeflang, E.P., L. Zhang, S. Tavare, R. Hubert, J. Srinidhi, M.E. MacDonald, R.H. Myers, M. de Young, N.S. Wexler, J.F. Gusella, and et al. 1995. Single sperm analysis of the trinucleotide repeats in the Huntington's disease gene: quantification of the mutation frequency spectrum. *Hum Mol Genet* 4: 1519-26.

Lim, H.N., H. Chen, S. McBride, A.M. Dunning, R.M. Nixon, I.A. Hughes, and J.R. Hawkins. 2000. Longer polyglutamine tracts in the androgen receptor are associated with moderate to severe undermasculinized genitalia in XY males. *Hum Mol Genet* 9: 829-34.

Lim, H.N., R.M. Nixon, H. Chen, I.A. Hughes, and J.R. Hawkins. 2001. Evidence that longer androgen receptor polyglutamine repeats are a causal factor for genital abnormalities. *J Clin Endocrinol Metab* 86: 3207-10.

Loy, C.J. and E.L. Yong. 2001. Sex, infertility and the molecular biology of the androgen receptor. *Curr Opin Obstet Gynecol* 13: 315-21.

Lubahn, D.B., D.R. Joseph, M. Sar, J. Tan, H.N. Higgs, R.E. Larson, F.S. French, and E.M. Wilson. 1988. The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. *Mol Endocrinol* 2: 1265-75.

Marcelli, M., W.D. Tilley, C.M. Wilson, J.E. Griffin, J.D. Wilson, and M.J. McPhaul. 1990. Definition of the human androgen receptor gene structure permits the identification of mutations that cause androgen resistance: premature termination of the receptor protein at amino acid residue 588 causes complete androgen resistance. *Mol Endocrinol* 4: 1105-16.

McMurray, C.T. 1999. DNA secondary structure: a common and causative factor for expansion in human disease. *Proc Natl Acad Sci U S A* 96: 1823-5.

McPhaul, M.J. 1999. Molecular defects of the androgen receptor. *J Steroid Biochem Mol Biol* 69: 315-22.

Mifsud, A., S. Ramirez, and E.L. Yong. 2000. Androgen receptor gene CAG trinucleotide repeats in anovulatory infertility and polycystic ovaries. *J Clin Endocrinol Metab* 85: 3484-8.

Mifsud, A., C.K. Sim, H. Boettger-Tong, S. Moreira, D.J. Lamb, L.I. Lipshultz, and E.L. Yong. 2001. Trinucleotide (CAG) repeat polymorphisms in the androgen receptor gene: molecular markers of risk for male infertility. *Fertil Steril* 75: 275-81.

Mirkin, S.M. and E.V. Smirnova. 2002. Positioned to expand. *Nat Genet* 31: 5-6.

Patrizio, P., D.G. Leonard, K.L. Chen, S. Hernandez-Ayup, A.O. Trounson. 2001. Larger trinucleotide repeat size in the androgen receptor gene of infertile men with extremely severe oligozoospermia. *J Androl* 22:444-8.

Puscheck, E.E., M.A. Behzadian, and P.G. McDonough. 1994. The first analysis of exon 1 (the transactivation domain) of the androgen receptor gene in infertile men with oligospermia or azoospermia. *Fertil Steril* 62: 1035-8.

Quigley, C.A., A. De Bellis, K.B. Marschke, M.K. el-Awady, E.M. Wilson, and F.S. French. 1995. Androgen receptor defects: historical, clinical, and molecular perspectives. *Endocr Rev* 16: 271-321.

Rajpert-De Meyts, E., H. Leffers, J.H. Petersen, A.G. Andersen, E. Carlsen, N. Jorgensen, and N.E. Skakkebaek. 2002. CAG repeat length in androgen-receptor gene and reproductive variables in fertile and infertile men. *Lancet* 359: 44-6.

Rebbeck, T.R., P.W. Kantoff, K. Krithivas, S. Neuhausen, M.A. Blackwood, A.K. Godwin, M.B. Daly, S.A. Narod, J.E. Garber, H.T. Lynch, B.L. Weber, and M. Brown. 1999. Modification of BRCA1-associated breast cancer risk by the polymorphic androgen-receptor CAG repeat. *Am J Hum Genet* 64: 1371-7.

Roy, A.K., R.K. Tyagi, C.S. Song, Y. Lavrovsky, S.C. Ahn, T.S. Oh, and B. Chatterjee. 2001. Androgen receptor: structural domains and functional dynamics after ligand-receptor interaction. *Ann N Y Acad Sci* 949: 44-57.

Sartor, O., Q. Zheng, and J.A. Eastham. 1999. Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. *Urology* 53: 378-80.

Simental, J.A., M. Sar, M.V. Lane, F.S. French, and E.M. Wilson. 1991. Transcriptional activation and nuclear targeting signals of the human androgen receptor. *J Biol Chem* 266: 510-8.

Takiyama, Y., S. Igarashi, E.A. Rogaeva, K. Endo, E.I. Rogaev, H. Tanaka, R. Sherrington, K. Sanpei, Y. Liang, M. Saito, and et al. 1995. Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Hum Mol Genet* 4: 1137-46.

Tut, T.G., F.J. Ghadessy, M.A. Trifiro, L. Pinsky, and E.L. Yong. 1997. Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. *J Clin Endocrinol Metab* 82: 3777-82.

Van Golde, R., K. Van Houwelingen, L. Kiemeney, J. Kremer, J. Tuerlings, J. Schalken, and E. Meuleman. 2002. Is increased CAG repeat length in the androgen receptor gene a risk factor for male subfertility? *J Urol* 167: 621-3.

Vogt, P.H. 1995. *Genetic aspects of artificial fertilization.* Oxford University Press, Heidelberg.

Vogt, P.H. 1999. Risk of neurodegenerative diseases in children conceived by intracytoplasmic sperm injection? *Lancet* 354: 611-2.

Wallerand, H., A. Remy-Martin, E. Chabannes, L. Bermont, G.L. Adessi, and H. Bittard. 2001. Relationship between expansion of the CAG repeat in exon 1 of the androgen receptor gene and idiopathic male infertility. *Fertil Steril* 76: 769-74.

Wang, Q., F.J. Ghadessy, A. Trounson, D. de Kretser, R. McLachlan, S.C. Ng, and E.L. Yong. 1998a. Azoospermia associated with a mutation in the ligand-binding domain of an androgen receptor displaying normal ligand binding, but defective trans-activation. *J Clin Endocrinol Metab* 83: 4303-9.

Wang, Q., F.J. Ghadessy, and E.L. Yong. 1998b. Analysis of the transactivation domain of the androgen receptor in patients with male infertility. *Clin Genet* 54: 185-92.

Wong, C.I., Z.X. Zhou, M. Sar, and E.M. Wilson. 1993. Steroid requirement for androgen receptor dimerization and DNA binding. Modulation by intramolecular interactions between the NH2-terminal and steroid-binding domains. *J Biol Chem* 268: 19004-12.

Yong, E.L., L.S. Lim, Q. Wang, A. Mifsud, J. Lim, Y.C. Ong, and K.S. Sim. 2000. Androgen receptor polymorphisms and mutations in male infertility. *J Endocrinol Invest* 23: 573-7.

Yoshida, K.I., M. Yano, K. Chiba, M. Honda, and S. Kitahara. 1999. CAG repeat length in the androgen receptor gene is enhanced in patients with idiopathic azoospermia. *Urology* 54: 1078-81.

Zar, J. 1996. *Biostatistical Analysis Third Edition.* Prentice-Hall International, Inc, Upper Saddle River, NJ.

CHAPTER FIVE MITOCHONDRIAL POLYMERASE GAMMA

5.1 Introduction

A reduction in cellular energy production as the result of an accumulation of mitochondrial DNA (mtDNA) mutation has been implicated in many diseases, it is also thought to be responsible for some cases of reduced semen quality; asthenozoospermia, in otherwise healthy males (Kao et al. 1995; Lestienne et al. 1997; Kao et al. 1998; Holyoake et al. 2001). As is the case with X-linked spermatogenesis genes, females can influence the fertility status of their male offspring yet remain unaffected, through transmission of her mitochondrial population in the cytoplasm of the oocyte. A heterozygous population of wild-type and abnormally functioning mitochondria that produce a reduced level of cellular energy may not noticeably affect somatic cell or oocyte function, but might however impair the motility of a male offspring's sperm cells, which have significantly fewer mitochondria per cell. These mtDNA mutations can be maintained in the gene pool because they are transmitted through the female lineage and since they are not harmful to females are not subject to negative selection (Chinnery et al. 2000).

5.1.1 The Mitochondrion.

Mitochondria are bounded by a double membrane. The outer membrane protects the organelle, and contains specialized transport proteins, such as porin to facilitate the passage of molecules into the intermembrane space. The inner membrane is folded into cristae that fill the inner matrix. Arranged in clusters on the cristae are the energy generating complexes of the electron transport chain. Cells that have a high energy demand, such as the flight muscle cells of birds, have greater numbers of mitochondrial cells, each with more cristae and electron transport chain clusters, in comparison, skin cells have few mitochondria (Ross 1999; Spurger 2003).

Most mammalian cells contain in the region of 10^3 to 10^4 copies of the same mtDNA genome (homoplasmy); oocytes contain an average of 200,000 mtDNA copies while

spermatocytes contain between 22 and 100 (Lightowlers et al. 1997; May-Panloup et al. 2003).

5.1.2 The Mitochondrial Genome

In most eukaryotes, mitochondrial DNA (mtDNA) is a small, 16-18kb circular duplex (Figure 5.1) arranged into structures called 'nucleoids', which usually contain four to five mtDNA copies, plus several associated proteins including the mitochondrial single-stranded DNA-binding protein (mtSSB), mitochondrial transcription factor A (Tfam), Twinkle (DNA helicase) and the mtDNA damage protection protein, Mgm101p (Meeusen et al. 1999; Garrido et al. 2003; Korhonen et al. 2003). Replication occurs semiconservatively beginning in the displacement (D) loop with the heavy-strand; light-strand synthesis begins once replication has proceeded approximately two-thirds of the way around (Anderson et al. 1981; Wallace 1999). Mammalian mtDNA encodes 13 polypeptides that are essential subunits of electron transport chain complexes (also known as OXPHOS proteins because of their role in oxidative phosphorylation) (Table 5.1). Dispersed amongst these genes are 22 tRNA genes and two rRNA genes (16S and 12S).

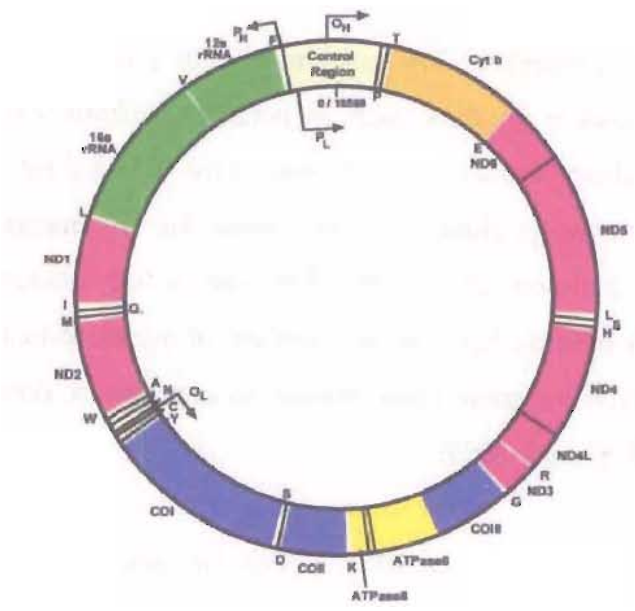


Figure 5.1. The mitochondrial genome. The 16,569 bp human mitochondrial genome encodes seven of the 43 Complex I subunits (*NADH dehydrogenase ND1, 2, 3, 4, 4L, 5* and *6*) shown in pink; one of the 11 subunits of Complex III (*Cytochome b, Cyt b*) in orange; three of the 13 Complex IV subunits (*COI, II* and *III*) in purple; and two of the 16 subunits of Complex V (*ATPase 6* and *8*) in yellow. The two rRNA genes are in green, and the tRNA genes are labelled with their amino-acid letter. The control region contains the heavy-strand origin of replication (O_H), and the heavy- and light-strand promoters (P_H and P_L). The O_L lies within several tRNA genes between *ND2* and *COI*. Adapted from Wallace (1999).

Even though the majority of the mitochondrial genome is devoted to producing OXPHOS proteins, a far greater number are produced by the nuclear genome and transported in (DiMauro and Schon 2003). This includes proteins required for mitochondrial protein synthesis, such as DNA and RNA polymerases, translation initiation and elongation factors, ribosomal proteins, aminoacyl tRNA synthetases, regulatory proteins such as Tfam, and the remaining 77 OXPHOS proteins (Table 5.1) (Anderson et al. 1981; Klug 1997; Chinnery et al. 2000). In addition, 60 additional nuclear-encoded proteins are required for electron transport chain functioning and assembly (DiMauro and Schon 2003).

Table 5.1. Mitochondrial and nuclear protein contribution to the electron transport chain.

No. of subunits encoded by:	Complex I (NADH: ubiquinone oxidoreductase)	Complex II (Succinate: ubiquinone oxidoreductase)	Complex III (Ubiquinol: cytochrome c oxidoreductase)	Complex IV (Cytochorome C oxidase)	Complex V (ATP synthase)	Total
mtDNA	7	0	1	3	2	13
nuclear	~39	4	10	10	~14	~77

Information from Wallace (1999) and DiMauro and Schon (2003).

There are several distinct differences between nuclear and mtDNA genetics. Unlike nuclear DNA, the majority of mtDNA is protein-coding; only the D-loop (approximately 1kb) is not, it is free of protective histone proteins, and is attached to the inner membrane (Chinnery et al. 2000).

5.1.3 Mitochondria are Maternally Inherited

Despite the fact that between 75 and 100 normally functioning paternal mitochondria enter the oocyte at fertilisation only maternal mtDNA is easily detectable in the developing embryo (Sutovsky et al. 2000). There is some evidence to suggest that paternal mtDNA is actively excluded from populating the developing embryo.

Sperm lose 90% of their mtDNA during spermatid elongation resulting in an average of 1.4 mtDNA copies per mitochondrion (Sutovsky et al. 2000; May-Panloup et al. 2003). This reduction is due to the down regulation of the mitochondrial transcription factor (Tfam) gene, that controls mtDNA copy number in mitochondria (Rantanen and Larsson 2000; Sutovsky et al. 2000). Thus, the number of paternal mtDNA received by the oocyte is a fraction of the estimated 100,000 maternal copies present (Ankel-Simons and Cummins 1996). Furthermore, this tiny paternal pool of mitochondria is then subjected to a mighty selection pressure as the number of mitochondria per cell is dramatically reduced before implantation in the uterus (Chinnery et al. 2000). Hence the chances of paternal mtDNA copies persisting are slim. However, should a paternal mitochondrion survive the bottle-neck it could potentially make up a significant proportion of the developing embryo's mitochondrial population.

Alternatively, or in concert with the above, paternal mtDNA may be targeted for active destruction before mtDNA replication resumes in the implanted embryo. Research suggests that a component of the sperm mitochondrial membrane is ubiquinated early in spermatogenesis (secondary spermatocyte or round spermatocyte stage) targeting the organelle for destruction by the oocytes 26S proteolytic machinery (Sutovsky et al. 2000; Thompson et al. 2003). This tag must be masked during passage through the epididymus since this is where abnormal sperm cells and histone proteins are ubiquinated and destroyed, and then uncovered in the oocyte (Sutovsky 2003). As yet is it unclear how paternal mitochondria are destroyed since they are significantly larger than the 26S proteasome. It is possible that only a few membrane proteins are degraded exposing signals that trigger a separate mechanism, such as autophagy of the entire paternal mitochondrion (Wojcik 2003).

It is hypothesised that the elimination of paternal mtDNA from embryos has developed as an evolutionary adaptation to prevent heteroplasmy and the inheritance of a potentially DNA-damaged mitochondrial genome (Cummins 2000; Sutovsky et al. 2000). Like the Y chromosome, mtDNA is subject to Müller's ratchet, hence the bottleneck in the pre-implantation embryo either removes highly mutated mtDNA genomes, or generates

individuals with them at high frequency, in which case the individual will probably be eliminated from the population by natural selection (Chinnery et al. 2000). This explains why offspring can be born with a higher proportion of mutated mtDNA than their mother, and why mtDNA mutations can go to fixation (via random genetic drift) within only a few generations (Chinnery et al. 2000).

5.1.4 The Mitochondrial Electron Transport Chain

Respiration can be divided into three main pathways: glycolysis, the mitochondrial tricarboxylic acid cycle (TCA), and the mitochondrial electron transport chain (Fernie et al. 2004). Glycolysis occurs in the cytosol and involves the oxidation of glucose to pyruvate. However, only a fraction (approximately $1/15^{th}$) of the potential energy from glucose is generated by glycolysis (Garett and Grisham 1995). Pyruvate is then oxidized by the mitochondrial TCA cycle producing the reduced coenzymes NADH and Succinate. Electrons are passed from the reduced coenzymes through four innermembrane protein complexes of the mitochondrial electron transport chain to oxygen, with the concomitant translocation of protons into the innermembrane space using the energy generated by oxidation of NADH and Succinate (Garett and Grisham 1995). This process creates a proton gradient across the innermembrane which serves as the energy reservoir to achieve ATP synthesis (chemiosmotic theory) (Garett and Grisham 1995; Fernie et al. 2004). The energy created by the difference in proton concentration is utilized by complex V, ATP synthase which catalyses the phosphorylation of ADP as protons flow through it down the concentration gradient and back into the mitochondrial matrix (Garett and Grisham 1995). Thus ATP synthase couples the oxidation of substrates to the phosphorylation of ADP. A portion of the energy generated by the proton gradient must be used to transport ATP out of the mitochondria, and ADP and phosphate in. The rate of oxidative phosphorylation depends on the availability of important substrates; the reducing agent, NADH; the terminal oxidising agent, O_2, and ADP. Approximately 15 ATP molecules are generated per pyruvate (Fernie et al. 2004).

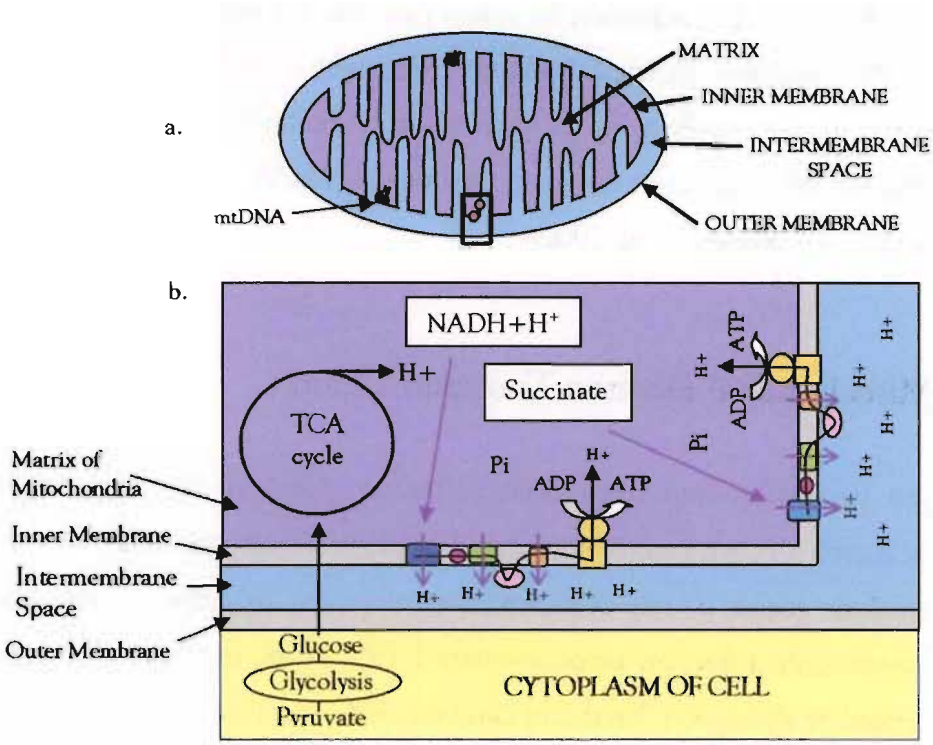


Figure 5.2. Schematic illustration of the electron transport chain. (a) A mitochondrion with double membrane, and two mtDNA molecules close to the inner membrane. Two electron transport chain clusters are indicated in red. (b) Enlargement of boxed section; Complex I (dark blue), II (light blue), CoQ (pink), III (green), Cyt C (light pink in intermembrane space), IV (orange), and ATP synthase (yellow box and circle). Electron movement through the chain is: NADH+H$^+$ to Complex I, CoQ, III, Cyt C, IV, V; and Succinate to Complex II, CoQ, III, Cyt C, IV, V. A very simplified version of Glycolysis in the cytoplasm of the eukaryotic cell and the TCA cycle are indicated on the left to show where the respiratory chain electrons originate. Adapted from DiMauro and Schon (2003).

5.1.5 Mitochondrial Disease

The mutation rate of mtDNA is estimated to be 10 times higher than that of nuclear DNA (Copeland et al. 2003), because it is largely unprotected (free of histones) and is located close to the site of reactive oxygen species (ROS) produced as by-products of oxidative phosphorylation, making it vulnerable to DNA damage (Croteau and Bohr 1997; Stevnsner et al. 2002). It was traditionally thought that mitochondria lacked any sort of DNA repair mechanism, but most have since been observed to some degree in various animal models including those for HCl-, H_2O_2-, and UV-induced damage, DNA glycosylases for base modification repair of 8-oxoG, thymidine glycol and uracil-DNA, a

small number of base excision repair (BER) enzymes, a *MutT* homolog, and some evidence of an ATP-dependent mismatch repair mechanism (Tomkinson et al. 1990; Croteau and Bohr 1997; Stierum et al. 1999; Stevnsner et al. 2002; Mason et al. 2003).

The electron transport chain is under the control of both the mitochondrial and nuclear genomes, which impacts on the way mitochondrial diseases are inherited, and the phenotypes produced. While mutations of the nuclear derived components of energy production result in mitochondrial diseases inherited in a Mendelian fashion, mitochondrial genetics is vastly different (DiMauro and Schon 2003). With few exceptions, mtDNA is inherited uniparentally through the maternal lineage, this means a mother passes her mtDNA complement to all her children, but only her daughters pass it on to theirs (Wallace 1999). Once an mtDNA mutation arises the mitochondrion is said to have a heteroplasmic mtDNA population (two types), as does the eukaryotic cell. As cells divide, mitochondria are randomly assorted to daughter cells, leading to potentially disproportionate quantities of mutant mitochondria in daughter cells. A cell and tissue can tolerate a proportion of their mitochondria functioning abnormally, but once cellular energy production drops below a minimum threshold level, symptoms of mitochondrial disease appear and usually become progressively worse. Different tissues have different tolerances, those highly dependent on high energy production such as the brain, heart, muscle, retina, and endocrine glands have lower threshold levels (Wallace 1999; DiMauro and Schon 2003).

