TRINUCLEOTIDE (CAG) REPEAT POLYMORPHISMS OF THE ANDROGEN RECEPTOR GENE IN HUMAN DISEASE

AMPARO MIFSUD GINER

NATIONAL UNIVERSITY OF SINGAPORE
2004

TRINUCLEOTIDE (CAG) REPEAT POLYMORPHISMS OF THE ANDROGEN RECEPTOR GENE IN HUMAN DISEASE

AMPARO MIFSUD GINER
(B. Science (Hons).Valencia)

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
DEPARTMENT OF OBSTETRICS AND GYNAECOLOGY
NATIONAL UNIVERSITY OF SINGAPORE

2004

This thesis is dedicated to Santhosh, Roger, Jolanda and my parents.

This thesis is mine as well as yours

ACKNOWLEDGEMENTS

Firstly and mostly I will like to thank my supervisor A/P E.L. Yong for giving me the opportunity to work in his laboratory, for his guidance, and for the chance to work in these interesting projects.

I would also like to acknowledge the help of:

A/P Koay and her staff team for allowing me to use the facilities in her laboratory.

Dr. Ghadessy for his help in the laboratory, interesting discussions, computing advice, and his unconditional friendship throughout the period he was a postdoctoral fellow in the laboratory.

To Dr. Loy for reviewing the thesis manuscript

I am also thankful to Dr. F. Dong of the department of Biostatistics for his help in statistical analysis.

My thanks to the Doctors of the Obstetrics and Gynaecology Department mainly for help in for collecting the blood specimens of the patients and for allowing me the access to the medical records of the patients included in the studies.
I am grateful to our collaborators at Baylor College of Medicine for their help in the US arm of our male infertility study.

I would like as well to say a word of appreciation to NUS for giving me the research scholarship to complete my Ph.D

TABLE OF CONTENTS

LIST OF FIGURES X

LIST OF TABLES XIII

LIST OF ABREVIATIONS XV

LIST OF PUBLICATIONS XVIII

SUMMARY XX

MAIN INTRODUCTION 1

1. The nuclear receptors 1
2. The androgen receptor 4
3. The polymorphic region of CAG repeats in the androgen receptor gene 9

CHAPTER 1: CAG REPEAT POLYMORPHISM IN THE ANDROGEN RECEPTOR GENE AND MALE INFERTILITY 17

INTRODUCTION 17
1. Role of testosterone in spermatogenesis 17
2. Genetic causes of male infertility 18

MATERIALS AND METHODS 23
1. Subjects included in the study 23
2. DNA extraction 23
3. DNA amplification 24
4. GeneScan analysis 25
5. Statistical analysis 25

RESULTS 28
1. Analyses of patient population from Baylor College of Medicine, Houston 28
   1.1 Characteristics of subjects 28
   1.2 Determination of the CAG repeat length 28
   1.3 Statistical analysis 39
   1.4 Clinical characteristics of patients with $\geq$ 26 CAG repeats 52
2. Analyses of patient population from The National University of Singapore 52

2.1 Characteristics of subjects 52
2.2 Statistical analyses 55
2.3 Clinical characteristics of patients with ≥ 26 CAG repeats 64
DISCUSSION 66

CHAPTER 2. CAG REPEAT POLYMORPHISM IN THE ANDROGEN RECEPTOR GENE AND POLYCYSTIC OVARIAN SYNDROME 77

INTRODUCTION 77
1. Oogenesis 77
2. Hormones involved in folliculogenesis 77
3. Definition of PCOS and selection criteria of patients 78
4. The polycystic ovary and the mechanism of anovulation. 80
5. Several markers of the condition that can be considered as diagnostic criteria 81
6. Genetic component in polycystic ovarian syndrome. Candidate genes studied 84
6.1 Genes coding for steroidogenic enzymes 85
6.2 Genes involved in the secretion and action of insulin 86
6.3 The follistatin gene 87
SUBJECTS AND METHODS 89
1. Study population 89
2. DNA extraction, DNA amplification and GeneScan analysis 89
3. Hormonal analysis 90
4. Data analyses 90
RESULTS 92
1. Subjects characteristics 92
2. Determination of CAG repeat length 92
3. Range of the number of CAG repeats and percentage of homozygous and heterozygous subjects 92
4. T-test to compare the means between patients and controls 96
5. Frequency distribution of the CAG repeat sizes of the short allele and long 99

allele.
6. Testosterone levels divide patients with PCOS into two subsets of AR-CAG length 99
7. Sensitivity test to evaluate the T cut-off point. 102
8. Correlation between the number of CAG repeats and the levels of testosterone. 108
9. CAG repeat number and the levels of LH and FSH 108
10. Ethnic differences in CAG length 108
DISCUSSION 112

CHAPTER 3: RELATIONSHIP BETWEEN PROSTATE-SPECIFIC ANTIGEN, SEX HORMONE BINDING GLOBULIN AND ANDROGEN RECEPTOR CAG REPEAT POLYMORPHISMS IN SUBFERTILE AND FERTILE MEN 119

INTRODUCTION 119
1. Androgenic parameters important for prostate cancer 120
1.1 Prostate specific antigen 120
1.2. Sex hormone-binding globulin 121
1.3. Testosterone and dihydrotestosterone 122
2. Relationship among the various androgenic parameters analyzed in this chapter 122
2.1 Association between the levels of testosterone and prostate specific antigen 122
2.2 Association between the levels of testosterone and sex hormone-binding globulin 125
2.3 The number of CAG repeats in the androgen receptor gene and prostate cancer 125
MATERIALS AND METHODS 127
1. Study population 127
2. Hormonal analysis and determination of the number of CAG repeats 127
3. Statistical analysis 130
RESULTS 131

1. Differences between fertile and subfertile men 131
2. Bivariate correlations 135
2.1 Sex hormone-binding globulin is highly correlated with testosterone 135
2.2 Levels of testosterone and prostate specific antigen 140
2.3 Number of CAG repeats and prostate specific antigen 140
2.4 Number of CAG repeats, testosterone, and sex hormone-binding globulin 140
DISCUSSION 147

CHAPTER 4: IN VITRO STUDIES OF THE TRANSACTIVATION ABILITY OF THE ANDROGEN RECEPTOR GENE CONTAINING DIFFERENT CAG REPEATS 155

INTRODUCTION 155
1.Aldehyde dehydrogenase gene 1 156
2. Prostate specific antigen gene 157
MATERIALS AND METHODS 159
1. Amplification of the prostate specific antigen promoter regions of 1.6Kb, 630bp, and 100bp 159
2. Amplification of the aldehyde dehydrogenase 1gene promoter region 160
3. Ligation reactions 161
4. Transformation 161
5. Purification of plasmid DNA 162
6. Sequencing the PSA-LUC construct 162
7. Transient transfection experiments 163
8. Harvesting the cells 164
9. Measurement of luciferase activity 164
10. Construction of the new PSA reporter system 165
11. Mammalian cell two-hybrid system experiments 165
RESULTS 166
1. Construction of the ALDH 1 reporter system 166
2. Construction of the PSA reporter systems containing the -1.6kb, -630bp and -100bp promoter regions 169

3. Transient co-transfection experiments 177
3.1 Transient co-transfection experiments to evaluate the responsiveness of the pM-ALDH-LUC reporter system to androgen action 177
3.2 Transient co-transfection experiments using the pM-PSA-LUC reporter systems 180
3.3 Inducibility of the PSA 1.6Kb fragment of the PSA promoter by the androgens DHT and MB 180
3.4 The activity of the 1.6 Kb prostate specific antigen promoter was evaluated in other cell lines 180
3.5 pGL3-PSA- LUC reporter system 187
3.6 Cotransfection studies with the reporter system PGL3-PSA-LUC 188
3.7 Optimisation of the relative amounts of AR and PGL3-PSA-LUC 188
3.8 Differences in the CAG length in the AR may transactivate the PGL3-PSA-LUC reporter system differently 188
4. TAD and LBD interaction using the mammalian cell two-hybrid system 189
DISCUSSION 193
CONCLUSIONS 199
REFERENCES 212

LIST OF FIGURES

MAIN INTRODUCTION		PAGE
Fig. 1	Schematic representation of the Androgen Receptor gene and protein.	6
Fig. 2	Schematic structure of the AR cDNA	11
CHAPTER 1: CAG REPEAT POLYMORPHISM IN THE ANDROGEN RECEPTOR GENE AND MALE INFERTILITY		
Fig. 1	Agarose gel electrophoresis with the PCR products comprising the polymorphic CAG repeats region in the AR gene.	32
Fig. 2	Gel image of the GeneScan polyacryamide gel.	34
Fig. 3	Confirmation of the fragment size that flanks the CAG region by sequence analysis. Sample 1	36
Fig. 4	Confirmation of the size of the fragment that flanks the CAG region by sequence analysis. Sample 2	38
Fig. 5	Q-Q plots to evaluate if the sample was normally distributed (Baylor)	41
Fig. 6	Allelic distribution of the AR(CAG)n in the fertile and infertile group of subjects	44
Fig. 7	Alellic distribution of the AR(CAG)n in the normal fertile and the oligospermic group of patients	45
Fig. 8	Alellic distribution of the AR(CAG)n in the normal fertile and azoospermic group	46
Fig. 9	Q-Q plots to evaluate if the sample was normally distributed	57
Fig. 10	Allelic distribution of the AR(CAG)n in the control fertile group and the infertile group	59
Fig. 11	Alellic distribution of the AR(CAG)n in the control fertile and the oligospermic group of patients	60
Fig. 12	Alellic distribution of the AR(CAG)n in the control fertile group and the azoospermic group of patients	61

CHAPTER 2. CAG REPEAT POLYMOEPHISM IN THE ANDROGEN RECEPTOR GENE AND POLYCYSTIC OVARIAN SYNDROME

Fig. 1	GeneScan analysis of the androgen receptor CAG length in women	94
Fig. 2	Frequency distribution of the short AR-CAG allele of PCOS patients and controls	100
Fig. 3	Frequency distribution of the long AR-CAG allele of PCOS patients and controls	101
Fig. 4	Frequency distribution of the short AR-CAG allele of PCOS patients with serum T levels below or above the normal laboratory mean value of 1.73nmol/L.	105

CHAPTER 3: RELATIONSHIP BETWEEN PROSTATE-SPECIFIC ANTIGEN, SEX HORMONE BINDING GLOBULIN AND ANDROGEN RECEPTOR CAG REPEAT POLYMORPHISM IN SUBFERTILE AND FERTILE MEN

Fig. 1	Frequency distribution of testosterone levels in fertile and subfertile subjects.	134
Fig. 2	Scatter graphs showing the positive correlation between the levels of SHBG and total T in the fertile group: A), and infertile group: B).	139
Fig. 3	Scatter graphs showing the correlation between the levels of total PSA and total T.	142
Fig. 4	Scatter graphs showing the correlation between the number of CAG repeats in the AR gene and the levels of total PSA.	144
Fig. 5	Scatter graphs showing the correlation between the number of CAG repeats in the AR gene and the levels of total T.	146
Fig. 6	Factors in the androgen economy in fertile and subfertile subjects	154

CHAPTER 4: IN VITRO STUDIES OF THE TRANSACTIVATION ABILITY OF THE ANDROGEN RECEPTOR GENE CONTAINING DIFFERENT CAG REPEATS

Fig. 1	Construction of the ALDH reporter system	168

Fig. 2	Cloning the 1.6Kb region of the PSA promoter gene into the pMAM*neo*-LUC reporter plasmid	171
Fig. 3	Screening for positive colonies after transformation	173
Fig. 4	Sequence analysis of the promoter region of the PSA gene	176
Fig. 5	Androgen regulation of the ALDH promoter in CV-1 and COS-7 cells	179
Fig. 6	Androgen regulation of the PSA promoter in CV-1 cells	182
Fig. 7	Dose response activities of the PSA and MMTV promoter in CV-1 cells.	184
Fig. 8	Co-transfection experiments with PSA reporter system in different cell type lines	186
Fig. 9	Interaction between the NTD and LBD in the mammalian two-hybrid system	192

LIST OF TABLES

CHAPTER 1: CAG REPEAT POLYMORPHISM IN THE ANDROGEN RECEPTOR GENE AND MALE INFERTILITY

Table 1	Clinical characteristics of the infertile group of patients (Baylor)	29
Table 2	Results of the Dunnett's test..	43
Table 3	Protective effect of short CAG repeats on the risk of male infertility (Baylor)	50
Table 4	Clinical characteristics of patients with $\geq$ 26 CAG repeats in the AR gene (Baylor)	53
Table 5	Clinical characteristics of the infertile group of patients (Singapore)	54
Table 6	Results of the Dunnett's test (Singapore)	58
Table 7	Different cut off points considered in the categorical analysis (Singapore)	63
Table 8	Clinical characteristics of infertile patients with $\geq$ 26 CAG repeats in the AR gene (Singapore)	65
Table 9	Studies to determine if the CAG polymorphic tract predisposes to male infertility	74

CHAPTER 2. CAG REPEAT POLYMOEPHISM IN THE ANDROGEN RECEPTOR GENE AND POLYCYSTIC OVARIAN SYNDROME

Table 1	Clinical and hormonal characteristics of PCOS patients	95
Table 2	CAG repeats mean values for the short, long and biallelic mean for patients and controls	98
Table 3	CAG repeats mean values were calculated and compared for patients with 'low T" and 'high T" respectively	104
Table 4	Sensitivity assay	107
Table 5	Comparisons of the CAG repeats number when patients were classified according to the hormonal levels of LH and FSH below and above the laboratory mean value	110

CHAPTER 3: RELATIONSHIP BETWEEN PROSTATE-SPECIFIC ANTIGEN, SEX HORMONE BINDING GLOBULIN AND ANDROGEN RECEPTOR CAG REPEAT POLYMORPHISM IN SUBFERTILE AND FERTILE MEN

Table1	Descriptive statistics of the various parameters measured in the study	133
Table 2	Results of the bivariate correlations established in the fertile group	136
Table 3	Results of the bivariate correlations established in the subfertile group	137

CONCLUSIONS

Table 1	Role of the AR-CAG length in diseases and processes regulated by androgens	209

LIST OF ABBREVIATIONS

aa	Aminoacid
ACT	Alpha-1-antichymotrypsin
ACTH	Adrenocorticotropic hormone
ALDH	Aldehyde dehydrogenase
AF-1	Activation function 1
AF-2	Activation function 2
AR	Androgen receptor
ARA 70	AR-associated protein 70
AR-CAG	CAG repeat tract in the androgen receptor gene
ARE	Androgen-response element
AZF	Azoospermia factor
BHP	Benign prostatic hyperplasia of the prostate gland
BMI	Body-mass index
bp	Base pairs
CAIS	Complete androgen insensitivity
CBAVD	Congenital bilateral absence of the vas deferens (CBAVD)
cAMP	Cyclic adenosine monophosphate
cDNA	Complementary DNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CHO	Chinese hamster ovary
CV	Coefficient of variation
CV-1	African green monkey kidney cells
CYP11a	Cholesterol side chain cleavage gene
CYP17	17-hydroxylase/17,20-lyase gene
DAZ	Deleted in azoospermia
DBD	DNA binding domain
DHT	Dihydrostestosterone
DM1	Myotonic dystrophy type 1
DMEM	Dubelcco's modified eagle media
DFFRY	Drosophila fat-facets related Y
DNA	Deoxyribonucleic acid

dNTP	Deoxynucleotide triphosphates
DRPLA	Dentatorubral-pallidoluysian atrophy
EDTA	Ethylenediamine tetra-acetic acid
fPSA	Free PSA
tPSA	Total PSA
FAI	Free androgen index
FBS	Fetal bovine serum
FSH	Follicle-stimulating hormone
GR	Glucocorticoid receptor
GTF	General transcription factors
Gln	Glutamine
GnRH	Gonadotropin-releasing hormone
HAT	Histone acetyltransferase
HD	Huntington disease
HRE	Hormone response elements
INSR	Insulin receptor gene
IGF-I	Insulin like growth factor
IGFBP-3	IGF binding protein 3
IU	International units
Kb	Kilobases
KDa	Kilo dalton
KLK	Kallikrein gene
KS	Klinefelter´s syndrome
LB	Luria broth
LBD	Ligand binding domain
LH	Luteinizing hormone
LUC	Luciferase
MB	Mibolerone
MJD	Machado-Joseph disease
MMTV	Mouse mammary tumour virus
NIDDM	Non-insulin-dependent diabetes mellitus
NR	Nuclear receptor
OR	Odds ratio

PAIS	Partial androgen insensitivity
PCOS	Polycystic ovarian syndrome
Poly-Gln	Polyglutamine
PBS	Phosphate buffered saline
PR	Progesterone receptor
PRL	Prolactin
PSA	Prostate specific antigen
RAR	All-trans retinoic acid
RBM	RNA-binding motif
RLUs	Relative light units
RIA	Radioimmunoassay
RXR	9-cisretinoic acid
SBMA	Spinal bulbar muscular atrophy
SCA	Spinocerebellar ataxia
SDS	Sodium dodecyl sulfate
SE	Standart error
SHBG	Sex hormone binding globulin
SRC	Steroid receptor coactivator
T	Testosterone
TAD	Transactivation domain
TBE	Tris-boric acid-EDTA
TFI	Testosterone free index
TIF 2	Transcription intermediary factor 2
TR	Thyroid hormone
TRIS	Tris-hydroximetyl-aminomethane
VDR	vitamin D3
VNTR	Variable number tandem repeats
U	Units
WHO	World health organization

LIST OF PUBLICATIONS RESULTING FROM THE THESIS

Full length articles

1. Amparo Mifsud, Sylvia Ramirez, E.L. Yong. (2000) Androgen receptor gene CAG trinucleotide repeats in anovulatory infertility and polycystic ovaries *Journal of Clinical Endocrinology and Metabolism* 85:3484-8.

2. Amparo Mifsud, Chris K.S. Sim, Holly Boettger-Tong, Sergio Moreira, Dolores J. Lamb, Larry I. Lipshultz, E.L. Yong. (2001) Trinucleotide (CAG) repeat polymorphism in the androgen receptor gene: Molecular markers of male infertility risk. *Fertility and Sterility* 75:275-81

3. Amparo Mifsud, Aw Tar Choon, Dong Fang, E.L. Yong (2001) Relationship between prostate-specific antigen, testosterone, sex-hormone binding globulin and androgen receptor CAG repeat polymorphisms in subfertile and normal men *Molecular Human Reproduction* 7:1007-13

4. Casella R, Maduro MR, Mifsud A, Lipshultz LI, Yong EL, Lamb DJ (2003) Androgen receptor gene polyglutamine length is associated with testicular histology in infertile patients *Journal of Urology* 169: 224-7

Review articles

1. Yong, E.L, Ghadessy, FJ, Qi, W, Mifsud, A, Ng, SC. (1998) Androgen receptor transactivation domain and control of spermatogenesis. *Reviews in Reproduction* 3: 141-144.

2. Yong, EL, Lim, J, Qi W, Mifsud, A, Ong, YC, Sim CKS. (2000) Genetics of male infertility: role of androgen receptor mutations and Y-microdeletions. *Annals of the Academy of Medicine of Singapore* 29: 396-400.

3. Yong, E.L., Lim, J., Qi, W., Mifsud, A. (2000) Molecular basis of androgen receptor diseases. *Annals of the Academy of Medicine of Singapore*, 32:15-22.

4. Yong, E.L., Lim, J., Qi, W., Mifsud, A., Lim J, Ong, Y.C., Sim (2000) Androgen receptor polymorphisms and mutations in male infertility. *Journal of Endocrinological* Investigation 23:573-577

Abstracts

1. Yong, EL, Wang Q, Liu PP, Mifsud A, Ghadessy FJ. The CAG repeat in the androgen receptor gene and its relationship to sperm production and prostate cancer.11th Asia-Oceania Congress of Endocrinology. February 1998, Seoul.

2. Amparo Mifsud, Farid J. Ghadessy, EL.Yong. Analysis of different promoters used in the study of androgen receptor structure-function relationships. European Congress of Endocrinology. May 1998, Seville.

3. A. Mifsud, S Ramirez, E.L. Yong. Non-hyperandrogenic polycystic ovarian syndrome is associated with short CAG repeats in the androgen receptor. 49th Annual Meeting of the American Society of Human Genetics. October 1999, San Francisco.

4. Dirk M. Hentschel, Amparo Mifsud, Eu Leong Yong, Joseph V. Bonventre, Steven I-H Hsu. C/EBPα and C/EBPβ compete with the androgen receptor for binding and transcriptional activation of the Human Prostate Specific Antigen (PSA) promoter. ASN 32 Annual Meeting. 1999, Singapore.

SUMMARY

The androgen receptor (AR) is a member of the superfamily of steroid hormone receptors. The AR binds to its ligands testosterone and 5α-dihydrotestosterone and mediates essential biological processes such as male sexual differentiation, spermatogenesis, development and maintenance of the prostate gland. The AR gene contains a polymorphic region comprising variable number of CAG trinucleotide repeats which encodes glutamine residues in its first exon. The average number of such repeats is 22.5± 2 with a range of 11 to 30 in the population of Singapore. Expansions exceeding the value of 40 repeats cause a fatal neuromuscular disease named Spinal bulbar muscular atrophy (SBMA) or Kennedy's disease.

Recently, besides the pathological expansions of polyglutamine lengths found in SBMA, variations in this CAG microsatellite tract, while remaining within the normal polymorphic range, have been inversely correlated with receptor activity. Thus short tracts are associated with high intrinsic AR activity whereas longer CAG tracts are associated with low AR activity.

In this thesis, the role of the polymorphic CAG repeats stretch in the activity of the AR was investigated in three diseases/conditions: male infertility, Polycystic Ovarian Syndrome (PCOS) in women, and prostate cancer. Subsequently, the mechanism by which the different CAG length confers different transactivation activity to the AR was investigated in cell culture studies by using human promoters.

To elucidate the relationships between CAG repeats and male infertility, we examined the distribution of AR-CAG alleles in infertile population from an American center, compared it to a different ethnic group from Singapore, and explored clinical phenotypes associated with long AR-CAG alleles. The mean AR-CAG length of infertile American patients was significantly longer than fertile controls. Logistic regression showed that each unit increase in CAG length was associated with a 20% increase in the odds of being azoospermic. The odds ratio for azoospermia was 7-fold higher for patients with $\geq$26 CAG repeats than those with <26 CAGs. Although mean CAG length in Singapore subjects was longer than the corresponding American samples, long AR-CAG alleles were significantly related to male infertility in both populations.

To investigate the role of the CAG repeat tract in PCOS, we measured its length in 91 patients and compared them to 112 fertile subjects. There were no differences in the mean CAG length between patients and controls when both alleles were considered together or separately. Since there is a subset of PCOS patients whose serum androgens are normal, we compared differences in CAG length between patients whose serum T levels were below the normal laboratory mean, to those that were higher. There was a difference in CAG length between patients with low and high T levels (20.38±0.51 vs 21.98±0.29), (p=0.004) when only the shorter allele of each individual was considered. Ethnic differences were also evident in our data, Indian subjects had a significantly shorter AR-CAG length compared to Chinese, being 22.08±0.50 and 23.16±0.17 respectively.

To understand the factors in the androgen economy that regulate Prostate Specific Antigen (PSA) levels (the most commonly used marker for prostate cancer), we measured levels of total and free T, Sex hormone binding globulin (SHBG), and the AR CAG repeat length. We then compared these to the total and free PSA levels in 91 fertile subjects, and 112 subfertile men with defective spermatogenesis. Our data suggested that, firstly, PSA correlated with T only in an environment of relatively low androgenicity, such as in subfertile patients. Secondly, in a low androgenic environment, short CAG tracts were associated with high PSA levels.

Lastly, to understand the molecular mechanism by which the size of the AR polyglutamine tract may modulate the activity of the AR, reporter systems containing androgen responsive human promoter were constructed and the activity of the AR tested in cell culture experiments.

In summary, the results from the three different studies in humans indicated that short CAG tracts were associated with high levels of androgen action leading to PCOS in females and high levels of PSA in males. Conversely, long CAG tracts were associated with low androgen receptor activity leading to male infertility.

MAIN INTRODUCTION

1. The nuclear receptors (NRs)

1.1 Description

The NRs are ligand-inducible transcription factors that specifically regulate the expression of many essential target genes involved in metabolism, development and reproduction. Their main function is to mediate the transcriptional response in target cells in response to ligands such as the sex steroids (progestins, estrogens and androgens), adrenal steroids (glucocorticoids and mineralocorticoids), vitamin D3, and thyroid and retinoid hormones. This superfamily encompasses the single largest class of eukaryotic transcription factors.

1.2 Classification

Phylogenetic analyses have identified three main subfamilies within this superfamily. Type I or steroid family includes the receptors for progestins, estrogens, androgens, glucocorticoids, and mineralocorticoids. In the absence of ligand they are bound to heat shock proteins and in this form they are inactive. In the presence of ligand they bind to palindromic repeats located in the promoter of targeted genes in a homodimeric head- to head arrangement. Type II receptors encompass those for thyroid hormone (TR), all-trans retinoic acid (RAR), 9-cisretinoic acid (RXR), and vitamin D3 (VDR). They are able to bind to their promoter DNA in the absence of ligand, and often exert a constitutive repressive effect upon the expression of the target genes. Ligand biding then relieves the repressive effect. The difference with the type I receptors is that they often form promiscuous dimers, mainly involving the presence of the RXR. Type III receptors contain the orphan receptors.

1.3 NRs structure

In the majority of the NRs, the receptor structure is comprised of six domains, named from A to F. Domains A/B are highly variable in sequence and length and they comprise the amino terminal activation function 1 (AF-1) that activates target genes presumably by interacting with components of the core transcriptional machinery, coactivators, or other transactivators. The domain C named DNA-binding domain (DBD) contains two zinc fingers, which are responsible for DNA recognition and dimerization. The domain D allows the protein to bend or change conformation and often contains a nuclear localization signal sequence. Domain E or carboxy-terminal ligand binding domain (LBD) includes the activation function 2 (AF-2). This domain has been studied in great detail due to the availability of crystal structures with different ligands and coactivators. In addition to its ligand-binding properties this region is important for dimerization, nuclear localization, transactivation and intermolecular silencing. The domain F is present only in certain receptors and its function remains unknown.

1.4 Mode of action of the NRs

Steroid and thyroid/retinoid hormones regulate transcription via enhancer elements that may be several kilobases from their target promoters, at which transcription is mediated by RNA polymerase II (Mc Kenna *et al*., 1999). The nuclear receptors and some of its specific coactivators interact directly or indirectly with the components of the basal transcription machinery. Direct protein-protein interaction between NR and GTFs (general transcription factors) had been reported in a number of studies. For instance Schulman *et al*. (1995) described the association between a portion of the TBP (TATA–binding protein) and the AF-2 region of the RXR. This interaction is a direct, specific and ligand-dependant.

Eukaryotic DNA is structured into nucleosomes, and this complex structure has to be dismantled in order to make the DNA accessible to the transcription machinery. One of the mechanisms to achieve the physical separation of the DNA from the histone proteins is through the acetylation of some core histone residues. The acquisition of negative charges as a result of acetylation makes the separation between histones and the negatively charged DNA possible, thereby creating an environment where the DNA is more accessible to transcription factors. In 1995 Brooks *et al.*, identified the first histone acetyltransferase (HAT)-A, a Tetrahymena protein that contain acetyltransferase activity. Their discovery was the first indication that the recruitment of histone acetylation activity by sequence-specific transcription factors might be involved in transcriptional regulation in eukaryotes. After this first discovery, various proteins with acetylation activity have been identified including some nuclear coactivators.

Numerous coactivators have been characterized in recent years and they have been classified into two major types. The Type I coactivators function primarily with the nuclear receptor at the target gene promoter to facilitate DNA occupancy, chromatin remodelling, or recruitment of general factors associated with the RNA polymerase II holocomplex. Examples of type I coactivators are those belonging to the steroid receptor coactivator (SCR) family such as the human SRC-1 (Li *et al.*, 1997). The Type II coactivators modulate the appropriate folding of the AR and its ligand, and facilitate the AR NH_2/COOH-terminal interaction. This category includes coactivators that stabilize the ligand-bound receptor, such as AR-associated protein 70 (ARA70) (Yeh *et al.*, 1996).

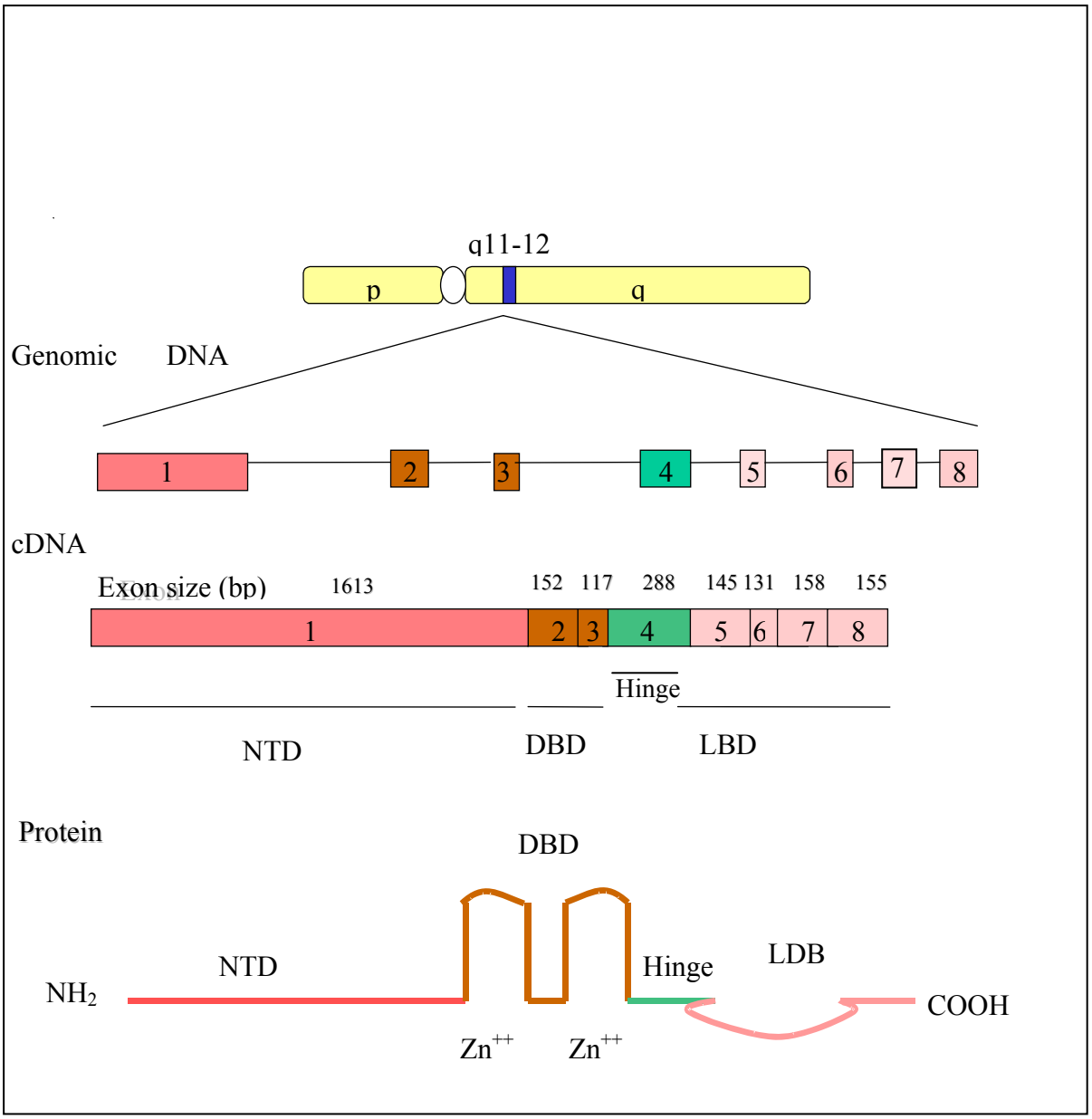
2. The androgen receptor (AR)

2.1 Definition

The AR is a member of the steroid family of NR. It is likely that all steroid receptors have evolved from a common ancestral gene. The human AR genomic DNA was first cloned by Lubahn *et al*. (1988) from a flow-sorted human X chromosome library by using a consensus nucleotide sequence from the DNA-binding domain of the family of nuclear receptors. It is located on the human X chromosome between the centromere and the q13. The gene comprises 75-90 Kb of genomic DNA. The cDNA has 8 exons. The protein has 910-919 amino acids (this variation in length is mainly due to a variable polyglutamine (poly-Gln) stretch in the transactivation domain (TAD), and a molecular weight of 110-114 kDa depending on the poly-Gln length. Like other steroid receptors, AR contains four major structural domains; the TAD, the DBD, a hinge region, and the LBD (Ligand Binding Domain). Sequence homology with other steroid receptors is present only in the last two domains. Fig.1 shows a schematic representation of the AR gene and protein.

Figure 1. Schematic representation of the Androgen Receptor gene and protein

The AR gene is located in the centromeric region of the long arm of the X-chromosome. The AR gene spans 75-90 Kb of genomic DNA and it contains eight exons (numbered boxes) separated by 7 introns (lines in between the boxes). The cDNA molecule comprises a coding region of approximately 2760 bp. The number of bp for each exon is indicated above the exon boxes. The exon 1 codes for the TAD, the exons 2 and 3 for the DBD, part of the exon 4 for the hinge region, and the 3'-portion of exon 4 and exons 5-8 for the LBD. Fig.1 is a schematic representation of the AR protein showing its four structural domains, TAD, DBD, hinge region and LBD.


q11-12
p
q
Genomic DNA
1
2
3
4
5
6
7
8
cDNA
Exon size (bp)
1613
152
117
288
145
131
158
155
1
2
3
4
5
6
7
8
Hinge
NTD
DBD
LBD
Protein
DBD
NTD
Hinge
LDB
NH2
COOH
Zn++
Zn++

2.2 The TAD domain

This domain comprises more than half of the AR, which is a unique attribute of the AR when compared to other steroid receptors. Its structural and functional characteristics are also different from the rest of the members of the steroid receptor superfamily. For instance, the TAD contains a number of homopolymeric amino acids stretches. The most amino-teminal of these is a polymorphic glutamine region with an average size of 21±2 Gln residues in the Caucasian population (Edward *et al*., 1992).

There is a shorter stretch of nine proline residues, located at amino acids 372-379, which does not vary in size. The third stretch is a polymorphic tract of an average size of about 24 glycines residues located at amino acid 449.

The main transactivation domain in the AR, named AF-1 is located in the TAD. The boundaries are not well defined but generally speaking the region between amino acids 51-211 is essential for the transactivation function of the AR (Jenster *et al.*, 1991).

2.3 DNA-binding-domain (DBD)

The DBD of the AR is encoded by exon 2 and 3, and comprises the amino acids 559 to 624. Similar to other steroid receptors it contains two zinc fingers motifs. There are three amino acids at the base of the first zinc finger (glycine 577, serine 578 and valine 581) which are essential for the interaction of the AR protein with the hormone response elements (HRE) in the DNA (Kufter *et al*., 1993).

2.4 The hinge region

This region is coded by the 5'portion of the exon 4 and contains the major part of the AR nuclear targeting signal formed by a cluster of basic amino acids (arginine, lysine, leucine, and lysine) at positions 629-633. The function of the hinge region is to facilitate the transfer of the AR from the cytoplasm to the nucleus (Jenster *et al*., 1991).

2.5 Ligand binding domain (LBD)

The C-terminal domain is the LBD, encoded by the exons four to eight. It encompasses residues 670 to 919. During the last two years the three-dimensional structure of the AR LBD has been determined. Despite substantial differences in the primary amino acid sequence between the AR LBD and other steroid hormone receptor, the three-dimensional structures of the LBDs of these molecules are quite similar. The LBDs of these receptors fold into 12 helices that form a ligand-binding pocket. When an agonist is bound, helix 12 folds over the pocket to enclose the ligand (Sack *et al*., 2001).

The principal function of the LBD is the specific and high affinity binding of androgens. This site is also the binding site for heat shock proteins that have an inhibitory effect while they bind to the AR (Smith, 1993).

The transcription activation function (AF-2) is located in the LBD. In comparison with others steroid receptors has a weak impact in the overall transactivation of the receptor. The AF-2 acts in a ligand dependent manner, and its activity is enhanced by numerous

coactivators such as the transcription intermediary factor 2 (TIF 2) (Jenster *et al*., 1991).

2.6 Mode of action of the AR

After binding to the androgens, the receptor dissociates from the accessory proteins (heat shock proteins), translocates into the nucleus, dimerizes and, through its DNA-binding domain, interacts with specific androgen-responsive elements (ARE) located in the promoter region of androgen responsive genes (Zhou *et al.*, 1994). The AR requires the presence of coactivators, such as SRC-1, p300, p/CAF, CBP, which are stabilized in a heterocomplex by protein-protein interaction with the AR. These complexes recruit general transcription factors to the TATA box, and also exert HAT activity by modifying the structure of histones and chromatin.

3. The polymorphic region of CAG repeats in the androgen receptor gene

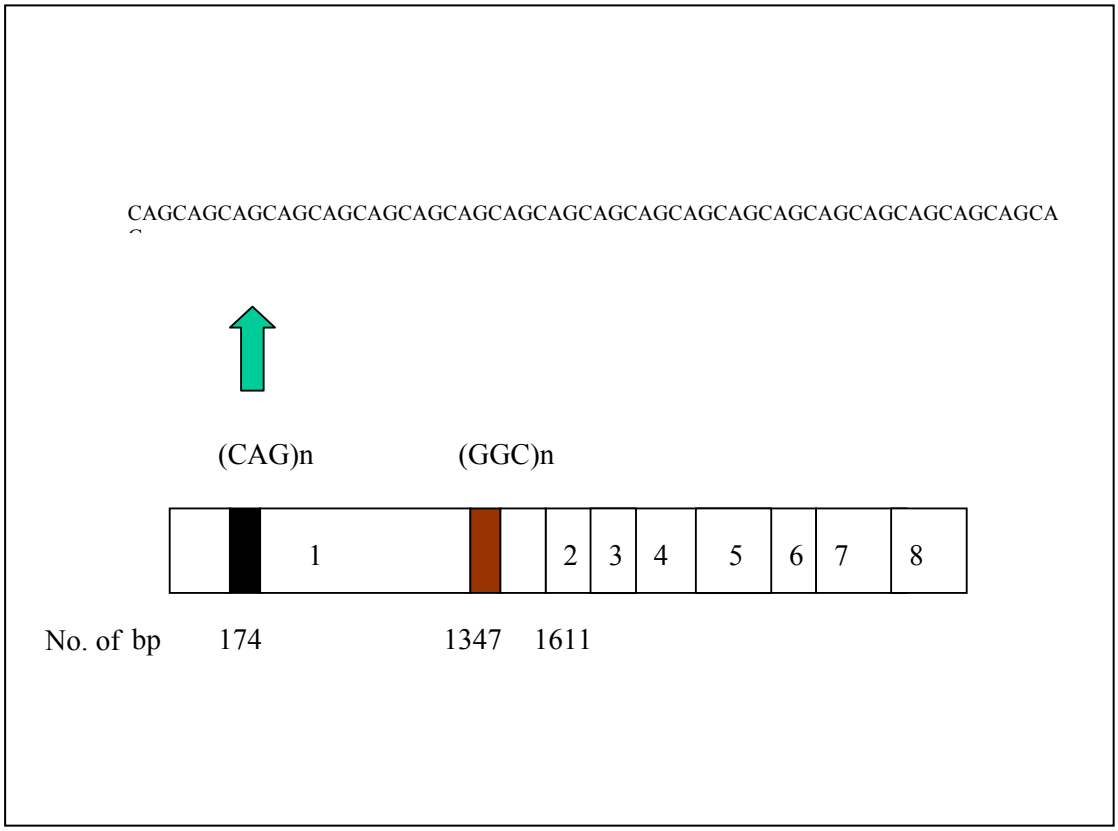
3.1 Description

The AR gene contains a polymorphic stretch of CAG triplets in the exon 1. The CAG repeats, which encodes glutamine residues, vary in number from 11 to 35 in the normal population. The average number of repeats is of 21±2 in the Caucasian population although this value is dependent on the population studied. For instance African-Americans have shorter CAG repeats tracts, an average of 19 CAGs, while Chinese have an average of 23 CAGs (Edwards *et al*., 1992). Fig. 2 shows the location of the CAG stretch in the AR gene.

Although poly-Gln stretches are present in other transcription-regulated proteins and they are believed to promote protein-protein interaction, their exact functions still remains unknown.

Figure 2. Schematic structure of the AR cDNA

The eight numbered boxes represent the exons of the AR gene. The exon 1 comprising more than half of the AR cDNA contains two polymorphic regions, the polymorphic region of CAG repeats at position 174 bp, and the GGC repeats stretch at position 1347 bp.


CAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA
(CAG)n
(GGC)n
1
2
3
4
5
6
7
8
No. of bp
174
1347
1611

Conservation of segments of the AR gene throughout evolution implicates these regions as being critical for the activity of the molecule. Choong *et al*. (1998) compared the AR DNA coding sequence from five primate species, humans, chimpanzee, baboon, macaque and collared brown lemur. The DBD and LBD domains were totally conserved. Nevertheless, a linear increase in trinucleotide repeat expansion of homologous CAG and GGC sequences occurs in the NH_2-terminal region and is proportional to the time of species divergence, suggesting that the CAG repeat expanded during divergence of the higher primate species.

3.2 Polyglutamine regions in the expanded pathological range are found in various genes

Polyglutamine regions are frequently found as polyglutamine tracts, encoded by CAG repeats, and despite this wide distribution, the functions of CAG repeats are often unclear. However, expansion of CAG repeats in genes has been implicated in the pathogenesis in a number of progressive neurodegenerative diseases including Huntington's disease, spinocerebellar ataxia, and dentatorubral-pallidoluysian atrophy, and in SBMA (Lieberman and Fischbeck, 2000). There are similarities shared among the diseases caused by poly-Gln stretches. With the exception of the X-linked SBMA, they all share the autosomal dominant mode of inheritance and anticipation on paternal transmission. The position of the poly-Gln in the coding region varies between the different proteins. For instance for AR and huntingtin, the poly-glutamine expansion are located within the amino terminal region, while for the atrophin is near in the C-terminus (Choong *et al*, 1998). An interesting observation is that there are no common domains shared between these proteins, only the poly-Gln stretch, and that the mutated proteins seem to have similar expression levels that the no mutated ones. A common

feature is the presence of insoluble intracellular aggregates in all of those disorders including SBMA. These aggregates are similar to those found in other neurodegenative disorders such as Alzheimer's, Parkinson's disease and prion diseases, suggesting that aggregates may be toxic for long living post-mitotic cells. It has been postulated that the expanded poly-Gln tracts exert a gain of toxic function (Brooks *et al*., 1997). Various hypotheses had been postulated regarding the mechanism by which expanded regions causes neuronal degeneration. Perutz *et al*. (1994) suggested that poly-Gln repeats might function as polar zippers, joining complementary proteins such as other transcription factors by hydrogen bonding. Longer stretches would result in stronger interactions. Choong *et al*. (1998) suggested that protein aggregation might occur after the proteolysis of the AR to smaller fragments encompassing an expanded CAG repeat. The AR with expanded CAG repeat would be preferentially digested by a protease, and with time, insoluble aggregates will form, triggering the apoptotic pathway.

The exact molecular mechanism leading to expansion of trinucleotide repeats remains unknown. It is likely to be related to the ability of repeat tracts to form unusual DNA secondary structures such as hairpins and slipped-strand DNA duplexes, which can interfere with aspects of DNA metabolism (Usdin *et al*., 1998). The amino acid repeats, particularly of uncharged polar amino acids can mediate or modulate protein-protein interactions, raising the possibility that changes in their length during evolution could result in changes in the strength of protein-protein interactions. As a number of studies have associated glutamine repeats with transcription factors, this could have implications for the evolution of gene regulatory networks. Evolutionary studies of a number of genes involved in the human triplet expansion diseases have indicated that

the repeats in these genes have arisen by gradual expansion of the tandem repeat, apparently resulting from replication slippage. Repeats are generally absent in rodent homologue genes and comparative studies indicate an increase in repeat length during primate evolution, with humans generally having the longest repeats (Rubinsztein *et al.,* 1994)

3.3 Kennedy's syndrome, a pathological expansion of CAG repeats in the AR gene beyond 40

Males with expansions of the CAG tract in the AR gene beyond 40 repeats develop Kennedy's disease, also known as X-linked spinal and bulbar muscular atrophy (SBMA). The X-linked SMBA was first reported by Kennedy in 1968. It is a rare inherited neurodegenerative disease characterized by progressive neuromuscular weakness due to the loss of motor neurons in the brain stem and spinal cord. Onset of this disorder is usually in the fourth or fifth decade, but may be as early as the mid-teens or as late as 60 years. La Spada *et al*., (1991) first reported that the origin of the disease was associated with expansion of the poly-Gln tract of the AR to values higher than 40 Gln. Motoneunoral cell death does not appear to be linked to a loss of function of the mutated AR, since neurodegeneration does not occur in patients with testicular feminization who lack of AR function.

The AR is expressed in many tissues including the central nervous and muscular system (Sar *et al*., 1990). The absence of any neuromuscular degeneration in patients with complete androgen insensitivity leads to the hypothesis that the AR with the expanded tracts is associated with a gain of neurotoxic function. The length of the CAG tract correlates inversely with the age of onset of the disease (La Spada *et al*., 1991) but it still uncertain if the length correlates with the severity of the disease

(Shimada *et al*., 1995). Endocrine abnormalities usually become manifest after neurological symptoms. More than half of patients with SBMA normally develop gynaecomastia, low sperm production, low testicular size, and one in three patients with SBMA has subnormal serum testosterone concentrations impairment (Harding *et al*., 1982; Arbizu *et al*., 1983)

3.3 AR mutations and CAG polymorphisms within the normal range

To date, the majority of the mutations in the AR gene have been found in the LBD and DBD domains. These mutations can cause a wide range of phenotypic manifestations, depending on the amino acid that is affected. The most severe case is complete androgen insensitivity syndrome (CAIS) due to a complete disruption of the AR, the affected individuals are healthy 46XY individuals with a female phenotype (Lim *et al*., 1997; Yong *et al*., 1994). Other mutations can cause partial androgen insensitivity (PAIS), which impairs the activity of the receptor without totally disrupting its function. Accordingly, the phenotype of the individual depends on how severe the activity of the AR has been impaired (Yong *et al.*, 1998).

With the above reviewed on CAIS and PAIS, both related to point mutations on the AR, as well as the link of SBMA with the CAG polymorphism within the AR, the aim of this thesis was to investigate if the length of the CAG tracts within the non-pathological range or normal have an effect on the intrinsic activity of the AR. To study this association different approaches were taken. The first one was to investigate whether expanded CAG tracts were implicated in reduced androgen action and therefore leading to a condition highly dependent on androgens such as defective spermatogenesis and male infertility. To investigate the relationship between AR-CAG length and male infertility, two retrospective case-controls studies were

performed. The first comprised of subjects from a predominantly Caucasian population recruited at Baylor College of Medicine; Houston, and the second included subjects of predominantly Chinese ethnic origin recruited at the National University Hospital in Singapore. The second question raised was whether having short CAG tracts would predispose to diseases characterized by high levels of androgen action. The disease investigated upon was the Polycystic Ovarian Syndrome (PCOS), in females. The third objective of this project was to determine whether the levels of Prostate Specific Antigen (PSA), a common oncogenic marker for prostate cancer, were related to the number of CAG repeats in the AR gene and consequently obtain a new molecular marker for prostate cancer. Lastly, the molecular mechanism by which the different CAG length modulate the activity of the AR was investigated in cell culture experiments using reporter systems containing the human promoters ALDH and PSA. In addition, the interaction between the LBD and TAD domain containing different CAG lengths was investigated in the mammalian cell of two-hybrid systems. A more detailed introduction section has been included in each of the four chapters of this manuscript.

CHAPTER 1: CAG REPEAT POLYMORPHISM IN THE ANDROGEN RECEPTOR GENE AND MALE INFERTILITY

INTRODUCTION

1. Role of testosterone (T) in spermatogenesis

Spermatogenesis is the process by which mature spermatozoa are produced and the number of chromosomes is reduced to the haploid state. One complete cycle takes approximately 60 days and it takes place within the coiled seminiferous tubules. Two types of cells facilitate the process of spermatogenesis: Sertoli and Leydig cells. Sertoli cells line the basal laminar of the seminiferous tubules and are attached to one other by specialized junctional complexes. They contribute to the formation of the blood-testis barrier, limiting the transport of certain molecules and fluid to the tubular lumen. Other functions are to provide an energy source for the developing spermatozoa and the phagocytosis of damaged cells. Outside the seminiferous tubules are the Leydig, which are the main source of the androgen T.

Various hormones take part in the process of spermatogenesis. One of the most critical hormones is T. The production of T occurs in the Leydig cells under the stimulation of the hormone LH. This binds to the AR and ultimately activates some of the genes involved in the process of spermatogenesis. T not only has androgenic effects of its own but can be converted to a much more potent androgen called dihydrotestosterone (DHT) by the action of the enzyme 5 alpha-reductase. In addition, it can also be aromatised to 17 beta-oestradiol in the testicular Leydig cells.