Mitochondrial disease can result from mutation of any of the proteins associated with energy production. The fact that mitochondria are in almost every cell in the body means that mtDNA mutations often result in multiple organ or system pathologies (central and peripheral nervous, and endocrine systems), although single organ symptoms do occur (Taivassalo et al. 1999; DiMauro and Schon 2003). The most common symptoms of mitochondrial disease are muscle weakness especially of those that control eye movement, degenerative disease of the central nervous system, and metabolic disorders commonly involving abnormally high levels of lactic acid (lactic acidosis) (Spurger 2003).

Some mtDNA mutations cause recognisable phenotypes, for instance, all Leber's hereditary optic neuropathy patients have mutations of the *ND* genes, yet others produce highly variable symptoms. This is compounded by the random nature of mitochondrial segregation to daughter cells; a mother can transmit variable amounts of mutant and wild-type mitochondria to her offspring resulting in all, some, or none developing mitochondrial disease, and leading to vast differences in mtDNA mutation phenotypes among members of the same family, with the same mtDNA mutation (Wallace 1999).

There is no effective treatment for mitochondrial disorders but therapies are being developed that aim to preferentially allow wild-type mitochondria to replicate, diluting out mutated mitochondria (Taivassalo et al. 1999; Mason et al. 2003), or targeting proteins to selectively destroy mutant mtDNA and deliver wild-type copies of mutated proteins (DiMauro and Schon 2003).

5.1.6 Mitochondrial Energy Production and Male Fertility

Substantial amounts of energy is used by sperm to swim up the female reproductive tract during fertilization (Kao et al. 1998). The relationship between a reduction in mitochondrial energy production and reduced sperm motility was first established by Ruiz-Pesini et al. (1998). While there had been some previous evidence to suggest a link between sperm mitochondria characteristics (number, location, volume) and quality of motility (Cardullo and Baltz 1991; Mundy et al. 1995), this study represented the first biochemical evidence linking mitochondrial energy generation (as electron transport chain complex activity) and sperm motility. Control samples exhibited higher levels of electron transport chain complex activity than that of asthenozoospermic samples, and in addition, a direct correlation was observed within the population as a whole between sperm motility and complex activity (Ruiz-Pesini et al. 1998).

Research screening the mitochondrial genome indicates that asthenozoospermic individuals have a higher incidence of mtDNA deletion which might account for a reduction in activity of mitochondrially-encoded electron transport complexes (Kao et al.

1995). Men with poor semen quality show a higher frequency of mtDNA nucleotide substitutions and deletions, produce greater levels of ROS, and have higher levels of ubiquination than normozoospermic men (Kao et al. 1995; Lestienne et al. 1997; Aitken et al. 1998; Holyoake et al. 2001; Sutovsky et al. 2004a). In addition, there is evidence to suggest that mtDNA copy number is increased per mitochondrion in men with poor sperm quality; men with two suboptimal semen parameters (sperm count, motility or morphology) had six times the number of mtDNA copies in their mitochondria than men with normal semen parameters (May-Panloup et al. 2003). It is hypothesised that this mechanism might act to compensate for the reduction in energy production by the abnormally functioning mitochondria where cells require more poorly performing mitochondria to produce a similar amount of energy as a normal mitochondrial cell (May-Panloup et al. 2003). However, the presence of large mtDNA deletions does not necessarily result in reduced semen quality (Cummins et al. 1998). Considerable levels of the 4977 'common' mitochondrial deletion have been observed in men with normal semen parameters (Cummins et al. 1998). It is possible that these deletions are occurring after spermiogenesis through oxidative damage by ROS (Cummins et al. 1998).

Paradoxically, while there is abundant evidence associating high levels of ROS production and defective sperm function (low motility and fertilising potential) (Aitken et al. 1989a; Aitken et al. 1989b; Aziz et al. 2004) low levels of oxidative stress actually enhance the ability of sperm to fuse with the oocyte (Aitken et al. 1998; Baker and Aitken 2004).

There are concerns that intracytoplasmic sperm injection (ICSI) facilitates the fertilisation of oocytes with sperm cells subjected to high levels of oxidative stress (Twigg et al. 1998). Direct injection of sperm cells into the oocyte bypasses the selection pressures acting in natural fertilisation because spermatocytes are not required to be motile or capable of fusing with the oocyte membrane. Although it is unlikely that paternal mtDNA will escape ubiquitin-mediated destruction, abnormally functioning nuclear respiratory proteins could be passed on (Lestienne et al. 1997; St John et al. 1997; Aitken et al. 1998; Twigg et al. 1998; Sutovsky et al. 2000; Sutovsky et al. 2004b).

5.1.7 The Mitochondrial DNA Polymerase Gamma

Human mtDNA is replicated by a single nuclear-derived DNA polymerase: POLγ, the catalytic subunit of which is encoded by the *POLγ* gene (15q24) (Ropp and Copeland 1996). Like other γ polymerases, POLγ performs high-fidelity DNA synthesis due to efficient 3'-5' exonucleolytic proof-reading, and association with a small beta subunit, POLG2 (17q21), results in enhanced polymerase activity (Longley et al. 1998a; Longley et al. 1998b; Longley et al. 2001). Several amino-acid residues within both the polymerase and exonuclease domains of POLγ are highly conserved among many polymerases including Pol I from *Escherichia coli,* Mip1p in yeast, and phage T7 DNA polymerase (Ropp and Copeland 1996). Substitution of two such residues, aspartate[198], and glutamate[200] in the exonuclease motif I abolishes the exonuclease proof-reading ability of POLγ while largely maintaining polymerase function (Figure 5.3) (Ropp and Copeland 1996). In cell culture, the substitution of aspartate[198] (D198A) alone was sufficient to result in an accumulation of mtDNA point mutations and a mutational load of 1:1700 over a three month culture period. The analogous substitution in the yeast Mip1p produces the same error-prone polymerase resulting in a 10^4-fold decrease in proof-reading ability and a 100-fold increase in the accumulation of mtDNA point mutations (Spelbrink et al. 2000). Furthermore, exonuclease-deficient *Mip1* mutations are partially dominant so that even in the presence of wild type Mip1p, mtDNA mutations accumulate (Spelbrink et al. 2000).

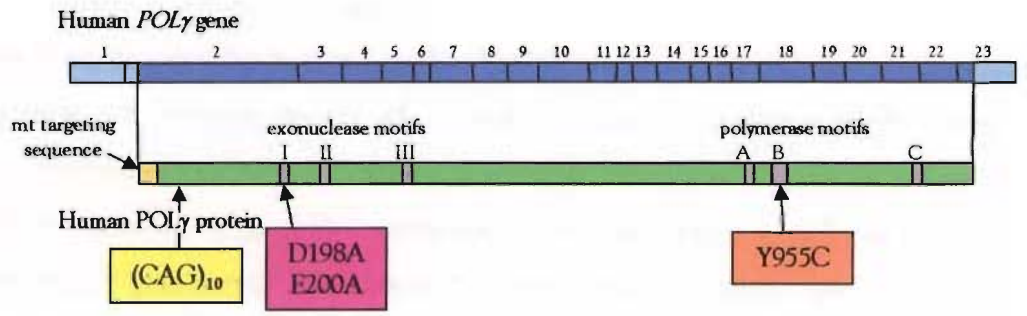


Figure 5.3. Schematic of the *POLγ* gene and protein with functional domains. The gene consists of 23 exons (not all of which are translated). The yellow box indicates the polyglutamine repeat region associated with male infertility, the pink and orange boxes indicate amino-acid substitutions that result in error-prone mtDNA synthesis by POLγ. Adapted from Copeland et al. (2003) with information from Ropp and Copeland (1996).

Similarly, three amino-acid residues in the polymerase motif B are highly conserved among DNA polymerases; these residues, arginine[943], tyrosine[955], and alanine[957], form the recognition site for the incoming nucleoside triphosphate (dNTP) (Figure 5.3) (Copeland et al. 2003). Substitutions of these residues are associated with progressive external ophthalmoplegia (PEO), a mitochondrial disease inherited in a Mendelian fashion (Lamantea et al. 2002). Heterozygous substitution of tyrosine[955] (Y955C) is commonly associated with a dominant form of PEO; Y955C retains a wild-type polymerase rate but shows a marked decrease in dNTP binding-affinity leading to error-prone mtDNA synthesis. This suggests one explanation for the accumulation of mtDNA point mutations and deletions observed in PEO patients (Ponamarev et al. 2002). No mutations of *POLG2* have been identified associated with PEO (Lamantea et al. 2002).

5.1.8 POLγ and Male Infertility

Unique to the human *POLγ* gene is a CAG trinucleotide repeat that is maintained at 10 repeat units in 88% of the population (Rovio et al. 1999). Recently it was observed that small variations in repeat number, above or below the common 10-unit repeat number is associated with reduced sperm production (oligozoospermia), and normozoospermic infertility (Rovio et al. 2001; Jensen et al. 2004). *In vivo* analysis of POLγ catalytic activity in the absence of the 10-unit CAG repeat indicated that the polymerase is unaffected by elimination of the polyglutamine region, in fact, cell lines displayed a 7 to 15-fold increase in polymerase activity (Spelbrink et al. 2000). While this might suggest that the polyglutamine repeat region does not impact on the enzymatic properties of POLγ, a similar increase in activity was noted in the androgen receptor protein upon total deletion of a polyglutamine repeat region (Kazemi-Esfarjani et al. 1995). Complete deletion of either repeat region has not been observed in any individual to date, rather, association with poor semen quality is due to small increases in repeat number. However, no studies have analysed POLγ activity with varying CAG repeat length alleles. Thus, it is unclear why a slight increase or decrease in repeat number on one or more alleles is significantly associated with a reduction in sperm production. Polyglutamine regions are known to be interfaces for protein interactions and so it is possible that deletion alters

interactions between *POLγ* and sperm-specific proteins resulting in suboptimal polymerase activity and the accumulation of mtDNA mutation (Spelbrink et al. 2000).

The accumulation of both large deletions and single nucleotide polymorphisms of mtDNA is associated with reduced sperm production and poor motility (Kao et al. 1995; Kao et al. 1998; Holyoake et al. 2001). These changes might be the result of oxidative DNA damage by reactive oxygen species produced in the mitochondrial innermembrane as by-products of cellular energy production (Aitken et al. 1998; Graziewicz et al. 2002). Alternatively, mutations of *POLγ* that reduce the exonucleolytic proof-reading ability could produce the high frequency of mtDNA point mutations observed in these individuals. However, the maternal inheritance of mitochondria and biparental inheritance of *POLγ* copies complicates matters since several genotype/phenotype situations are possible. A mother might have one mutated error-prone POLγ enzyme that increases the frequency of mtDNA mutation in her mitochondrial population. It is possible therefore, for her son to inherit two normal *POLγ* copies and her heterozygous mitochondrial population, or he could also inherit her error-prone polymerase which, even though a wild-type copy exists, will keep producing mtDNA replication errors. To date no studies have been conducted evaluating the prevalence of *POLγ* exonuclease domain mutation within a human study population.

5.1.9 Aim of Research

I aim to establish the level of *POLγ* mutation in my study population by firstly establishing the variation in CAG repeat allele distribution in relation to semen quality, and then screening for mutation within the three exonuclease motifs. The frequency of single nucleotide polymorphisms in several of the mtDNA OXPHOS genes is known for 109 individuals of my study population through research conducted by previous students (Holyoake et al. 2001), thus if exonuclease domain mutations are found they can be further analysed to establish whether *POLγ* exonuclease domain mutation is correlated to nucleotide substitution levels in mtDNA. This data will be added to X and Y

chromosome mutation analysis to evaluate the relationship between mutational load and semen quality.

5.2 Method

5.2.1 PCR Conditions for *POLγ* CAG Repeat Amplification

The PCR primer pair used to amplify the *POLγ* CAG repeat region consisted of the reverse primer previously published by Rovio et al. (1999) and a forward primer designed to bind more 5' in the *POLγ* sequence (version AF497906.1; NCBI; refer Chapter Two Section 2.3.2.1) to produce a larger PCR product for manual sequencing (Table 5.2) (Appendix G).

Table 5.2. Primer pair used to amplify the *POLγ* CAG repeat region.

Primer Name	Nucleotide location	Sequence	GC content	Tm (1M Na^+)	Product Size (bp)
POLGF*	2364-2383	ccaaagccaggtgttctgac	55%	70°C	295
MIP32	2632-2658	ctctcgagagcatctggatgtccaatg	51%	78°C	

POLGF* is the forward primer and MIP32 the reverse. Primer sequences are given in the 5'-3' direction. Refer Appendix G for nucleotide sequence.

Only the final primer concentration (0.32 μM) was different from concentrations given in Chapter Two. The PCR cycling programme consisted of an initial denaturation at 96°C for 2 min, followed by 30 cycles of 95°C for 1 min, 62°C for 45 s, and 72°C for 2 min, with a final extension step of 72°C for 5 min.

5.2.2 DNA Sequencing of the *POLγ* CAG Repeat Region

The purified PCR products were α^{33}P-dCTP internally labelled and DNA sequenced using a Thermosequenase sequencing kit (Amersham Pharmacia Biotech) as per Chapter Four Section 4.2.2.2. Sequencing was carried out using the reverse primer, Mip32, and the 295 bp product was obtained in a single run. CAG repeat number was determined by

visualisation of only the ddATP and ddTTP lanes for the final 104 samples, to increase sample throughput.

5.2.3 Genotype Categorisation.

Each sample was assigned a genotype based on the number of *POLγ* CAG repeats in each allele as outlined by Rovio et al. (2001). A WT (Wild-type) allele is defined as one carrying the 10-unit CAG repeat; hence a homozygous WT individual carries two WT alleles. Heterozygotes carry one WT and one 'other' allele. 'Other' alleles have either greater than or fewer than ten CAG repeat units. Homozygous mutants carry two 'other' alleles, either two copies of a single 'other' allele length, or two different alleles each with fewer or greater than 10 CAG units.

5.2.4 Statistical Analysis of *POLγ* CAG Repeat Variation

A G-test of independence was conducted to test the null hypothesis that semen quality is independent of genotype (CAG repeat variation) against the alternate hypothesis that they are related (Sokal and Rohlf 1995; Zar 1996).

5.2.5 PCR Amplification of the Three Exonuclease Motifs

Three PCR primer pairs were designed to amplify each of the three motifs that make up the *POLγ* exonuclease domain, using the nucleotide sequence AF497906 (NCBI) (Appendix G) (Table 5.3).

Sample size was reduced for the analysis of exonuclease motifs II and III in order to conduct an initial screening and evaluation of mutation frequency in these motifs.

Table 5.3. Primer pairs used to amplify exonuclease domains I, II, and III.

Exo Motif	Primer Name	Nucleotide Location	Sequence	GC content	Tm (1M Na^+)	Product size (bp)
Motif I	ExoF2	2811-2830	gacaacctggaccagcactt	55%	70°C	250bp
	T3exoR3	3043-3060	taatacgactcactataggg-cattggcggtggccatat	47%	83°C	
Motif II	Exo2F	7254-7273	gtggaagagcgttactcttg	50%	68°C	182bp
	Exo2R	7417-7435	tgatccaggtagggttcct	53%	63°C	
Motif III	Exo3F	8773-8791	agaactgtttgtgaagggc	47%	65°C	278bp
	Exo3R	9031-9050	ctaccgctcttcttggagag	55%	62°C	

F denotes the forward primer, R, the reverse. Nucleotide position is based on the AF497906 sequence. The DNA sequence before the hyphen in T3exoR3 denotes the T3 tag. Refer Appendix G for nucleotide sequences.

The PCR reaction mixes used standard concentrations of buffer, dNTP's, *Taq* polymerase and DNA (as per Chapter Two); the final primer concentration was 0.6 μM. The cycling program for exonuclease motif I consisted of a 2 min at 95°C, 30 cycles of 95°c and 58°C for 1 min each, and 72°C for 2 min, followed by a final 7 min 72°C extension. Exonuclease motifs II and III were amplified with shorter cycling times; 95°C for 2 min, 30 cycles of 95°C and 58°C for 40 s, and 72°C for 1 min, and a final 72°C for 5 min.

5.2.6 DNA Sequencing of the Three *POLγ* Exonuclease Motifs.

5.2.6.1 Radioactive Sequencing of Exonuclease Motif I.

Sequencing of 16 exonuclease motif I samples was conducted using γ ^{33}P-dATP end-labelled sequencing with a Thermosequenase Cycle Sequencing Kit (as per Chapter Four substituting [^{33}P]-dATP). Sequencing was conducted from both the forward and reverse primers, ExoF2 and T3exoR3 in a single electrophoresis run.

5.2.6.2 Fluorescent Sequencing of Exonuclease Motif I.

The majority of exonuclease I samples were DNA sequenced by fluorescent labelling of the T3-tagged PCR product (generated by T3-tagging of the reverse primer, T3exoR3 in

the PCR reaction), using the Thermosequenase sequencing kit, and electrophoresis through a 4% polyacrylamide gel overnight (4% acrylamide, 1 x TBE, 8.3 M UREA; Appendix F) at 2250 V, 30.6 mA, 68.8 W). The reactions were prepared in a darkened room; the method is identical to that described in Chapter Four with the exception that 2 pmol of T3 fluorescent primer is added in place of radioactive label. Sequencing reactions were initially tested at half volume and were reduced to quarter volumes. Reactions were stored and transported in a black box with watery ice in between cycling and electrophoresis. The cycle sequencing program consisted of 95°C for 4 min, 30 cycles of 95°C for 15 s, 50°C for 30 s, and 72°C for 1 min, followed by 10 cycles of 95°C for 15 s, and 72°C for 1 min.

Samples were loaded in the dark and 2 μl of each sample was dotted onto parafilm before loading in order to prevent oil interfering with sample loading.

5.2.6.3 Radioactive Sequencing of Exonuclease Motifs II and III.

DNA sequencing of exonuclease motifs II and III was accomplished using γ ^{33}P-dATP primer-labelled reactions as per Chapter Four. Initially only one primer for each domain was labelled; exo2R and exo3F, but sequencing was conducted in the opposite direction for samples that displayed changes to the reference sequence (AF497906). In addition, several exonuclease III motif gels (12 samples each) were repeated to capture the end of the motif if electrophoresis ran too long (Figure 5.7).

5.3 Results - *POLγ* CAG Repeat Allele Distribution

5.3.1 PCR Analysis of the *POLγ* CAG Trinucleotide Repeat

The 295 bp *POLγ* CAG repeat region was successfully amplified by PCR in 275 samples (Figure 5.4) (Appendix D).

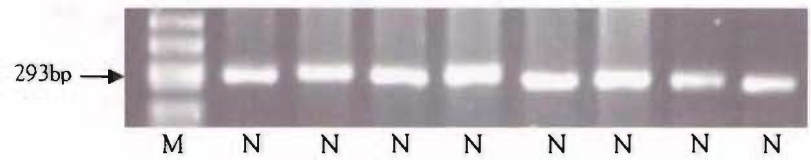


Figure 5.4. *POLγ* CAG repeat PCR product of eight men with normozoospermia. The marker is pBS(HaeIII).

5.3.2 DNA Sequencing Analysis and Genotype Categorisation.

The *POLγ* CAG repeat region was manually sequenced in 275 samples (182 men with poor semen quality, and 93 men with normal semen parameters) and genotype categories assigned (Table 5.4) (Figure 5.5). Because the sequencing product was a mixture of both *POLγ* alleles, wild-type alleles ran together through the gel while alleles of greater than or fewer than 10 CAG repeats displaced the upstream sequence by multiples of three base pairs depending on the number of extra CAG repeats (Figure 5.5). These displaced bases were half as bright as the rest of the sequence.

DNA sequencing using just the ddATP and ddTTP lanes still allowed discrimination of allele repeat number and doubled the throughput (Figure 5.5).

Table 5.4. Distribution of *POLγ* CAG trinucleotide repeat alleles.

Semen Classification	Homozygous Wild-type 10 and 10	Heterozygote 10 and ≠10	Homozygous Mutant ≠10 and ≠10	n
NormoZS	57 (61.3%)	33 (35.5%)	3 (3.2%)	93
Total Non-normozs	134 (73.6%)	41 (22.5%)	7 (3.8%)	182
AZS	10 (91%)	1 (9%)	-	11
Severely OligoasthenoZS	35 (74.5%)	10 (21.3%)	2 (4.2%)	47
Severely OligoZS	14 (93.3%)	1 (6.7%)	-	15
OligoasthenoZS	22 (75.9%)	5 (17.2%)	2 (6.9%)	29
OligoZS	18 (66.7%)	8 (29.6%)	1 (3.7%)	27
AsthenoZS	35 (66%)	16 (30.2%)	2 (3.8%)	53

ZS signifies 'zoospermia'. Non-normozs includes all individuals with reduced semen quality. The percentage in parentheses represents the proportion of the total number of individuals of that group. n indicates sample size. ≠10 refers to 'other' alleles with greater than or fewer than 10 CAG repeats.

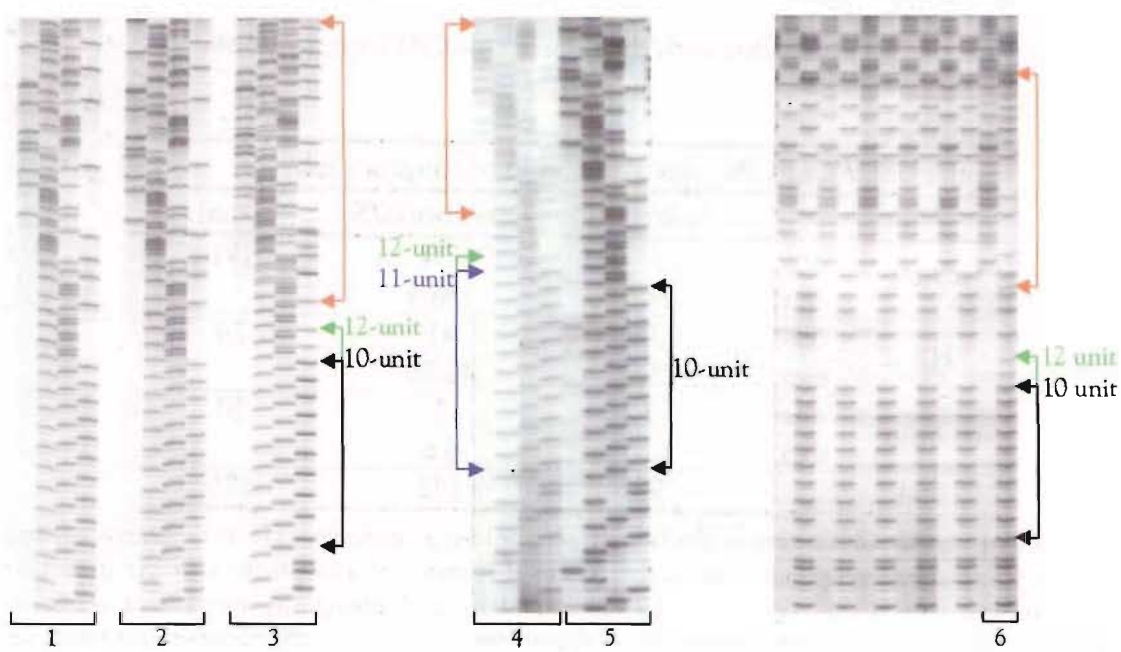


Figure 5.5. Autoradiography images of the *POLγ* CAG repeat region. Sample 1, 2, and 5 are all homozygous WT; with 2 copies of the 10-unit CAG repeat. Sample 3 is a heterozygote; two different alleles, one with the 10 unit repeat and one with 12 (the two fainter bands indicated by green arrow), producing a displacement of the repeat sequence 6 bps upwards (orange brackets). Sample 4 is a homozygous mutant; encoding one 11-unit allele (purple brackets) and one 12-unit allele (green) (up until 11 repeats the intensity of the bands is the same, the 12th is fainter). Sample 6 is heterozygous (the rest are homozygous wild-type), only the ddATP and ddTTP lanes have been electrophoresed. The orange brackets indicate the upstream displacement that helps distinguish the allele CAG repeat number.