T is required to maintain quantitatively normal spermatogenesis, and decreasing T levels in the seminiferous tubules results in defective sperm production (Zirkin *et al*, 1989)

FSH plays an important role too. It binds to specific receptors on the Sertoli cells and stimulates cAMP production resulting in the activation of a number of proteins responsible for the trafficking of nutrients from the Sertoli cells to the germ cells.

2. Genetic causes of male infertility

In 40-60% of cases, the aetiology of male infertility remains unknown and has to be classified as idiopathic (De Kretser, 1997). Known factors of male infertility include varicocele, infections, irradiation chemotherapy, testicular torsion, cryptorchidism, sexual disorders, hyperprolactinaemia and hypogonadotrophic hypogonadism. With the advancement of molecular biology techniques it seems that more and more of the so-called idiopathic male infertility cases with impaired spermatogenesis appear to be of genetic origin. (Skakkebaek *et al.*, 1994). Currently the main genetic defects that are known to lead to male infertility are:

2.1. Structural chromosomal disorders

Some genetic causes may be due to structural chromosomal disorders, the most frequent being Klinefelter's syndrome (KS). KS accounts for 14% of the cases of azoospermia. It is present in 1 in 500 live male births and is the most common abnormality of sexual differentiation. Clinical symptoms of KS besides azoospermia are gynaecomastia and small firm testis. The chromosomal constitution of these patients is 47, XXY in 90% of the cases, and 46, XY/47, XXY in the other 10%.

Other types of chromosomal disorder are caused by autosomal translocations, and they are thought to be responsible for defective spermatogenesis in up to 7% of infertile males. They might interrupt important genes involved in spermatogenesis, gonadal differentiation, androgen biosynthesis, androgen action, and sperm quality.

2.2. Y chromosome deletions

In humans the Y chromosome is essential for normal testicular differentiation and spermatogenesis. Cytogenetic studies on azoospermic subjects two decades ago led to the postulation of the existence of a gene locus in the distal euchomatic part of Yq (Yq11), and it was defined as the azoospermia factor (AZF). With the advancement of molecular biology techniques numerous studies (Liow *et al.*, 1998; Simoni *et al.*, 1997) have reported microdeletions in this gene region of infertile men. They represent the aetiological factor of 10-20% of idiopathic azoospermia and severe oligospermia, depending on the population studied.

Multiple non-overlaping spermatogenesis loci in this region have been identified, suggesting the involvement of more than one gene in Yq spermatogenesis. Some of the candidate genes identified are the RBM (RNA-binding motif) of which there are approximately 30 copies and are located on both arms of the Y chromosome; the DAZ (deleted in azoospermia), which like the RBM gene also has multiple copies in the Y chromosome; and the DFFRY (Drosophila fat-facets related Y). Out of these three genes, the DAZ gene represents the most frequently deleted region in infertile males. Such deletions appear to remove the entire DAZ gene cluster and have been associated with a wide range of spermatogenesis alterations, ranging from azoospermia, due to

Sertoli cell only, to oligospermia with different testicular phenotype (Reijo *et al.*, 1995; Liow *et al*., 1998).

2.3. Androgen receptor mutations and polymorphisms

Androgens play an important function throughout the different stages of human development and reproduction. It is well known that sperm production is dependent on androgens (Zirkin *et al.*, 1989) and suboptimal androgen action could lead to defective spermatogenesis (Aiman and Geiffin, 1982). Androgen action is mediated by the AR, so abnormalities of androgen action can be due to mutations on the AR gene that impairs its optimal transactivation capability. To date, the majority of the mutations have been found in the LBD and DBD domain of the AR. These mutations can cause a wide range of phenotypic manifestations, depending on the aa that is affected. The most severe case is complete androgen insensitivity syndrome (CAIS) that occurs when the mutation completely disrupts the function of the receptor. Patients with CAIS have a female phenotype: clitoris, labia minora and majora, but with an absence of pubic and axilar hair, a blind-ending vagina, and no uterus. Postpubertal individuals with CAIS have testis, increased levels of LH, and normal male levels of T. (Lim *et al.*, 1997; Yong *et al*. 1994). Other mutations can cause partial androgen insensitivity (PAIS) when they impair the activity of the receptor without totally disrupting its function. Accordingly, the phenotype of the individual depends on how severe the activity of the AR has been impaired. (Yong *et al.*, 1998).

The majority of the AR mutations are found in the LBD perhaps because the screening of the TAD is quite difficult due to its high CG content and large size. However, Wang *et al.* (1998) described three G to A transitions in codons 210, 211 and 214 after

screening one hundred and fifty-three patients with various degrees of defective spermatogenesis. The transitions in codon 210 and 211 were polymorphisms and did not change the sense of the codon. They were associated with ethnic origin and were present in 15% of Indians and 10% of individuals from the Middle-East, but were absent in the Chinese population. The third mutation, in codon 214, resulted in a non-conservative glycine to arginine substitution, and was associated with a 20% decrease in AR transactivation compared to the wild type.

In the exon 1 of the AR gene is present a polymorphic trinucleotide repeat segment, (CAG)n; this segment encodes a poly-Gln tract which normally ranges from 11 to 31. This tract has been the source of extraordinary interest in recent years because as mention in the main introduction, section 3.3 spinal (page 14), it was found that CAG expansion to beyond 40 repeats leads to SBMA, a fatal neuromuscular disease. Patients with SBMA have evidence of reduced androgen receptor function in the form of low virilization, oligospermia or azoospermia, testicular atrophy, and reduced fertility (Amato *et al*., 1993; Arbizu *et al*., 1983). Besides the pathological expansions of CAG lengths found in SBMA, short alleles were found to be associated with the androgen-dependent tumour prostate cancer. Accordingly, short CAG repeats increase AR androgenicity and are associated with earlier age of onset, increased tumour grade and increased risk of extra-prostatic extension in prostate cancer (Hardy *et al*., 1996; Giovannucci *et al*., 1997). If short CAG lengths lead to prostate cancer, we wanted to determine whether at the other end of the spectrum, long CAG repeats might cause reduced AR androgenicity, and a condition that is sensitive to reduction of AR activity such as impaired spermatogenesis. Preliminary data by Tut *et al.* (1997) indicated that long CAG repeats while still in the normal range could reduce the activity of the AR

leading to male infertility. In this first study, 153 patients with varying degrees of impaired spermatogenesis were analyzed and compared to 71 fertile individuals. The racial composition was mainly of Asian origin and made up of about 60% Chinese subjects. They concluded that patients with more than 28 Gln in their AR had more than a four fold increased risk of reduced spermatogenesis. There was also a trend whereby the more severe the spermatogenesis defect, the higher the number of glutamines in the AR protein. No data was available for a Caucasian population. In this project, to further understand the relationship between CAG repeats and male infertility a large U.S population of mainly Caucasian ethnic background was analyzed and compared to a population from Singapore of a Chinese ethnic background. In addition the phenotypes of the AR gene alleles with long CAG tracts were studied in great detail.

MATERIALS AND METHODS

1. Subjects included in the study

Patients were recruited from the Division of Male Reproductive Medicine and Surgery of the Scott Department of Urology, Baylor College of Medicine, Houston; and the Infertility Clinics, Department of Obstetrics and Gynecology, National University Hospital, Singapore. A complete history and physical examination were performed, and the use of any medications or previous surgery was recorded. The size of both testes was measured with a Prader orchidometer and presence of any associated varicoceles noted. Patients who had hypopituitarism, hyperprolactinemia, infective or obstructive syndromes of the genital tract were excluded from this study. All patients included in the study were 46XY males. Patients with mutations in the AR gene or Y-microdeletions were not included in the study. Sperm parameters were assessed according to standard criteria and were calculated by taking the mean of at least two analyses done 3 months apart. Azoospermia was defined as the absence of any spermatozoa despite centrifugation of the semen specimen, and oligospermia defined as the mean sperm density being <20 mil/ml. Control subjects of matched age were men of proven fertility with no previous infertility history or treatment, and without any genetic disease. The project was approved by the Ethical Review Committees of both Institutions.

2. DNA extraction

DNA was extracted from the peripheral blood leucocytes of patients and control subjects. 10 mL of blood was collected in a EDTA tube and centrifuged for 15 minutes at 3000 rpm. The buffy coat was transferred to a clean tube and 8.5 mL of extraction

buffer was added (10mM Tris Cl, [pH 8.0], 0.1 M EDTA [pH 8.0], 0.5% SDS). This mixture was incubated for 1 hour at 37°C with occasional mixing. Afterwards proteinase K was added to a final concentration of 100µg/ml and the mixture transferred to a shaking waterbath at 55°C for 3 hours. Then the tubes were placed on ice to cool the solution and an equal volume of water saturated phenol added. The two aqueous and organic layers where vigorously mixed until a fine emulsion was formed. The two phases were afterwards separated by centrifugation at 3000 rpm for 15 min at 4°C. The aqueous top phase were the DNA is present was transfered to a clean tube. To precipitate the DNA, 0.2 volumes of 10M ammonium acetate and 2 volumes of ethanol was added at room temperature. The tube was carefully inverted a few times until it was possible to see the swirled DNA. The DNA was taken out of the solution with the help of a pasteur pipette and dissolved in 1mL of sterile distilled water. The following day the DNA was quantified by measuring its absorvance at 260 nm

3. DNA amplification

The CAG repeat segment was amplified using the sense (5'-TCCAGAATCTGTTCCAGAGCGTGC) and antisense (5'-GCTGTGAAGGTTGCTGTTCCTC) primers. The 30 µl reaction mix contained 500ng of genomic DNA, 50 µM of dNTPs (Epicentre Technologies), 0.25µM of each primer, the fluorescent-labeled dNTP (R6G) (Perking-Elmer) and 0.5 units of Taq polymerase (Pharmacia). A 2-step 30-cycle amplification protocol was used in which the denaturing temperature was 95°C for 45 sec, and the combined annealing and extension temperature was 68°C for 1.5 min. In the first cycle the sample was denatured for 5 min.

To verify the PCR amplified the expected fragment, the amplicons were resolved on a 2% agarose gel with the 100pb marker (Promega).

4. Genescan analysis

4.1 Preparation of the 4% Acrylamide-Urea gel.

50 ml of the gel contained: 18 g of urea (Bio-Rad Laboratories), 5 ml of (19:1) acrylamide stock solution (Bio-Rad Laboratories), 1% of ion-exchange resin (ICN Biomedicals). The mixture was dissolved in 20 ml of water (some heat was necessary to dissolve the urea crystals). Afterwards the solution was filtered through a 0.22 μm filter and the de gas was removed by exposure to vacuum suction for 10 minutes. Afterwards 10 ml of 5X TBE buffer were added to the mixture and the volume made up to 50 ml with water. Before casting the gel, 250μl of freshly made 10% ammonium persulfate (Sigma-Aldrich) and 35μl of TEMED (Sigma-Aldrich) were added. The gel took about two hours to polymerize.

4.2 Preparation of the PCR products for GeneScan analysis

The amplified samples were mixed with formamide, loading buffer and the Rox(500) size standards (Perking-Elmer) in a ratio 1:10:2:2 respectively. This mixture was denatured for 5 min at 95°C and then loaded into the 4% denaturing polyacrylamide gel. The sizes of the samples were analyzed on a 377 DNA Sequencer running Genescan 672 software (Perking-Elmer).

5. Statistical Analysis

Although the variable number of CAG repeats is a discrete variable the wide range of values, from 8 to 33 repeats allowed us to consider it as a continuous one. The Q-Q

plot is a graph used to determine if a sample is approximately normally distributed. It plots a set of quantiles or percentiles from the sample versus the corresponding set of quantiles from a normal distribution. If the graph appears close to a straight line then the assumption that the sample is approximately normally distributed can be made. In modern goodness of fit analysis this quantile graphs are preferred to formal goodness of fit tests because the test tend to be always significant for large samples. Results from the Q-Q plot indicated that it was possible to make the assumption that the data was normally distributed. Of course it can not be completely normal because the number of CAG repeats is a discrete value, but it is close enough to be able to perform the subsequent statistical tests.

The *t*-test and Anova procedures were the initial steps of the analysis. The *t*-test or Student's test is specifically designed to compare the mean values of two groups. The Anova test tells whether there are overall differences between the various groups of patients and the control group. If the Anova test shows that there are differences among those groups then a Dunnett's test is performed. The Dunnett's test is a special case of means comparison in which the only comparison that need to be tested is between a set of treatments and a single control treatment. In a general multiple comparison between K treatments and one control there are (K+1) (K)/2 comparisons performed. In the Dunnett's procedure we restrict ourselves to only K comparison and hence obtain a more powerful test.

In this study, firstly, the mean number of CAG repeats from infertile patients were compared to fertile controls using a two-sample independent *t*-test. Infertile patients were subsequently divided in azoospermic and oligospermic patients. Multiple

comparisons were then performed comparing the mean number of CAG repeats of azoospermic and oligospermic groups with the fertile control group using ANOVA and then Dunnett's test.

The number of CAG repeats was then assessed as a continuous variable. Logistic regression analysis was performed using the number of CAG repeats as the predictor and clinical infertility as the outcome. Logistic regression is a standard method for regression analysis for cases when the response variable is not quantitative but dichotomous. It is a test commonly used in the field of health sciences.

For the categorical analysis, the number of 26 CAG repeats was chosen as a cut-off point because this size is about two standard deviation above the mean. Having $\geq$26 CAG repeats was taken as a reference group and examined if the risk of infertility decreased progressively with shorter CAG repeat lengths. From this analysis the odds ratios and corresponding confidence intervals were calculated.

Statistical analyses were performed using the SPSS Version 9.01 (SPSS) computer software. Statistical significance was defined as obtaining a two-sided p-value less than 0.05 and data reported as mean±SE.

RESULTS

To investigate the relationship between AR-CAG length and male infertility, two case-controls studies were designed. The first one comprised of subjects from a predominantly Caucasian population recruited at Baylor College of Medicine; Houston, and the second one included subjects of predominantly Chinese ethnic origin recruited at the National University Hospital in Singapore.

1. Analyses of patient population from the Baylor College of Medicine, Houston

1.1 Characteristics of subjects

The study examined 95 infertile subjects with various degrees of infertility from Baylor College of Medicine. 55 individuals of proven fertility were recruited as controls. The age of patients and controls was comparable, averaging 37 and 43 years respectively. The racial composition of controls and patients was similar and consisted of mainly Caucasians (68% and 71% respectively). The other two races were individuals of Hispanic and Arabic origin. The clinical characteristics of the patients are summarized in Table 1.

1.2 Determination of the CAG repeat length

The region comprising the CAG polyglutamine tract was amplified as described previously in the materials and methods, chapter 1; section 3 (page 24). Fragments were firstly resolved in a 2% agarose gel electrophoresis along with the 100bp DNA ladder marker for about 1 hour in order to verify the PCR reaction. Despite this being a difficult region to amplify due to the high CG content, most reactions yield a substantial and unique PCR product.

Table 1. Clinical characteristics of the infertile group of patients.

Number of infertile patients	95
Number of patients with azoospermia	23
Number of patients with oligospermia	72
Age (years)	37± 2.3
Mean sperm density for oligospermic patients (Normal : ≥ 20 mil/ml)	6.34±0.7
Percentage of patients with varicocele	41%
Mean testicular volume (Normal : 17.82 ±0. 61 ml)	14.42+0.61

The size of the amplified fragment was dependent on the length of the CAG stretch, in the range from approximately 260 to 320 bp (Fig. 1).

The precise size of the CAG repeat to a base pair level was achieved by GeneScan technology. Each PCR product was then mixed with the size standards of known sizes and resolved in the 4% acrylamide gel. The advantage of this analysis is that each sample has its own internal standard. Both the sample and the internal standards migrated across the gel at the same time and under the same electrophoresis conditions.

The software analyzes the data and creates a gel file. In this study the PCR products were labeled with the green fluorescent dye and the commercially available internal markers with red color (Fig. 2). When the electropherogram was displayed in a horizontal way, it was observed not to be composed of a single peak but of several overlapping peaks. This is due to the stuttering effect, a well-known phenomenon in the amplification of trinucleotides repeat regions, believed to be due to PCR slippage. To elucidate which one of the three peaks corresponded to the correct size of the sample, some samples were sequenced and compared to the GeneScan profile. It was determined that the correct size was given by the peak corresponding to the largest number of CAG repeats, the third one in the GeneScan electropherogram (Fig. 3 and 4).

Figure 1. Agarose gel electrophoresis with the PCR products comprising the polymorphic CAG repeats region in the AR gene.

A 2% agarose gel was run for about 1 hour at 100 V to verify the PCR reaction. The first lane corresponded to the 100 bp DNA ladder marker, and the rest of the lanes to the different PCR products. Each PCR yielded only one PCR product. Their sizes were located between the 200 and 300 bp lanes of the DNA marker as they were expected.

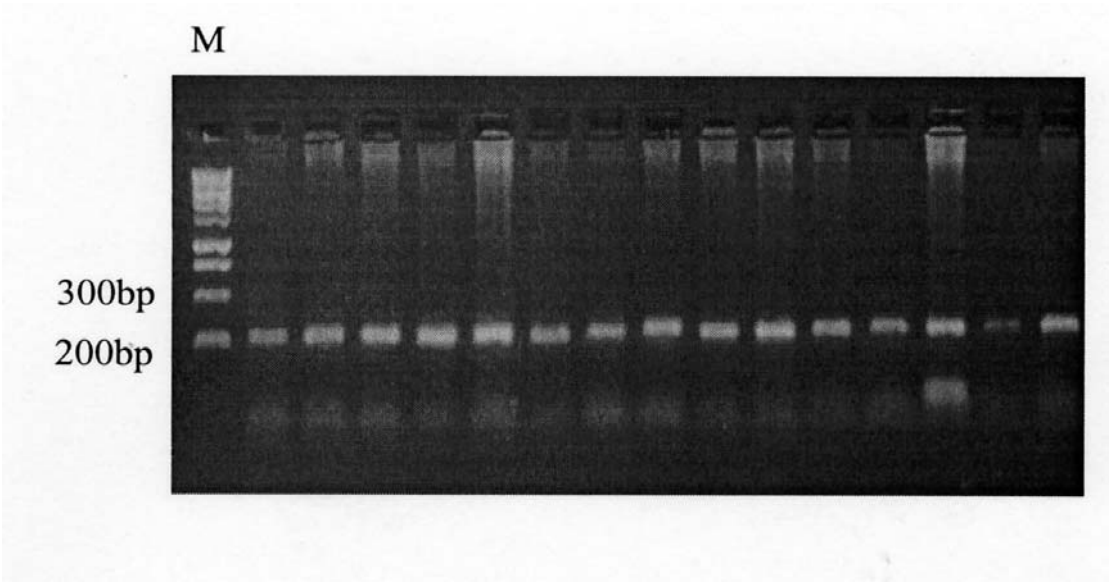

M
300bp
200bp

Figure 2. Gel image of the GeneScan polyacryamide gel.

The gel contained 36 lanes (not all the lanes are shown in the picture). Each lane was loaded with the fluorescently labeled PCR products (green color) along with the commercially available size standards GeneScan 500 (ROX), (red color). The gel image provided a qualitative overview of the run.

The picture in left corner shows the image of the first sample loaded on the gel (lane 1).

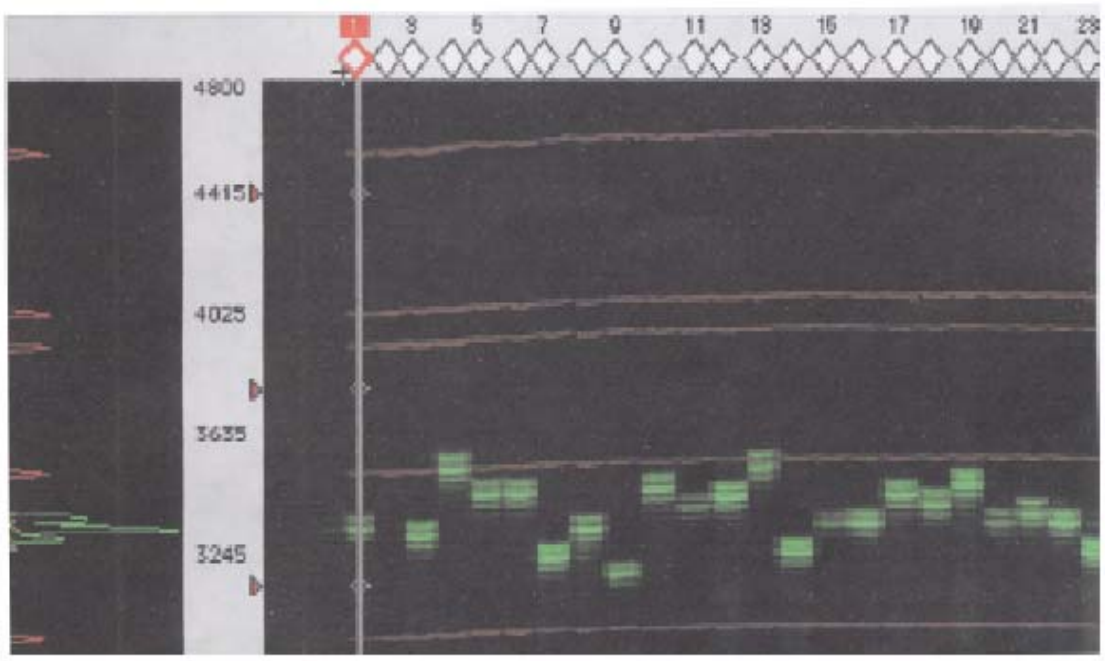
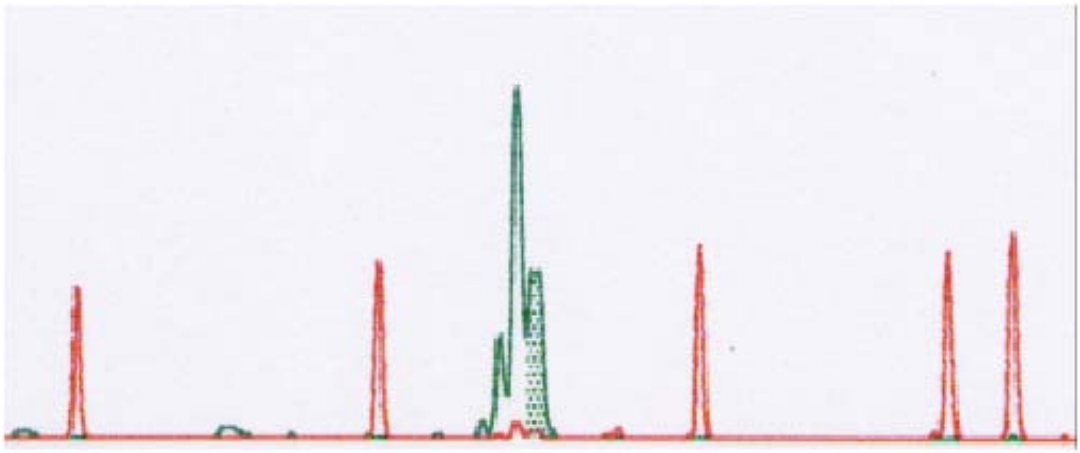

3
5
7
9
11
13
15
17
19
21
4800
4415
4025
3635
3245

Figure 3. Confirmation of the fragment size that flanks the CAG region by sequence analysis. Sample 1

A) GeneScan profile of the sample, and B) sequence

To confirm the accuracy of the GeneScan analysis for some of the samples the same region was sequenced. The size indicated by the third peak of the GeneScan electropherogram, highlighted peak, (in this sample of 273 bp) match with the number of base pairs obtained from the sequencing reaction. B) Shows the sequence of the same area containing 18 CAG repeats.

A)



B)

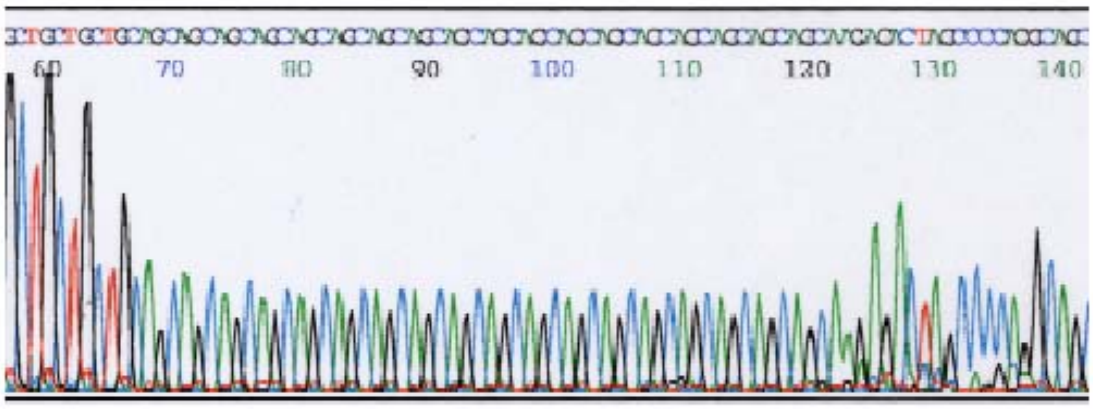
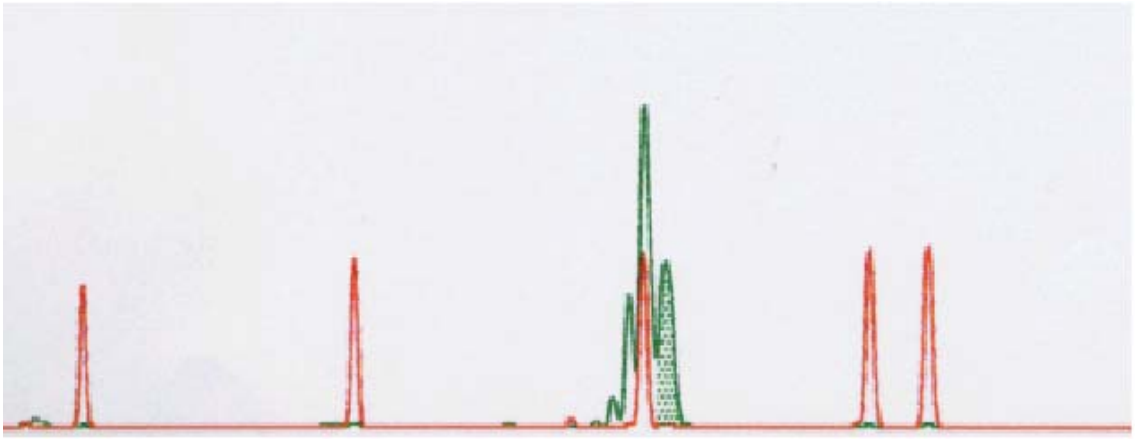


Figure 4. Confirmation of the size of the fragment that flanks the CAG region by sequence analysis. Sample 2

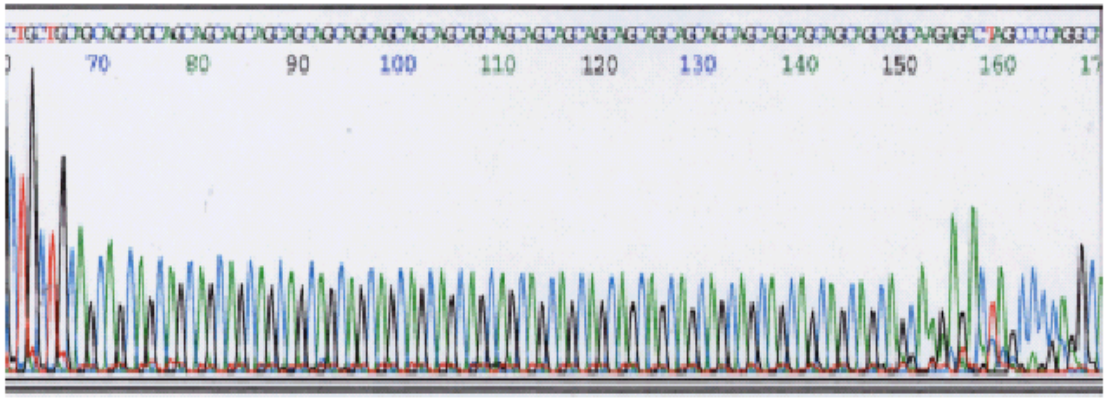
A): GeneScan profile of the sample, and B): sequence.

To confirm the accuracy of the GeneScan analysis some samples were sequenced. The size indicated by the third peak of the GeneScan profile, highlighted peak (303 bp) corresponds to the number of base pairs obtained from the sequence. B) Shows the area sequenced containing 28 CAG repeats.

A)



B)



1.3 Statistical analysis

The number of CAG repeats ranged in infertile males from 14 to 31 whereas the number in fertile controls ranged from 8 to 27.

A Q-Q plot was performed to assess if the data could be analyzed as normally distributed because despite the fact that the number of CAG repeats is a discrete variable it takes values over a reasonably large range and so it may be taken as a continous variable. The Q-Q plots in Fig. 5 indicate the assumption that the data is normally distributed could be made for each one of the three groups, normal fertile, oligospermic and azoospermic patients. The normal fertile controls appear less normally distributed that the other two groups of infertile patients but since it not too far away from normality is was possible to apply the statistical analysis based on ANOVA and *t*-test for all the groups

Mean values for the number of CAG repeats were obtained and a *t*-test was performed to compare them. Results indicated that the mean AR-CAG length of infertile subjects was significantly longer than fertile controls, 21.95±0.31 *versus* 20.72±0.52 respectively ($p=0.034$).

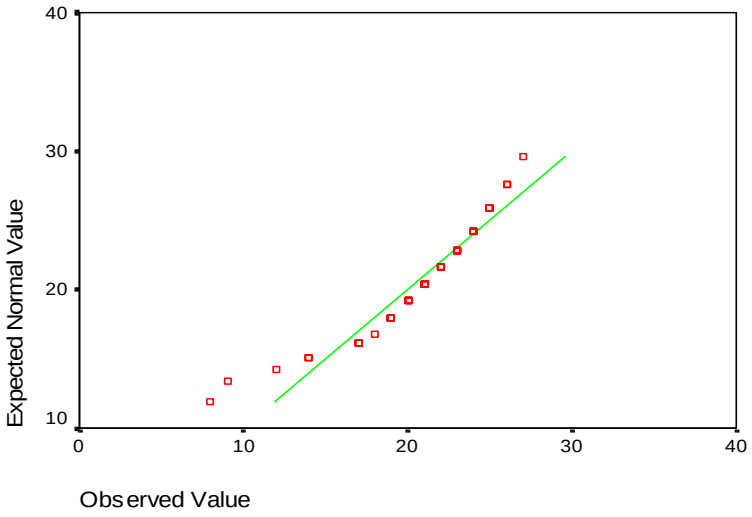
For further analysis, infertile patients were then classified according to their sperm count into two categories, oligospermic (<20 million sperm per ml) and azoospermic (no spermatozoa present). The oligospermic category included those cases with severe oligospermia and mild oligospermia. For this study they were combined because the number of patients with severe oligospermia was 16, quite low, and the mean value of the number of CAG repeats fairly similar to those patients with mild oligospermia, (21.12 *versus* 21.80).

Figure 5. Q-Q plots to evaluate if the sample was normally distributed.

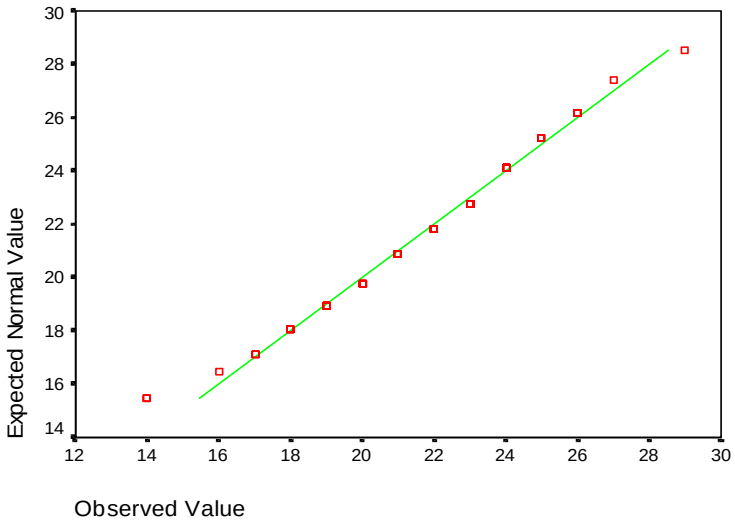
To perform the Q-Q plot analysis subjects were divided into the categories: normal fertile, oligospermic and azoospermic.

The observed quantiles values for the number of CAG repeats in each category were plotted against the expected quantile values. The straight line resulting from the plot indicates that the variable number of CAG repeats is normally distributed for each category.

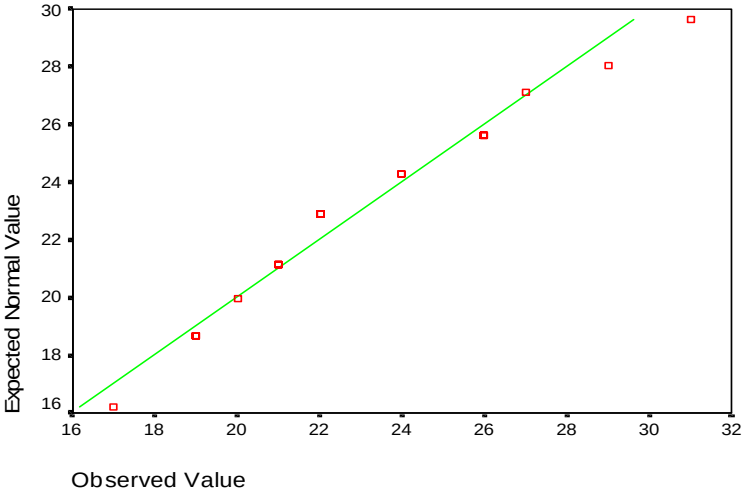
NORMAL FERTILE CONTROLS



OLIGOSPERMIC PATIENTS



AZOOSPERMIC PATIENTS



Anova and Duneett's tests were performed to establish multiple comparisons between the azoospermic and the oligospermic group against the fertile control group. Results are summarized in Table 2. Longer CAG repeat tracts corresponded to greater severity of spermatogenic defects, with the CAG repeat length increasing in the order: fertile controls <oligospermic < azoospermic. Azoospermic patients had the higher number of CAG repeats and the mean value was statistically different to that of the fertile control group (22.91 *versus* 20.72, p=0.019). Oligospermic patients also had an increased number of CAG repeats in their AR compared to the fertile population (21.65 *versus* 20.72, p=0.226), although this difference did not reach statistical significance.

1.3.1 Distribution of the CAG repeat length

Percentage values of the CAG repeat length in the infertile population were calculated and compared to those in the normal fertile group (Fig. 6, 7 and 8). The histogram in Fig. 8 shows the two samples, normal and azoospermic patients. The distribution of the azoospermic patients is shifted to the right of the distribution of normal patients. Moreover the distribution of azoospermic appears to be bimodal, whereas the sample of normal patients appears unimodal. The first mode of the azoospermic patients is close to the mode of the normal patients, whereas the second mode is at a higher number of CAG repeats

Table 2. Results of the Dunnett's test.

Comparison of the number of CAG repeats were established between the oligospermic and azoospermic group against the fertile control one.

BAYLOR COLLEGE	No. of cases	Mean±SE (CAG length)	Range	P-value [a]
Fertile controls	55	20.72±0.52	8-27	
Oligospermic [b]	72	21.65±0.34	14-29	0.226
Azoospermic	23	22.91±0.73	17-31	0.019

[a] Dunnett's test , in comparison to fertile control subjects

[b] Sperm counts less than 20 mil/ml

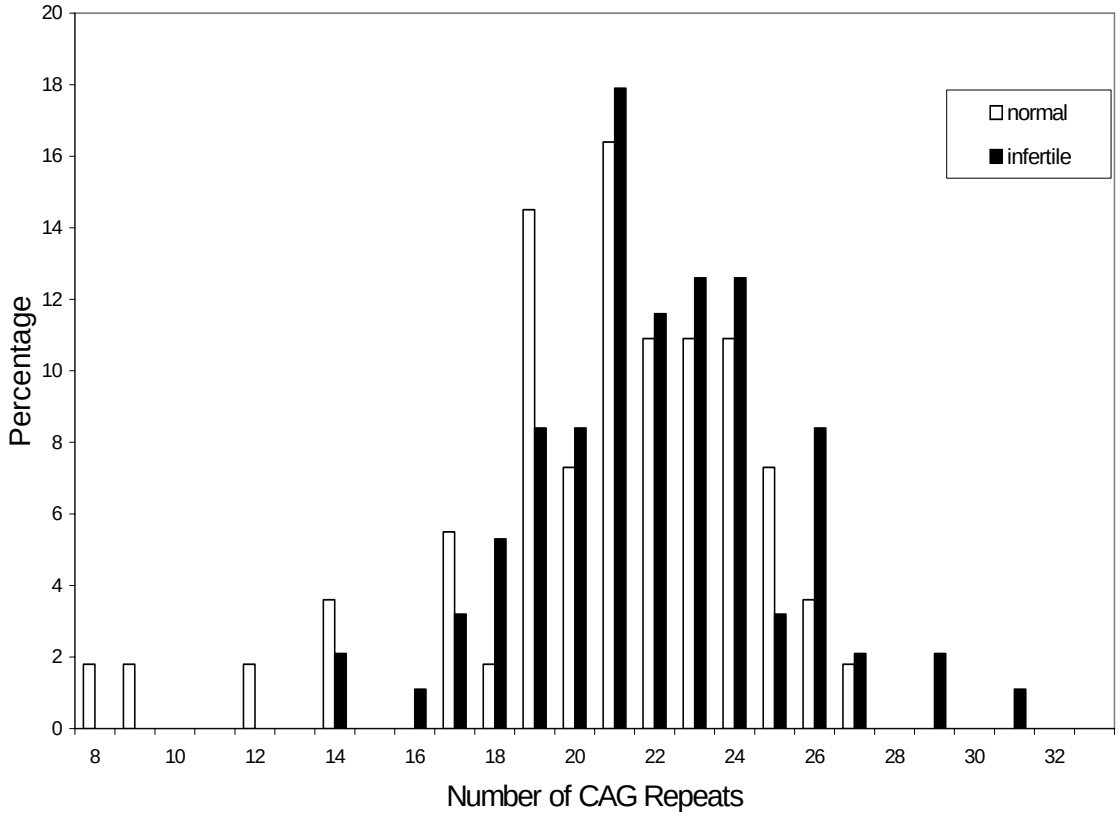


Figure 6. Allelic distribution of the AR(CAG)n in the fertile and infertile group of subjects.

Representation of the CAG repeat number allelic distribution in the normal fertile group (white bars), and the infertile group comprised of azoospermic and oligospermic patients (black bars).

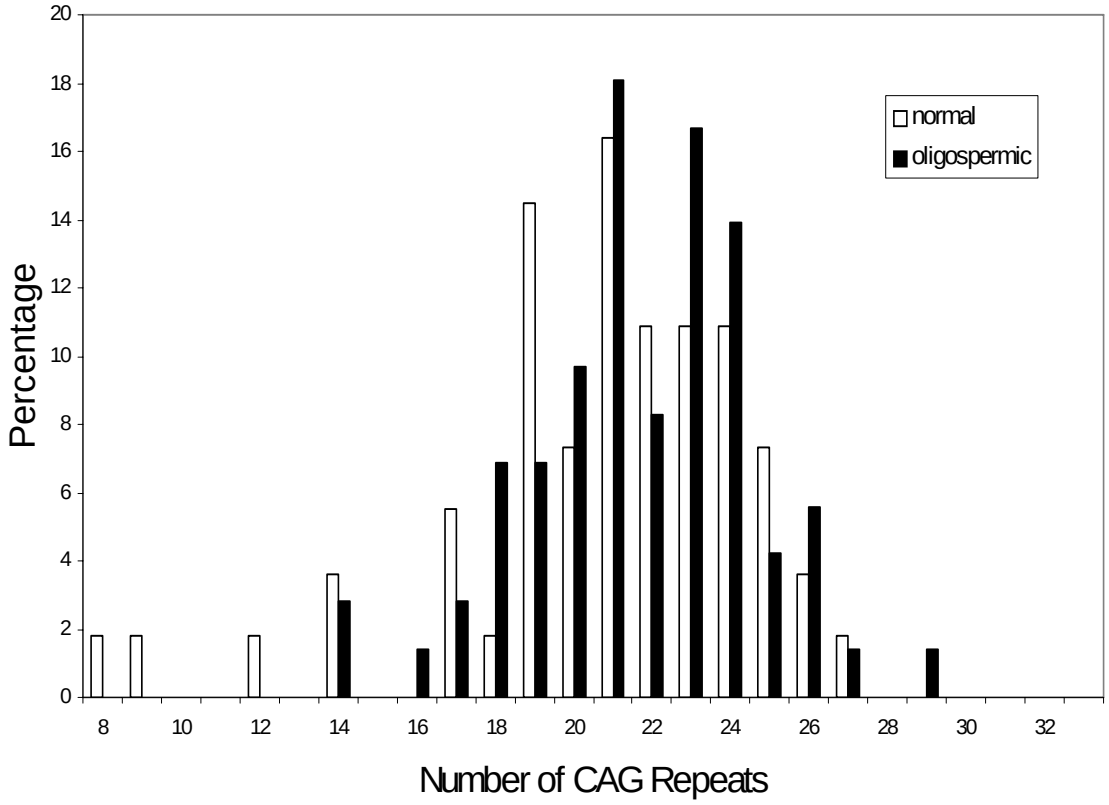


Figure 7. Allelic distribution of the AR(CAG)n in the normal fertile and the oligospermic group of patients.

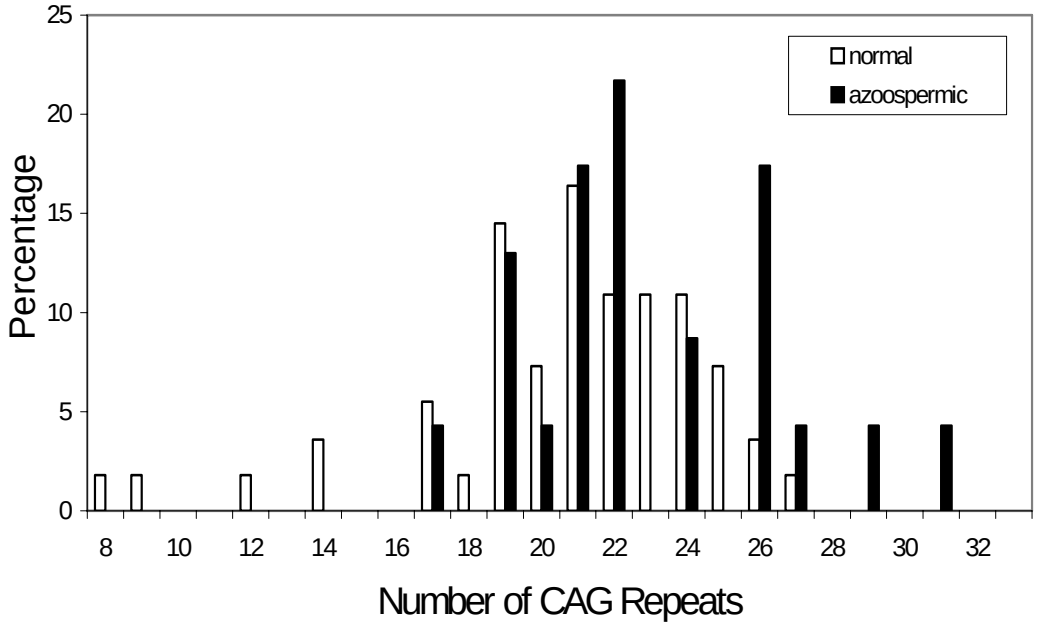


Figure 8. Allelic distribution of the AR (CAG)n in the normal fertile and azoospermic group.

The azoospermic group of patients (black bars) shows a bimodal distribution, whereas in the normal fertile group (white bars) the allelic distribution appears unimodal

1.3.2 Logistic regression analysis using the number of CAG repeats as the predictor, and clinical infertility or azoospermia as the outcome.

When logistic regression was performed using the number of CAG repeats as a continuous predictor variable, the odds ratio for each unit increase in CAG was exp(β)= 1.11 (95% CI:1.05-1.23; p=0.038) for the outcome of clinical infertility, defined as a composite of oligospermic and azoospermic subjects. When the outcome was the azoospermic group of patients, exp(β) increased to 1.20 (95% CI:1.02-1.40; p=0.0277). Thus, each unit increase in CAG length was associated with a 20% increase in the odds of being azoospermic.

1.3.3 Number of CAG repeats assessed as a categorical variable.

For the categorical analysis, the threshold of 26 CAG repeats was chosen because this size was about two standard deviation above the mean. When the subjects were divided using 26 CAG repeats as the cut off point normal, and the infertile group of patients compared no significant differences were encountered (OR=2.75, 95%CI: 0.747-10.108, p=0.128). However, when the normal *versus* the azoospermic groups were compared the differences between them were significant. Patients with ≥26 CAG repeats were 7-fold more likely to be azoospermic compared to those with < 26 CAG repeats (OR= 7.52, 95%CI: 1.75-32.78, p=0.006). The results imply that only those patients with azoospermia have significantly longer CAG repeats.

1.3.4 Protective effect of having short CAG repeats towards azoospermia.

From our previous tests we conclude that long CAG repeats increase the odds of being infertile. Another way to analyze this data is to study the protective effect towards infertility of having short CAG repeats. Using CAG repeats ≥ 26 as a reference group,

Table 3A compares the proportion of azoospermic *versus* normal patients between two ranges of CAG repeats. The first range is CAG repeats $\geq$ 26 compared to CAG repeats $\leq$ C, where the values of C are taken as 22, 21, 20, and 19. As we reduce C we observed smaller OR and smaller p values for the chi square test.

As we can see from Table 3A the odds of being azoospermic decreased with the number of CAG repeats. For instance, if the number of CAG repeats is $\leq$ 19, the OR of being azoospermic was 0.100 compared to those with $\geq$ 26 CAG repeats, (p=0.009). Similar results were obtained when the infertile population as a whole was considered and compared to the fertile group although the values did not reach statistical significance (Table 3B).

Table 3. Protective effect of short CAG repeats on the risk of male infertility.

A) Only the group of azoospermic patients was considered.

B) The whole group of infertile males was considered.

Results in both tables show the same pattern, the risk of infertility decreases as it does the number of CAG repeats. Only when azoospermic patients were considered (Table 3A) values were statistically significant.

A)

CAG LENGTH	≥26	≤22	≤21	≤20	≤19
OR [a]	1	0.194	0.150	0.102	0.100
CI [b]		(0.043-0.887)	(0.031-0.723)	(0.019-0.540)	(0.017-0.572)
P value		0.018	0.018	0.007	0.009

B)

CAG LENGH	≥26	≤22	≤21	≤20	≤19
OR [a]	1	0.35	0.37	0.30	0.26
CI [b]		(0.09-1.32)	(0.09-1.41)	(0.07-1.18)	(0.06-1.06)
P value		0.122	0.144	0.084	0.06

[a] Odds Ratio

[b] 95% Confidence Intervals

1.3.5 Varicocele and male infertility

The relationship between the presence of varicocele and male infertility has been studied, but the causes that originate varicoceles still remain controversial. The proportion of patients with varicocele was exceptionally high because Baylor is a referral center for this condition and therefore it was possible to study if there was any relationship between the number of CAG repeats and the presence of varicocele.

In order to address this question the following t-tests and logistic regression analysis were performed.

1. Mean CAG repeat values were compared between patients with and without varicocele. The mean value for the patients without varicocele was 22.07±0.39 and for those with varicocele was 21.97±0.53 (p=0.875), suggesting that the number of CAG repeats does not play a role in the aetiology of varicocele in infertile patients.

2. When the same test was performed among azoospermic patients, again no significant differences in the number of CAG repeats were found between patients with and without varicocele (23.75 *versus* 22.88, p=0.652). However, the number of azoospermic patients with varicocele was very small, only four.

3. When the number of CAG repeats was assesed as a continuous variable and the condition of varicocele as the outcome, a logistic regression test showed no significant correlation between the number of CAG repeats and the presence of varicocele, OR= 0.97, (95% CI:0.87-1.08, p=0.6028).

In line with other studies (Famiglietti *et al*., 2002; Schoor *et al*., 2001) a strong correlation was found between the presence of varicocele and the condition of

subfertility, being the OR= 2.55, (95% CI:1.13-5.791,p=0.027). When the azoospermic group was analyzed the OR increased to 7.79, (95%CI: 2.13-28.49; p=0.002).