Among the 'other' alleles the most common in heterozygotes was an 11-unit CAG repeat allele (79%), then 12 (16%), 9 (3%), and 8 (1%). Of homozygous mutant genotypes, just over half (six individuals; 55%) displayed CAG repeat alleles of 12+11, two individuals (18%) had 11+11, a further two (18%) had 8+11, and one (9%) had 9+11. This data suggests the *POLγ* CAG trinucleotide repeat is more likely to undergo slight expansions in repeat number than contractions, although since the number of repeat units was not established in the parents it is impossible to determine whether the expansion occurred within these individuals or was inherited.

5.3.3 Statistical Analysis

A 2x3 contingency table (Table 5.5) was created to determine how likely it is that the two sample subpopulations (normozoospermic and non-normozoospermic) were drawn from the same overall population with the same *POLγ* CAG repeat variation.

Table 5.5. *POLγ* CAG repeat variation contingency table.

		NormoZS	Non normoZS	Total
HOWT	obs	57	134	191
	exp	64.6	126.4	
HE	obs	33	41	74
	exp	25.0	48.98	
HOMut	obs	3	7	10
	exp	3.38	6.6	
Total		93	182	275

ZS is zoospermia. HOWT refers to the homozygous wild-type genotype, HE is heterozygous, and HOMut is homozygous mutant. Obs is the observed number of individuals with the particular genotype, and exp is the expected value generated by multiplying the fraction of the total population with that genotype (row total/total population), by the total number in that group (column total) to produce the expected number in that sample if it came from the overall population (for example, exp for normozs HOWT is 191/275 x 93 = 64.6)

The hypotheses were: H_O: there is no association between semen quality and genotype against, H_A: semen quality and genotype are related. A G-test of independence was calculated with William's correction to lower the observed value of G slightly and give more confidence in the test (Sokal and Rohlf 1995; Zar 1996).

Using the data in Table 5.5 and $\alpha=0.05$, $G_{adj}=4.962$ and, therefore, because $P(\chi^2 \geq X^2)$ is equal to $(5.99 \geq 4.962)$, H_O is true; there is no association between semen quality and *POLγ* CAG repeat genotype in this New Zealand population of men (Appendix H for details of calculations) (Sokal and Rohlf 1995; Zar 1996).

5.4 Results - *POLγ* Exonuclease Motif Analysis

5.4.1 PCR Analysis of Exonuclease Motifs I, II, and III

The first exonuclease motif was successfully amplified in 274 samples (83 normozoospermic and 191 non-normozoospermic), while motifs II and III were amplified in only a test sample of the total population; 126 samples in motif II (61 normozoospermic and 65 non-normozoospermic) and 124 in motif III (60 normozoospermic and 64 non-normozoospermic) (Figure 5.6).

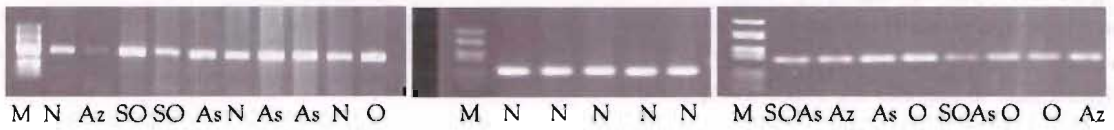


Figure 5.6. PCR images of *POLγ* exonuclease motifs I, II, and III. The PCR images are left to right: exonuclease motif I, II, and III; with products sizes of 250bp, 182bp, and 278bp respectively. The marker (M) is pBS(HaeIII).

5.4.2 DNA Sequencing Analysis

DNA sequencing of PCR products of all three exonuclease motif regions detected no change from the reference sequence AF497906 (Figure 5.7) (Appendix G). Re-sequencing of twenty sequences (16 motif II, and 4 motif III) confirmed the sequence to be the same as the reference.

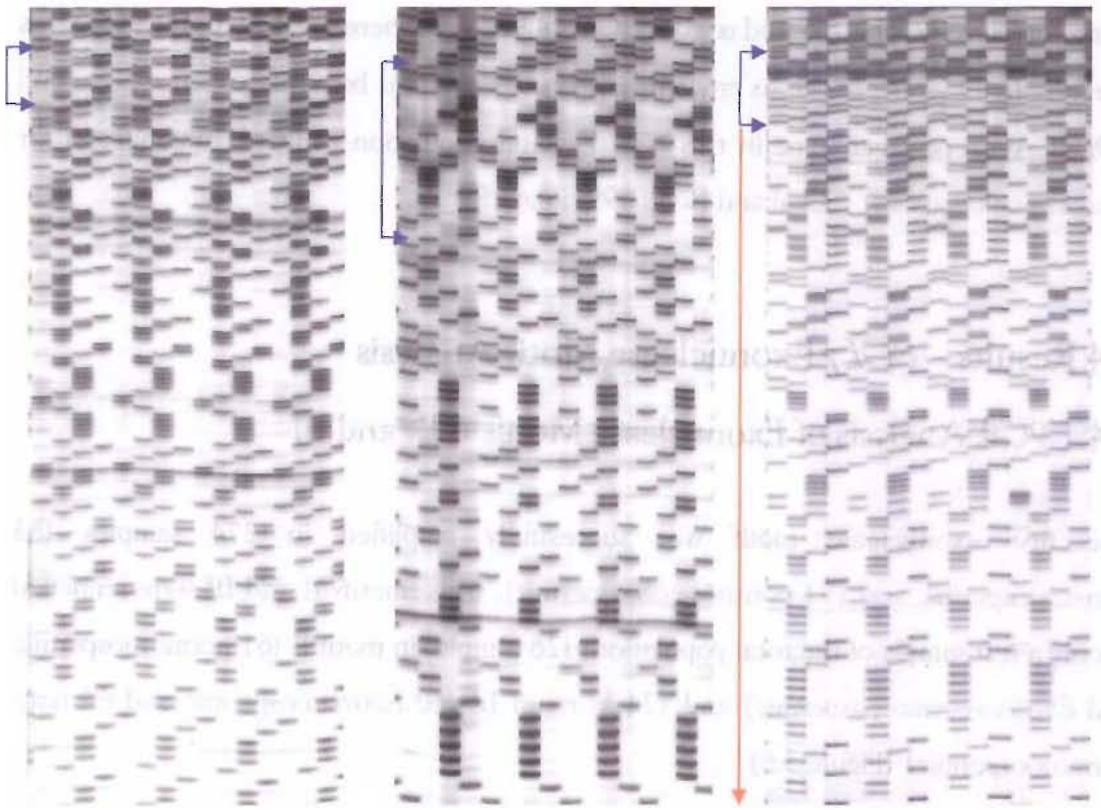


Figure 5.7. Autoradiography images of exonuclease motifs I, II, and III DNA sequences. Purple arrow brackets indicate exonuclease motifs. The orange arrow indicates the intron sequence separating the two sides of exonuclease III motif, the other end is just off the bottom of the image. Nucleotide sequences are listed in Appendix G.

5.5 Discussion

5.5.1 *POLγ* CAG Repeat Allele Distribution

The CAG repeat length variation was not statistically related to semen quality ($P>0.05$). Homozygous wild-type individuals predominated, making up 60% or higher of each semen quality grouping. Individuals with greater than or fewer than 10 CAG repeats on both alleles (homozygous mutants) were rare, and were not limited to classes of men with reduced sperm quality. The highest proportion of homozygous mutants was observed in men with both reduced sperm count and motility (oligoasthenozoospermia; 7%), although this actually represented only two individuals, and was not statistically significant.

There was not a great range of allele sizes observed, consistent with the idea that this is a stable trinucleotide repeat region (Rovio et al. 1999). The smallest CAG repeat length of eight repeats was found in three individuals, and the largest repeat length consisting of twelve was found in eighteen individuals.

5.5.1.1 Comparison with Other Research

The 10-unit allele is typically observed at a frequency of 80-90% in ethnically diverse populations (Rovio et al. 1999). The 10-unit allele was observed at a frequency of 0.83 (83%) in my study population suggesting it is not genetically unusual.

Statistical analysis indicated that semen quality is not related to CAG repeat variation (genotype) in the present study, a contrast to the findings of two other studies which have found associations between the homozygous absence of the 10-unit allele and oligozoospermia (Rovio et al. 2001) and with normozoospermic infertility (Jensen et al. 2004).

Interestingly, the differences between findings in the present study and Jensen et al. (2004) may not be as disparate as they first appear. Increasingly, population studies

investigating infertility are seeking to categorise participants based on their actual fertility status despite what semen category their sperm count, motility and morphology places them in according to World Health Guidelines. The study conducted by Jensen and colleagues found a significant proportion (14.3%) of their normozoospermic-yet-infertile men had two alleles with greater or fewer than 10 CAG repeats (homozygous mutants), compared to their group of unselected controls (men with unknown fertility; 2.3%) and recent fathers (0.9%). One of the difficulties with this New Zealand population is that their actual fertility status is unknown and thus they have been categorised according to semen parameters. Therefore, it is not known how many men are infertile but with normozoospermia.

5.5.2 *POLγ* Exonuclease Domain Analysis

No nucleotide changes were observed within any of the three *POLγ* exonuclease motifs, nor any heterozygosity noted. This suggests that the sequence integrity of this domain is important. Screening of this region in PEO (progressive external ophthalmoplegia) patients and their families has identified recessive mutations in the DNA regions broadly surrounding the exonuclease domain II (T251I and R309L), however both fall well outside the region screened (Copeland et al. 2003). The absence of mutation within the *POLγ* exonuclease domain may not necessarily be surprising since while cell culture might tolerate a mutator POLγ enzyme, it is unlikely that the production of an error-prone polymerase and consequent high mtDNA mutational load in every cell of the body would support a functional human being. This idea is somewhat supported by observations in PEO patients, where the same, or similar mutations produced a much slower rate of mtDNA mutation and was not accompanied by loss of mtDNA copy number than was reported in cell culture (Lamantea et al. 2002).

While 50 of the 109 individuals in the sample population had known polymorphisms within mtDNA genes (the other 59 had no changes in the gene sequences screened) a mutated POLγ was not the underlying cause (Appendix E). Homoplasmic polymorphisms are likely to have been inherited however without sequencing of the maternal mtDNA

genome it cannot be determined whether heteroplasmic silent single nucleotide polymorphisms arose in these men or were inherited.

Interestingly, while PEO is associated with an accumulation of mtDNA mutation, infertility is not mentioned as being associated with the disease although other symptoms characteristic of mitochondrial diseases are (Ponamarev et al. 2002). This could be for a number of reasons, even simply that not enough patients had attempted to have children, however, it is probable that a proportion of them would have reduced fertility given that an accumulation of mtDNA mutation is associated with poor semen quality (Kao et al. 1995; Lestienne et al. 1997; Holyoake et al. 2001).

References:

Aitken, R.J., J.S. Clarkson, and S. Fishel. 1989a. Generation of reactive oxygen species, lipid peroxidation, and human sperm function. *Biol Reprod* 41: 183-97.

Aitken, R.J., J.S. Clarkson, T.B. Hargreave, D.S. Irvine, and F.C. Wu. 1989b. Analysis of the relationship between defective sperm function and the generation of reactive oxygen species in cases of oligozoospermia. *J Androl* 10: 214-20.

Aitken, R.J., E. Gordon, D. Harkiss, J.P. Twigg, P. Milne, Z. Jennings, and D.S. Irvine. 1998. Relative impact of oxidative stress on the functional competence and genomic integrity of human spermatozoa. *Biol Reprod* 59: 1037-46.

Anderson, S., A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, and I.G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-65.

Ankel-Simons, F. and J.M. Cummins. 1996. Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. *Proc Natl Acad Sci U S A* 93: 13859-63.

Aziz, N., R.A. Saleh, R.K. Sharma, I. Lewis-Jones, N. Esfandiari, A.J. Thomas, Jr., and A. Agarwal. 2004. Novel association between sperm reactive oxygen species production, sperm morphological defects, and the sperm deformity index. *Fertil Steril* 81: 349-54.

Baker, M.A. and R.J. Aitken. 2004. The importance of redox regulated pathways in sperm cell biology. *Mol Cell Endocrinol* 216: 47-54.

Cardullo, R.A. and J.M. Baltz. 1991. Metabolic regulation in mammalian sperm: mitochondrial volume determines sperm length and flagellar beat frequency. *Cell Motil Cytoskeleton* 19: 180-8.

Chinnery, P.F., D.R. Thorburn, D.C. Samuels, S.L. White, H.M. Dahl, D.M. Turnbull, R.N. Lightowlers, and N. Howell. 2000. The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? *Trends Genet* 16: 500-5.

Copeland, W.C., M.V. Ponamarev, D. Nguyen, T.A. Kunkel, and M.J. Longley. 2003. Mutations in DNA polymerase gamma cause error prone DNA synthesis in human mitochondrial disorders. *Acta Biochim Pol* 50: 155-67.

Croteau, D.L. and V.A. Bohr. 1997. Repair of oxidative damage to nuclear and mitochondrial DNA in mammalian cells. *J Biol Chem* 272: 25409-12.

Cummins, J.M. 2000. Fertilization and elimination of the paternal mitochondrial genome. *Hum Reprod* 15 Suppl 2: 92-101.

Cummins, J.M., A.M. Jequier, R. Martin, D. Mehmet, and J. Goldblatt. 1998. Semen levels of mitochondrial DNA deletions in men attending an infertility clinic do not correlate with phenotype. *Int J Androl* 21: 47-52.

DiMauro, S. and E.A. Schon. 2003. Mitochondrial respiratory-chain diseases. *N Engl J Med* 348: 2656-68.

Fernie, A.R., F. Carrari, and L.J. Sweetlove. 2004. Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol* 7: 254-61.

Garett, R.H. and C.M. Grisham. 1995. *Biochemistry.* Saunders College Publishers, Fortworth.

Garrido, N., L. Griparic, E. Jokitalo, J. Wartiovaara, A.M. van der Bliek, and J.N. Spelbrink. 2003. Composition and dynamics of human mitochondrial nucleoids. *Mol Biol Cell* 14: 1583-96.

Graziewicz, M.A., B.J. Day, and W.C. Copeland. 2002. The mitochondrial DNA polymerase as a target of oxidative damage. *Nucleic Acids Res* 30: 2817-24.

Holyoake, A.J., P. McHugh, M. Wu, S. O'Carroll, P. Benny, I.L. Sin, and F.Y. Sin. 2001. High incidence of single nucleotide substitutions in the mitochondrial genome is associated with poor semen parameters in men. *Int J Androl* 24: 175-82.

Jensen, M., H. Leffers, J.H. Petersen, A. Nyboe Andersen, N. Jorgensen, E. Carlsen, T.K. Jensen, N.E. Skakkebaek, and E. Rajpert-De Meyts. 2004. Frequent polymorphism of the mitochondrial DNA polymerase gamma gene (POLG) in patients with normal spermiograms and unexplained subfertility. *Hum Reprod* 19: 65-70.

Kao, S., H.T. Chao, and Y.H. Wei. 1995. Mitochondrial deoxyribonucleic acid 4977-bp deletion is associated with diminished fertility and motility of human sperm. *Biol Reprod* 52: 729-36.

Kao, S.H., H.T. Chao, and Y.H. Wei. 1998. Multiple deletions of mitochondrial DNA are associated with the decline of motility and fertility of human spermatozoa. *Mol Hum Reprod* 4: 657-66.

Kazemi-Esfarjani, P., M.A. Trifiro, and L. Pinsky. 1995. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. *Hum Mol Genet* 4: 523-7.

Klug, W.S., and Cummings, M.R. 1997. *Concepts of Genetics.* Prentice Hall International, Inc., Upper Saddle River, New Jersey.

Korhonen, J.A., M. Gaspari, and M. Falkenberg. 2003. TWINKLE Has 5' -> 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem* 278: 48627-32.

Lamantea, E., V. Tiranti, A. Bordoni, A. Toscano, F. Bono, S. Servidei, A. Papadimitriou, H. Spelbrink, L. Silvestri, G. Casari, G.P. Comi, and M. Zeviani. 2002. Mutations of mitochondrial DNA polymerase gammaA are a frequent cause of autosomal dominant or recessive progressive external ophthalmoplegia. *Ann Neurol* 52: 211-9.

Lestienne, P., P. Reynier, M.F. Chretien, I. Penisson-Besnier, Y. Malthiery, and V. Rohmer. 1997. Oligoasthenospermia associated with multiple mitochondrial DNA rearrangements. *Mol Hum Reprod* 3: 811-4.

Lightowlers, R.N., P.F. Chinnery, D.M. Turnbull, and N. Howell. 1997. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. *Trends Genet* 13: 450-5.

Longley, M.J., D. Nguyen, T.A. Kunkel, and W.C. Copeland. 2001. The fidelity of human DNA polymerase gamma with and without exonucleolytic proofreading and the p55 accessory subunit. *J Biol Chem* 276: 38555-62.

Longley, M.J., R. Prasad, D.K. Srivastava, S.H. Wilson, and W.C. Copeland. 1998a. Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro. *Proc Natl Acad Sci U S A* 95: 12244-8.

Longley, M.J., P.A. Ropp, S.E. Lim, and W.C. Copeland. 1998b. Characterization of the native and recombinant catalytic subunit of human DNA polymerase gamma:

identification of residues critical for exonuclease activity and dideoxynucleotide sensitivity. *Biochemistry* 37: 10529-39.

Mason, P.A., E.C. Matheson, A.G. Hall, and R.N. Lightowlers. 2003. Mismatch repair activity in mammalian mitochondria. *Nucleic Acids Res* 31: 1052-8.

May-Panloup, P., M.F. Chretien, F. Savagner, C. Vasseur, M. Jean, Y. Malthiery, and P. Reynier. 2003. Increased sperm mitochondrial DNA content in male infertility. *Hum Reprod* 18: 550-6.

Meeusen, S., Q. Tieu, E. Wong, E. Weiss, D. Schieltz, J.R. Yates, and J. Nunnari. 1999. Mgm101p is a novel component of the mitochondrial nucleoid that binds DNA and is required for the repair of oxidatively damaged mitochondrial DNA. *J Cell Biol* 145: 291-304.

Mundy, A.J., T.A. Ryder, and D.K. Edmonds. 1995. Asthenozoospermia and the human sperm mid-piece. *Hum Reprod* 10: 116-9.

Ponamarev, M.V., M.J. Longley, D. Nguyen, T.A. Kunkel, and W.C. Copeland. 2002. Active site mutation in DNA polymerase gamma associated with progressive external ophthalmoplegia causes error-prone DNA synthesis. *J Biol Chem* 277: 15225-8.

Rantanen, A. and N.G. Larsson. 2000. Regulation of mitochondrial DNA copy number during spermatogenesis. *Hum Reprod* 15 Suppl 2: 86-91.

Ropp, P.A. and W.C. Copeland. 1996. Cloning and characterization of the human mitochondrial DNA polymerase, DNA polymerase gamma. *Genomics* 36: 449-58.

Ross, J. 1999. 7th May 1999 (last update). Online. Available on world wide web .jdaross.mcmail.com/mitochon.htm. Accessed 2nd March 2004. In.

Rovio, A., V. Tiranti, A.L. Bednarz, A. Suomalainen, J.N. Spelbrink, N. Lecrenier, A. Melberg, M. Zeviani, J. Poulton, F. Foury, and H.T. Jacobs. 1999. Analysis of the trinucleotide CAG repeat from the human mitochondrial DNA polymerase gene in healthy and diseased individuals. *Eur J Hum Genet* 7: 140-6.

Rovio, A.T., D.R. Marchington, S. Donat, H.C. Schuppe, J. Abel, E. Fritsche, D.J. Elliott, P. Laippala, A.L. Ahola, D. McNay, R.F. Harrison, B. Hughes, T. Barrett, D.M. Bailey, D. Mehmet, A.M. Jequier, T.B. Hargreave, S.H. Kao, J.M. Cummins, D.E. Barton, H.J. Cooke, Y.H. Wei, L. Wichmann, J. Poulton, and H.T. Jacobs. 2001. Mutations at the mitochondrial DNA polymerase (POLG) locus associated with male infertility. *Nat Genet* 29: 261-2.

Ruiz-Pesini, E., C. Diez, A.C. Lapena, A. Perez-Martos, J. Montoya, E. Alvarez, J. Arenas, and M.J. Lopez-Perez. 1998. Correlation of sperm motility with mitochondrial enzymatic activities. *Clin Chem* 44: 1616-20.

Sokal, R.R. and F.J. Rohlf. 1995. *Biometry : The Principles and Practise of Statistics in Biological Research.* W. H. Freeman and Company, New York.

Spelbrink, J.N., J.M. Toivonen, G.A. Hakkaart, J.M. Kurkela, H.M. Cooper, S.K. Lehtinen, N. Lecrenier, J.W. Back, D. Speijer, F. Foury, and H.T. Jacobs. 2000. In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *J Biol Chem* 275: 24818-28.

Spurger, L. 2003. 5th December 2003 (last update). Online. University of Texas Medical Branch, for the cell biology graduate programme. Available on world wide web cellbio.utmb.edu/cellbio/mitoch2.htm. Accessed 2nd March 2004. In.

St John, J.C., I.D. Cooke, and C.L. Barratt. 1997. Mitochondrial mutations and male infertility. *Nat Med* 3: 124-5.