1.4 Clinical characteristics of patients with ≥ 26 CAG repeats.

To determine if any particular phenotype would be related to long CAG repeats, the clinical characteristics of patients whose CAG length was 26 or more were analyzed. About 27% of Baylor patients had ≥ 26 CAGs, the point at which risk of infertility significantly increased. None of the patients with long CAGs had any neurological deficit or abnormal secondary sexual development. In particular, none were noted to have hypospadius, microphallus, gynaecomastia or abnormal sexual hair distribution. Among patients having ≥ 26 CAGs, 54% were azoospermic. The mean sperm counts of non-azoospermic patients were 6.5 mil/ml. Varicoceles were present in 46% of Baylor patients. This parameter was however not significantly associated with CAG length. The high prevalence of varicoceles in Baylor patients was due to the clinic being a quaternary referral centre for urological problems. The testicular volume of infertile patients with ≥ 26 CAG repeats was significantly smaller than that of the fertile controls and the mean FSH values were above the normal range. Table 4 summarises these results.

2. Analyses of patient population from the National University of Singapore

2.1 Characteristics of subjects

To compare differences in CAG repeat length between ethnic groups, 120 infertile and 87 control males from the Singapore Hospital were analyzed. The racial composition was predominantly of Chinese origin. The clinical characteristics of the patients are described in Table 5.

Table 4. Clinical characteristics of the patients with ≥26 CAG repeats in the AR gene.

Number of patients = 13	
Azoospermia	54%
Non-azoospermic patients	
Sperm Density (N: ≥20 mil/ml)	6.5±3.0
Motility (Normal: ≥50%)	38.4±9.9
Testicular volume (Normal:17.82±0.61ml).	13.0±1.6
Varicocele	46%
Mean FSH (Normal: 0.8-4.7 IU/L)	16.4±2.9

Table 5. Clinical characteristics of the infertile group of patients.

Number of infertile patients	120
Number of patients with azoospermia	33
Number of patients with oligospermia	87
Age (years)	35
Mean sperm density for oligospermic patients (Normal : $\geq$ 20 mil/ml)	5.02 mil/ml

2.2 Statistical analyses

The number of CAG repeats ranged from 14 to 33 CAG repeats in the infertile group and from 11 to 29 in the fertile controls.

To assess if the variable number of CAG repeats was normally distributed we performed a Q-Q plot. As in the case of the Baylor population the assumption that it was normally distributed was made. Results of the Q-Q plot are shown in Fig. 9

The mean value of the number of CAG repeats in the infertile population was 23.13 ± 0.28, slightly higher that the mean value of the normal fertile group which was 22.38 ± 0.32. In contrast with the American population this difference was not statistically significant, ($p=0.088$), although it was quite close.

Multiple comparisons using the Anova and Dunnett's tests were established to evaluate the differences among the different groups and to compare the oligospermic and azoospermic groups to the control one. Table 6 summarises these results. As in the case of the American population, no significant differences in the number of CAG repeats were encountered between oligospermic patients and fertile controls and were only observed when the azoospermic group of patients was compared to the control fertile group.

2.2.1 Distribution of the CAG repeat length

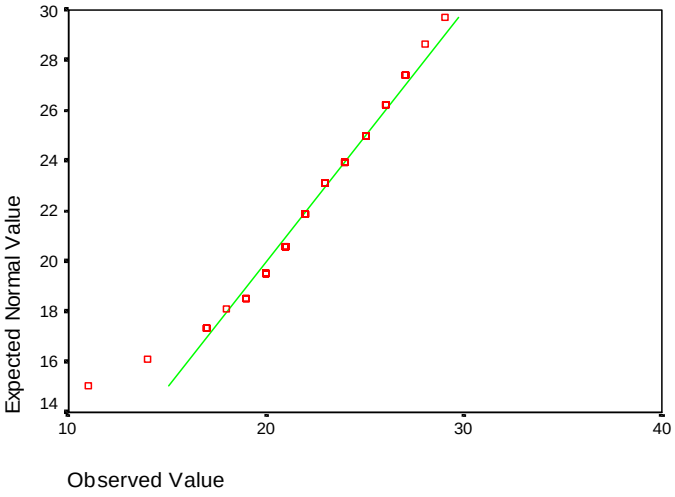
Percentage values of the CAG repeat length in the infertile and normal populations are shown in Fig. 10, 11 and 12.

Figure 9. Q-Q plots to evaluate if the sample was normally distributed.

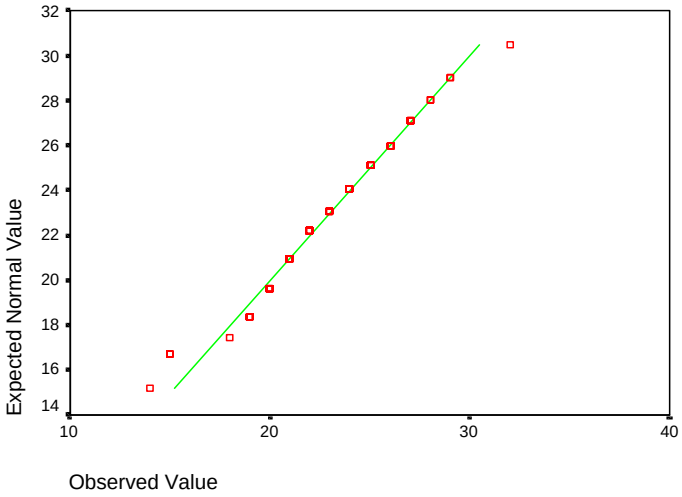
To perform the Q-Q plot analysis subjects were divided into the categories: normal fertile, oligospermic and azoospermic.

The observed quantiles values for the number of CAG repeats in each category were plotted against the expected quantile values. The straight line resulting from the plot indicates that the variable number of CAG repeats is normally distributed for each category.

FERTILE CONTROLS



OLIGOSPERMIC PATIENTS



AZOOSPERMIC PATIENTS

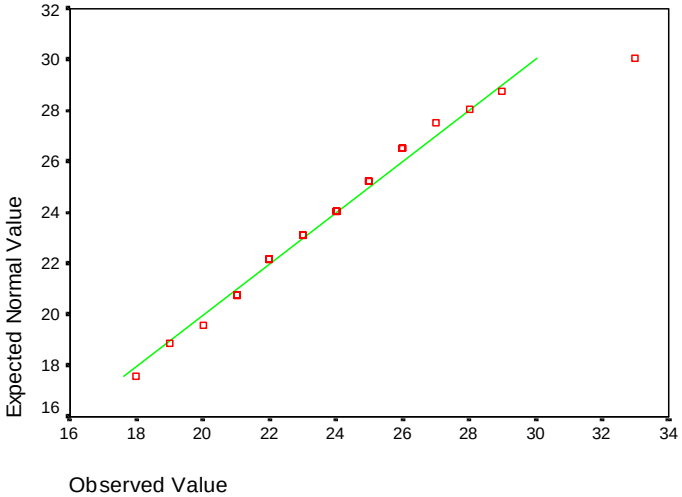


Table 6. Results of the Dunnett's test.

Comparison of the number of CAG repeats were established with the oligospermic and azoospermic group against the fertile control one.

SINGAPORE	No.of cases	Mean±SE (CAG length)	Range	P-value[a]
Fertile controls	87	22.38±0.32	11-29	
Oligospermic [b]	87	22.85±0.33	14-32	0.514
Azoospermic	33	23.82±0.52	18-33	0.043

[a] Dunnett's test after ANOVA, in comparison to fertile control subjects

[b] Sperm counts less than 20 mil/ml

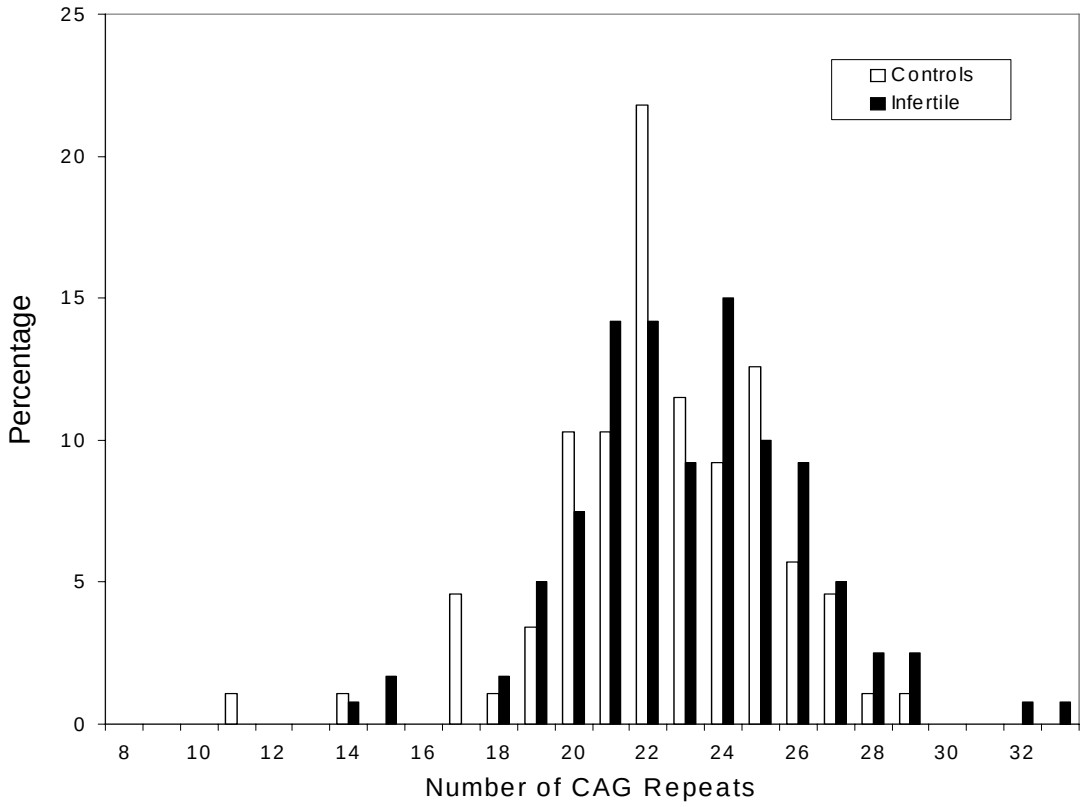


Figure 10. Allelic distribution of the AR(CAG)n in the control fertile group and the infertile group

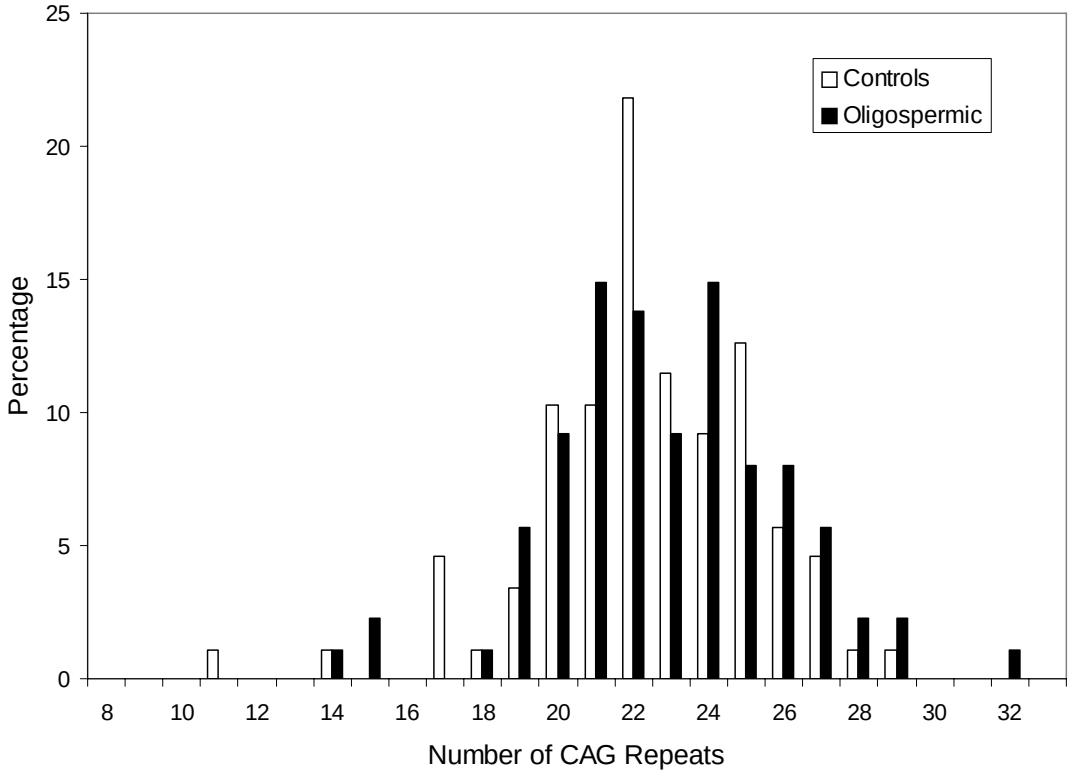


Figure 11. Allelic distribution of the AR(CAG)n in the control fertile and the oligospermic group of patients.

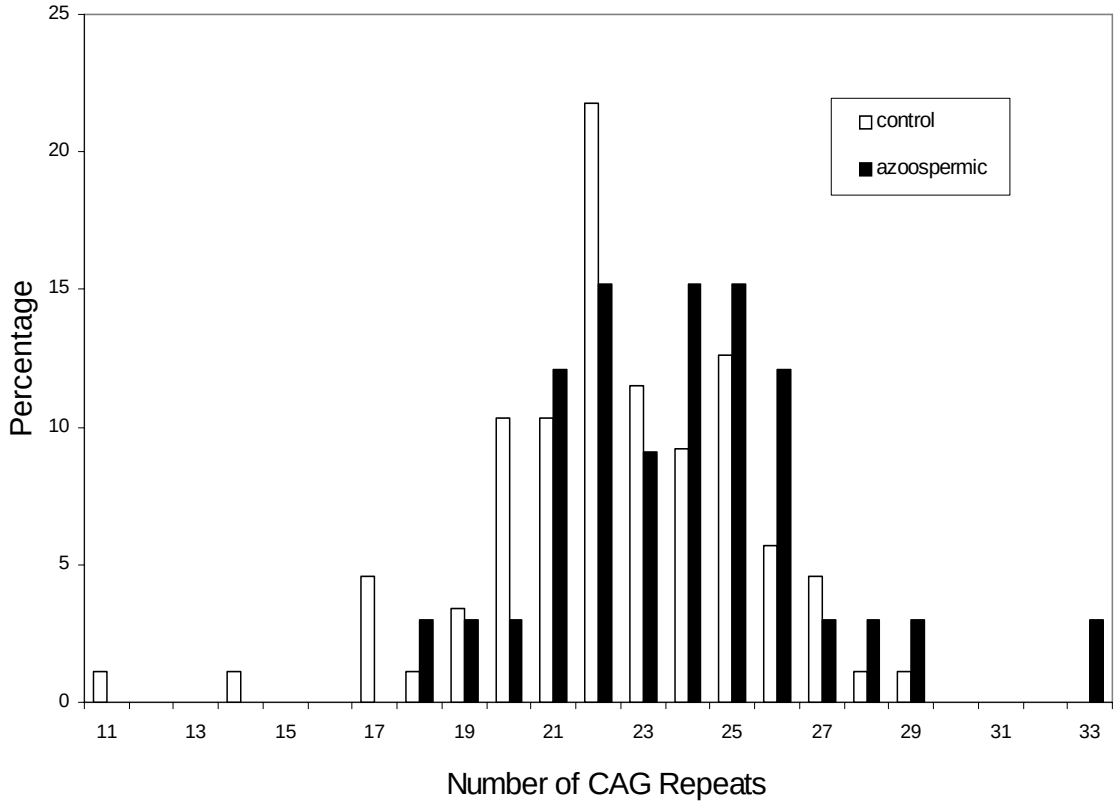


Figure 12. Allelic distribution of the AR(CAG)n in the control fertile group and the azoospermic group of patients.

There was a shift towards the right in the azoospermic group (black bars) respect to the control fertile sample (white bars), although the shift was less obvious than that of the American population.

Fig.12 shows that there is a shift towards the right of the azoospermic sample with respect to the normal sample, although the observed shift is smaller than the shift that was observed for the Baylor population. The bimodality of the azoospermic sample was also less pronounced than that observed for the Baylor population.

2.2.2 Logistic regression analysis using the number of CAG repeats as the exposure and clinical infertility as the outcome.

When we performed the logistic regression method to establish any possible correlation between the number of CAG repeats and the condition of infertility the value of exp(β) was 1.083 (95% CI:0.99-1.19; p=0.09). In contrast with the American population, when the group of infertile patients were analyzed as a whole, no significant relationship was observed between the number of CAG repeats and the condition of infertility. The logistic regression analysis when only the azoospermic patients was considered showed a value of exp(β) of 1.18 (95% CI:1.01-1.37;p=0.032). This result indicates a relationship between the number of CAG repeats and the condition of azoospermia and it is possible to predict that there is an increased risk for azoospermia of 18% for each unit increase in CAG length.

2.2.3 Categorical analysis

Results from the categorical analysis indicated that for the Singaporean population there was no single cut-off point at which the risk of being azoospermic increases significantly. Table 7 summarizes the results obtained from the various cut-off points considered in the categorical analysis.

Table 7. Different cut off points considered in the categorical analysis. For the Singapore study there was not any statistically significant CAG repeat number cut off point at which the risk of being azoospermic was significant

Cut off point	Exp(β)	95% CI	P-value
n25	1.75	0.73 - 4.14	0.20
n26	1.91	0.67 - 5.45	0.23
n27	1.90	0.50 - 7.25	0.34
n28	4.34	0.69 - 7.25	0.12

2.3 Clinical characteristics of patients with ≥ 26 CAG repeats.

20% Singapore patients had ≥26 CAG repeats. None of the patients had any abnormal secondary sexual development. In particular none of them had gynaecomastia, hypospadius, microphallus or abnormal hair distribution. Amongst patients having 26 CAG repeats or more 33% were azoospermic. The mean sperm counts of non-azoospermic patients were 4.1 mil/ml for Singapore patients and the mean testicular volume was significantly reduced to 13 mls. Varicoceles were present in 25% of Singapore patients, this parameter was however not significantly associated with CAG length as described previously. The mean FSH value of these patients was also slightly raised to 5.2±0.9 IU/ml. These results are summarised in Table 8.

Testicular biopsies of 14 patients with long CAG tracts were available. Histology was non-specific and included incomplete spermatocytic arrest, paucity of mature sperm, Sertoli cell only syndrome, tubular atrophy, hyalinization and fibrosis of seminiferous tubules. In one patient, all the above were observed in seminiferous tubules biopsied from different parts of the testes. Of the 7 azoospermic Singapore patients with biopsies, three had complete absence of spermatogenesis, while 4 others had some evidence of spermatogenesis. Azoospermia associated with long CAGs therefore does not preclude the possibility of spermatogenesis in their tubules.

Table 8. Clinical characteristics of infertile patients with ≥26 CAG repeats in the AR gene.

Number of patients = 24	
Azoospermia	33%
Non-azoospermic patients	
Sperm density (N: ≥20 mil/ml)	4.1±0.7
Motility (Normal: ≥50%)	32.0±5.6
Testicular volume (Normal:17.82±0.61ml).	13.1±0.8
Varicocele	25%
Mean FSH (Normal: 0.8-4.7 IU/L)	5.2±0.9

DISCUSSION

Analyses of over 350 patients and controls show a significant direct correlation between CAG length in the AR gene and severity of spermatogenic defect. The major observation in this study is that there is a subgroup of azoospermic men than have significantly longer AR-CAG alleles compared to fertile controls. The CAG length was measured accurately with the use of internal size standards in every lane, and allelic size assigned objectively by the GeneScan software. Subjects with $\geq$26 CAGs have a 7-fold higher risk of azoospermia compared to those with < 26 CAGs. Conversely patients with short AR-CAG alleles had a lower risk of male infertility. The shorter the tract, the lower the risk. Thus men with $\leq$19 CAGs had only 10% the risk for infertility compared to those with $\geq$26 CAGs. No clear phenotype was observed in patients with long AR-CAG alleles. About 25% of Singapore patients with long AR-CAG alleles had moderate to large varicoceles, consistent with occurrence rates reported for infertile men (De Kretser *et al*., 1999). A high incidence of varicoceles where observed in the Baylor population, both for patients with long and normal CAG lengths, because the Baylor clinic is a referral center for such conditions. Overall our data indicate that long CAG tracts were not significantly associated with varicoceles. Serum FSH was raised slightly in the Singapore population and more so in the Baylor population, these differences most likely arising because of referral patterns to the two centres. Although about 33-46% of patients with long CAGs were azoospermic, spermatozoa were present in the seminiferous tubules in the majority of azoospermic men, suggesting that basal spermatogenic function was preserved in even the most severely affected patients. It is tempting to speculate that boosting intra-testicular levels of androgens could be one method to improve spermatogenesis in these patients (Yong *et al*., 1994).

Ethnic differences in CAG length are well known; for instance, African-Americans have shorter CAG repeats compared to White-Americans (Edwards *et al*., 1992).

The present study compared AR-CAG allele distribution in two different ethnic groups, one predominantly Caucasian and the other Chinese. There was a clear ethnic difference in CAG length, with the mean CAG lengths in the Singapore subjects being 0.91 to 1.66 longer than the one corresponding American subjects. However, long CAG tracts were significantly associated with azoospermia in both populations, indicating that these differences were independent of ethnicity.

It is well known that prostate cancer is an androgen-driven tumour and androgen-ablation therapy is a commonly used treatment modality. There is evidence that spermatogenesis requires high concentrations of androgens (Zirkin *et al*., 1989). Patients with AR mutations that reduce intrinsic receptor activity have impaired spermatogenesis (Tsukada *et al*., 1994, Wang *et al*., 1998, Ghadessy *et al*., 1999). Conceptually prostate cancer and defective spermatogenesis may represent opposite ends of a spectrum of AR activity. In combination, these studies suggest that polymorphic changes in CAG length inversely regulate the function of the AR. This hypothesis would imply that lower risk of defective spermatogenesis would be associated with increased severity and earlier age of onset of prostate cancer, and vice-versa. Long-term cohort studies are required to validate this model of AR function.

Long AR-CAG alleles encoding long AR polyglutamine tracts are associated with low intrinsic AR activity in reporter gene assays (Mhatre *et al*., 1993, Chamberlain *et al*., 1994). A possible molecular mechanism as to how a change in polyglutamine length can affect the activity of the receptor has been suggested with the identification a novel

nuclear G-protein, Ras-related nuclear protein/ARA24, that acts as a coactivator with the AR and can bind differentially with different lengths of AR polyglutamines (Hsiao *et al*., 1999). AR-CAG/ARA24 interactions become stronger as the number of glutamines decreases, thereby increasing coactivation and AR transactivation capability. Recent studies have provided more insights into the molecular mechanism of how the CAG tract affects the functional competence of AR. The transactivation domain of AR is important for normal AR function, providing the essential AF-1 function and coactivator interactions. Irvine *et al* showed that with increasing CAG repeats length, p160-mediated coactivation (such as AIB1, SRC-1a, GRIP1) of AR is repressed (Irvine *et al.*, 2000). It was further demonstrated that the partial loss of AR function (with 65 CAG repeats) is due to decreased mutant AR protein that has been preferentially targeted for degradation via the ubiquitin-proteasome pathway (Lieberman *et al.*, 2002). The mutant receptor also undergoes altered post-translational modifications such as hyperacetylation and phosphorylation, thereby targeting AR for ubiquitylation and degradation (Lin *et al.,* 2002).

Although the likely increase in AR intrinsic activity with each reduction in AR-CAG length is relatively small, these effects are genetically determined and therefore exert its effects over the entire lifetime of the individual. Small changes can over time have significant cumulative pathological effects.

Collectively, the evidence supports the hypothesis that the glutamine repeat has a role in AR function by fine-tuning the balance between excess and deficient receptor function. The highly polymorphic nature of the glutamine repeat would imply a subtle

gradation of AR function among individuals, possibly allowing alleles with evolutionary advantages to be rapidly selected and transmitted to future generations.

Recently, a number of studies have addressed the question of the importance of an expanded CAG repeat track and male infertility. Not all their conclusions are in agreement with the findings of the study presented in this thesis. Strikingly different is that of Komory *et al*. (1999) in which their researchers described that infertile males have higher chances of having less than 16 CAG repeats in the AR gene than the normal fertile individuals. This study was done on a small number of patients, only 59, without specifying how many of the infertile patients were azoospermic or oligospermic. Recently Sasagawa *et al*. (2001) reported in a Japanese population that no significant differences were found in the number of CAG repeats, when they compared the AR of 30 men with idiopathic azoospermia and 51 fertile males.

Other studies performed by Dadze *et al*. (2000), conducted on a Caucasian population, did not find significant differences in the number of CAG repeats between the infertile and the fertile group, although the sample size was reasonable large (119 patients and 22 controls). This study included only 22 controls and therefore may not had the power to prove significance. Likewise von Eckardstein *et al.* (2001) did a study in a German population and no differences where encountered in the number of CAG repeats between infertile and fertile. Nevertheless in a multivariate analysis the CAG repeat length was significantly correlated to sperm concentrations with a coefficient of –0.25 only in the fertile group. In a Swedish population (Giwercman *et al.* 1998) there was no association between the number of CAG repeats and impaired sperm production, neither did a study performed in the Netherlands by Van Golde *et al*., (2002). It is

worth mentioning that in this last study the number of azoospermic patients was only of ten, from a total of seventy-five subfertile patients. Combined data from previously published European studies also did not find any difference in the CAG repeat length distribution between fertile and infertile males (Rajpert-De Meyts *et al.*, 2002). However there were discrepancies in the various studies analyzed. For instance two studies of German populations (Dadze *et al*., 2000; Hiort *et al*., 1999) reported quite markedly different means of CAG repeat length in their small control groups (20.8 and 24.0). Since the significant differences between patients and controls were usually one to two CAG lengths, it is possible that differences in analytical techniques and patient selection have contributed to the discrepancies observed between the European meta-analyses (Rajpert-De Meyts *et al.*, 2002) and our data. Other investigators suggest that differences observed could be due to the ethnic origin (Van Golde *et al*., 2002;), since the majority of the studies performed on European populations did no found an association between the CAG repeat length and subfertility.

Nevertheless, the data presented in this thesis suggest that the differences encountered are not so much dependent on the ethnicity but on the design of the study. We had obtained significant results from our two ethnically diverse populations, one a Caucasian population from the US and the other a predominantly Chinese one from Singapore. The selection criteria employed to choose the cohort of patients in our studies was very selective. We excluded patients whose cause of infertility might be other that the one being investigated. In the study presented in this thesis, the exclusion criteria were patients diagnosed with chromosomal disorders, deletions in the Y-chromosome, AR mutations, infective or obstructive syndromes of the genital tract,

hypopituitarism and hyperprolactinemia. The significant association between long CAG and azoospermia was found in both US and Singapore subjects.

In line with our study, is that of Dowsing *et al*. (1999) performed on an Australian population composed of mainly Caucasians. The researchers also found significant differences in the number of CAG repeats between infertile patients and the fertile controls (24.0±0.7 *versus* 20.5±0.3) despite the sample size being very small, composed of only 30 patients with idiopathic infertility and 32 controls. The large difference found in the number of CAG repeats between the fertile and infertile group could be due to the fact that the infertile population was composed of 23 patients with severe oligospermia and azoospermia, while only 7 of those infertile patients had oligospermia. In a previous study, Tut *et al.* (1997) in a predominantly Chinese population from Singapore demonstrated that the longer the CAG tract the greater the severity of the spermatogenic defect. This is also in concordance with our study, which demonstrated that longer CAG tracts were mainly associated with azoospermia. Likewise, a study performed on a Japanese population by Yoshida *et al*. (1999) showed that patients with azoospermia had longer CAG tracts in their AR gene. The study included 41 patients with idiopathic azoospermia and 48 normal fertile males, and concluded that the mean values were 26.5±3.5 and 23±2.9 for infertile and fertile males respectively, (p=0.0013). They also reported that 4 of the patients with idiopathic azoospermia had 34 CAG repeats and in contrast none of the fertile males in the control group had more than 31 CAG repeats in their AR gene. Patrizio *et al*., (2001) performed a study in a Caucasian population from US and also reported that there were significant differences in the number of CAG repeats when the fertile and infertile groups were compared (p=0.03). Likewise longer CAG repeats tracts were

encounter in infertile Israeli men when compared to their fertile counterparts (18.6 ± 3.0 *vs* 16.6 ± 2.7). It is important to notice that out of 61 patients 46 were azoospermic and 15 were severe oligospermic. In a French study by Wallerand *et al*. (2001), the authors found an association between the CAG length and infertility. The selection criteria to include patients in the study was very strict and excluded those with hypogonadotropic hypogonadism, hyperprolactinemia, obstructives syndromes of the genital tract, microdeletions of the long arm of the Y chromosome, mutations of the CFTR gene and karyotypic abnormalities. Moreover the proportion of azoospermic males was of 76% from a total of thirty-seven infertile males. In a Spanish population Mengual *et al*., (2003) measured the AR CAG repeat length of 102 patients with azoospermia and compared it to that of 96 fertile controls being significantly longer in azoospermic patients than in the control group. In line with our study, these investigators indicated that the number of 26 CAG repeats could be used as a cut-off point and showed that men with more than 26 CAG repeats have 4.09 greater risk of being azoospermic. Other investigators such as Kukuvitis *et al*., (2002), in a Greek population, performed a study on the contribution of genetic polymorphisms in the oestrogen receptor alpha gene (XbaI) as well as the AR CAG tract on male infertility. Although they did not find a difference in the number of CAG repeats between the fertile and azoospermic population, they did encountered a higher frequency of larger CAG repeats in patients with idiopathic oligospermia and azoospermia than those of the fertile group. The authors suggested that a possible synergy might exist between unfavourable genotypes between these two genes (ER and AR) in male infertility. In a recent report by Casella *et al*., (2003) in a predominantly white population, the CAG repeat tract was analyzed in seventy severely infertile patients without genital obstruction who were undergoing testicular biopsy (37 with Sertoli-cell-only

syndrome, 15 with maturation arrest, and 18 with hypospermatogenesis) Results indicated that only patients with hypospermatogenesis had longer CAG repeats than the fertile controls, suggesting a contributory role of these expanded polyglutamine tracts in the aetiology of hypospermatogenesis. Strikingly, longer AR CAG repeats are also associated with moderate to severe undermasculinization in male infants (undermasculinized: n=78, median 25, range 23-26; control: n=850, median 23, range 22-26, P = 0.002) (Lim *et al.*, 2000). The last piece of data from an entirely different paediatric patient population unrelated to male infertility strongly suggests the concept that CAG repeat length has a role in the intrinsic activity of the androgen receptor. Table 9 displays a list of publications that investigated the role of the CAG tract length and male infertility.

It is likely that the genetic background and other environmental influences may contribute to AR activity and sperm production. Spermatogenesis is a very dynamic biological process and can be affected by a wide number of factors such as reactive oxygen species (ROS), which could damage not only the sperm plasma membrane but also sperm DNA (Aitken *et al.,* 1998), exposure to toxins, fertilizers and herbicides (Kenkel *et al*., 2001). Expansions of unstable trinucleotide repeats in a number of other gene loci also lead to many other non-reproductive hereditary disorders. Examples include CGG or CCG expansion in fragile X syndrome (O'Donnell *et al*., 2002), CTG expansion in myotonic dystrophy type 1 (DM1), CAG expansions in Huntington disease (HD), dentatorubral-pallidoluysian atrophy (DRPLA), spinocerebellar ataxias (SCAs), and Machado–Joseph disease (MJD).

Table 9. Studies to determine if the CAG polymorphic tract predisposes to male infertility

STUDY GROUP	No. CASES (infertile /control)	MEAN CAGs INFERTILE Mean (range)	MEAN CAGs CONTROL Mean (range)	SIGNIFI-CANCE? (p VALUE)	REFERENCE	COMMENTS
Singapore-Chinese	153/72	No determined	No determined	No determined	Tut et al., 1997	Patients with ≥28 CAG had 4 fold increased risk of azoospermia
Swedish	33/294	21.9(16-27)	23.2(8-30)	NS	Giwercman et al.; 1998	
Australian	35/32	23.2(15-34)	20.5(17-25)	S (0.0001)	Dowsing et al., 1999	
Belgian	223/181	21(15-30)	21(14-29)	NS	Legius et el., 1999	
Japanese	41/48	26.5(20-34)	23.9(17-30)	S (0.0013)	Yoshida et al., 1999	
German	119/22	22(16-34)	20.8(15-26)	NS	Dadze et al., 2000	
UK	78/850	25(23-26)	23(22-26)	S (0.002)	Lim et al., 2000	
US	95/55	21.95(14-31)	20.72(8-27)	S (0.034)	Mifsud et al., 2001	
Sinagporean-Chinese	120/87	23.13	22.38	NS	Mifsud et al., 2001	azoospermic vs control was significant
German	62/62	20.7(17-27)	19.4(15-28)	NS	Eckardstein et al., 2001	CAG repeats inversely correlates with sperm count in normal males
Japanese	30/51	23.4(19-30)	23.7(17-28)	NS	Sasagawa et al., 2001	
Spanish	102/96	23.25	22.42	S	Mengual et al., 2003	Azoospermic vs control
French	37/50	23.91(13-28)	22.2(17-27)	S (0.008)	Wallerand et al., 2001	
Danish	119/110	21.5(15-29)	21.8(14-33)	NS	Rajpert-De Meyts et al., 2002	
Dutch	75/70	22.2	21.7	NS	Van Golde et al., 2002	
Greek	109/64			NS	Kukuvitis et al., 2002	Higher frequency of long CAG in infertile

Although exactly how these repeat expansions cause disease is not clear, a hallmark of these diseases is `anticipation'; i.e. the severity is greater and/or the age of onset becomes earlier in successive generations within the family (Timchenko and Caskey, 1996). The clinical severity of these diseases is positively correlated with the length of the expanded trinucleotide repeats. Since the repeat length increases as it is transmitted, it is thought that the molecular basis of anticipation is meiotic instability of the trinucleotide repeat. In addition to instability during gametogenesis, these trinucleotide repeats are also unstable in somatic cells, resulting in somatic heterogeneity (Timchenko and Caskey, 1996).

Taking these observations into consideration, Pan *et al*. (2002) examined the number of repeats at the loci DM1, MJD, DRLA, AR, and SCA8 of 48 patients with azoospermia, and compared them to 47 control subjects. Results indicated that the median CAG length in the AR gene was significantly longer in azoospermic patients than in controls (23 *vs* 21, p=0.001) and in addition increased (CTG/CAG) length were encountered at DM1 and MJD loci in a subset of azoospermic patients. Thus, it is noteworthy to evaluate whether offspring of these azoospermia patients, if born by assisted reproductive technologies, have an increased risk of trinucleotide repeat diseases.

The important role of the genetic component in and male infertility has in recent years become clearer. Congenital bilateral absence of the vas deferens (CBAVD) is a form of male infertility in which mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene have been identified. The molecular basis of CBAVD is not completely understood. Although patients with cystic fibrosis have mutations in both

copies of the CFTR gene, most patients with CBAVD have mutations in only one copy of the gene. Cystic fibrosis is important in the European population accounting for at least 6 percent of cases of obstructive azoospermia and being responsible for 1 to 2 percent of cases of infertility in men (Jequier *et al*., 1985)

Y-chromosome microdeletions were detected and confirmed in 6.8% of azoospermic and in 3.5% of severely oligozoospermic patients in our local population (Liow *et al*., 1998). However these patients with Y-deletions and long CAG patients in our study did not overlap. Thus our studies on CAG repeat length adds to the armamentarium of tests available to pinpoint the diagnoses in cases previously labelled as "idiopathic".

The additional data increases information for patients and enable them to understand the cause of male infertility and low sperm counts in subjects with azoospermia or severe oligospermia. Modern assisted reproductive technologies, utilising sperm obtained by testicular biopsies and the intra-cytoplasmic sperm injection procedure (ICSI), could result in viable pregnancies and the transmission of this trait. Screening for long AR-CAG tracts and appropriate genetic counselling would be important for these individuals.

CHAPTER 2: CAG REPEAT POLYMORPHISM IN THE AR GENE AND POLYCYSTIC OVARIAN SYNDROME (PCOS)

INTRODUCTION

1. Oogenesis

Reproduction in humans depends on the production of fertilizable oocytes by oogenesis. This is a discontinuous process that begins in the foetus and ends once puberty is reached. Oogenesis takes place in the ovary. The ovary is composed of three separate functional units, the stroma, the follicles and the corpus luteum. Follicles at all stages of development are present in the ovaries at any time until menopause. They contain the oocytes and serve as the vehicles for ovulation as well as producing steroid hormones. The mature Grafian follicle is composed of several layers of cells surrounding the oocyte. Two type of cells are present: the granulosa cells that maintain the conditions in the antral cavity which permit development of the oocyte, such as the transfer of nutrients and hormones, and the theca cells located in the outermost layers. The theca cells are supplied by a rich network of capillaries and synthesise mainly androgens.

2. Hormones involved in folliculogenesis

Under stimulation by LH, the theca cells synthesize mainly androgens. Granulosa cells of mature follicles have a very high aromatase activity and convert most of the androgens produced by theca cells into oestrogens. In this way the intrafollicular enviroment of the follicle is kept highly oestrogenic. The activity of the aromatase enzyme is stimulated by FSH. LH and FSH are both secreted by the same cell in the anterior pituitary gland in response to stimulation by the gonadotropin-releasing

hormone (GnRH). The secretion of GnRH by the hypothalamic neurons is intermittent and hence the release of gonadotrophins is pulsatile in nature. The rising level and increased frequency of LH discharge that occurs throughout the follicular phase of the cycle is responsible for stimulating this increase in estrogen secretion from the dominant follicle. The levels of FSH decrease as the pre-ovulatory follicle matures. When the dominant follicle has reached maturity, the secretion of oestradiol is sufficient to induce a positive feed back effect and a massive discharge of pituitary LH occurs. LH then induces a series of changes in the structure of the follicle as well as the termination of the meiosis by the oocyte.

3. Definition of PCOS and selection criteria of patients

Polycystic ovary syndrome (PCOS) is a heterogeneous endocrine disorder, which is considered to be the most common cause of anovulatory infertility and hirsutism (Adams *et al*., 1986). It occurs in approximately 4% of women of reproductive age. It was first described by Stein and Leventhal in 1935, although the association between masculinization and abnormal carbohydrate metabolism was first described by Achard and Triers in 1921, who called it “The diabetes of bearded women”.

There are many clinical, biochemical and metabolic manifestations of this disorder that make the criteria for disease diagnosis a very complex issue. The most widely accepted definition of (PCOS) is the association of anovulation (manifesting itself as irregular menses, oligomenorrhoea or amenorrhoea), hyperandrogenism, (with variable degrees of cutaneous symptoms such as acne, seborrhea and hirsutism) and ultrasonic evidence of polycystic ovaries. The last criteria has become important and many researchers now use it as a powerful tool to establish a diagnosis. Sonographically diagnosed

polycystic ovaries are enlarged, contain an increased number of follicles, or exhibit an increased amount or density of stroma (Takahashi *et al.,* 1993). Patients with PCOS may not show all the symptoms described above and normal healthy individuals may have some of these characteristics.

Since the definition of PCOS is very controversial, epidemiological studies on PCOS patients should clearly specify the selection criteria followed when selecting the patients as well as the ethnic background. It appears that there is some variability of clinical manifestations among races. For instance obesity and hirsutism are not pronounced in Japanese PCOS patients (Taketani, 1990).

Some specialists have attempted to establish subgroups among PCOS patients to explain the variety of symptoms encountered. For instance, Balen *et al*. (1995) proposed two main groups based on the clinical phenotype that encompasses about 85% of the wide range of manifestations. Among the first group the main clinical manifestations are hyperinsulinaemia, oligomenorrhea, obesity and hyperandrogennaemia. The women in the second group have high levels of LH, tend to be slim, with regular menses, but they have increased infertility and rate of miscarriage. Several authors failed to divide their PCOS patients into these two distinct groups (Grulet *et al*., 1993; Tropeano *et al*., 1994) because the gonadothrophic and metabolic disturbances are inter-linked. For instance, *in vitro*, the hormone insulin has been shown to stimulate androgen secretion directly and to enhance LH-mediated responses in isolated theca cells to a greater degree than in normal ovaries (Bergh *et al*., 1993). *In vivo*, the frequent coexistence of elevated LH and insulin concentrations leads to more severe expression of the syndrome (Fulghesu *et al*., 1999).

4. The polycystic ovary (PCO) and the mechanism of anovulation.

The polycystic ovary (PCO), on macroscopic appearance, typically contains 10 or more follicles ranging in size between 2-10 mm. In addition to the increased number of antral follicles there is twice the number of primary and secondary follicles in the PCO compared with the normal ovary. This suggests that there is an underlying disorder of folliculogenesis with increased follicular recruitment, but an arrest of follicular development at small antral stages

The mechanisms of anovulation in PCOS still remain obscure in spite of being the most common cause of anovulatory infertility. Anovulation is characterised by the arrest of antral follicle development at the 5-10 mm stage and the failure to enter the preovulatory phase of the cycle (Franks *et al*., 1988). Spontaneous ovulation can occasionally occur and the disorder can be reversed in most cases by treatment that increases serum concentrations of FSH, such as antiestrogens or exogenous gonadotrophins (Hamilton-Fairley *et al*., 1991).

The granulosa cells of follicles from anovulatory women with PCOS remain steroidogenically active and actually show increased aromatase activity when compared with folicles of similar size from women with normal ovaries. The production of oestradiol is therefore increased. Steroidogenesis by theca cells is abnormal in both anovulatory and ovulatory PCOS subjects. They produce higher levels of androestenedione, 17α hydroxyprogesterone and progesterone (Gilling-Smith *et al*., 1994).

Other forms of chronic anovulation not associated with hyperandrogenism can result in cystic ovaries, like multifollicular ovaries. The differences are that such ovaries do not have the peripheral distribution of ovarian cysts or the increased stroma characteristic of PCOS.

There are also differences between the presence of ultrasound-detected multicystic ovaries and PCOS (Hamilton-Fairley *et al*., 1991). The former occurs when characteristic ultrasound features are detected in a woman with a regular menstrual cycle and there are no symptoms or signs of hyperandrogenism. Actually up to 22% of the population may show ultrasound evidence of PCO although the majority will have regular ovulatory cycles. The latter exists when the ovarian ultrasound appearance is detected in a woman with menstrual irregularities, hyperandrogenism and perhaps obesity. Moreover, approximately 10% of women with all the features of the endocrine syndrome have normal ovaries as evidenced by ultrasound examination.

5. Several markers of the condition that can be considered as diagnostic criteria

5.1 Excess of androgens

The ovaries and the adrenal glands secrete androgens in response to their respective hormones, LH and ACTH. Ovarian hyperandrogenism is found in about 70% of patients with PCOS while adrenal hyperandrogenism is found in about 50% of the patients. If they are elevated serum T levels are between 2.4 and 4.1 nmol/L. The most likely cause of increased production of androgens from both glands seems to be the abnormal regulation of the 17-hydroxylase and the 17,20-lyase activities of P-450c17, the rate-limiting enzyme in the biosynthesis of androgens. The deregulation does not appear to be the result of a mutation in the gene as was suggested by Carey *et al*.

(1993), and it is likely to be due to hyperinsulinaemia, elevated LH levels and/or serine/threonine phosphorylation of the receptor itself, which increases its 17, 20-lyase activity (Ehrmann *et al*., 1995).

Hyperandrogenism manifests itself in the presence of acne, excessive hair growth and sometimes acanthosis nigrans, which is characterized by symmetric gray-brown hyperpigmentation of the skin affecting the nape of the neck and axillae. The presence or absence of hirsutism depends on whether androgens are converted peripherally by 5-alpha reductase to the more potent androgen dihydrotestosterone (DHT) and 3 alpha diol-G .

5.2 Changes in concentrations of LH and FSH and prolactin.

Altered gonadotrophin secretion, which in the past was considered to be the causal factor in the pathogenesis of the syndrome (Mais *et al*., 1986), is no longer considered a universal finding in PCOS although it helps physicians in establishing the diagnosis. The basal levels of LH and the frequency and amplitude of LH pulses are increased to levels usually above 20 IU/L. In contrast, the level of FSH is in the low-normal range. Actually additional amounts of FSH given to PCOS patients are sometimes sufficient to induce ovulation. It has been found that a LH/FSH ratio greater than 3, provided the LH level is not lower than 8mIU/ml, may be used for the diagnosis of women with clinical features of PCOS. Women with PCOS have mildly elevated levels of prolactin (20-30ng/ml), possibly related to the increased pulsatility of GnRH.

5.3 Pituitary abnormalities

Abnormalities of GnRH secretion exist in a significant subset of patients and leads to persistent hypersecretion of LH, which seems to be an important component of the syndrome, particularly in non-obese patients (Marshall *et al*., 1999). Suppression of the pituitary-ovarian axis by the chronic administration of agonist and analogues of GnRH results in a fall in circulating androstenedione and testosterone.

5.4 Metabolic disturbances. Hyperinsulinemia

In addition to the abnormalities of the pituitary ovarian axis, PCOS is characterized by significant metabolic abnormalities. These include fasting and glucose-stimulated hyperinsulinaemia, peripheral insulin resistance (affecting predominantly muscle and adipose tissue), abnormalities of energy expenditure and dyslipidaemia. Hyperinsulinemia in PCOS is the result of insulin resistance as well as decreased insulin clearance. These abnormalities, together with obesity, explain the increased prevalence of glucose intolerance in PCOS patients.

Obesity is present in about 25-50% of the PCOS population, while hyperinsulinemia associated with insulin resistance is present in more than 50% of PCOS patients. It is well known that obesity, and especially the accumulation of abdominal fat (android fat distribution), is associated with insulin resistance and when the pancreatic function is intact, insulin resistance will result in compensatory hyperinsulinaemia. PCOS represents a major risk factor for non-insulin dependent diabetes mellitus (NIDDM). The prevalence of impaired glucose tolerance or frank diabetes in young obese women with PCOS ranges (depending on the population studied) between 11% and 38% (Dunaif *et al*., 1993). Controversy exists as to whether women with PCOS have an

increased prevalence of gestational diabetes mellitus and larger studies are required to address this issue (Lanzone *et al*., 1995).

6. Genetic component in PCOS. Candidate genes studied

If a woman in a family has PCOS, the risk of the syndrome in her sister is about 50% as compared to a prevalence of only 5-10% in a population for such risk. This familial clustering suggests a major genetic component to the aetiology of PCOS, although it seems unlikely that there is a single cause for the syndrome (Franks *et al*., 1988). There are of course obvious problems, which make genetic studies of PCOS difficult to perform. For instance, the heterogeneity and the lack of universally acceptable clinical or biochemical diagnostic criteria are a major problem. Another issue is that PCOS is a disorder that primarily affects women of reproductive age and it is therefore very difficult to analyze more than one generation. In addition to this, there is no commonly accepted male phenotype, although premature balding has been suggested as the likely manifestation of affected status in males (Lunde *et al*., 1989).

To date, there has been no consensus about the mode of inheritance of PCOS. Certain segregation analyses gave results that were consistent with an autosomal dominant mode of inheritance (Lunde *et al*., 1989; Carey *et al.,* 1993) however one study suggested an X-linked mode (Jahanfar *et al*., 1995). Only a few clinical studies have been performed over the last 20 years. Evidence that any of the candidate genes researched play a role in PCOS has not been replicated extensively. These uncertainties are common in “complex” genetic diseases, where identifying the contributing genes is difficult because of the multiple aetiologies and the

environmental contributions (Urbanek *et al*., 1999). A description of the genes that have been studied is listed below.

6.1 Genes coding for steroidogenic enzymes

6.1.1 The 17-hydroxylase/17,20-lyase gene (CYP17)

This gene encodes P450c17α, which is a rate-limiting enzyme in the biosynthesis of androgens. Carey *et al*. (1994), identified a single base change (a T to C substitution) at –34 bp from the starting point of translation. This variant allele includes a restriction site for the enzyme *Msp-1,* thus allowing the DNA screening by restriction fragment length polymorphism (RFLP) analysis. Preliminary case-control data suggested an association between the variant allele of CYP17 and PCOS. These findings were based on a small population of subjects (71 patients and 33 controls) and other research groups have been unable to replicate this association (Gharani *et al*., 1996; Techatraisak *et al*., 1997; Franks *et al*, 1997).

6.1.2 Cholesterol side chain cleavage gene, CYP11a

Based on the observation that ovarian theca cells in culture produce an excess of both androgens and progesterone (Gilling-Smith *et al*., 1994; Franks *et al*., 1996), Gharani *et al*. (1997) examined the CYP11a gene as a possible candidate gene for abnormal steroidogenesis. The CYP11a gene encodes P450 side chain cleavage (P450scc), an enzyme which catalyzes the conversion of cholesterol to pregnenolone, the first rate-limiting step at the start of the steroid hormone biosynthesis pathway. The promoter region of the CYP11a gene contains a polymorphic region, a pentanucleatide repeat (tttta)n located at –528bp from the AGT start of translation site. They examined the segregation of CYP11a in 20 families and performed association studies in

premenopausal European women with PCOS. Patients and controls were classified according to the presence or absence of this repeat (at least one copy). Results showed that variation at the CYP11a gene were associated with both PCOS and serum testosterone levels.