Stevnsner, T., T. Thorslund, N.C. de Souza-Pinto, and V.A. Bohr. 2002. Mitochondrial repair of 8-oxoguanine and changes with aging. *Exp Gerontol* 37: 1189-96.

Stierum, R., G. Dianov, and V. Bohr. 1999. Single-nucleotide patch base excision repair of uracil in DNA by mitochondrial protein extracts. *Nucleic Acids Research* 27: 3712-9.

Sutovsky, P. 2003. Ubiquitin-dependent proteolysis in mammalian spermatogenesis, fertilization, and sperm quality control: killing three birds with one stone. *Microsc Res Tech* 61: 88-102.

Sutovsky, P., R. Hauser, and M. Sutovsky. 2004a. Increased levels of sperm ubiquitin correlate with semen quality in men from an andrology laboratory clinic population. *Hum Reprod* 19: 628-38.

Sutovsky, P., R.D. Moreno, J. Ramalho-Santos, T. Dominko, C. Simerly, and G. Schatten. 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol Reprod* 63: 582-90.

Sutovsky, P., K. Van Leyen, T. McCauley, B.N. Day, and M. Sutovsky. 2004b. Degradation of paternal mitochondria after fertilization: implications for heteroplasmy, assisted reproductive technologies and mtDNA inheritance. *Reprod Biomed Online* 8: 24-33.

Taivassalo, T., K. Fu, T. Johns, D. Arnold, G. Karpati, and E.A. Shoubridge. 1999. Gene shifting: a novel therapy for mitochondrial myopathy. *Hum Mol Genet* 8: 1047-52.

Thompson, W.E., J. Ramalho-Santos, and P. Sutovsky. 2003. Ubiquitination of prohibitin in Mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. *Biol Reprod* 69: 254-60.

Tomkinson, A.E., R.T. Bonk, J. Kim, N. Bartfeld, and S. Linn. 1990. Mammalian mitochondrial endonuclease activities specific for ultraviolet-irradiated DNA. *Nucleic Acids Res* 18: 929-35.

Twigg, J.P., D.S. Irvine, and R.J. Aitken. 1998. Oxidative damage to DNA in human spermatozoa does not preclude pronucleus formation at intracytoplasmic sperm injection. *Hum Reprod* 13: 1864-71.

Wallace, D.C. 1999. Mitochondrial diseases in man and mouse. *Science* 283: 1482-8.

Wojcik, C. 2003. David vs Goliath: can proteosomes chop entire mitochondria? *Commentary on Thompson et al. (2003) Ubiquitination of prohibitin in Mammalian sperm mitochondria: possible roles in the regulation of mitochondrial inheritance and sperm quality control. Posted 27th March 2003* www.update.bmn.com.

Zar, J. 1996. *Biostatistical Analysis Third Edition.* Prentice-Hall International, Inc, Upper Saddle River, NJ.

CHAPTER SIX – GENERAL DISCUSSION AND CONCLUSIONS OF THESIS RESEARCH

6.1 Conclusions of Thesis Research

Of the three different gene regions analysed, only the Y chromosome AZF region showed an association between mutation and reduced semen quality; no mutation was observed in the nuclear *POLγ* gene, neither was X chromosomal *AR* gene variation correlated with semen quality. While, all three chromosomal regions play an important role in the development of fertility, *AR* and *POLγ* mutation are likely to have a causal role in only rare cases of impaired spermatogenesis. While these results make it more difficult to establish the importance of trinucleotide repeat variation in both the *AR* and *POLγ* genes to infertility, it is consistent with the idea that many cases of infertility are due to an accumulation of minor changes in many fertility genes that cumulatively produce a reduction in semen quality. This idea suggests the presence of threshold levels such that a gradient of impaired sperm production exists with different frequencies of mutation. It also does not rule out the likelihood that some mutations have more direct and severe consequences on sperm production.

None of the 51 individuals previously identified with at least one mtDNA polymorphism, (several of which were heteroplasmic) had mutations within one or more of the three exonuclease motif regions of *POLγ* to account for this (Appendix E).

6.1.1 Analysis of the Three Genes Collectively

Of the 321 individuals in the sample population, a total of 91 individuals (44 normozoospermic and 47 men with poor semen quality) were screened for mutation in all five regions (*AR* CAG, Yq11 AZF, *POLγ* CAG and exonuclease motifs I, II, and III) (Table 6.1). Although sample size is low, there appears to be a general trend towards more severely impaired spermatogenesis in individuals with more than one high-risk criterion: moderately expanded *AR* CAG repeat region, Yq micro-deletion, HE or HOMut *POLγ* alleles, and mtDNA SNPs. All five samples with both a Yq micro-deletion and at least one mtDNA SNP belong to the azoospermic or severely

oligoasthenozoospermic semen classifications (Table 6.1); a sixth severely oligoasthenozoospermic sample has a Yq micro-deletion but has not been analysed for the presence of mtDNA SNPs. Two other individuals with less severely reduced semen quality have Yq micro-deletions but no mtDNA SNPs. A trend also seems to exist between mtDNA SNPs and asthenozoospermia, and two samples have both moderately expanded *AR* CAG repeat numbers (24 and 31) and more than one mtDNA SNP (Table 6.1).

As is often the case, associations between particular genetic changes and semen quality are not absolute since the presence of Yq micro-deletions and SNPs does not account for every azoospermic or severely oligoasthenozoospermic sample and several normozoospermic samples also have mtDNA SNPs (10/44; Table 6.1). However, it is important to note that the actual fertility status of this sample population is unknown. Each sample has been assigned a semen classification based on semen parameters but it is possible that a proportion of men in the normozoospermic group may have difficulty conceiving children, and similarly, inclusion within the non-normozoospermic group does not necessarily signify infertility. Without knowledge of actual fertility status, it becomes difficult to distinguish between harmless polymorphisms and important mutations. Yet, the fact that even increased frequencies of silent mtDNA SNPs are associated with reduced sperm production suggests that genetic background is an important determinant of fertility status (Holyoake et al. 2001).

These observations are consistent with the idea that infertility is caused by multiple factors. Rare mutations such as some Yq AZF micro-deletions seem to act in a dominant fashion to result in severely impaired spermatogenesis, but that most infertility is due to the accumulation of multiple nucleotide substitutions and deletions that act synergistically to reduce semen quality.

Table 6.1. Summary of individuals with all five gene regions analysed.

Semen Classification	Sample	Yq11	*AR* CAG	*POLγ* CAG	*POLγ* EXO	mtDNA Mutation*
Az	1454	Yq	18	HOWT	I, II, III	*ND1*/4
Az	2473	Yq	21	HOWT	I, II, III	*ND6*/4
Az	3008	Yq	24	HOWT	I, II, III	*ND1, 2, 3*/4
Az	4071	sY149	24	HOWT	I, II, III	*ND1*/3
Az	4449	sY158	21	HE 10+11	I, II, III	*ND6*/3
SOAs	1028	Yq	21	HOWT	I, II, III	0/2
SOAs	2128	Yq	21	HE 10+11	I, II, III	0/5
SOAs	2565	Yq	21	HOMut 11+12	I, II, III	*ND5*/3
SOAs	2866	Yq	28	HOMut 11+12	I, II, III	0/4
SOAs	4110	Yq	19	HE 10+12	I, II, III	*ND1*/3
SOAs	4290	sY149	21	HOWT	I, II, III	*ND1*/2
SOAs	4646	Yq	25	HOWT	I, II, III	0/2
SOAs	4823	Yq	20	HE 10+11	I, II, III	0/2
SOAs	6290	sY149	24	HOWT	I, II, III	-
SOAs	BT	sY146	13	HOWT	I, II, III	*COII*/3
SOAs	CP	sY146,156	20	HE 10+11	I, II, III	*ND5*/5
SOAs	SF96-333	Yq	21	HOWT	I, II, III	*ND6, COIII*/7
SO	2578	Yq	24	HOWT	I, II, III	*ND5*/3
SO	4406	Yq	21	HOWT	I, II, III	-
SO	SF96-470	Yq	22	HOWT	I, II, III	*ND1*/3
OAs	2275	Yq	21	HOWT	I, II, III	-
OAs	2908	Yq	20	HOWT	I, II, III	-
OAs	2935	Yq	29	HE 8+10	I, II, III	0/1
OAs	3072	Yq	20	HOWT	I, II, III	0/3
OAs	3449	Yq	22	HOMut 8+11	I, II, III	-
OAs	AM235Q	Yq	31	HOWT	I, II, III	*ATPase6, COIII*/6
OAs	AT734U	Yq	18	HOWT	I, II, III	0/4
OAs	Mr X	Yq	20	HOMut 8+11	I, II, III	-
OAs	SF96-1667	Yq	18	HE 10+11	I, II, III	0/1
O	1699	Yq	26	HE 10+11	I, II, III	0/3
O	2410	Yq	21	HOWT	I, II, III	0/5
O	2686	Yq	24	HOWT	I, II, III	0/1
O	5597	Yq	28	HOMut 11+12	I, II, III	0/1
O	JN	Yq	18	HOWT	I, II, III	*ND1,COII*/6
O	SF94-380	sY149	22	HOWT	I, II, III	0/5
As	1644	Yq	19	HE 10+11	I, II, III	-
As	3138	Yq	27	HE 10+11	I, II, III	0/2
As	94-110	Yq	24	HOWT	I, II, III	0/5
As	94-149	Yq	24	HE 10+11	I, II, III	*ATPase6, 8*/5
As	AM433B	Yq	19	HOWT	I, II, III	*ND1, COII*/4
As	AT665C	Yq	18	HOWT	I, II, III	*ND2, COII*/8

As	CCR	Yq	21	HOWT	I, II, III	*ND6, COII*/8
As	RJ	Yq	21	HOWT	I, II, III	0/1
As	SF94-303	Yq	17	HOWT	I, II, III	0/2
As	SF95-590	Yq	22	HOWT	I, II, III	0/2
As	SF96-166	Yq	25	HOWT	I, II, III	*ND1*/7
As	WT	sY156	24	HE 10+12	I, II, III	0/4
	47					
N	1469	Yq	24	HE 10+11	I, II, III	0/3
N	1535	Yq	21	HE 10+11	I, II, III	0/3
N	1541	Yq	25	HE 9+10	I, II, III	0/4
N	1738	Yq	21	HOWT	I, II, III	0/4
N	1751	Yq	21	HOWT	I, II, III	0/1
N	2016	Yq	24	HE 10+11	I, II, III	0/2
N	2097	Yq	25	HOWT	I, II, III	0/2
N	2223	Yq	26	HOWT	I, II, III	0/2
N	2404	Yq	21	HE 10+11	I, II, III	*ND1*/2
N	2450	Yq	15	HOWT	I, II, III	0/4
N	2631	Yq	20	HOWT	I, II, III	0/4
N	2905	Yq	20	HE 10+11	I, II, III	*ND1*/5
N	3002	Yq	20	HOWT	I, II, III	0/1
N	3182	Yq	23	HOWT	I, II, III	0/1
N	3408	Yq	23	HOWT	I, II, III	*ND1*,3
N	3598	Yq	19	HOWT	I, II, III	-
N	3725	Yq	19	HOWT	I, II, III	0/2
N	3967	Yq	9	HE 10+11	I, II, III	-
N	3999	Yq	20	HOMut 11+11	I, II, III	-
N	4022	Yq	24	HOWT	I, II, III	0/1
N	4817	Yq	20	HE 10+12	I, II, III	0/3
N	5016	Yq	20	HE 10+12	I, II, III	0/2
N	5222	Yq	19	HOWT	I, II, III	0/3
N	5239	Yq	21	HOWT	I, II, III	0/2
N	5477	Yq	38	HOWT	I, II, III	-
N	6109	Yq	23	HE 10+11	I, II, III	0/2
N	6169	Yq	20	HE 10+11	I, II, III	0/2
N	95-27	Yq	22	HE 10+11	I, II, III	*ND1, 6, COII*/7
N	AM422U	Yq	22	HOWT	I, II, III	0/5
N	AM447I	Yq	21	HE 10+11	I, II, III	0/4
N	AM467O	Yq	21	HE 10+11	I, II, III	0/5
N	AM498B	Yq	21	HOWT	I, II, III	*COIII*/8
N	AT689M	Yq	26	HOWT	I, II, III	*ND1, COIII*/8
N	AT697N	Yq	24	HE 10+11	I, II, III	*ATPase6*/8
N	QD747V	Yq	22	HOWT	I, II, III	*ATPase8, COII*/2
N	QD845B	Yq	26	HE 10+11	I, II, III	0/5
N	RL1911	Yq	23	HE 10+11	I, II, III	-
N	SF155	Yq	20	HOWT	I, II, III	0/3
N	SF289	Yq	21	HOWT	I, II, III	-

N	SF437	Yq	23	HOWT	I, II, III	0/1
N	SF94-399	Yq	21	HOWT	I, II, III	*ATPase6*/4
N	SF95-122	Yq	25	**HOMut 9+11**	I, II, III	*ND1, ATPase6, COIII*/7
N	SF96-229	Yq	18	**HOMut 11+11**	I, II, III	0/4
N	SF96-310	Yq	20	**HE 10+12**	I, II, III	0/2
	44					

This table is a summary of the information in Appendices D and E. *The mtDNA mutations column indicates known mtDNA mutations, if present, out of the number of mtDNA genes analysed, gathered from research previously conducted by other researchers. Dashes indicate the sample has not been analysed for mtDNA mutation. Bolded cells in the Yq11, *AR* CAG, and *POLγ* CAG columns indicate results that in previous studies were associated with reduced semen quality; only the Yq11 micro-deletions were significantly associated with reduced semen quality in the current study.

The three men exposed to mutagens; chemotherapy (individual 8257, azoospermic) and occupational chemicals (RI2012 and SF96-341, both normozoospermic) did not have an increased frequency of mutation, both orchardists were completely within the normal range for the three genes studied: 22 and 21 *AR* CAG repeats, no Yq micro-deletions, two 10-unit *POLγ* alleles, and no mtDNA SNPs. While it is unknown what the fertility status or semen quality of 8257 was before chemotherapy, the cause of his azoospermia is not due to an accumulation of Yq11 micro-deletions. A fourth individual (7234, azoospermic) is identified as having spermatid maturation arrest. Analysis conducted in this study suggests this impairment of sperm production is not due to Yq micro-deletion; however, the sample did display significant expansion of the *AR* CAG repeat region (27 repeats).

6.2 Discussion

This study represents the first research effort analysing multiple fertility gene regions in a single study population. Three earlier studies evaluated the contribution of both androgen receptor (AR) trinucleotide repeat variation and Y chromosome micro-deletion to reduced semen quality in the same population (Yong et al. 2000; Madgar et al. 2002; Dhillon and Husain 2003) however, it is common practise to screen the Y chromosome

for micro-deletion since infertility-causing mutations are routinely observed in approximately six percent of men with reduced semen quality, and may therefore be an alternative reason for the observed infertility than mutation within the gene of interest.

Although the sample size of the group analysed for mutation in all three genes plus mitochondrial single nucleotide polymorphism (SNP) analysis (Table 6.1) is small, the indication is there that the presence of more than one fertility gene mutation increases the risk of reduced semen quality. This outcome suggests that studies like this on a larger scale would produce valuable insight into threshold levels for semen quality, particularly of Y chromosome micro-deletion and mitochondrial SNPs. Despite the amount of research in these areas the level of tolerance for SNPs or micro-deletion is unclear. Normozoospermic individuals have been observed both with large deletions of the mitochondrial genome (Cummins et al. 1998) and micro-deletions of AZF (Pryor et al. 1997; Kent-First et al. 1999; Kerr et al. 2000), the difference between these individuals and those with reduced semen quality has not yet been conclusively established. It is generally agreed that the genetic background of these individuals is more robust, and are possibly even super-fertile and can therefore tolerate a greater amount of mildly deleterious mutations. Alternatively it is hypothesised that mutations observed in men with both normal and non-normal semen classifications are not dominant acting; and thus the reduction in semen quality is due to a dominant mutation or accumulation of weaker mutations in genes elsewhere in the genome; either undiscovered or not analysed (De Kretser and Baker 1999).

6.2.1 Use of Human Study Populations to Investigate Male Infertility

It is currently acknowledged that more meaningful associations between gene mutation and semen quality come from a sample population that is highly selected based on semen characteristics. This sounds quite obvious since it stands to reason that the same types of pathology will be produced from the same mutation or mutations of the same gene, however, in reality this has proven to be harder than it appears. Ideally, research would be based on a sample population of generous numbers with highly similar semen

characteristics, DNA would be available from the mother and at least one paternal relation (ideally brother and father) to determine the origin of the mutation, and in addition to the gene of interest, the Yq11 AZF region, and mitochondrial genome would be analysed for mutation, given the known mutation rate of these areas. Studies that have all of these points are rare because it requires an exceptional level of organisation and participation. Furthermore, fertility seems to be far more complex than linking one abnormally functioning protein with reduced sperm production. While there do seem to be dominant acting mutations associated with impaired spermatogenesis, such as micro-deletions of AZF, even these can produce a frustrating amount of phenotypic variability in different individuals.

Mouse models are more useful in the study of male infertility since the animals are more amenable to interference in mating choice (inbreeding and backcrossing to grandparents is commonplace), and chemical mutagenesis in order to create infertility where it wasn't previously.

6.2.2 Future Research Direction

We do not currently have a very good understanding of the numerous pathways and gene interactions involved in the development and maintenance of fertility; and in general most cases of fertility (where a genetic origin is suspected) go unexplained. This has repercussions for infertile couples since if the cause involves germ cells there is the potential to pass on infertility-causing mutations to offspring conceived via artificial reproductive techniques.

Part of the problem is the lack of information on the number and function of genes and gene functions involved. In addition to gametogenesis defects in pathways contributing to sexual differentiation, fertilisation, and embryonic development can cause infertility or miscarriage (Matzuk and Lamb 2002).

Therefore, further research analysing multiple fertility regions in men with reduced semen quality is required before we can begin to understand the complexity of male infertility and develop effective treatments. Genetics and molecular medicine have an increasing need for rapid genotyping and mutational analysis. It is likely that future research will increasingly employ DNA technologies that allow analysis of genetic changes and expression patterns rapidly, with high throughput and enable parallel analysis of very large number of genes (Matzuk and Lamb 2002; Carella et al. 2003).

For examples, microarray technology has the potential to be used for a variety of applications including genotyping and diagnosis using a set of known mutations or polymorphisms, gene expression, predicting gene function and linking cell pathways (based on expression patterns), and drug discovery and drug target validation (Carella et al. 2003). Once genes have been identified in relation to disease the next stage is to understand gene function and interaction with other cellular components and processes. Microarrays (oligonucleotide and cDNA arrays) can provide information regarding the localisation of gene expression how the expression pattern of one gene relates to those of others (Carella et al. 2003). A relatively new technology is protein arrays which allow high-throughput detection of proteins in biomedical samples. In addition, since most biological activities are mediated by protein-protein interactions, the technology has been adapted to detect macromolecular interactions, such as those between proteins and potential drugs (Carella et al. 2003).

Two notable studies using this type of technology to specifically search for genes expressed in mouse male germ cells have identified 19 novel premeiotic testis-specific transcripts, nine of which were X-linked (Wang et al. 2001), and 1,652 transcripts whose expression increased during or after meiosis (Schultz et al. 2003). These genes represent candidate infertility genes and are also possible targets for male contraceptives (Schultz et al. 2003).

References:

Carella, M., S. Volinia, and P. Gasparini. 2003. Nanotechnologies and microchips in genetic diseases. *J Nephrol* 16: 597-602.

Cummins, J.M., A.M. Jequier, R. Martin, D. Mehmet, and J. Goldblatt. 1998. Semen levels of mitochondrial DNA deletions in men attending an infertility clinic do not correlate with phenotype. *Int J Androl* 21: 47-52.

De Kretser, D.M. and H.W. Baker. 1999. Infertility in men: recent advances and continuing controversies. *J Clin Endocrinol Metab* 84: 3443-50.

Dhillon, V.S. and S.A. Husain. 2003. Cytogenetic and molecular analysis of the Y chromosome: absence of a significant relationship between CAG repeat length in exon 1 of the androgen receptor gene and infertility in Indian men. *Int J Androl* 26: 286-95.

Holyoake, A.J., P. McHugh, M. Wu, S. O'Carroll, P. Benny, I.L. Sin, and F.Y. Sin. 2001. High incidence of single nucleotide substitutions in the mitochondrial genome is associated with poor semen parameters in men. *Int J Androl* 24: 175-82.

Kent-First, M., A. Muallem, J. Shultz, J. Pryor, K. Roberts, W. Nolten, L. Meisner, A. Chandley, G. Gouchy, L. Jorgensen, T. Havighurst, and J. Grosch. 1999. Defining regions of the Y-chromosome responsible for male infertility and identification of a fourth AZF region (AZFd) by Y-chromosome microdeletion detection. *Mol Reprod Dev* 53: 27-41.

Kerr, N.J., J. Zhang, F.Y. Sin, P. Benny, and I.L. Sin. 2000. Frequency of microdeletions in the azoospermia factor region of the Y-chromosome of New Zealand men. *N Z Med J* 113: 468-70.

Madgar, I., L. Green, M. Kent-First, R. Weissenberg, R. Gershoni-Baruch, B. Goldman, and E. Friedman. 2002. Genotyping of Israeli infertile men with idiopathic oligozoospermia. *Clin Genet* 62: 203-7.

Matzuk, M.M. and D.J. Lamb. 2002. Genetic dissection of mammalian fertility pathways. *Nat Cell Biol* 4 Suppl: s41-9.

Pryor, J.L., M. Kent-First, A. Muallem, A.H. Van Bergen, W.E. Nolten, L. Meisner, and K.P. Roberts. 1997. Microdeletions in the Y chromosome of infertile men. *N Engl J Med* 336: 534-9.

Schultz, N., F.K. Hamra, and D.L. Garbers. 2003. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proc Natl Acad Sci U S A* 100: 12201-6.

Wang, P.J., J.R. McCarrey, F. Yang, and D.C. Page. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet* 27: 422-6.

Yong, E.L., J. Lim, Q. Wang, A. Mifsud, Y.C. Ong, and C.K. Sim. 2000. Genetics of male infertility: role of androgen receptor mutations and Y-microdeletions. *Ann Acad Med Singapore* 29: 396-400.

APPENDIX A - ABBREVIATIONS

A_{260}	light absorbance at 260nm
A_{280}	light absorbance at 280nm
aa	amino-acid
APS	ammonium persulfate
AR	androgen receptor
As	Asthenozoospermia
AZ	Azoospermia
AZF	Azoospermic Factor
bp	base pairs of DNA
CAIS	complete androgen insensitivity syndrome
cpm	counts per minute
dATP	deoxyadenosine triphosphate
DAZ	Deleted in Azoospermia
DBD	DNA-binding domain
dCTP	deoxycytidine triphosphate
ddH_2O	deionised, distilled water
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
ds	double-stranded
dTTP	deoxythymidine triphosphate
DTT	dithiothreiotol
EDTA	etyhylenediaminetetra acetic acid
GC	Guanine - Cytosine base pairing
HCl	Hydrogen Chloride
ICSI	Intracytoplasmic sperm injection
IVF	*In vitro* fertilisation
kcal	Kilocalories
LBD	ligand-binding domain
M	molar
mA	milliamps (a measure of current)
Mg^{2+}	Magnesium
min	minutes
MSY	male specific region of Y
mtDNA	mitochondrial DNA
NaCl	Sodium Chloride
nm	nanometer
NormoZS	Normozoospermia
PAGE	Polyacrylamide Gel Electrophoresis
PAIS	partial androgen insensitivity syndrome
PCR	Polymerase Chain Reaction
PEO	progressive external ophthalmoplegia
POLγ	Polymerase gamma
O	Oligozoospermia
OAs	Oligoasthenozoospermia
OD	optical density
OXPHOS	oxidative phosphorylation
ROS	reactive oxygen species
s	seconds
SBMA	Small bulbar muscular atrophy (Kennedy's Disease)
SNP	single nucleotide polymorphism
SO	Severe Oligozoospermia
SOAs	Severe Oligoasthenozoospermia
STS	sequence tagged site
TAD	transcription-activating domain
TAE	Tris/acetate electrophoresis buffer
Taq	*Thermus aquaticus*
TBE	Tris/borate electrophoresis buffer
TEMED	tetramethylethylenediamine
temp	temperature
UV	ultraviolet
V	Volts (a measure of voltage)
W	Watts (a measure of power)
w/v	weight per volume
ZS	Zoospermia

Units:	g	grams
	l	litres
	m	meters
Prefixes:	c	centi ($1x10^{-2}$)
	m	milli ($1x10^{-3}$)
	μ	micro ($1x10^{-6}$)
	n	nano ($1x10^{-9}$)
	p	pico ($1x10^{-12}$)
	f	femto ($1x10^{-15}$)

APPENDIX B – CHEMICAL SUPPLIERS

Product	Supplier
Acrylamide:bisacrylamide 19:1 (40%)	BioRad
Acrylamide:bisacrylamide 29:1 (30%)	BioRad
Agarose	Pure Sciences
α^{33}P-dCTP, γ^{32}P-dATP	NEN, PerkinElmer Life Sciences Inc.
Ammonium acetate	BDH
Ammonium persulfate (APS)	Boehringer Mannheim (Roche)
Blocking reagent	Roche
Bromophenol blue	Sigma
Chloroform	BDH
Dexonucleoside triphosphates (dNTPs)	Eppendorf
1, 4-Dithiothreitol (DTT)	Boehringer Mannheim (Roche)
Ethanol	BDH
Ethidium Bromide	Sigma
Ethylene diaminetetra-acetic acid (EDTA)	Scharlau (Scientific Supplies Ltd)
Isoamyl alcohol	BDH
Isopropanol	BDH
Mineral oil	Sigma
Orange G	Sigma
Orthoboric acid	BDH
Phenol	Scharlau (Scientific Supplies Ltd)
Proteinase K	Roche
SigmaCote	Sigma
Sodium acetate	BDH
Sodium chloride	Scharlau (Scientific Supplies Ltd)
Sodium dodecyl sulfate (SDS)	Gibco BRL (Invitrogen)
Sodium hydroxide	Panreac (Scientific Supplies Ltd)
Sodium citrate	BDH
N,N,N',N'-Tetramethyl-ethylenediamine (TEMED)	Sigma
Tris Base	Boehringer Mannheim (Roche)
Urea	BDH
Xylene cyanol FF	Sigma

APPENDIX C - SAMPLE POPULATION CHARACTERISTICS

Semen Classification	Sample Code	Sperm Count (10^6/ml)	Motility (%)	Date of Birth	Comments
AZOOSPERMIA					
Az	798	0.00	0	-	
Az	1454	0.00	0	1963	SNP'S
Az	2473	0.00	0	1966	SNP'S
Az	2762	0.00	0	-	
Az	3008	0.00	0	1959	SNP'S
Az	4071	0.00	0	1959	SNP'S, Vasectomy Reversal
Az	4449	0.00	0	1971	SNP'S
Az	5753	0.00	0	-	SCO*
Az	7234	0.00	0	-	Maturation Arrest
Az	7422	0.0	0	1971	T.V.=0.6mL
Az	8055	0.00	0	-	
Az	8061	0.00	0	-	Orchidopexy
Az	8135	0.00	0	-	Orchidopexy
Az	8246	0.00	0	-	Orchidopexy
Az	8257	0.00	0	1964	T.V.=2.3mL, Post Chemotherapy
Az	8708	0.00	0	1974	
Az	8712	0.00	0	1972	T.V.=0.4mL
Az	AT752B	0.00	0	1959	SNP'S
Az	SF96-238	0.00	0	-	
	19				
SEVERE OLIGOASTHENOZOOSPERMIA					
SOAs	1020	1.40	33	1962	
SOAs	1028	0.15	0	1963	
SOAs	1336	0.30	33	1961	Vasectomy Reversal
SOAs	1696	0.10	0	1960	
SOAs	2128	0.30	10	1965	
SOAs	2565	4.30	20	1965	SNP'S
SOAs	2866	0.70	29	1968	
SOAs	2940	0.40	36	1936	
SOAs	3261	3.60	45	1967	Count rose to 24 M/mL over 5 yrs
SOAs	3318	0.13	40	1961	
SOAs	3993	0.07	30	1968	
SOAs	4110	1.50	31	1966	SNP'S
SOAs	4290	0.28	18	1967	SNP'S
SOAs	4300	4.00	45	1962	SNP'S
SOAs	4646	2.30	29	1963	
SOAs	4823	0.80	9	1959	
SOAs	5002	0.46	30	1964	SNP'S, Obstruction
SOAs	5080	1.00	8	1952	
SOAs	5271	1.30	37	1961	
SOAs	5349	1.90	5	1972	

SOAs	5585	2.50	37	1962	SNP'S
SOAs	5715	0.60	13	1962	
SOAs	6290	0.01	36	1963	
SOAs	6485	0.30	29	1905	
SOAs	6585	3.50	37	1973	
SOAs	6611	4.60	45	1960	
SOAs	7284	0.9	44	1971	T.V.=3.8mL
SOAs	7496	0.6	20	1963	T.V.=1.6mL
SOAs	7521	0.17	17	1958	T.V.=2.4mL
SOAs	7859	0.046	18	1966	
SOAs	7900	3.9	31	1968	T.V.=2.6mL
SOAs	7983	4.8	28	-	
SOAs	8101	3.4	27	1968	T.V.=5.6mL
SOAs	8200	0.09	6	1960	T.V.=2.0mL
SOAs	8629	<5.0	*	1968	*Rare sperm seen, 2 motile
SOAs	8697	1.30	0	1944	
SOAs	95-30	0.1	0	-	
SOAs	AM435D	2.60	47	1963	
SOAs	BT	3.70	31	1960	SNP'S
SOAs	CD	0.65	8	1965	SNP'S
SOAs	CP	1.60	10	-	SNP'S
SOAs	QR522G	0.80	38	1966	
SOAs	SF96-333	0.50	33	1966	SNP'S
SOAs	SF96-389	0.25	10	1955	
SOAs	TI	1.35	22	1951	SNP'S
SOAs	TM	3.40	23	1961	SNP'S, Cryptorchidism
SOAs	ZC	0.50	6	1957	
SOAs	ZK	3.30	40	1961	
	48				

SEVERE OLIGOZOOSPERMIA

SO	2070	3.20	54	1962	SNP'S
SO	2578	2.20	68	1950	SNP'S
SO	2583	4.70	54	1964	
SO	4406	1.20	50	1965	
SO	5364	3.81	61	-	SNP'S
SO	5453	1.41	52	1963	
SO	6562	3.50	55	1963	
SO	7381	0.9	56	1975	T.V.=3.1mL
SO	7431	0.8	61	-	T.V.=2.2mL
SO	7704	0.833	69	1973	
SO	8117	4.5	65	1970	T.V.=3.0mL
SO	8448	0.20	56	1962	T.V.=4.2mL
SO	77976	4.9	55	1966	
SO	95-15	3.10	52	1956	
SO	SF 96-1503	1.70	57	1951	
SO	SF96-470	0.03	100	1968	SNP'S
SO	SF97-2255	1.00	63	1961	
	17				

OLIGOASTHENOZOOSPERMIA					
OAs	1528	10.10	38	-	
OAs	1991	12.00	36	1937	
OAs	2050	6.80	10	1958	
OAs	2126	5.50	30	1962	
OAs	2275	8.50	7	1958	
OAs	2703	13.40	42	1960	
OAs	2764	16.00	25	1959	
OAs	2908	17.10	3	1968	
OAs	2935	10.00	41	1959	
OAs	3072	12.00	26	1949	
OAs	3089	8.70	38	1952	
OAs	3332	8.40	33	1964	
OAs	3392	5.00	30	1954	Vasectomy Reversal
OAs	3449	6.50	35	1953	
OAs	4046	6.80	22	1970	
OAs	4421	18.30	28	1963	
OAs	4859	17.50	46	1970	SNP'S
OAs	95-17	5.30	26	-	SNP'S
OAs	95-34	7.2	47	1959	SNP'S
OAs	AM235Q	9.90	32	1958	SNP'S
OAs	AT734U	13.10	29	1966	
OAs	Mr X	5.70	43	1963	
OAs	SF156	15.00	18	1948	
OAs	SF94-87	14.60	47	-	
OAs	SF96-1580	6.20	42	1968	
OAs	SF96-1667	17.00	17	1966	
OAs	SF96-283	8.40	17	1958	
OAs	SF96-434	17.00	26	1964	
	28				

OLIGOZOOSPERMIA					
O	1337	8.20	62	1955	
O	1699	10.70	67	1963	
O	1948	8.40	65	1961	
O	2307	12.30	57	1968	
O	2410	15.00	54	1959	
O	2686	14.90	61	1962	
O	2705	9.20	55	1960	SNP'S
O	2882	9.80	56	1938	
O	3208	6.90	69	1962	
O	3535	11.00	51	1967	
O	5378	16.20	57	1963	SNP'S
O	5597	16.80	63	-	
O	5611	8.00	68	-	
O	5869	5.20	83	1970	
O	95-22	6.70	57	1954	SNP'S
O	95-23	12.00	80	1962	

O	AM472M	19.80	57	1949	
O	EV	7.70	53	-	SNP'S
O	EZDW2	19.10	61	1954	SNP'S
O	JN	15.00	61	1957	SNP'S
O	OP883E	5.00	58	1967	SNP'S
O	OP921M	17.00	55	1955	
O	QN148Q	13.30	61	1962	
O	SF94-380	15.80	66	1959	
O	SF95-196	16.40	56	-	
O	SM	6.40	53	-	
O	TP	8.50	51	-	
	27				

ASTHENOZOOSPERMIA

As	1625	61.00	26	1965	
As	1644	36.00	43	1969	
As	2049	180.00	45	1958	
As	2256	204.00	41	1958	
As	2339	49.00	48	1946	SNP'S
As	2362	110.00	35	1963	SNP'S
As	2368	71.00	25	1968	
As	2534	99.00	9	1945	
As	2600	26.00	1.5	1938	
As	2609	36.00	28	1959	
As	2634	127.00	41	1965	
As	2736	85.00	11	1957	
As	2772	24.00	0.30	1960	
As	2841	114.00	48	1960	
As	2924	49.00	13	1971	SNP'S
As	3047	29.00	29	1968	
As	3138	114.00	31	1968	
As	3330	175.00	46	1959	
As	3505	27.70	49	-	SNP'S
As	4020	50.00	35	1965	
As	4057	80.00	7	1955	
As	4126	122.00	21	1963	
As	4170	27.00	28	1968	
As	4616	88.00	46	1963	
As	4652	49.00	48	1961	
As	4950	201.00	35	1966	
As	5146	22.00	11	1967	
As	5288	233.00	0	1963	SNP'S
As	5415	161.70	41	1959	
As	5554	159.00	25	1961	SNP'S
As	5621	109.90	29	1954	
As	94-110	446.00	42	1955	
As	94-149	34.00	40	1961	SNP'S
As	94-150	26.00	19	1951	
As	95-24	28.00	28	-	SNP'S

As	95-54	26.00	35	1965	
As	95-79	20.00	37	-	SNP'S
As	AM433B	49.00	36	1958	SNP'S
As	AT665C	51.00	29	1959-1969	SNP'S
As	CCR	152.00	1	-	SNP'S
As	KT	36.30	3	1967	
As	PIER1	40.00	35	-	
As	QD720I	54.00	31	-	SNP'S
As	QN230K	44.00	31	1957	
As	QN242P	105.00	48	1947	
As	QU3711	100.00	35	-	
As	RJ	43.00	2	1957	
As	RS	45.00	20	1965	SNP'S
As	SF362	24.00	43	1965	SNP'S
As	SF94-303	22.10	6	1959	
As	SF94-86	85.70	35	1979	SNP'S
As	SF95-590	80.00	35	1967	
As	SF96-166	186.00	38	1950	SNP'S
As	SF96-220	67.00	47	1958	
As	SF96-279	65.00	30	1960	SNP'S
As	WT	110.00	26	-	
	56				

NORMOZOOSPERMIC CONTROLS

N	1469	265.00	61	1961	
N	1535	33.00	75	1949	
N	1541	54.00	74	1955	
N	1738	147.00	62	1962	
N	1751	76.00	57	1962	
N	1929	63.00	77	1959	
N	1939	64.00	69	1962	
N	2016	23.00	68	1944	
N	2086	163.00	65	1962	
N	2097	55.00	54	1965	
N	2151	43.00	72	1953	
N	2223	57.00	55	1964	
N	2371	23.10	56	1954	
N	2404	92.00	55	1960	SNP'S
N	2407	95.00	54	1960	
N	2450	113.00	70	1964	
N	2520	107.00	67	1967	
N	2607	100.00	68	1968	SNP'S
N	2631	109.00	52	1960	
N	2661	39.00	82	1965	
N	2905	135.00	64	1960	SNP'S
N	2945	59.00	62	1956	SNP'S
N	3002	49.20	59	1965	
N	3099	234.00	60	1958	SNP'S
N	3182	194.00	52	1964	

N	3309	81.00	58	1965	
N	3350	91.00	59	1968	
N	3408	102.50	62	1969	SNP'S
N	3486	47.90	77	1958	
N	3489	230.90	55	-	
N	3545	30.00	78	1950	
N	3598	40.00	83	1971	
N	3725	80.00	60	1961	
N	3967	21.00	50	1958	
N	3999	348.00	66	1962	
N	4022	73.00	71	1969	
N	4510	51.30	74	1967	
N	4817	57.00	63	1973	
N	4948	104.00	77	1962	
N	5016	22.00	57	1965	
N	5222	97.00	81	1967	
N	5239	36.70	89	1968	
N	5310	121.00	78	1968	
N	5311	46.30	60	1961	
N	5477	100.00	57	1959	
N	5864	146.00	62	1960	
N	5956	33.50	75	1959	
N	6109	75.00	67	1965	
N	6122	166.00	64	1968	
N	6169	70.40	56	-	
N	6197	132.00	65	1968	
N	6353	175.00	68	1965	
N	6522	25.00	61	1965	
N	6570	75.00	79	1965	
N	6635	99.00	80	1961	
N	94-108	345.00	78	1961	
N	95-05	24.00	75	1956	SNP'S
N	95-27	79.00	69	1954	SNP'S
N	95-56	245.00	69	-	
N	AM232N	39.60	62	1949	
N	AM422U	114.00	55	1956	
N	AM447I	73.00	79	1960	
N	AM467O	290.00	52	1960	
N	AM468P	124.00	55	1958	
N	AM477R	208.00	89	1965	
N	AM498B	22.70	64	1971	SNP'S
N	AT635U	148.00	80	1957	
N	AT689M	41.00	81	1957	SNP'S
N	AT697N	25.00	73	1963	SNP'S
N	JH	20.00	87	1959	
N	MR	365.00	59	1961	
N	OP927S	135.00	69	-	SNP'S
N	QD747V	200.00	57	1962	SNP'S
N	QD788L	85.00	50	1957	SNP'S

N	QD845B	101.00	58	1956	
N	QH516G	262.00	61	1966	
N	QH529M	80.00	51	1962	
N	QN237R	136.00	68	1965	
N	QR509H	222.00	57	1966	SNP'S
N	QU447K	48.60	85	1967	
N	RL0204B	97.60	73	-	
N	RL0506B	45.30	55	-	SNP'S
N	RL1006B	178.80	70	-	
N	RL1911	41.60	90	1973	
N	RL2012	140.00	82	1959	Chemical exposure
N	RL2203	40.00	63	1967	Infection
N	SF155	129.00	70	-	
N	SF289	106.00	60	1969	
N	SF437	125.00	66	1964	
N	SF94-399	81.00	60	1947	SNP'S
N	SF94-97	86.00	72	-	SNP'S
N	SF95-122	26.00	50	1964	SNP'S
N	SF96-229	75.00	50	1959	
N	SF96-305	140.00	67	1955	
N	SF96-310	152.00	56	1955	
N	SF96-341	126.00	50	1970	Chemical exposure
N	SF96-400	62.00	53	1958	
	97				
	292				

SAMPLES EXHAUSTED DURING RESEARCH

Az	2158	0.00	0	1965	Orchidopexy
Az	5544	0.00	0	-	Obstruction
Az	QN149E	0.00	0	1964	SNP'S
SOAs	1124	1.70	35	1959	
SOAs	2696	0.40	13	1957	
SOAs	3285	0.60	21	1964	
SOAs	QR568B	0.30	20	1968	SNP'S
SOAs	SF96-309	1.10	16	1962	
SO	1398	0.15	83	1957	Cryptorchidism
SO	4397	0.50	50	1974	
SO	6518	2.00	64	1968	
SO	AM264B	0.90	58	1958	
OAs	3077	16.00	47	1959	
OAs	4191	11.80	23	1960	
OAs	4501	12.50	35	1970	
O	4970	15.00	51	-	SNP'S
O	AM302J	17.90	61	1955	
O	SF96-1386	9.40	52	1961	
As	4149	187.00	9	-	
As	AM241P	112.00	42	1960	

N	3334	30.00	62	1968	
N	3726	67.00	66	1968	
N	4658	96.00	65	-	
N	5014	70.00	59	1959	
N	AM436E	70.00	53	1959	SNP'S
N	AT440E	173.00	59	1966	
N	OP988O	143.00	83	1959	
N	SF96-201	87.00	66	1962	
N	SF96-202	78.70	64	1958	
	29				

Notes: Semen classification is shortened to AZ (Azoospermia), SOAs (Severe Oligoasthenozoospermia), SO (Severe Oligozoospermia), OAs (Oligoasthenozoospermia), O (Oligozoospermia), As (Asthenozoospermia), and N (Normozoospermia). Sample codes are colour referenced to the comments column; blue denotes individuals with mitochondrial DNA single nucleotide polymorphisms (SNP's), pink indicates there are epigenetic factors that might impact on semen quality, for example, men who have been exposed to harmful chemicals through their work are highlighted in green. SCO* refers to Sertoli Cell Only Syndrome. T.V stands for testicular volume. Sperm count is expressed in millions/ml, and motility is expressed as a percentage of that. An unknown date of birth is indicated with a dash. The thirty samples with insufficient quantity to complete all analyses are grouped at the bottom of the table.