6.2 Genes involved in the secretion and action of insulin

Numerous metabolic studies have revealed abnormalities of both insulin secretion and action in women with PCOS (Holte, 1996). The results of such studies raise the possibility that genes implicated in the secretion and action of insulin may have a role in the aetiology of PCOS. Some of them are described below.

6.2.1 The insulin receptor gene

There have been sporadic reports of a PCOS-like phenotype occurring in patients with severe insulin resistance associated with defects of the insulin receptor gene (Moller and Flier, 1988). Conway *et al*., (1994) were unable to detect any abnormalities of the tyrosine kinese domain of the insulin receptor gene in a population of 22 hyperinsulinaemic women with PCOS. In another study, Talbot *et al*. (1996) scanned the entire coding region of the insulin receptor gene on DNA samples from 24 well-characterized women with PCOS. Likewise, no missense or nonsense mutations were found concluding that mutations of the insulin gene were rare in women with PCOS.

6.2.2. The insulin gene

Abnormalities of the insulin secretion have been reported in studies of women with PCOS, with or without a family history of NIDDM (Ehrmann *et al*., 1995; Holte *et al*., 1994; Dunaif and Fineggod, 1996). At the 5’ end of the insulin gene there is a

minisatelite region implicated in the regulation of insulin secretion. At this locus there is a bimodal distribution of repeats, class I alleles being short (average of 40 repeats) and class III alleles being much longer (average 157). Waterworth *et al*. (1997) looked for an association between the insulin gene VNTR (variable number tandem repeats) and PCOS in three different populations of European origin. They found that class III alleles were associated with PCOS in the three populations studied, and this association was stronger when only the cases with anovulatory PCOS were considered. They concluded that the VNTR of the insulin gene is a major susceptibility locus for PCOS, particularly anovulatory PCOS, and may contribute to the mechanism of hyperinsulinaemia and to the high risk of NIDDM in women with PCOS.

6.3 The follistatin gene.

Urbanek *et al*. (1999) studied 37 candidate genes for PCOS. The study was carried out in 150 nuclear families, mainly of European origin, with at least one affected case. The analysis included certain genes previously studied as candidate genes by others, such as the insulin gene, CY11a, and the insulin receptor gene (INSR). They could not replicate the previously described association between the allele III of the insulin gene and PCOS. Neither could they replicate the association between the allele 5 in the promoter region of the CYP11a gene and PCOS after correction for multiple analyses, although without correcting for multiple analyses the association was found. Regarding the INSR, they found no mutations in the INSR coding region of patients with PCOS. The only association found was between the gene coding for follistatin and PCOS. Follistatin is an activin-biding protein which neutralizes the biological activity of activin. Both proteins are expressed in various tissues such as the ovary, pituitary, adrenal cortex, and pancreas. Activin promotes ovarian follicular

development, inhibits theca-cell androgen production, and increases pituitary follicle-stimulating hormone production and insulin secretion by the pancreas (Shibata *et al*., 1996). An increase in the level or functional activity of follistatin will overcome all those effects caused by activin. In fact overexpression of follistatin in transgenic mice results in suppression of serum levels of follicle stimulation hormone and arrested ovarian folliculogenesis (Guo *et al*., 1998).

SUBJECTS AND METHODS

1. Study population

In this prospective study 91 patients with PCOS were recruited from the Subfertility and Reproductive Endocrinology clinics of the National University Hospital, Singapore. The women must not have taken any hormonal medication for at least 2 months prior to the study. Inclusion criteria for patients were the presence of polycystic ovaries, infrequent periods with intermenstrual interval of >35 days, and involuntary infertility. An ovary with the ultrasound appearance of more than 10 subcapsular follicles (< 10 mm in diameter) in the presence of prominent ovarian stroma was deemed polycystic. Patients with hyperprolactinaemia, thyroid and adrenal diseases; 21-hydroxylase deficiency and androgen-secreting tumours were excluded. The weight and height of all subjects were recorded. Hirsutism was not used as an inclusion criteria because was relatively rare (21%) in our predominantly Chinese patient population. Hormonal parameters measured in the patient include LH, FSH, PRL and T. 112 normal controls of proven fertility, no history of subfertility treatment and with normal menstrual cycles every 25-32 days were recruited from the contraceptive clinic. Ethical committee approval was received and informed consent was obtained from all subjects and controls.

2. DNA extraction, DNA amplification and GeneScan analysis

The detailed procedures for DNA extraction, DNA amplification and GeneScan analysis are described in the materials and methods section 3 of Chapter 1 (page 24). The only difference with the male infertility study is that in this study the participants were women. Since the AR gene is located in the X chromosome, women have two

alleles if they are heterozygous for this locus. The number of CAG repeats for each allele was recorded.

3. Hormonal analysis

Radioimmunoassays for serum LH, FSH, PRL and T were performed using standard reagents supplied by the WHO matched-reagent program (Yong *et al*., 1992). Free T was measured with the Coat-A-Count free T solid phase ^{125}I radioimmunoassay (Diagnostic Products Corp). ^{125}I-labeled T analog competes for a fixed time with free T in the patient sample for sites on testosterone-specific antibody immobilized to the wall of a polypropylene tube. The tube is then decanted, to isolate the antibody-bound fraction, and counted in a gamma counter, the counts being inversely related to the concentration of free T in the sample. The Coat-A-count Free T procedure is a direct or single-tube assay: the results are not calculated as a function of total T and SHBG or some other parameter, but interpolated from a standard curve calibrated in free T concentrations. In this respect it differs from conventional equilibrium dialysis methods and from so-called testosterone free index (TFI) determinations. It is also a single-stage assay, requiring neither a pre-incubation step nor preliminary isolation of the free fraction by dialysis, filtration or column chromatography. The assay can detect as little as 0.15pg/mL. The inter-and intra-assay coefficients of variation were less than 15%.

4. Data analyses

The mean values of AR-CAG alleles lengths of cases and the controls were compared using a two independent samples *t*-test. Since two X-linked CAG alleles are present in every subject, comparisons between subject groups were performed by firstly,

considering the mean of the two AR-CAG alleles (biallelic mean) in each subject and secondly, the short and long alleles of each subject separately.

To investigate whether CAG repeats are associated with any subset of anovulatory subjects, patients were categorised into “high T” and “low T” cases using a cut-off T value of 1.73nmol/L which corresponded to the mean value for fertile controls. Mean values of the number of CAG repeats of both subgroups were obtained and similarly compared. The mean AR-CAG lengths of patients with levels of FSH and LH above or below the respective means for normal subjects were similarly subjected to a *t*-test.

Pearson’s correlation coefficient was calculated to assess the relationship between CAG length and T.

The contribution of ethnic differences in AR-CAG alleles was evaluated by analysing the distribution of AR-CAG alleles among the subjects of the two ethnic groups of our study population.

Statistical analyses were performed using the SPSS Version 9.01 (SPSS) computer software. Statistical significance was defined as a two-sided p-value of less than 0.05 and data was reported as mean ±SE.

RESULTS

1. Subjects characteristics

203 participants were included in the study, comprising 91 patients (83% Chinese and 17% Indians) and 112 healthy controls (92% Chinese, 8% Indians). Patients with hyperprolactinemia were excluded from the study. The clinical characteristics of the patients are summarized in Table 1. Patients had a higher mean BMI of 26.9±0.73 kg/m^2 compared to 22.08±0.43 kg/m^2 for control subjects. Hirsutism was not common, occurring in 21% of our patients.

2. Determination of CAG repeat length.

The precise number of CAG repeats was determined by GeneScan technology in the same way as in the male infertility study. The only difference was that since the AR is located on the X chromosome each female has two AR alleles instead of one when women are heterozygous for this locus. Fig. 1 shows the two peaks corresponding to the two alleles from two subjects heterozygous for this locus.

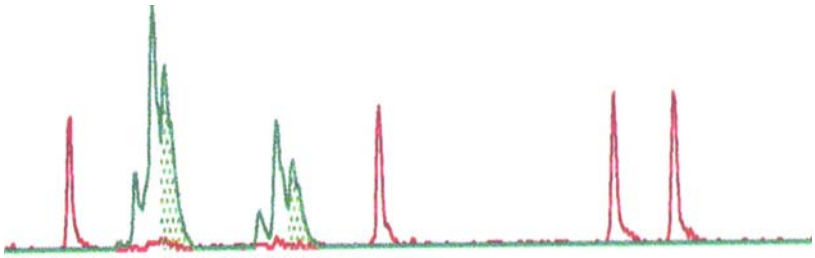
3. Range of the number of CAG repeats and percentage of homozygous and heterozygous subjects.

There were 28 AR-CAG different alleles in our population, ranging from 11 to 33 CAGs in patients, and 14 to 38 in controls. The AR is located on the X chromosome, therefore each female has two AR alleles. Women can be then homozygous or heterozygous for this locus.

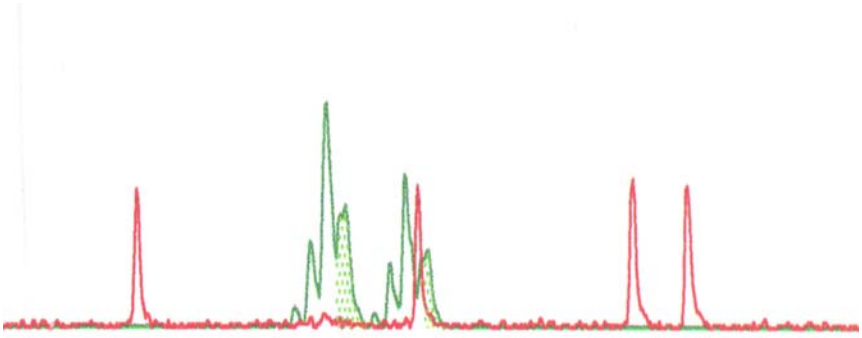
Figure 1. GeneScan analysis of the androgen receptor CAG length in women.

The polymorphic CAG repeat tract in exon 1 of the androgen receptor gene was amplified and internally labelled with the green fluorescent marker (dCTP, R6G) represented by the composed peaks in the picture. Its length was measured with the red internal size standard (ROX, 500) represented by the single peaks in the picture. The electropherograms from two different women show the two alleles with A) 15 and 22 CAGs, and B) 22 and 27 CAGs.

A)



B)



Deleted: <sp>

Deleted: <sp>

Table 1. Clinical and hormonal characteristics of PCOS patients

	Mean ± SE	Range
Age (years)	31.7 ± 0.66	19-45
BMI (Kg/m^2)	26.93 ± 0.73	18.2-39.9
LH (IU/L)	9.50 ± 0.80	1-30.4
FSH (IU/L)	2.95 ± 0.34	0.6-24.2
Total T (nmol/L)	2.32 ± 0.04	0.79-8.30

The number of homozygous patients in our population was 12 (13.18%), ranging from allele 21 to 26. In contrast for fertile controls the number of homozygous was of 21(18.75%), ranging from allele 19 to 28. Among heterozygous the difference in the number of CAG repeats between alleles was quite similar in the control and patient population, being the average of the difference 3.21 (±2.81. SD) CAGs, and 3.15 (±2.59 SD) CAGs respectively.

4. T-test to compare the means between patients and controls

The biallelic mean was calculated for patients and controls, being 22.97±0.24 and 23.09±0.23 respectively. A two-sided t-test was performed and the result indicated no significant differences between the biallelic means of patients and controls.

Since all females have two AR genes and one of them is inactivated (Lyon *et al*., 1988), analyses were performed on the short and long alleles separately. There were no significant difference on the long allele length between patients and controls. The respective mean lengths being 21.38±0.27 and 21.49±0.25 respectively. Similarly, there were no significant differences on the long allele between patients and controls, their mean lengths being 24.53±0.25 and 24.7±0.67 respectively. Table 2 summarizes these results and also includes the corresponding values when both races (Chinese and Indian) were analyzed independently. No significant differences in the number of CAG repeats between patients are controls were found when the analysis was performed separately for each race.

Table 2. CAG repeats mean values for the short, long, and biallelic mean for patients and controls

The number of CAG repeats was calculated and their values compared between patients and controls using a *t*-test. The corresponding P-values are indicated in the last column of the table. Firstly, the biallelic mean was calculated and subsequently the two alleles (short and long) were analyzed separately. The analysis was done for each race independently (Chinese and Indian) and as a whole group (total).

	Patients	*Controls*	P-Value
	CAG Repeat Biallelic Mean Value (SE)		
Chinese	23.24±0.25	23.06±0.24	0.63
Indians	22.43±0.57	21.5±0.93	0.38
Total	22.97±0.24	22.97±0.23	0.68
	CAG Repeat Short Allele Mean Value (SE)		
Chinese	21.49±0.28	21.65±0.26	0.67
Indian	20.87±0.83	19.67±1.18	0.40
Total	21.38±0.27	21.49±0.25	0.78
	CAG Repeat Long Allele Mean Value (SE)		
Chinese	24.64±0.29	24.82±0.29	0.18
Indian	24.00±0.57	23.33±0.97	0.53
Total	24.53±0.25	24.70±0.28	0.67

5. Frequency distribution of the CAG repeat sizes of the short allele and long allele.

The frequency distribution when both alleles, short and long, were analyzed separately did not differ much between the patient and control groups. Fig. 2 and 3.

6. Testosterone levels divide patients with PCOS into two subsets of AR-CAG length.

The clinical manifestations of patients with PCOS are very wide and since there is a subset of PCOS patients who are non-hyperandrogenemic (Conway *et al*., 1989), the relationship between AR-CAG length and serum androgen levels was investigated. As the mean T value in the normal population was 1.73nmol/L, the patients group was divided into two groups: "low T" subjects with T<1.73nmol/L and "high T" subjects with T>1.73nmol/L respectively. There were 57 patients with T>1.73nmol/L, and 34 patients with T<1.73nmol/L; no patient had a T of exactly 1.73 nmol/mL. Their CAG ranges were 11 to 29, and 17 to 33 respectively. Total T was used in the analysis since there was a close correlation between total and free T in our patients ($r= 0.82$, $p<0.0001$) and because free T assays are also less readily available than total T. Furthermore there is evidence in the rabbit (Franks *et al*., 1995) and human (Allen *et al*., 1997) that at least some of the albumin and sex-hormone globulin bound T fractions may be bio-available.

There was a trend for "low T" patients to have a lower average CAG biallelic mean compared to "high T" cases, being 22.47±0.36 and 23.25±0.29 respectively although

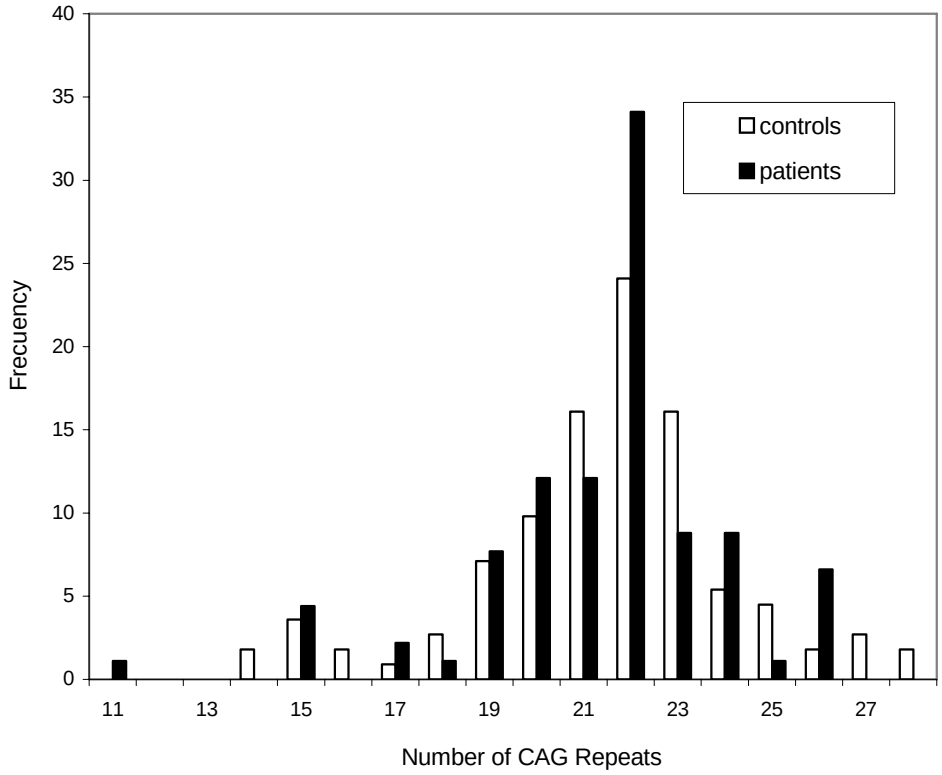


Figure 2. Frequency distribution of the short AR-CAG allele of PCOS patients and controls. No major differences were encountered in the allelic distribution of the number of CAG repeats between patients (black bars) and controls (white bars) when only the short allele was taken for the analysis

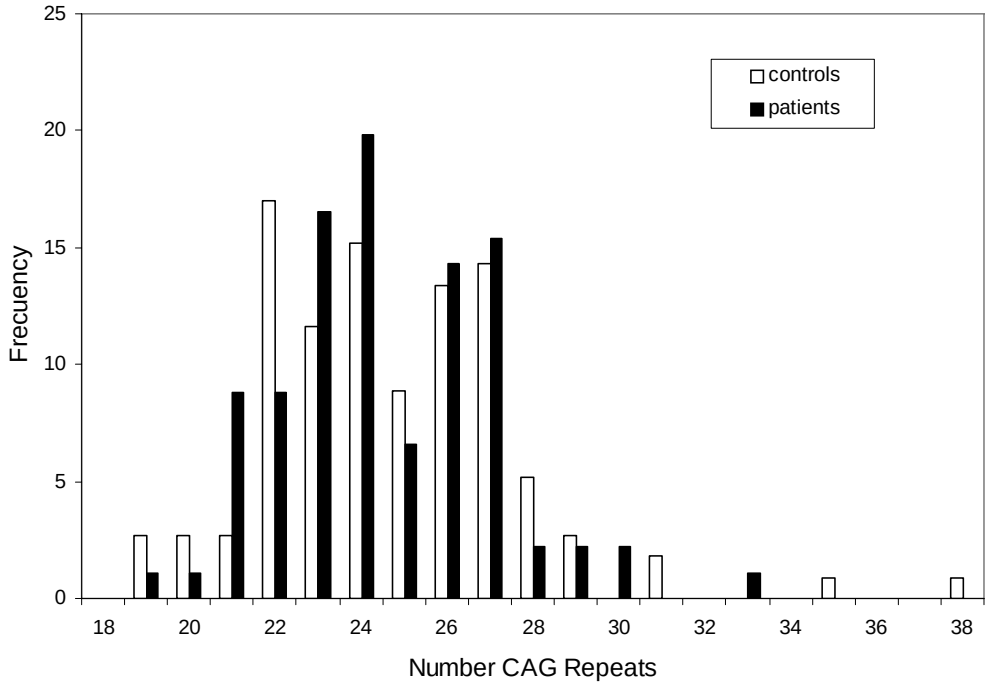


Figure 3. Frequency distribution of the long AR-CAG allele of PCOS patients and controls No major differences were encountered in the AR-CAG allelic distribution between the patient (black bars), and control group (white bars) when only the long allele was considered in the analysis.

this difference did not reach statistical significance (p=0.09) but was marginally significant. This was not the case when only the short allele was considered. Results indicated that the mean length of the shorter allele for patients with "low T" was significantly lower than for patients with "high T", being 20.38±0.51 versus 21.98±0.29 respectively (p=0.004). Conversely, no significant differences were observed when the long allele was analyzed. The mean length value of the long allele for patients with "low T" was of 24.56 ±0.38 CAGs, similar than that for patients with "high T", being 24.53± 0.35 (p=0.95). Table 3 summarizes the mean values of the number of GAG repeats when patients were divided according to their levels of T. Values are expressed considering the two racial groups separately as well as in the combined data.

When the shorter allele was evaluated, differences in frequency distribution were observed, with 53% of patients with "low T" having ≤21 CAGs compared to 33% for "high T" ones (Fig. 4).

7. Sensitivity test to evaluate the T cut-off point.

Different cut-off points above and below 1.73nmol/L were considered when the short allele was taken into account for analysis. Table 4 summarizes the results of such test. For instance, when the cut-off was increased to 1.91nmol/L, the p-value decreased to 0.058, value of borderline significance, which may be a result of the reduction in power with the reduced sample size in the "high T" group.

Table 3. CAG repeats mean values were calculated and compared for patients with “low T” and “high T” respectively.

Patients were divided into two groups depending on their T levels, named “low T” and “high T". The number of CAG repeats was compared among these groups with a *t*-test. The p-values are indicated in the last column of the table. Firstly the biallelic mean was calculated and subsequently the two alleles (short and long) were analyzed separately. The analysis was done for each race (Chinese and Indian) and as a whole group (total).

	Low T	*High T*	*P-Value*
CAG Repeat Biallelic Mean (SE)			
Chinese	22.75±0.39	23.23±0.32	0.37
Indians	21.36±0.90	23.27±0.60	0.079
Total	22.47±0.36	23.25±0.29	0.098
CAG Repeat Short Allele Mean (SE)			
Chinese	20.70±0.53	21.91±0.32	0.041
Indian	19.14±1.42	22.37±0.59	0.046
Total	20.38±0.51	21.98±0.29	0.004
CAG Repeat Long Allele Mean (SE)			
Chinese	24.81±0.43	24.55±0.39	0.66
Indian	23.57±0.78	24.37±0.84	0.50
Total	24.56±0.38	24.53±0.35	0.95

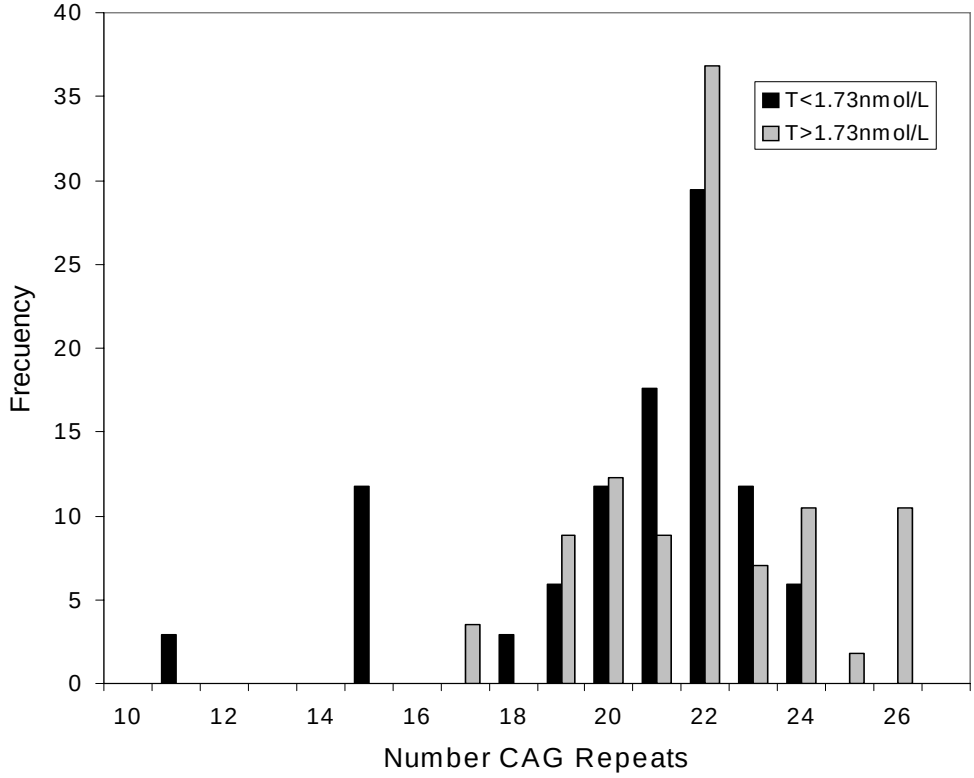


Figure 4. Frequency distribution of the short AR-CAG allele of PCOS patients with serum T levels below or above the normal laboratory mean value of 1.73nmol/L. There were differences in the AR-CAG allelic distribution of patients with low (black bars), and high levels (grey bars) of T when the short allele was analysed.

Table 4. Sensitivity assay.

Sensitivity assay showing the differences in the number of AR-CAGs (column 3) when the cut-off point values for T varied (column 1). The cut-off values considered were bellow and above the original cut off point of 1.73nmol/L (indicated in bold characters). The corresponding P-values after performing a *t*-test are indicated in column 4.