APPENDIX D - RESEARCH DATA

AR CAG	Yq M.DEL	*POLγ*CAG	EXO I	EXO II	EXO III	SNPs	Sample Code	Semen Classification	Comments
AZOOSPERMIA									
	Yq		I			-	798	Az	
18	Yq	HOWT	I	II	III	1/4	1454	Az	
21	Yq	HOWT	I	II	III	1/4	2473	Az	
	Yq		I			-	2762	Az	
24	Yq	HOWT	I	II	III	3/4	3008	Az	
24	sY149	HOWT	I	II	III	1/3	4071	Az	Vasectomy Reversal
21	sY158	HE(10+11)	I	II	III	1/3	4449	Az	
22	Yq		I			-	5753	Az	*SCO Maturation Arrest
27	Yq		I			-	7234	Az	
	Yq	HOWT	I			-	7422	Az	T.V.=06mL
23	Yq		I			-	8055	Az	
27	Yq		I			-	8061	Az	Orchidopexy
19	Yq		I			-	8135	Az	Orchidopexy
24	Yq		I			-	8246	Az	Orchidopexy
	Yq	HOWT	I			-	8257	Az	T.V.=2.3mL, Post Chemotherapy
	Yq	HOWT	I			-	8708	Az	
	Yq	HOWT	I			-	8712	Az	T.V.=0.4mL
22	Yq	HOWT	I			1/8	AT752B	Az	
	Yq	HOWT	I			0/2	SF96-238	Az	
12	19	11	19	5	5	7	19		
SEVERE OLIGOASTHENOZOOSPERMIA									
21	Yq	HE 10+12	I			0/3	1020	SOAs	
21	Yq	HOWT	I	II	III	0/2	1028	SOAs	
28	Yq	HE 10+11	I			0/1	1336	SOAs	Vasectomy Reversal
	Yq	HE 10+11	I			0/1	1696	SOAs	
21	Yq	HE 10+11	I	II	III	0/5	2128	SOAs	
21	Yq	HOMut 11+12	I	II	III	1/3	2565	SOAs	
28	Yq	HOMut 11+12	I	II	III	0/4	2866	SOAs	
	Yq	HOWT	I			0/3	2940	SOAs	
18	Yq	HOWT	I			0/1	3261	SOAs	Count rose to 24 mill/mL over 5 yrs
24	Yq	HOWT	I			0/2	3318	SOAs	
21	Yq	HOWT	I			0/2	3993	SOAs	
19	Yq	HE 10+12	I	II	III	1/3	4110	SOAs	
21	sY149	HOWT	I	II	III	1/2	4290	SOAs	
	Yq	HE 10+11	I	II	III	1/2	4300	SOAs	
25	Yq	HOWT	I	II	III	0/2	4646	SOAs	
20	Yq	HE 10+11	I	II	III	0/2	4823	SOAs	
24	Yq	HOWT	I			1/1	5002	SOAs	Obstruction
21	Yq	HOWT	I			0/3	5080	SOAs	

22	Yq	HOWT	I			0/3	5271	SOAs	
	Yq	HOWT	I	II	III	0/2	5349	SOAs	
21	Yq	HE 10+11	I			1/1	5585	SOAs	
22	Yq	HE 10+11	I			0/2	5715	SOAs	
24	sY149	HOWT	I	II	III	-	6290	SOAs	
20	Yq	HOWT	I			0/2	6485	SOAs	
13	Yq	HOWT	I			0/2	6585	SOAs	
20	Yq	HOWT	I			0/2	6611	SOAs	
	Yq	HOWT	I			-	7284	SOAs	T.V.=3.8mL
	Yq	HOWT	I			-	7496	SOAs	T.V.=1.6mL
	Yq	HOWT	I			-	7521	SOAs	T.V.=2.4mL
	Yq	HOWT	I			-	7859	SOAs	
	Yq	HOWT	I			-	7900	SOAs	T.V.=2.6mL
	Yq	HOWT	I			-	7983	SOAs	
	Yq	HOWT	I			-	8101	SOAs	T.V.=5.6mL
	Yq	HOWT	I			-	8200	SOAs	T.V.=2.0mL
	Yq	HOWT				-	8629	SOAs	
	Yq	HOWT	I			-	8697	SOAs	
	Yq		I			0/4	95-30	SOAs	
31	Yq	HOWT	I			0/7	AM435D	SOAs	
13	sY146	HOWT	I	II	III	1/3	BT	SOAs	
	Yq	HOWT	I			1/7	CD	SOAs	
20	sY146/156	HE 10+11	I	II	III	1/5	CP	SOAs	
22	Yq	HOWT	I			0/5	QR522G	SOAs	
21	Yq	HOWT	I	II	III	2/7	SF96-333	SOAs	
19	Yq	HOWT	I			0/4	SF96-389	SOAs	
22	Yq(not 86,127)	HOWT		II	-VE PCR	2/7	TI	SOAs	
20	Yq	HOWT	I			1/7	TM	SOAs	Cryptorchidism
	sY146	HOWT	I	II	III	-	ZC	SOAs	
20	Yq	HOWT	I			0/6	ZK	SOAs	
31	48	47	46	16	15	36	48		

SEVERE OLIGOZOOSPERMIA

21	Yq			II	III	1/4	2070	SO	
24	Yq	HOWT	I	II	III	1/3	2578	SO	
19	Yq		I	II	III	0/2	2583	SO	
21	Yq	HOWT	I	II	III	-	4406	SO	
24	Yq	HOWT	I			1/3	5364	SO	
25	Yq	HOWT	I			0/3	5453	SO	
28	Yq	HOWT	I			0/2	6562	SO	
	Yq	HOWT	I			-	7381	SO	T.V.=3.1mL
	Yq	HOWT	I			-	7431	SO	T.V.=2.2mL
	Yq	HOWT	I	II	III	-	7704	SO	
	Yq	HOWT	I			-	8117	SO	T.V.=3.0mL
	Yq		I			-	8448	SO	T.V.=4.2mL
	Yq	HE 10+12	I			-	77976	SO	
26	Yq	HOWT	I			0/6	95-15	SO	
28	Yq	HOWT	I			1/2	SF96-1503	SO	
22	Yq	HOWT	I	II	III	1/3	SF96-470	SO	
24	Yq	HOWT	I			0/3	SF97-2255	SO	
11	17	14	16	6	6	10	17		

OLIGOASTHENOZOOSPERMIA

22	Yq	HOWT	I			0/1	1528	OAs	
20	Yq	HE 10+11	I			-	1991	OAs	

24	Yq	HOWT	I			0/3	2050	OAs
25	Yq	HOWT	I			0/1	2126	OAs
21	Yq	HOWT	I	II	III	-	2275	OAs
20	Yq	HOWT	I			0/1	2703	OAs
18	Yq	HOWT	I			-	2764	OAs
20	Yq	HOWT	I	II	III	-	2908	OAs
29	Yq	HE 8+10	I	II	III	0/1	2935	OAs
20	Yq	HOWT	I	II	III	0/3	3072	OAs
20	Yq	HOWT	I			0/3	3089	OAs
22	Yq	HOWT	I			-	3332	OAs
20	Yq	HOWT	I			-	3392	OAs Vasectomy Reversal
22	Yq	HOMut 8+11	I	II	III	-	3449	OAs
23	Yq	HE 10+11	I			-	4046	OAs
22	Yq	HOWT	I			0/3	4421	OAs
21	Yq	HOWT	I			1/3	4859	OAs
	Yq	HOWT	I	II	III	2/5	95-17	OAs
	Yq	HOWT	I	II	III	2/4	95-34	OAs
31	Yq	HOWT	I	II	III	2/6	AM235Q	OAs
18	Yq	HOWT	I	II	III	0/4	AT734U	OAs
20	Yq	HOMut 8+11	I	II	III	0/1	Mr X	OAs
20	Yq	HOWT	I			0/4	SF156	OAs
31	Yq	HOWT	I			0/3	SF94-87	OAs
22	Yq	HOWT	I			0/1	SF96-1580	OAs
18	Yq	HE 10+11	I	II	III	0/1	SF96-1667	OAs
20	sY156	HOWT	I			0/3	SF96-283	OAs
23	Yq	HOWT	I			0/4	SF96-434	OAs
26	28	28	28	11	11	20	28	

OLIGOZOOSPERMIA

23	Yq	HE 10+11	I			0/1	1337	O
26	Yq	HE 10+11	I	II	III	0/3	1699	O
18	Yq	HOWT	I			-	1948	O
18	Yq	HOWT	I			0/3	2307	O
21	Yq	HOWT	I	II	III	0/5	2410	O
24	Yq	HOWT	I	II	III	0/1	2686	O
21	Yq	HOWT	I			1/1	2705	O
21	Yq	HE 10+11	I			0/1	2882	O
22	Yq	HOWT	I			0/1	3208	O
24	Yq	HOWT	I			0/2	3535	O
	Yq	HE 10+11	I	II	III	1/3	5378	O
28	Yq	HOMut 11+12	I	II	III	0/1	5597	O
23	Yq	HE 10+11	I			0/1	5611	O
26	Yq	HOWT	I			0/2	5869	O
21	Yq	HOWT	I			1/4	95-22	O
22	Yq	HOWT	I			0/3	95-23	O
18	Yq	HOWT	I			0/3	AM472M	O
	Yq	HOWT	I	II	III	1/2	EV	O
20	Yq	HE 10+12	I			1/8	EZDW2	O
18	Yq	HOWT	I	II	III	2/6	JN	O
	Yq	HOWT	I	II	III	2/3	OP883E	O
19	Yq	HOWT	I			0/2	OP921M	O
21	Yq	HOWT	I			0/4	QN148Q	O
22	sY149	HOWT	I	II	III	0/5	SF94-380	O
19	Yq	HE 10+11	I			0/3	SF95-196	O

23	Yq	HOWT	I			0/1	SM	O
20	Yq	HE 10+11	I			0/3	TP	O
24	27	27	27	9	9	26	27	

ASTHENOZOOSPERMIA

20	Yq	HOWT	I			0/5	1625	As
19	Yq	HE 10+11	I	II	III	-	1644	As
22	Yq	HOWT	I			0/2	2049	As
25	Yq	HE 10+11	I			0/3	2256	As
19	Yq	HE 10+11	I			1/1	2339	As
21	Yq	HOWT	I			1/3	2362	As
20	Yq	HOWT	I			0/3	2368	As
22	Yq	HE 10+11	I			0/2	2534	As
20	Yq	HE 10+11	I			-	2600	As
14	Yq	HOWT	I			0/3	2609	As
22	Yq	HE 10+11	I			0/4	2634	As
23	Yq	HE 10+12	I			0/3	2736	As
8	Yq	HOWT	I			0/3	2772	As
24	Yq	HOWT	I			0/4	2841	As
27	Yq	HOWT	I			1/3	2924	As
22	sY149		I	II	III	0/3	3047	As
27	Yq	HE 10+11	I	II	III	0/2	3138	As
25	Yq		I			0/1	3330	As
	Yq	HOWT	I	II	III	1/3	3505	As
20	Yq	HOWT	I			-	4020	As
20	Yq	HOWT	I			-	4057	As
25	Yq	HOWT	I			0/2	4126	As
23	Yq	HOWT	I			-	4170	As
24	Yq	HOWT	I			-	4616	As
23	Yq	HOWT	I			0/2	4652	As
21	Yq	HOWT	I			-	4950	As
22	Yq	HOWT	I			0/2	5146	As
	Yq	HOWT	I	II	III	1/2	5288	As
21	Yq	HOWT	I			0/2	5415	As
18	Yq	HOWT	I			2/3	5554	As
20	Yq	HOMut 11+12	I			-	5621	As
24	Yq	HOWT	I	II	III	0/5	94-110	As
24	Yq	HE 10+11	I	II	III	2/5	94-149	As
27	Yq	HOWT	I			0/8	94-150	As
28	Yq	HE 10+12	I			1/7	95-24	As
23	Yq	HOWT	I			0/2	95-54	As
18	Yq	HE 10+11	I			1/3	95-79	As
19	Yq	HOWT	I	II	III	2/5	AM433B	As
18	Yq	HOWT	I	II	III	2/8	AT665C	As
21	Yq	HOWT	I	II	III	2/8	CCR	As
23	Yq	HOWT	I			0/5	KT	As
22	Yq	HE 9+10	I			0/4	PIER1	As
21	Yq			II	III	2/7	QD720I	As
26	Yq	HOWT	I			0/5	QN230K	As
23	Yq	HE 10+11	I			0/4	QN242P	As
23	Yq	HOWT	I			1/6	QU3711	As
21	Yq	HOWT	I	II	III	0/1	RJ	As
18	Yq	HE 10+11	I	II	-VE PCR	1/5	RS	As
19	Yq	HOWT	I			1/4	SF362	As
	Yq	HOMut 11+12	I	II	III	0/2	SF94-86	As

17	Yq	HOWT	I	II	III	2/7	SF94-303	As
22	Yq	HOWT	I	II	III	0/2	SF95-590	As
25	Yq	HOWT	I	II	III	1/7	SF96-166	As
25	Yq	HE 10+12	I			0/3	SF96-220	As
28	Yq	HOWT	I			1/6	SF96-279	As
24	sY156	HE 10+12	I	II	III	0/4	WT	As
53	56	53	55	18	18	48	56	

NORMOZOOSPERMIA

24	Yq	HE 10+11	I	II	III	0/3	1469	N
21	Yq	HE 10+11	I	II	III	0/3	1535	N
25	Yq	HE 9+10	I	II	III	0/4	1541	N
21	Yq	HOWT	I	II	III	0/4	1738	N
21	Yq	HOWT	I	II	III	-	1751	N
20	Yq	HOWT	I			0/3	1929	N
24	Yq	HOWT	I			0/4	1939	N
24	Yq	HE 10+11	I	II	III	0/2	2016	N
	Yq	HOWT	I			0/1	2086	N
25	Yq	HOWT	I	II	III	-	2097	N
24	Yq	HOWT	I			-	2151	N
26	Yq	HOWT	I	II	III	0/2	2223	N
23	Yq	HOWT	I			0/2	2371	N
21	Yq	HE 10+11	I	II	III	1/2	2404	N
20	Yq	HOWT	I			0/2	2407	N
15	Yq	HOWT	I	II	III	0/4	2450	N
24	Yq		I	II	III	0/2	2520	N
	Yq	HE 10+11	I	II	III	1/3	2607	N
20	Yq	HOWT	I	II	III	0/4	2631	N
24	Yq	HE 10+11	I			-	2661	N
20	Yq	HE 10+11	I	II	III	1/5	2905	N
	Yq	HOWT	I	II	III	2/3	2945	N
20	Yq	HOWT	I	II	III	0/1	3002	N
22	Yq	HOWT	I			1/1	3099	N
23	Yq	HOWT	I	II	III	1/1	3182	N
18	Yq	HE 10+11	I	II	-VE PCR	0/1	3309	N
22	Yq	HOWT	I			0/3	3350	N
23	Yq	HOWT	I	II	III	1/3	3408	N
20	Yq	HOWT	I			0/3	3486	N
20	Yq	HOWT	I			0/1	3489	N
24	Yq	HOWT	I			-	3545	N
19	Yq	HOWT	I	II	III	-	3598	N
19	Yq	HOWT	I	II	III	0/2	3725	N
9	Yq	HE 10+11	I	II	III	-	3967	N
20	Yq	HOMut 11+11	I	II	III	-	3999	N
24	Yq	HOWT	I	II	III	0/1	4022	N
	Yq	HE 10+11	I	II	III	0/3	4510	N
20	Yq	HE 10+12	I	II	III	0/3	4817	N
25	Yq	HOWT	I			0/3	4948	N
20	Yq	HE 10+12	I	II	III	0/2	5016	N
19	Yq	HOWT	I	II	III	0/3	5222	N
21	Yq	HOWT	I	II	III	0/2	5239	N
25	Yq	HE 10+11	I			0/1	5310	N
24	Yq	HOWT	I			1/3	5311	N
38	Yq	HOWT	I	II	III	-	5477	N
24	Yq	HE 10+11	I			0/2	5864	N
26	Yq	HE 10+11	I			-	5956	N

23	Yq	HE 10+11	I	II	III	0/2	6109	N
20	Yq	HE 10+11	I			0/3	6122	N
20	Yq	HE 10+11	I	II	III	0/2	6169	N
23	Yq	HE 10+11	I			0/2	6197	N
19	Yq	HE 10+11				0/3	6353	N
22	Yq	HOWT	I			0/2	6522	N
20	Yq	HOWT				0/3	6570	N
20	Yq	HE 10+11	I			0/3	6635	N
20	Yq	HOWT	I			0/4	94-108	N
	Yq	HOWT	I	II	III	2/4	95-05	N
22	Yq	HE 10+11	I	II	III	3/7	95-27	N
19	Yq	HOWT	I			0/4	95-56	N
18	Yq	HOWT	I			0/5	AM232N	N
22	Yq	HOWT	I	II	III	0/5	AM422U	N
21	Yq	HE 10+11	I	II	III	0/4	AM447I	N
21	Yq	HE 10+11	I	II	III	0/5	AM467O	N
17	Yq		I	II	III	0/5	AM468P	N
25	Yq	HE 10+11				0/5	AM477R	N
21	Yq	HOWT	I	II	III	1/8	AM498B	N
21	Yq	HOWT	I			0/5	AT635U	N
26	Yq	HOWT	I	II	III	2/8	AT689M	N
24	Yq	HE 10+11	I	II	III	1/8	AT697N	N
25	Yq					0/4	JH	N
20	Yq	HOWT	I			0/3	MR	N
	Yq	HOWT	I	II	III	1/2	OP927S	N
22	Yq	HOWT	I	II	III	2/2	QD747V	N
23	Yq	HOWT		II	III	1/4	QD788L	N
26	Yq	HE 10+11	I	II	III	0/5	QD845B	N
24	Yq	HE 10+11		II	III	0/3	QH516G	N
	Yq			II	III	0/2	QH529M	N
	Yq	HE 10+12	I	II	III	0/4	QN237R	N
20	Yq	HOWT		II	III	2/5	QR509H	N
19	Yq	HOMut 11+12		II	III	0/2	QU447K	N
21	Yq	HOWT				0/3	RL0204B	N
20	Yq	HOWT	I			1/3	RL0506B	N
20	Yq	HOWT	I			0/3	RL1006B	N
23	Yq	HE 10+11	I	II	III	-	RL1911	N
22	Yq	HOWT				0/3	RL2012	N
20	Yq		I			0/3	RL2203	N
20	Yq	HOWT	I	II	III	0/3	SF155	N
21	Yq	HOWT	I	II	III	-	SF289	N
23	Yq	HOWT	I	II	III	0/1	SF437	N
24	Yq	HOWT		II	III	1/4	SF94-97	N
21	Yq	HOWT	I	II	III	1/4	SF94-399	N
25	Yq	HOMut 9+11	I	II	III	2/5	SF95-122	N
18	Yq	HOMut 11+11	I	II	III	0/4	SF96-229	N
14	Yq	HOWT		II	III	2/7	SF96-305	N
20	Yq	HE 10+12	I	II	III	0/2	SF96-310	N
21	Yq	HOWT		II	III	0/4	SF96-341	N
21	Yq	HOWT	I			0/4	SF96-400	N
89	97	92	83	61	60	84	97	
246	292	272	274	126	124	231	292	

Chemical exposure
Infection
Chemical exposure

SAMPLES EXHAUSTED DURING RESEARCH

AR CAG	Yq M.DEL	*POLγ*CAG	EXO I	EXO II	EXO III	SNPs	Sample Code	Semen Classification	Comments
19						-	2158	Az	Orchidopexy
20	sY149*					-	5544	Az	Obstruction
21						1/5	QN149E	Az	
16	Yq*					0/1	1124	SOAs	
	sY156*					0/3	2696	SOAs	
21	Yq*					0/1	3285	SOAs	
18						1/3	QR568B	SOAs	
27	Yq*					0/4	SF96-309	SOAs	
	Yq					0/1	1398	SO	Cryptorchidism
22	Yq*	HOWT				0/1	4397	SO	
23	Yq*					-	6518	SO	
21	Yq*					0/3	AM264B	SO	
25	Yq*	HE 10+11				0/1	3077	OAs	
20	Yq*					-	4191	OAs	
21	Yq*					0/3	4501	OAs	
20	Yq*					1/1	4970	O	
	sY158*					0/2	AM302J	O	
21						0/1	SF96-1386	O	
27	Yq*					0/2	4149	As	
23	Yq*					0/3	AM241P	As	
20	Yq*					0/1	3334	N	
20	Yq*	HOWT				0/2	3726	N	
23	Yq*					1/2	4658	N	
19	Yq*					1/1	5014	N	
23	Yq*					2/5	AM436E	N	
22	Yq*					0/5	AT440E	N	
19						0/2	OP988O	N	
24	Yq*					0/3	SF96-201	N	
25						0/4	SF96-202	N	
26	22	3	0	0	0	25	29		

Notes: Blanks indicate that that factor wasn't investigated in the sample. The *AR*-CAG column displays numbers of Androgen Receptor trinucleotide repeats. A "Yq" in the Y micro-deletion column denotes no deletions were found of any of the 14 STS markers analysed. An STS (in red) noted in the column indicates a deletion. An asterisk (*) signifies only partial screening; only the nine AZFc and the SRY positive STS's. "HOWT" in the *POLγ* CAG column indicates a homozygous wild-type individual (two copies of the *POLγ* 10-unit allele). "He" refers to heterozygote individuals (one 10-unit and one 'other' allele), and "HOMut" denotes a homozygous mutant individual with two 'other' alleles (CAG-repeat number is given under the phenotype). The "Exo" columns indicates what exonuclease domains were sequenced. The SNP column indicates the number of mtDNA genes analysed previously, and how many of them have polymorphisms. Semen classification is shortened as per Appendix C. The sample codes are colour referenced to the comments column to indicate other epigenetic factors that might account for the semen quality.

APPENDIX E - *POLγ* EXONUCLEASE DOMAIN SAMPLES

*POLγ*CAG	*POLγ*EXO	Sample Code	mtDNA Mutation*	Semen Classification	Count (10^6/ml)	Motility (%)
HOWT	I, II, III	1454	ND1/4	Az	0.00	0
HOWT	I, II, III	2473	ND6/4	Az	0.00	0
HOWT	I, II, III	3008	ND1, 2, 3/4	Az	0.00	0
HOWT	I, II, III	4071	ND1/3	Az	0.00	0
HE 10+11	I, II, III	4449	ND6/3	Az	0.00	0
HOMut 11+12	I, II, III	2565	ND5/3	SOAs	4.30	20
HE 10+12	I, II, III	4110	ND1/3	SOAs	1.50	31
HOWT	I, II, III	4290	ND1/2	SOAs	0.28	18
HE 10+11	I, II, III	4300	ND1/2	SOAs	4.00	45
HOWT	I, II, III	BT	COII/3	SOAs	3.70	31
HE 10+11	I, II, III	CP	ND5/5	SOAs	1.60	10
HOWT	I, II, III	SF96-333	ND6, COIII/7	SOAs	0.50	33
HOWT	II	TI	ND1, ATPase8/7	SOAs	1.35	22
	II, III	2070	ND1/4	SO	3.20	54
HOWT	I, II, III	2578	ND5/3	SO	2.20	68
HOWT	I, II, III	SF96-470	ND1/3	SO	0.03	100
HOWT	I, II, III	95-17	ATPase6, 8/5	OAs	5.30	26
HOWT	I, II, III	95-34	ATPase6, 8/4	OAs	7.2	47
HOWT	I, II, III	AM235Q	ATPase6, COIII/6	OAs	9.90	32
HE 10+11	I, II, III	5378	ND1/3	O	16.20	57
HOWT	I, II, III	EV	ATPase6/2	O	7.70	53
HOWT	I, II, III	JN	ND1,COII/6	O	15.00	61
HOWT	I, II, III	OP883E	ND5, ATPase6/3	O	5.00	58
HOWT	I, II, III	3505	ND1/3	As	27.70	49
HOWT	I, II, III	5288	ND1/2	As	233.00	0
HE 10+11	I, II, III	94-149	ATPase6, 8/5	As	34.00	40
HOWT	I, II, III	AM433B	ND1, COII/4	As	49.00	36
HOWT	I, II, III	AT665C	ND2, COII/8	As	51.00	29
HOWT	I, II, III	CCR	ND6, COII/8	As	152.00	1
	II, III	QD720I	ATPase8, COII/7	As	54.00	31
HE 10+11	I, II	RS	ND1/5	As	45.00	20
HOMut 11+12	I, II, III	SF94-86	ND1, ATPase8/7	As	85.70	35
HOWT	I, II, III	SF96-166	ND1/7	As	186.00	38
HE 10+11	I, II, III	2404	ND1/2	N	92.00	55
HE 10+11	I, II, III	2607	ND1/3	N	100.00	68
HE 10+11	I, II, III	2905	ND1/5	N	135.00	64
HOWT	I, II, III	2945	ND1,6/3	N	59.00	62
HOWT	I, II, III	3408	ND1,3	N	102.50	62
HOWT	I, II, III	95-05	ATPase6, 8/4	N	24.00	75
HE 10+11	I, II, III	95-27	ND1, 6, COII/7	N	79.00	69
HOWT	I, II, III	AM498B	COIII/8	N	22.70	64
HOWT	I, II, III	AT689M	ND1, COIII/8	N	41.00	81
HE 10+11	I, II, III	AT697N	ATPase6/8	N	25.00	73
HOWT	I, II, III	OP927S	ND1/2	N	135.00	69

HOWT	I, II, III	QD747V	ATPase8, COII/2	N	200.00	57
HOWT	II, III	QD788L	COII/4	N	85.00	50
HOWT	II, III	QR509H	ND1, COII/5	N	222.00	57
HOWT	II, III	SF94-97	ND1/4	N	86.00	72
HOWT	I, II, III	SF94-399	ATPase6/4	N	81.00	60
HOMut 9+11	I, II, III	SF95-122	ND1, ATPase6, COII/7	N	26.00	50
HOWT	II, III	SF96-305	ND1, ATPase6/7	N	140.00	67
		51				

*POLγ*CAG	*POLγ*EXO	Sample Code	mtDNA Mutation	Semen Classification	Count (10^6/ml)	Motility (%)
HOWT	I, II, III	1028	0/2	SOAs	0.15	0
HE 10+11	I, II, III	2128	0/5	SOAs	0.30	10
HOMut 11+12	I, II, III	2866	0/4	SOAs	0.70	29
HOWT	I, II, III	4646	0/2	SOAs	2.30	29
HE 10+11	I, II, III	4823	0/2	SOAs	0.80	9
HOWT	I, II, III	5349	0/2	SOAs	1.90	5
	I, II, III	2583	0/2	SO	4.70	54
HE 8+10	I, II, III	2935	0/1	OAs	10.00	41
HOWT	I, II, III	3072	0/3	OAs	12.00	26
HOWT	I, II, III	AT734U	0/4	OAs	13.10	29
HE 10+11	I, II, III	SF96-1667	0/1	OAs	17.00	17
HE 10+11	I, II, III	1699	0/3	O	10.70	67
HOWT	I, II, III	2410	0/5	O	15.00	54
HOWT	I, II, III	2686	0/1	O	14.90	61
HOMut 11+12	I, II, III	5597	0/1	O	16.80	63
HOWT	I, II, III	SF94-380	0/5	O	15.80	66
	I, II, III	3047	0/3	As	29.00	29
HE 10+11	I, II, III	3138	0/2	As	114.00	31
HOWT	I, II, III	94-110	0/5	As	446.00	42
HOWT	I, II, III	RJ	0/1	As	43.00	2
HOWT	I, II, III	SF94-303	0/2	As	22.10	6
HOWT	I, II, III	SF95-590	0/2	As	80.00	35
HE 10+12	I, II, III	WT	0/4	As	110.00	26
HE 10+11	I, II, III	1469	0/3	N	265.00	61
HE 10+11	I, II, III	1535	0/3	N	33.00	75
HE 9+10	I, II, III	1541	0/4	N	54.00	74
HOWT	I, II, III	1738	0/4	N	147.00	62
HOWT	I, II, III	1751	0/1	N	76.00	57
HE 10+11	I, II, III	2016	0/2	N	23.00	68
HOWT	I, II, III	2097	0/2	N	55.00	54
HOWT	I, II, III	2223	0/2	N	57.00	55
HOWT	I, II, III	2450	0/4	N	113.00	70
	I, II, III	2520	0/2	N	107.00	67
HOWT	I, II, III	2631	0/4	N	109.00	52
HOWT	I, II, III	3002	0/1	N	49.20	59
HOWT	I, II, III	3182	0/1	N	194.00	52
HE 10+11	I, II	3309	0/1	N	81.00	58
HOWT	I, II, III	3725	0/2	N	80.00	60

HOWT	I, II, III	4022	0/1	N	73.00	71
HE 10+11	I, II, III	4510	0/3	N	51.30	74
HE 10+12	I, II, III	4817	0/3	N	57.00	63
HE 10+12	I, II, III	5016	0/2	N	22.00	57
HOWT	I, II, III	5222	0/3	N	97.00	81
HOWT	I, II, III	5239	0/2	N	36.70	89
HE 10+11	I, II, III	6109	0/2	N	75.00	67
HE 10+11	I, II, III	6169	0/2	N	70.40	56
HOWT	I, II, III	AM422U	0/5	N	114.00	55
HE 10+11	I, II, III	AM447I	0/4	N	73.00	79
HE 10+11	I, II, III	AM467O	0/5	N	290.00	52
	I, II, III	AM468P	0/5	N	124.00	55
HE 10+11	I, II, III	QD845B	0/5	N	101.00	58
HE 10+11	II, III	QH516G	0/3	N	262.00	61
	II, III	QH529M	0/2	N	80.00	51
HE 10+12	I, II, III	QN237R	0/4	N	136.00	68
HOMut 11+12	II, III	QU447K	0/2	N	48.60	85
HOWT	I, II, III	SF155	0/3	N	129.00	70
HOWT	I, II, III	SF437	0/1	N	125.00	66
HOMut 11+11	I, II, III	SF96-229	0/4	N	75.00	50
HE 10+12	I, II, III	SF96-310	0/2	N	152.00	56
HOWT	II, III	SF96-341	0/4	N	126.00	50
HOWT	I, II, III	6290		SOAs	0.01	36
HOWT	I, II, III	ZC		SOAs	0.50	6
HOWT	I, II, III	4406		SO	1.20	50
HOWT	I, II, III	7704		SO	0.833	69
HOWT	I, II, III	2275		OAs	8.50	7
HOWT	I, II, III	2908		OAs	17.10	3
HOMut 8+11	I, II, III	3449		OAs	6.50	35
HOMut 8+11	I, II, III	Mr X		OAs	5.70	43
HE 10+11	I, II, III	1644		As	36.00	43
HOWT	I, II, III	3598		N	40.00	83
HE 10+11	I, II, III	3967		N	21.00	50
HOMut 11+11	I, II, III	3999		N	348.00	66
HOWT	I, II, III	5477		N	100.00	57
HE 10+11	I, II, III	RL1911		N	41.60	90
HOWT	I, II, III	SF289		N	106.00	60
		75				

Notes: The Table lists samples previously screened for mtDNA mutation. As per Appendix D the *POLγ* CAG results and exonuclease domains sequenced are listed against each sample; a blank indicates the sequencing was repeatedly not successful or not attempted. The mtDNA mutation column indicates which genes SNP's were identified in out of how many mitochondrial genes were screened, for example, ND1/4 indicates there was a polymorphism identified in ND1 of the four mtDNA genes screened, while 0/3 denotes no mutation found in any of the three mtDNA genes sequenced.

APPENDIX F - RECIPES

5% PAGE - Visualisation of Yq Mulitplex PCR Groups 3 & 4

Two glass plates (10cm x 7cm (short plate) and 10cm x 8cm (long plate)) should be cleaned with 96% ethanol and the gel apparatus assembled in the casting tray before mixing the following in a small (100ml) volumetric flask:

1.66 ml	30% acrylamide:bisacrylamide (29:1)
1.0 ml	10x TBE
7.27 ml	ddH_2O
0.07 ml	10% ammonium persulfate (APS)
0.035 ml	TEMED

The PAGE solution was swirled several times to mix and then pipetted between the plates using a Pasteur pipette until full, before adding the 15-well 1.0mm comb. The gel was left for 40 min to polymerise before sample loading. A 1x TBE buffer (300ml) was used for electrophoresis.

Pre-hybridisation Solution for Southern Blot Hybridisation.

The final amount of solution required is $0.1ml/cm^2$ of membrane, but make excess. For 30ml of pre-hybridisation solution mix the following in a 100ml volumetric flask:

9.0 ml	20x SSC
18.0 ml	ddH^2O

Heat at 65°C (along with the frozen Denhardts solution) 20 min, and then add:

1.5 ml	Denhardts
1.5 ml	10% SDS
0.15 g	blocking reagent

Stir on a moderate heat for 2 hrs.

6% Polyacrylamide Gels – for DNA Sequencing

In a 250 ml beaker measure:

35g	UREA
7.0 ml	10x TBE
10.5 ml	40% acrylamide:bisacrylamide (19:1)

Make up to 60 mls with ddH_2O, and dissolve with moderate heat and stirring.
When completely dissolved make up to 70mls and pour into a vacuum flask for de-gassing 10 min.

Thoroughly clean one long (42 x 34) and one short (40 x 34) glass plate with water and finally 96% ethanol. The short plate was siliconised with Sigmacote® (Sigma Chemical Co.) so that the gel stays on the long plate when separated after electrophoresis. Plates are placed together separated by 0.4mm spacers and clamped tight around three sides excluding the top.

Add 700μl of 10% ammonium persulfate (APS) and 14μ TEMED to the flask, swirl and pour into a plastic measuring cylinder for gel pouring. The acrylamide solution was poured between the plates raised on an angle, and left to polymerise for 40 min. Excess acrylamide and UREA should be cleaned off the outside and edges of the plates before assembling in the gel rig for electrophoresis.

4% Polyacrylamide gels - LICOR Fluorescently-tagged DNA Sequencing.

In a 250 ml beaker measure:

33.6g	UREA
8.0 ml	10x TBE
6.4 ml	40% acrylamide:bisacrylamide (19:1)

Make up to 70 mls with ddH_2O, and dissolve with moderate heat and stirring.
When completely dissolved make up to 80mls and pour into a vacuum flask for de-gassing 10 min.

Thoroughly clean the LICOR glass plates with water and 96% ethanol. Assemble the plates, spacers and clamps, and tighten.

After de-gas add 500μl of 10% APS and 40μl TEMED, swirl to mix, and transfer to a 100ml measuring cylinder to pour between tilted plates. Remove all bubbles and insert well comb. Leave to polymerise for 40 min. Thoroughly clean the outside of both plates, paying particular attention to the bottom 10cm where the laser runs across. Prepare 500ml of 1x TBE buffer.

APPENDIX G – NUCLEOTIDE SEQUENCES

Androgen Receptor nucleotide sequence indicating the CAG trinucleotide repeat region, PCR primers, and product.

991 gcttcccgca ggtgggcagc tagctgcagc gactaccgca tcatcacagc ctgttgaact

cttctgagca agagaagggg aggcggggta agggaagtag gtggaagatt cagccaagct

caaggatgga agtgcagtta gggctgggaa gggtctaccc tcggccgccg tccaagacct

ANDRO RF1 →

accgaggagc tttccagaat ctgttccaga gcgtgcgcga agtgatccag aacccgggcc

ccaggcaccc agaggccgcg agcgcagcac ctcccggcgc cagtttgctg ctgctgcagc

agcagcagca gcagcagcag cagcagcagc agcagcagca gcagcagcag cagcagcagc

agcaagagac tagccccagg cagcagcagc agcagcaggg tgaggatggt tctccccaag

cccatcgtag aggccccaca ggctacctgg tcctggatga ggaacagcaa ccttcacagc

← ANDRO RR2

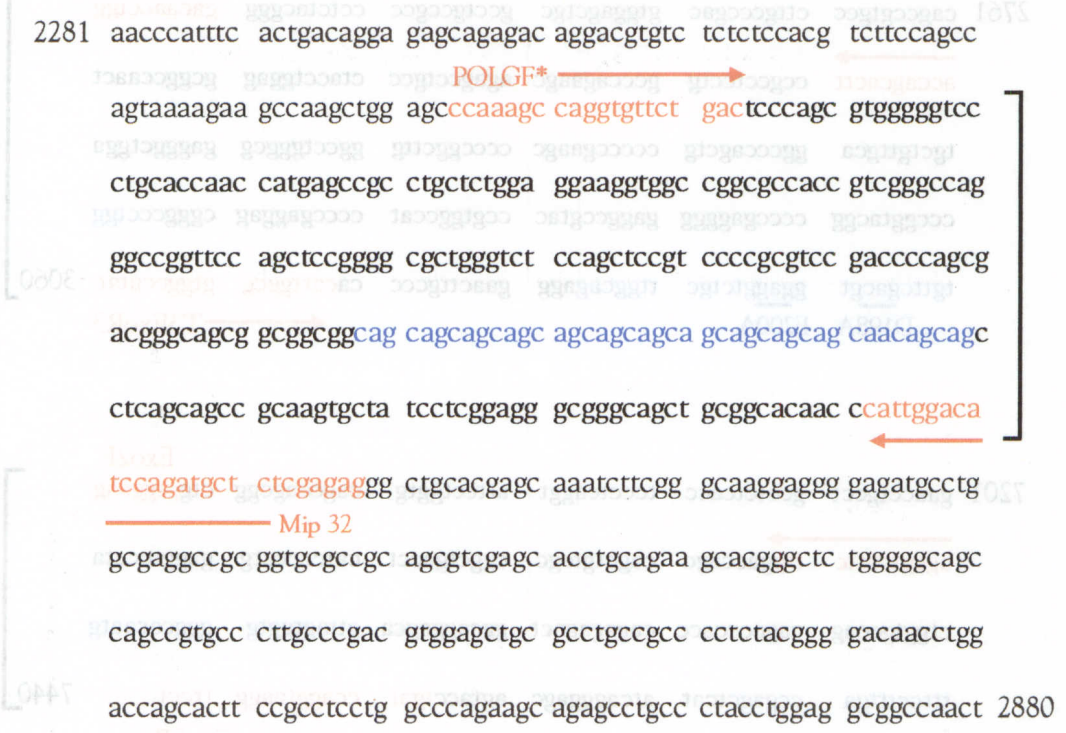
cgcagtcggc cctggagtgc caccccgaga gaggttgcgt cccagagcct ggagccgccg

tggccgccag caaggggctg ccgcagcagc tgccagcacc tccggacgag gatgactcag

ctgccccatc cacgttgtcc ctgctgggcc ccactttccc cggcttaagc agctgctccg 1650

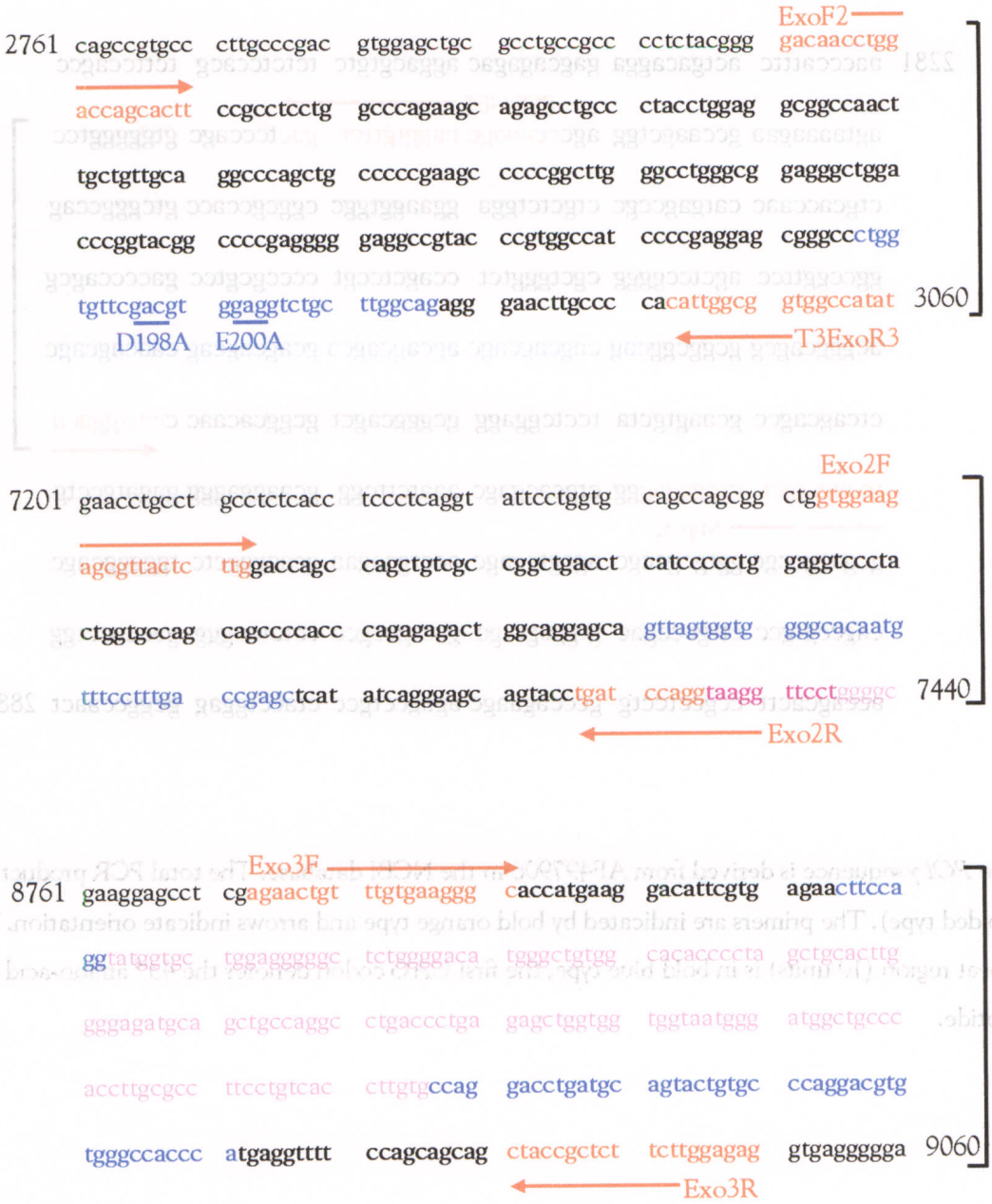
Total PCR product size (bolded nucleotide bases, including primers) is 288bp. The forward and reverse primers (Andro RF1 and Andro RR2, respectively) are indicated in orange, and the CAG trinucleotide repeat region in blue. The sequence is encoded in the first exon within the transactivation domain of the androgen receptor gene.

POLγ nucleotide sequence indicating CAG repeat and PCR primers.



The *POLγ* sequence is derived from AF497906 in the NCBI database. The total PCR product is 295bp (bolded type). The primers are indicated by bold orange type and arrows indicate orientation. The CAG repeat region (10 units) is in bold blue type, the first CAG codon denotes the 43rd amino-acid of the peptide.

POLγ nucleotide sequence indicating exonuclease motifs I, II, and III, D198A and E200A substitutions, and PCR primers.



The *POLγ* sequence is derived from AF497906 in the NCBI database. The exonuclease domains are marked in bold blue type and PCR primers in bold orange. The reverse PCR primer, Exo2R (motif II) extends into an intron (bold pink) and exonuclease motif III is interrupted by a 145bp intron (light pink). The amino-acid substitutions D198A and E200A that abolish proof-reading are indicated beneath the nucleotide sequence in exonuclease motif I. The PCR products generated are 250bp, 182bp, and 278bp in length respectively.

APPENDIX H – STATISTICAL CALCULATIONS

Table 5.5. *POLγ*CAG repeat variation contingency table.

		NormoZS	Non normoZS	Total
HOWT	obs	57	134	191
	exp	64.6	126.4	
HE	obs	33	41	74
	exp	25.0	48.98	
HOMut	obs	3	7	10
	exp	3.38	6.6	
Total		93	182	275

Obs is the observed number of individuals with the particular genotype, and exp is the expected value generated by multiplying the fraction of the total population with that genotype (row total/total population), by the total number in that group (column total) to produce the expected number in that sample if it came from the overall population (for example, exp for normozs HOWT is 191/275 x 93 = 64.6)

$\alpha=0.05$, H_O= Genotype is independent of semen quality.
H_A=Genotype is related to semen quality.

$G = 4.60517[\Sigma\Sigma fij\log fij - \Sigma Ri\log Ri - \Sigma Cj\log Cj + n\log n]$

$G = 4.60517\,[57(1.7559) + 33(1.5185) + 3(0.4771) + 134(2.1271) + 41(1.6128) + 7(0.8451) - 93(1.9685) - 182(2.2601) - 191(2.2810) - 74(1.8692) - 10(1) + 275(2.4393)]$

$G = 4.60517\,[100.085 + 50.111 + 1.431 + 285.032 + 66.124 + 5.916 - 183.069 - 411.333 - 435.677 - 138.323 - 10 + 670.816]$

$G = 4.60517[1.113]$

$G = 5.126$ with $df = (r-1)(c-1) = 2$

Williams' correction for a R x C table is: $q = 1+\frac{[275(1/191 + 1/74 + 1/10) - 1][275(1/93 + 1/182) - 1]}{6(275)(2)(1)}$

$q = 1 + \frac{[31.656][3.468]}{3300}$

$q = 1.033$

$G_{adj} = G/q$

$G_{adj} = 5.126/1.033$

$G_{adj} = 4.962$

Since $\chi^2_{0.05,2}=5.991 \geq G_{adj}=4.962$ we accept H_O that genotype is independent of semen quality (Sokal and Rohlf 1995; Zar 1996).

Acknowledgements

At the end of this research I find I can look back and truthfully say I've enjoyed it. During the past three and a bit years I have discovered a passion for fertility research that has given me a career direction and the drive to be involved at the cutting-edge.

Frank, I have thoroughly enjoyed working with you, it is so much fun to have a conversation with someone who is equally as passionate about fertility research. Your enthusiasm is infectious, and your experience and guidance kept me on the right track and slowed me down when I needed it, thank you. I'll also fondly remember your generosity and those fantastic wine and cheese evenings.

Neil, thank you for challenging me to do the best research I could. I really appreciate the time you gave to me when you were rather stretched for spare minutes; your phenomenal memory for relevant references is a frequent source of amazement to me. Thank you also for what must have been a rather good reference.

On a less geeky note, to my family, I love the way you think I'm special when sometimes I really feel just one of the bunch; it spurs me to do the best I can. Your support and love grounds me and it is a continual source of happiness for me to spend time and laugh with you. It must truly be love to have sat through rather one-sided conversations about the latest fertility research with me talking a mile-a-minute. To my new family, it has been so wonderful to have been engulfed by a group of people that care about how I'm doing, are interested enough to ask, your caring has allowed me to concentrate on writing whenever inspiration hit, and your friendship has given me plenty of happy time-out moments, thank you.

Lastly, Dale, I love the way you bought me coffee and biscuits when I was writing, and comforted me through my stress-attacks and tired bad-moods, you know me so well, I love you.

ABSENCE OF ASSOCIATION OF ANDROGEN RECEPTOR TRINUCLEOTIDE EXPANSION AND POOR SEMEN QUALITY

Tanya Erasmuson, Iris L. Sin* and Frank Y.T. Sin

Department of Zoology, University of Canterbury, Christchurch, New Zealand.

* New Zealand Centre for Reproductive Medicine, Hiatt Chambers, New Zealand, and Department of Obstetrics and Gynaecology, University of Otago, Dunedin, New Zealand

Keywords: CAG repeat, Androgen receptor, semen quality, male infertility

Correspondence: Frank Y.T. Sin, Department of Zoology, University of Canterbury, Private Bag 4800, Christchurch 1, New Zealand.

E-mail: f.sin@canterbury.ac.nz

ABSTRACT:

This study investigated the relationship between variation in the polymorphic CAG trinucleotide repeat (TNR) region of the human androgen receptor (AR) gene and semen quality in a Caucasian sample population. These men were patients attending the New Zealand Centre for Reproductive Medicine in Christchurch. The AR TNR region was amplified by PCR and then DNA sequenced to determine exact numbers of CAG repeats for each sample. In addition, the samples were screened for microdeletions within the AZFc region of the Y-chromosome. A total of 105 men with poor semen quality were compared to a group of 93 men with normal semen quality. Men with poor semen quality had similar CAG repeat number to men with normal semen quality (21.46 ± 0.30 vs 20.99 ± 0.28, $p = 0.126$). Y-chromosome microdeletions were only detected in men with suboptimal semen parameters (7.4%). However, the presence of a deletion was not related to CAG repeat number. The CAG repeat number in the men with normal semen quality in the present study is similar to the Australian and German samples, but lower than those reported for the Swedes, Dutch and Danes. These results argued against the hypothesis that higher CAG repeats are associated with infertility in men, but strongly suggest that different populations may show different number of CAG repeats in addition to racial variation reported in previous studies.

INTRODUCTION:

It has been suggested that mutations in the androgen receptor (AR) gene causes reduced sperm production in 10-20% of men who are otherwise healthy (Yong et al., 1998). The AR protein triggers differentiation of male secondary sexual characteristics and initiates and maintains sperm production in response to stimulation by the male sex androgens, testosterone and dihydrotestosterone (Yong et al., 2000; Mifsud et al., 2001).

The AR-CAG repeat region in exon 1 of the AR gene is unstable and undergoes expansions and contractions in repeat number during meiotic DNA replication (Kovtun et al., 2001). Thus the number of CAG repeats is polymorphic in different populations and between different races, with a range between 9 and 36 repeats (Dadze et al., 2000). African American men have shorter repeat lengths (19 repeats), Asian men have longer repeat lengths (23-24 repeats), and Caucasian men have around 21 repeats (Hardy et al., 1996; Sartor et al., 1999; Yoshida et al., 1999; Mifsud et al., 2001). It is known that patients with Kennedy's disease, a fatal neurodegenerative disease, have over 40 CAG repeats in the AR TNR and are subfertile resulting from reduced sperm production (Puscheck et al., 1994).

Recent studies have suggested that expansion increasing the number of AR-CAG repeats above the mean is a significant risk determinant of reduced sperm production in otherwise healthy males (Dowsing et al., 1999; Yoshida et al., 1999; Mifsud et al., 2001; Patrizio et al., 2001). In vitro studies have demonstrated a reduction in AR-transcription factor interaction, and transcription-activating potential with increases in repeat number within the normal polymorphic range (Kazemi-Esfarjani et al., 1995; Tut et al., 1997; Hsaio et al., 1999; Irvine et al., 2000). This presumably alters rates of transcription of genes involved in spermatogenesis. This increased risk of defective sperm production is predicted to be four times greater when repeat numbers near the extremes of the normal range, greater than or equal to 28 CAG repeats (Tut et al., 1997).

This present study analysed the AR-CAG repeat lengths of 93 Caucasian men with normal semen quality and 105 men with poor semen quality. We showed that there was no significant variation in the CAG repeat lengths between the two groups of men in the Christchurch population.

METHODS:

Patients and Semen Classification

Semen samples were obtained with written consent from couples attending the New Zealand Centre for Reproductive Medicine. Ethical approval was obtained from the Canterbury Health Ethics Committee.

Samples were assigned a fertility category based on the World Health Organisation criteria (WHO, 1999). Men who had greater than 20 million spermatozoa/ml and over 50% of which were motile were classified as normozoospermic (n=93). Asthenozoospermic samples had greater than 20 million spermatozoa/ml but less than 50% motility (n=53). All other patients were classified as having poor semen quality. Semen samples with less than 20 million spermatozoa/ml were classified as oligozoospermic (between 5 and 20 million sperm/ml) (n=52), severely oligozoospermic (0.1-5 million sperm/ml) (n=45), or azoospermic (no sperm was found in the ejaculate) (n=8). None of these individuals had hypogonadism, cryptorchidism, obstructions or infection that might have otherwise explained the observed poor semen parameters. In addition, 98% of samples were examined for deletions within the AZFc region of the Y-chromosome (Human STS's sY146, 149, 156, 157, 158, 239, 254, 269, 277) by PCR using conditions described previously (Kerr et al., 2000).

AR gene analysis

Sperm DNA was isolated from 200 μl of semen samples using a standard phenol/chloroform extraction protocol (Holyoake et al., 2001). 200 ng of DNA was used to amplify the AR-CAG repeat region by polymerase chain reactions (PCR) using primers specific to DNA flanking the repeat (forward primer 5'tccagaatctgttccagagcgtgc 3' and reverse primer 5' gctgtgaaggttgctgttcctcat 3', Yoshida et al., 1999). PCR analysis was conducted on a MJ Research Inc. PTC-100 thermo-cycler. The programme consisted of an initial denaturation at 94°C for four minutes followed by 30 cycles of 94°C, 58°C, and 72°C for one minute each, and a final ten-minute extension at 72°C. For DNA sequencing all 265 samples were either end-labelled with γ dCTP 32P (approximately 3000 Ci/mmol) or internally labeled with α dCTP 33P (approximately 3000 Ci/mmol), and sequenced using a Thermosequenase (Amersham Pharmacia Biotech) or Amplicycle (Applied Biosystems) sequencing kit respectively. Samples were run through a 6%

denaturing polyacrylamide gel at a constant 70W power (approximately 1500V and 45mA) and a direct count of CAG repeat number taken.

Statistical Analysis

The mean ± standard error of mean (SE), mode, median, interquartile range (IQR), and range were calculated for each group of men. In addition, the numbers of CAG repeat lengths falling outside 1.5 x IQR ($\leq$ 15 and $\geq$ 28 repeats based on an IQR of 3 and quartiles of 20 and 23 repeats calculated for the total subfertile population) were compared between men with poor semen quality and normozoospermic populations. Multiple comparisons were performed to test whether groups of men with poor semen quality had significantly greater mean AR-CAG repeat lengths compared to normozoospermic men using both one- and two-tailed two-sample t-tests (Zar et al. 1996). The null hypothesis was Ho:μsubfertile $\leq$ μnormozs against Ha:μsubfertile $>$ μnormozs (onetailed) and Ho: μsubfertile = μnormozs against Ha: μsubfertile $\neq$ μnormozs (two-tailed), $p < 0.05$. Asthenozoospermic individuals were not included in this comparison.

RESULTS:

All mean CAG repeat lengths were within one repeat of one another, and of the average normal repeat length for Caucasian men (21 repeats).

Statistical analysis of AR-CAG repeat length is summarised in Table 1. Mean repeat length of each group of men with poor semen quality, and the group as a whole (total subfertile population) was compared to the group of normozoospermic men using one-tailed and two-tailed two-sample t-tests (Table 1). This indicated that men with poor semen quality did not have a significantly different distribution of CAG repeat numbers than men with normal semen quality. Azoospermic men had the highest mean CAG repeat length (21.75 ± 0.86), and men with normal semen quality the lowest (20.99 ± 0.28). As the severity of reduced sperm production increased (oligozoospermy through to azoospermy), the mean CAG repeat numbers remained the same (21.42 ± 0.41 to 21.75 0.86, $p > 0.05$). However, the group with poor semen quality had some members with longer CAG repeat lengths (>26 repeats) than the maximum length

observed for normozoospermic men. These individuals represented 5% of the total population of men with poor semen quality (Table 1).

AZFc microdeletions were detected in 7.4% (12/163) of men with suboptimal semen parameters (two asthenozoopsermic men, four oligozoospermic men, five severely oligozoospermic, and one azoospermic men) but were not detected in normozoospermic men. The presence of Y-chormosome microdeletions was not related to AR CAG repeat length. The number of AR CAG repeats observed in these men ranged between 12 and repeats with a mean of 20.22 ± 1.13 and median of 21 repeats.

DISCUSSION

Since the correlation between expanded, yet normal AR TNR lengths was suggested to be a significant risk factor for reduced sperm production by Tut et al. (1997) and Dowsing et al. (1999) many other studies have sought to clarify this association in other populations. We found no significant association between poor semen quality and moderate AR-CAG repeat expansion in our analysis of a group of Christchurch men who identify themselves as Caucasian. Comparison with other studies indicates that for men with reduced sperm count there is little variation in mean CAG repeat number between Caucasian populations (approximately 21.5 – 22 CAG repeats), where sample sizes are high (Dadze et al., 2000; Mifsud et al., 2001; Wallerand et al., 2001; Rajpert-De Meyts et al., 2002; Van Golde et al., 2002) (Table 2). However, two separate studies of Japanese men revealed slight variation in CAG repeats amongst Azoospermic men (Yoshida et al., 1999; Sasagawa et al., 2001) (Table 2). This might also suggest that for those Caucasian populations where the mean CAG repeat number is greater than 22 repeat units the sample size might not be large enough. It appears that different populations may have different CAG repeat length distributions. For example, a mean of 22 repeats in subfertile men vs 20.8 in fertile is not significant in a German Caucasian population (Dadze et al., 2000) but 21.95 repeats vs 20.72 is significant in a Texan Caucasian population (Mifsud et al., 2001). Also, the ranges of AR-CAG repeat lengths observed are remarkably similar between fertile and subfertile populations, usually spanning much of the normal polymorphic range (9-36 repeats).

Additionally several studies indicate that the population of men with poor semen quality are not always the group with the greatest proportion of long CAG repeat lengths (≥28),

proposed by Tut et al. (1997) as being at especially high risk of reduced sperm production (Giwercman et al., 1998; Komori et al., 1999).

In light of these observations it seems unlikely that a small expansion in the AR CAG repeat number within the normal polymorphic range represents a significant risk factor for reduced sperm production in healthy men.

It is widely accepted that AR-CAG repeat length distribution varies slightly between ethnicities, and for this reason recent studies have attempted to analyse predominantly homogeneous sample populations. With this in mind we believe it is possible that there may be a geographic influence due to genealogy on AR-CAG repeat distribution. Comparison of mean CAG repeat length in normozoospermic men suggests that the mean length in our New Zealand Caucasian population is low compared to that in European Caucasian populations of similar sample size (Table 2). This might reflect a
Polynesian ethnic influence in our sample group's genealogy, although our patients identified themselves as of European descent. Other studies have reported a higher mean CAG repeat length in Caucasian azoospermic men than observed in their fertile control group. One further problem is that most studies had a small sample size for the azoospermic group (e.g. Dowsing et al., 1999). We believe that a larger sample size may provide further insights into the relationship between the AR-CAG repeat lengths and human male infertility.

ACKNOWLEDGEMENTS:

We wish to thank Mary Whyte and Tina Newsome of the New Zealand Centre for Reproductive Medicine for providing the semen quality analysis. Tanya Erasmuson was supported by a University of Canterbury PhD scholarship.

Table 1. Statistical Characteristics of AR-CAG length

Fertility Status	n	Mean± SE	Mode (Median)	*p* Value	Range	Q3-Q1= IQR	%≤15[a] CAG's	%≥28[a] CAG's
NormoZS	93	20.99±0.28	20(21)		9-26	23-19=4	3%	0%
Total Subfertile	105[b]	21.46±0.30	20(21)	0.126[c] 0.252[d]	12-30	23-20=3	2%	5%
AZS	8	21.75±0.86	21(21.5)	0.212[c] 0.424[d]	17-25	23-20=3	0%	0%
Severely OligoZS	45	21.40±0.48	21(21)	0.207[c] 0.414[d]	12-28	24-20=4	4%	4%
OligoZS	52	21.42±0.41	20(21)	0.183[c] 0.366[d]	17-30	23-20=3	0%	6%
AsthenoZS	53	21.70±0.47	21(22)	0.095[c] 0.190[d]	8-26	24-20=4	4%	4%

ZS refers to 'zoospermic' for example, Normozoospermic.

15[a] and 28[a] refer to CAG repeat lengths that for the boundaries 1.5 x 3 (IQR of the total infertile population), considered extreme lengths for the distribution. Values represent the proportion of samples that fall outside these values. Q1 and Q3 are the lower and upper quartiles, IQR is the interquartile range. Total Subfertile Population size[b] does not include Asthenozoospermic samples (n=53) because their sperm count is normal (more than 20 x 106 sperm/ml).

P values were calculated using either a one-tailed[c] or two-tailed[d] two-sample t-test with Ho: μinfertile $\leq$ μfertile against Ha: μinfertile $>$ μfertile[c], or Ho: μinfertile $=$ μfertile against Ha: μinfertile $\neq$ μfertile[d]. Values were considered significant if $p < 0.05$.

Table 2. Comparison of population subgroup statistics on AR-CAG length variability, and its association with infertility in other studies.

Fertility Status	Population Ethnicity	Mean ± SE (N)	Range CAG's	p VALUE[a]	REFERENCE
AZS	Caucasian*	21.75 ± 0.86 (8)	17-25	Not significant	This Study
	Caucasian*	23.0 ± 0.63 (16) 25.40 ± 0.77 (27)	19-26 20-39	Not significant p =0.0005	Patrizio *et al.*, 2001 (<1x10^6 spm/ml)
	Caucasian*	23.91 ±0.50(37)		Significant	Wallerand et al., 2001
	German	21.4 ± 0.64 (18) 21.9 ± 0.34 (59)	16-26 16-29	Not significant Not significant	Dadze *et al.*, 2000 (<1x10^6 spm/ml)
	Caucasian*	24.70 ± 1.0 (10)	20-29	.001< p <.0025	Dowsing *et al.*, 1999
	US (Texan) Chinese	22.91 ± 0.73 (23) 23.82 ± 0.52 (33)	17-31 18-33	p =0.019 p =0.043	Mifsud *et al.*, 2001
	Japanese	23.4 ± 0.52 (30)	19-30	Not Significant	Sasagawa et al., 2001
	Japanese	26.5 ± 0.54 (41)	20-34	p = 0.0013	Yoshida *et al.*, 1999
Severely OligoZS	Caucasian*	21.40 ± 0.48 (45)	12-28	Not significant	This Study
	Caucasian*	22.20 ± 0.45 (26)	18-26	Not significant	Patrizio *et al.*, 2001
	German	22.3 ± 0.80 (29)	16-34	Not significant	Dadze *et al.*, 2000
	Caucasian*	23.0 ± 1.2 (13)	15-34	.025< p <.05	Dowsing *et al.*, 1999
OligoZS	Caucasian*	21.42 ± 0.42 (52)	17-30	Not significant	This Study
	German	23 ± 0.94 (13)	17-28	Not significant	Dadze *et al.*, 2000
	Caucasian*	22.0 ± 1.0 (7)	20-27	Not significant	Dowsing *et al.*, 1999
	US (Texan) Chinese	21.65 ± 0.34 (72) 22.85 ± 0.33 (87)	14-29 14-32	Not significant Not significant	Mifsud *et al.*, 2001
	Japanese	21.2 ± 0.55 (59)	14-32	Not significant	Komori *et al.*, 1999
Total Subfertile Population	Caucasian*	21.46 ± 0.30 (105)	12-30	Not significant	This Study
	Dane	21.5 ± 0.26 (119)	15-29	Not significant	Rajpert-De Meyts *et al.*, 2002
	Dutch	22.2 ± 0.36 (75)	13-36	Not significant	Van Golde *et al.*, 2002
	Caucasian*	23.5 ± 0.41 (69)	18-39	p =0.03	Patrizio *et al.*, 2001
	German	22 ± 0.29 (119)	16-34	Not significant	Dadze *et al.*, 2000
	Swedish	21.9 (33)	16-27	Not significant	Giwercman *et al.*, 1998
	US (Texan) Chinese	21.95 ± 0.31 (95) 23.82 (120)	17-31 18-33	p =0.034 p =0.032	Mifsud *et al.*, 2001
NormoZS	Caucasian*	20.99 ± 0.28 (93)	9-26		This Study
	Dane	21.8 ± 0.32 (110)	14-33		Rajpert-De Meyts *et al.*, 2002
	Dutch	21.7 ± 0.41 (70)	15-31		Van Golde *et al.*, 2002
	Caucasian*	22.0 ± 0.42 (45)	12-30		Patrizio *et al.*, 2001

Caucasian*	22.2 ± 0.40 (50)		Wallerand et al., 2001
German	20.8 ± 0.70 (22)	15-26	Dadze *et al*, 2000
Caucasian*	20.5 ± 0.3 (32)	17-25	Dowsing *et al.*, 1999
Swedish	23.2 (294)	8-30	Giwercman *et al.*, 1998
US (Texan)	20.72 ± 0.52 (55)	8-27	Mifsud *et al.*, 2001
Chinese	22.38 ± 0.32 (87)	11-29	
Japanese	23.7 ± 0.44 (51)	17-28	Sasagawa et al., 2001
Japanese	21.4 ± 0.58(36)	16-31	Komori *et al.*, 1999
Japanese	23.9 ± 0.42 (48)	17-30	Yoshida *et al.*, 1999

(N) indicates the sample size of each fertility classification.
* indicates that the sample population is predominantly Caucasian in composition.

REFERENCES:

Dadze, S., Wieland, C., Jakubiczka, S., Funke, K., Schroder, E., Royer-Pokora, B., Willers, R. & Wieacker, P. F. (2000) The size of the CAG repeat in exon 1 of the androgen receptor gene shows no significant relationship to impaired spermatogenesis in an infertile Caucasoid sample of German origin. *Molecular Human Reproduction.* 6(3), 207-214.

Dowsing, A. T., Yong, E. L., Clark, M., McLachlan, R. I., de Kretser, D. M. & Trounson, A. O. (1999) Linkage between male infertility and trinucleotide repeat expansion in the androgen-receptor gene. *The Lancet.* 354, 640-643.

Giwercman, Y. L., Xu, C., Arver, S., Pousette, A. & Reneland, R. (1998) No association between the androgen receptor gene CAG repeat and impaired sperm production in Swedish men. *Clinical Genetics.* 54, 465-436.

Hardy, D. O., Scher, H. I., Bogenreider, T., Sabbatini, P., Zhang, Z. F., Nanus, D. M. & Catterall, J. F. (1996) Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. Journal *of Clinical and Endocrinological Metabolism.* 81(12), 4400-4405.

Holyoake, A. J., McHugh, P., Wu, M., O'Carroll, S., Benny, P., Sin, I. L. & Sin F.Y.T. (2001) High incidence of single nucleotide substitutions in the mitochondrial genome is associated with poor semen parameters in men. *International Journal of Andrology,* 24, 175-182.

Hsaio, P., Lin, D., Nakao, R. & Chang, C. (1999) The linkage of Kennedy's neuron disease to ARA24, the first identified androgen receptor polyglutamine region-associated coactivator. *Journal of Biological Chemistry.* 274(29), 20229-20234.

Irvine, R. A., Ma, H., Yu, M. C., Ross, R. K., Stallcup, M. R. & Coetzee, G. A. (2000) Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. *Human Molecular Genetics.* 9(2), 267-74.

Kazemi-Esfarjani, P., Trifiro, M. A. & Pinksy, L. (1995) Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n – expanded neuropathies. *Human Molecular Genetics.* 4(4), 523-527.

Kerr, N. J., Zhang, J., Sin, F. Y. T., Benny, P., Sin, I. L. (2000) Frequency of microdeletions in the azoospermia factor region of the Y-chromosome of New Zealand men. *The New Zealand Medical Journal.* 113 (1121), 460-462.

Komori, S., Kasumi, H., Kanazawa, R., Sakata, K., Nakata, Y., Kato, H. & Koyama, K. (1999) CAG repeat length in the androgen receptor gene of infertile Japanese males with oligozoospermia. *Molecular Human Reproduction.* 5(1), 14-16.

Kovtun, I. V., Goellner, G. & McMurray, C. T. (2001) Structural features of trinucleotide repeats associated with DNA expansion. *Biochemical Cell Biology.* 79, 325-336.
Mifsud, A., Sim, C. K. S., Boettger-Tong, H., Moreira, S., Lamb, D. J., Lipshultz, L. I. & Yong, E. L. (2001) Trinucleotide (CAG) repeat polymorphisms in the androgen receptor gene: molecular markers of the risk for male infertility. *Fertility & Sterility.* 75(2), 275-281.

Patrizio, P., Leonard, D. G. B., Chen, K. L., Hernandez-Ayup, S. & Trounson, A. O. (2001) Larger trinucleotide repeat size in the androgen receptor gene of infertile men with extremely severe oligozoospermia. *Journal of Andrology.* 22(3), 444-448.

Puscheck, E. E., Ali Behzadian, M. & McDonough, P. G. (1994) The first analysis of exon 1 (the transactivation domain) of the androgen receptor gene in infertile men with oligospermia or azoospermia. *Fertility & Sterility.* 62(5), 1035-1038.

Rajpert-De Meyts, E., Leffers, H., Petersen, J. H., Andersen, A-G., Carlsen, E., Jorgensen, N. & Skakkebæk, N. E. (2002) CAG repeat length in androgen-receptor gene and reproductive variables in fertile and infertile men. *The Lancet.* 359, 44-46.

Sartor, O., Zheng, A. & Eastham, J. A. (1999) Androgen receptor gene CAG repeat length varies in a race-specific fashion in men without prostate cancer. *Urology.* 53(2), 378-380.

Sasagawa, I., Suzuki, Y., Ashida, J., Nakada, T., Muroya, K., Ogata, T. (2001) CAG repeat length analysis and mutation screening of the androgen receptor gene in Japanese men with idiopathic azoospermia. *Journal of Andrology.* 22(5), 804-808.

Tut, T. G., Ghadessy, F. J., Trifiro, M. A., Pinsky, L. & Yong, E. L. (1997) Long polyglutamine tracts in the androgen receptor are associated with reduced transactivation, impaired sperm production, and male infertility. *Journal of Clinical Endocrinology & Metabolism.* 82(11), 3777-3783.

Van Golde, R., Van Houwelingen, K., Kiemeney, L., Kremer, J., Tuerlings, J., Schalken, J. & Meuleman, E. (2002) Is increased CAG repeat length in the androgen receptor gene a risk factor for male subfertility. *Journal of Urology.* 167, 621-623.

Wallerand, H., Remy-Martin, A., Chabannes, E., Bermont, L., Adessi, G. L., Bittard, H. (2001) Relationship between expansion of the CAG repeat in exon 1 of the androgen receptor gene and idiopathic male infertility. *Fertility & Sterility.* 76(4), 769-774.

World Health Organisation (1999) *WHO Laboratory Manual for the Examination of Human Semen and Sperm-Cervical Mucus Interaction.* Cambridge University Press, Cambridge.
Yong, E. L., Ghadessy, F., Wang, Q., Mifsud, A. & Ng, S. C. (1998) Androgen receptor transactivation domain and control of spermatogenesis. *Review of Reproduction.* 3(3), 141-144.

Yong, E. L., Lim, L. S. E., Wang, Q., Mifsud, A., Lim, J., Ong, Y. C. & Sim, K. S. (2000) Androgen receptor polymorphisms and mutations in male infertility. *Journal of Endocrinological Investigation.* 23, 573-577.

Yoshida, I., Yano, M., Chiba, K., Honda, M. & Kitahara, S. (1999) CAG repeat length in the androgen receptor gene is enhanced in patients with idiopathic azoospermia. *Urology.* 54(6), 1078-1081.

Zar, J. H. (1996) *Biostatistical Analysis Third Edition.* Prentice-Hall International, INC, Upper Saddle River, N. J.