T cut-off point (nmol/L)	Number of cases	Number of CAG repeats	P-value
1.21	<1.20 = 7 ≥1.21 = 85	20.14 21.49	0.191
1.38	<1.37 = 15 ≥1.38 = 76	20.27 21.61	0.069
1.56	< 1.55 = 28 ≥ 1.56 = 63	20.18 21.92	0.003
1.73	**< 1.72 = 57** **≥ 1.73 =37**	**20.38** **21.98**	**0.004**
1.91	< 1.90 = 40 ≥ 1.91 = 51	20.80 21.84	0.058
2.07	< 2.06 = 43 ≥ 2.07 = 48	20.88 21.83	0.083
2.25	< 2.24 = 49 ≥ 2.25 = 42	21.14 21.67	0.345
2.42	< 2.41 = 56 ≥ 2.42 = 35	21.23 21.63	0.483

8. Correlation between the number of CAG repeats and the levels of testosterone.

Bivariate correlation analysis between CAG repeats and the T levels for each allele did not show any correlation when each of the alleles were analysed separately, values for the shorter and longer allele being ($r=0.063$, $p=0.551$) and ($r=-0.113$, $p=0.285$) respectively. This result suggests that the relationship is not a dose-response or linear one, but is more closely approximated by a threshold response.

9. CAG repeat number and the levels of LH and FSH

In order to elucidate if the differences in the number of CAG repeats when patients were classified according to their levels of testosterone were to be due to chance, patients were similarly classified according to their "high" or "low" levels of the hormones FSH and LH and their CAG number analyzed. The cut off point used in the analysis was the mean value obtained for the control group, being in the case of LH of 3.5(IU/L) and of 6.1(IU/L) for FSH. Results are summarised in Table 5. No differences in the number of CAG repeats were detected when those with low or high FSH and LH were analyzed, however when the hormone testosterone was used to divide the patients into two categories the difference in the number of CAG repeats was highly significant and not likely to be due to chance.

10. Ethnic differences in CAG length

The average biallellic mean CAG length in Chinese subjects (patients and controls) was longer than for Indians, being 23.16 ± 0.17 and 22.08 ± 0.5 respectively ($p=0.035$). This result agrees with other studies where it has been reported than Chinese individuals have longer number of CAG repeats than those of other races (Edwards *et al*., 1992).

Table 5. Comparisons of the CAG repeats number when patients were classified according to the hormonal levels of LH and FSH below and above the laboratory mean value.

The average number of CAG repeats in the shorter androgen receptor allele in patients with hormonal levels of LH and FSH below or above the normal laboratory mean was calculated and compared to the values obtained for T. *Independent 2-sample *t*-test was used for calculation of statistical significance (last column). Statistical differences in the number of CAG repeats were observed only when patients were classified according to the levels of T. This was not the case when patients were classified according to their levels of LH and FSH respectively.

	Normal Laboratory Mean	Cases with hormone levels below Laboratory Mean		Cases with hormone levels above Laboratory Mean		
		N	Mean CAG No. ± SE	N	Mean CAG No.±SE	P*
T	1.73nmol/L	34	20.38 ± 0.51	57	21.98 ± 0.29	0.004
LH	6.1 IU/L	43	20.97±0.42	48	21.71±0.31	NS
FSH	3.6 IU/L	68	21.51±0.32	23	20.76±0.42	NS

The mean length of the short allele was also different between these two races, being 21.5±0.19 and 20.4±0.68 for Chinese and Indians respectively (p=0.045). There was also a difference between “low T” Chinese and Indians when the short allele alone was considered, being 20.7±053 and 19.1±1.42 respectively, although this did not reach statistical significance because of the small numbers.

DISCUSSION

PCOS is a heterogeneous endocrine disorder manifesting itself in a wide variety of symptoms that affects approximately 4% of women of reproductive age. Due to its heterogeneous nature, establishing a definition, which encompasses all the clinical and biochemical manifestations was a difficult task. The most widely accepted definition of PCOS is the association of anovulation (manifesting itself as irregular menses, oligomenorrhoea or amenorrhoea), hyperandrogenism, (with variable degrees of cutaneous symptoms such as acne, seborrhea and hirsutism) and ultrasonic evidence of polycystic ovaries (Takahashi *et al*., 1993).

Due to the heterogeneity of symptoms encountered in PCOS patients the selection criteria when recruiting the patients was a very important issue. In this study to account for such a variety of clinical manifestations and to encompass the various subsets of patients of this complex disorder, the patients were selected using a broad criteria of menstrual cycles >35 days and ultrasound-diagnosed polycystic ovaries. Emphasis was not placed on hyperandrogenemia as we wanted to include the subset of patients whose serum androgens levels were not abnormally high and whose symptoms could possibly be due to intrinsic hyperactivity of the receptor apparatus. Moreover the Chinese patients that comprised the majority of our patients did not have high levels of hirsutism (a commonly used marker for hyperandrogenemia).

Familial clustering of PCOS patients suggests a major genetic component involved in the aetiology of PCOS. As described previously in the literature review, a reduced number of genes have been implicated in the aetiology of PCOS, some of them related

to the insulin metabolism and others to the estrogenicsteroidogenic pathway, although the conclusions for the majority of genes studied remain controversial and further studies are needed.

One of the genes that has been implicated in the excess of androgen action characteristic of PCOS patients is the CYP11a gene, which codes for the enzyme P450scc, a rate-limiting enzyme in the steroid hormone pathway and could possibly lead to an increase in androgen production. Variations at the promoter region of this locus were associated with an increase of total T levels and the development of PCOS (Gharani *et al*., 1997). Results of this study were replicated by another study by Urbanek *et al*. (1999) where the researchers analysed 37 candidate genes implicated in the aetiology of PCOS and concluded that only the CYP11a (before the correction for multiple analysis) and the follistatin gene were accountable for some of the symptoms characteristic of the disease.

Taking into account the excessive androgen action characteristic of PCOS patients, the AR gene was selected as a candidate gene for our study. The mode of action of androgens is through binding to the AR followed by a cascade of events that ultimately leads to the activation of target genes. Therefore, not only the levels of circulating androgens are important but also the activity of the complex AR and hormone as a whole and the question raised was if the AR-CAG microsatellite tract, located in the exon 1 of the AR gene, contributes to the hyperandrogenism in patients with PCOS.

Since the AR is located in the X chromosome, each woman has two AR alleles, which may be of different sizes possibly due to different AR-CAG repeat lengths. We

therefore proceeded to measure both CAG repeat tracts in women with PCOS and compared them to those of normal fertile women. The average biallelic mean AR–CAG length did not differ between controls and patients. Further we found no statistically significant differences when the long and short alleles of patients and controls were analyzed separately and compared. However it was ascertained that the short allele contained slightly fewer tracts in the patient group as opposed to the fertile.

Subsequently, patients were divided into two subgroups, those with T levels below 1.73nmol/L, ("low T") and those with T values above 1.73 nmol/L ("high T") (the cut-off point of 1.73nmol/L was selected because it was the mean T value obtained in our laboratory for the normal fertile group). It was highly significant that patients with “low T” levels had shorter CAG repeats tract compared to patients with “high T” when the short allele was taken for the analysis. 37% have serum T levels below the normal laboratory mean and these cases have a significantly lower mean CAG length compared to patients whose T levels were above the laboratory mean. The association between short AR-CAG alleles and patients with “low T” was regardless of ethnic origin, and found in both Chinese and Indian subjects.

A similar analysis was performed by dividing the patients into those with “high”, and “low” levels of LH and FSH, respectively. The cut-off point was the mean value of the control population for these hormones. The aim of this analysis was to determine if the differences found in the number of CAG repeats between patients with “low T” and “high T” were due to chance. No differences in the number of CAG repeats were observed between patients with “low” and “high” LH or FSH indicating that the differences appear only when the hormone T was used in the analysis.

More than half of “low T” patients have at least one AR-CAG allele of ≤21 (Fig. 4), section 7 of results (page 105), and could potentially be at risk of hyperandrogenism due to the greater intrinsic activity of the short AR-CAG allele. The data supports the conclusion that there is a subset of PCOS patients with normal circulating androgen levels that have shorter CAG stretches in the AR gene, at least in one of the alleles, compared to patients with high levels of androgens. One will tend to think that due to the inactivation of one of the X chromosomes, this subset of patients could have active the AR allele containing the shorter number of CAG repeats in important target organs, such as the ovaries, and therefore responding to a greater degree to androgen action. This study is consistent with the findings of a recent study, which did not find any difference in the number of CAG repeats between hirsute patients and controls (Vottero *et al*., 1999). The authors suggested that significant skewing of X-chromosome inactivation occurred, allowing the shorter of the two AR-CAG alleles to be expressed in cases of non-hyperandrogenic hirsutism. If so, the effects of the short AR allele would be accentuated, increasing the chances of hirsutism in those with both long and short alleles in their X-chromosomes.

Other studies support the hypothesis that short CAG tracts in the AR gene may account for the development of PCOS. For instance Carey *et al*. (1993) studied the mode of inheritance of PCOS in families, and the affected status was assigned on the basis of ultrasound evidence of polycystic ovaries in the women, and male pattern baldness, before age 30 in male. Coincidentally, Sawaya *et al*. (1988) found an association between shorter CAG tracts in the AR and the condition of male pattern boldness. In another study by Legro *et al*. (1994) on a Hispanic population, the researchers found

that normoandrogenic patients with idiopathic hirsutism had an inverse relation between the Ferriman Gallway score and the number of CAG repeats (r=-0.60, p=0.014). This correlation was not observed when the patients with high levels of T were taken into consideration. Furthermore low androgen action associated with long CAG repeat tracts had been correlated with increased incidence of breast cancer. Using a sample of women who inherited germline mutation BRCA1, Rebbeck *et al*. (1999) found that these women were at significantly higher risk of developing breast cancer if they carried at least one AR allele with $\geq$28 CAG repeats, possibly due to the lack of androgen inhibition in the tumor.

Ethnic differences were evident in our data, Indian subjects having a shorter biallelic mean AR-CAG length compared to Chinese. This ethnic difference was also observed when only the shorter AR allele was analyzed. Data from Table 3, section 7 of results (page 104), shows that Indian women with “low T” have an average of 19.14$\pm$1.42 CAG repeats in their AR gene compared to 22.37$\pm$0.59 CAG repeats in those with “high T”. This difference in the number of CAG repeats was not as acute in the Chinese population where women with "low T" have an average value of 20.70$\pm$0.53 CAG repeats *versus* 21.91$\pm$0.32 in the “high T” group. These results provide, at least in part, a physiological basis for the observation that the prevalence of PCOS in Indian women is very high, about 52% compared to <10% in other populations (Rodin *et al*., 1998). Interestingly Indians have a higher age-standardized incidence rate of the androgen-driven tumor prostate cancer, compared to Chinese (11.0 *versus* 7.6 per 100,000/year) (Chia *et al*., 1996). It would be very interesting in the future to include a higher number of Indian patients since in this study the majority of patients and controls were of Chinese ethnic origin.

An ethnic difference in AR-CAG distribution was also observed when patients and controls were grouped together. The most common allele length for our Chinese subjects was 23, similar to the Asian population described by Edwards *et al*. (1992). The AR-CAG allele has well-established population differences. The most frequent allele in African-Americans is 18 CAGs compared to 21 for White-Americans, and 23 for the Chinese population.

In summary, this study has identified a subset of PCOS patients with normal testosterone levels and shorter CAG tracts, in at least one of the alleles of the AR gene. PCOS is a complex genetic disease and it is likely that more than one gene would be involved. The complexity of this disease makes it difficult to define exactly how this locus interacts with other genetic factors Besides the genetic components, environmental influences such as fat and carbohydrate consumption, exercise level, peripubertal stress and hormonal exposure may contribute to the development of this complex disorder (Kashar-Miller *et al*., 1999). Nevertheless, this study adds to a growing list of conditions that are related to variations in the AR-CAG tract. We have examined a subpopulation of Singaporean PCOS subjects with normal testosterone levels whose disease was previously puzzling. Our data suggest the novel concept that the cause of hyperandrogenism in these subjects could be related to intrinsic hypersensitivity of the AR due to short CAG tracts. If this hypothesis is confirmed in other populations, our study would have made a significant contribution to the understanding of the disease etiology in the previously inexplicable subset of subjects with all the features of PCOS except for raised androgen levels. Increased

understanding of the mechanism of pathogenesis could lead to improved therapeutic modalities for this common and distressing condition.

CHAPTER 3: RELATIONSHIP BETWEEN PROSTATE-SPECIFIC ANTIGEN, SEX-HORMONE BINDING GLOBULIN AND ANDROGEN RECEPTOR CAG REPEAT POLYMORPHISMS IN SUBFERTILE AND FERTILE MEN

INTRODUCTION

Prostate cancer is the most diagnosed malignant tumour, the second leading cause of cancer death in American men, (Boring *et al*., 1994.) and the sixth commonest cancer among men in Singapore (Koh *et al*., 2001). It is as well-known fact that androgens are required for the development of a normal prostate and also for the development of prostate cancer. The depletion of androgens by surgical or chemical treatments slows cancer growth (Cunha *et al*., 1987). Serum prostate specific androgen (PSA) levels are a common oncogenic marker for prostate cancer because the serum PSA levels parallel oncogenic growth in the initial androgen-dependent and in early stages of androgen independent phase of the cancer.

The AR activates the transcription of the PSA. The hormones playing an important role in prostate cancer are T, its more potent form dihydrotestosterone (DHT), and sex hormone binding globulin (SHBG) The levels of these hormones and PSA are biologically related to each other. A preliminary description of PSA, SHBG and the androgens T and DHT in addition to a short review on how these factors are related to each other is listed below.

1. Androgenic parameters important for prostate cancer

1. 1 Prostate specific antigen

Prostate specific antigen (PSA) is a single chain glycoprotein enzyme (serine protease) of 33-Kd. In males it is found within the cytoplasm of prostatic duct cells and in the lumina of prostatic ducts. It is present in urine, seminal plasma, and serum (Warhol & Longtine, 1985). PSA concentration in male serum is a valuable tumour marker for diagnosis and management of prostate cancer. The expression of the PSA protein is an AR-mediated process (Young *et al*., 1991).

PSA was believed to be a highly specific marker for normal or cancerous prostatic tissue, but recent studies have also found PSA in female tissues and fluids such as milk breast, ovary, and amniotic fluid, (Diamandis *et al*., 1995). The protein production of PSA in nonprostatic tissues is under regulation of steroid hormones, such as androgens, glucocorticoids, and progestins. Estrogen itself seems to have no effect on PSA regulation, but it can impair androgen-induced PSA production (Diamandis *et al*. 1997).

PSA occurs in three major forms in blood. The major immunodetectable form is PSA complexed with the serine protease inhibitor, alpha-1-antichymotrypsin (PSA-ACT). Uncomplexed, or free PSA, is another immunodetectable form of PSA in serum. The majority of free PSA in serum appears to be an inactive form that cannot complex with protease inhibitors and may be either a PSA zymogen or an enzymaticaly-inactive cleaved form of PSA. A third form of PSA, a complex with alpha-2-macroglobulin, is not detectable with current immunoassays due to the engulfment and subsequent

masking of PSA epitopes by the alpha-2-macroglobulin molecule (McCormack *et al*., 1995). The sum of the first and second forms of PSA constitute the total PSA.

It has been postulated that PSA is more extensively complexed in prostate cancer cells than in nondiseased prostatic or benign prostatic hyperplasia cells. A lower ratio of free PSA to total PSA (fPSA/tPSA) in patients with prostate cancer has been demonstrated in numerous studies (Stephan *et al*., 2000); (Basso *et al*., 2000). This ratio has been considered as a helpful tool for distinguishing between prostate cancer and benign prostatic hyperplasia (BHP).

1.2 Sex hormone-binding globulin

Sex hormone-binding globulin (SHBG) is a plasma glycoprotein that binds certain estrogens and androgens with high affinity. SHBG is synthesized in the liver. It binds T and DHT with high affinity and estradiol with a lower affinity. The function of SHBG still remains controversial, and one view is that only the non-protein bound steroid hormones are capable of entering the cells from the bloodstream, and thereby initiating their biological responses. Another biological function is that SHBG binds specifically to cell membranes through its membrane receptor, and stimulate the production cyclic adenosine monophosphate (cAMP). When the steroid binds to the SHBG, this inhibits the binding site for the SHBG receptor (Rosner *et al*., 1999)

T circulates primarily protein-bound, principally to SHBG, but also to albumin and cortisol binding globulin. Since variations in the carrier protein levels may affect the concentration of T in circulation, SHBG levels are commonly measured as a

supplement to total T determinations (Cunningham *et al*., 1985). The free androgen index (FAI) is the ratio of total T to SHBG and is sometimes an indicator of abnormal androgen status.

1.3 Testosterone and dihydrotestosterone

Androgens affect a multitude of physiological parameters, including muscle mass, bone density, central nervous system function, bone marrow production and sexual function.

Testosterone (T) is the predominant circulating androgen in men. It is secreted by the Leydig cells of the testis in response to the hormone LH. T circulates almost entirely bound to transport proteins and normally less than 1% is T free. The principal transport protein for T is SHBG. Albumin and cortisol binding globulin (CBG) are the other T transport proteins.

Dihydrotestosterone (DHT), a reduced form of T and a much more potent androgen, is essential for normal growth and development of the prostate gland. Greater than 95% of intracellular T is converted to DHT within the prostate gland by the enzyme 5α-reductase.

2. Relationships among the various androgenic parameters analyzed

2.1 Association between the levels of testosterone and prostate specific antigen.

A number of studies have shown that androgens are necessary for the development of prostate cancer, at least in the initial stages, before the tumour escapes its androgenic regulation. In this initial androgen-dependent stage appears to be a

correlation between the levels of T and those of PSA. However, there is limited data on the correlation between T and PSA in a normal healthy population, and people with low androgen status such as hypogonadism, during male puberty, and in women.

Monath *et al*. (1995) measured the levels of T and PSA of 150 males without previous history of prostate cancer. They found no correlation between T and PSA ($r= 0.054$; $P = 0.515$), even when corrected for age and weight. No significant correlation was found between T and age, but a significant correlation was found between increasing age and PSA. In another study, MacIndoe *et al*. (1995) reported that serum PSA levels did not change significantly in 11 healthy young men receiving exogenous T administration. They concluded that the expression of PSA appeared to be dependent on the presence of androgens in patients with prostate cancer. On the contrary, in men without prostate cancer or with benign hyperplasia of the prostate gland (BHP) a correlation between the expression of PSA and the administration of exogenous T was not found. Similarly, Copper *et al*. (1998) studied 31 young healthy males following 15 weeks of T injection. They concluded that no detectable significant change occurred in their serum PSA levels or prostate volume after therapy.

In the study by Kim *et al*. (1999) 77 normal boys in Tanner stages I to V, from 10 to 18 years, were analyzed for PSA levels, SHBG, insulin like Growth factor (IGF-I), IGF binding protein 3 (IGFBP-3) and T. A positive correlation was found between PSA and T during puberty ($r=0.86$, $p<0.001$). Behre *et al* (1994) studied

78 hypogonadal males undergoing T treatment for 6 months. They reported an increase in PSA levels and prostate volume as a response to the T treatment.

Similarly, Douglas *et al*. (1995) studied the effect of exogenous T replacement on PSA levels in 10 hypogonadal men ranging in age from 40 to 76 years. These men had no clinical signs or symptoms of BHP, prostatitis, or carcinoma of the prostate and had normal PSA levels. Surprisingly they found no significant correlation between PSA and total T levels. The researchers found this lack of correlation unusual, and attributed it to the short duration of their study or to the small sample size used. Weber *et al*. (1989), reported a strong correlation between serum PSA and serum T levels in seven BHP patients undergoing a period of androgen deprivation. Aus *et al*. (1994), demonstrated no correlation between serum T and serum PSA in 120 men undergoing transurethal prostatectomy for symptomatic BHP.

Since the discovery that women also have detectable levels of PSA, Escobar-Morreale *et al*. (1998) studied the relationship between hirsutism and elevated levels of serum PSA. Their study included 37 patients and 11 controls. Serum PSA concentrations were clearly elevated in hirsute women compared to those in the control group. When patients and controls were studied as a whole, serum PSA values showed positive correlation with basal T, and a negative correlation with basal SHBG levels and age. In another study, Goh (1999) used as a model of PSA regulation female transsexual patients whom were given exogenous doses of T. Serum levels of PSA were significantly raised in female transsexuals after long-term androgen therapy. His study showed that high levels of T were able to up-

regulate PSA production in transsexual women in a time and dose dependent manner.

2.2 Association between the levels of testosterone and sex hormone-binding globulin

The correlation between the levels of T and SHBG remains a controversial subject. Carlstrom *et al*. (1990) measured the levels of T and SHBG in 80 healthy males aged from 25 to 80 years and found that there was a positive correlation for all different age groups analyzed. Similarly, Gann *et al*. (1996), described a strong correlation (r=0.55) between the levels of T and SHBG measured from the plasma of 190 healthy individuals. On the contrary, Plymate *et al*. (1983) described that the levels of SHBG decreased in normal males after treatment with external doses of T. Cell culture experiments, using the human hepatocarcinoma cell line HepG2 demonstrated that the secretion of SHBG was stimulated by increasing levels of T (Edmunds *et al*., 1990).

2.3 The number of CAG repeats in the androgen receptor gene and prostate cancer.

Some studies have analyzed whether there is a relationship between the number of CAG repeats in the AR gene and the development of prostate cancer. Hardy *et al*. (1996) reported that there was a significant correlation between CAG repeat length and age of onset of prostate cancer, suggesting shorter CAG repeats may be associated with the development of prostate cancer at a young age. The authors failed to find a correlation between the number of CAG repeats and other parameters such as PSA levels once the tumour was diagnosed. In a larger study comprising of 587 cases of prostate cancer and 588 control males, Giovannucci *et*

al. (1997) demonstrated that men with shorter number of CAG repeats in the AR gene had higher grade and advanced stages of prostate cancer at diagnosis, metastasis of the tumour, and mortality from the disease. Stanford *et al*. (1997) evaluated the number of CAG repeats of 301 cases of prostate cancer and 277 controls in a population-based case-control study. Their results also supported the hypothesis that the number of CAG repeats is a predictor of risk for prostate cancer. In a recent study by Hsing *et al*. (2000), the researchers measured the number of CAG repeats in the AR gene from 190 prostate cancer patients and 304 healthy controls in a population of Chinese ethnic origin. Results indicated that the median value of CAG repeats for their subjects was of 23, and that men with a CAG length of less than 23 repeats had a 65% increased risk of prostate cancer.

In this study we wanted to gain information about how some of the androgenic factors relating to prostate cancer, such as the serum levels of free and total T, free PSA, total levels of PSA, and SHBG change with respect to each other in our population. Specifically, this study aimed to determine whether the levels of PSA were related to the number of CAG repeats in the AR gene, and therefore obtain a new molecular marker for prostate cancer.

MATERIALS AND METHODS

1. Study population

112 subfertile males were recruited from the infertility and andrology clinics of National University Hospital, Singapore. A complete history and physical examination were performed, and the use of any medications or previous surgery was recorded. Patients who had chromosomal abnormalities, hypopituitarism, hyperprolactinemia, undescended testes, obstructive syndromes of the genital tract, and hypogonadism secondary to surgery, trauma, or chemotherapy were excluded from this study. Sperm parameters were assessed according to standard criteria and were the mean of at least two analyses done 3 months apart. Azoospermia was defined as the absence of any spermatozoa despite centrifugation of the semen specimen, severe oligospermia was sperm density <5 mil/ml, and oligospermia defined as the mean sperm density being <20 mil/ml. 91 control men of proven fertility with no previous infertility history or treatment, and without any genetic disease were included in the study. Ethical committee approval was received and informed consent was obtained from all subjects

2. Hormonal analysis and determination of the number of CAG repeats

Assays for serum total and free T, SHBG, total and free PSA were performed.

Free T was measured with the Coat-A-Count free T solid phase ^{125}I radioimmunoassay (Diagnostic Products). ^{125}I-labeled T analog competes for a fixed time with free T in the patient sample for sites on testosterone-specific antibody immobilized to the wall of a polypropylene tube. The tube is then decanted to isolate the antibody-bound fraction, and counted in a gamma counter. The counts are inversely related to the concentration of free T in the sample. The Coat-A-count Free T procedure is a direct or single-tube assay. That means the results are not calculated as a function of total T and

SHBG, but interpolated from a standard curve calibrated in free T concentrations. In this respect it differs from conventional equilibrium dialysis methods and from so-called testosterone free index (TFI) determinations. It is also a single-stage assay, requiring neither a pre-incubation step nor preliminary isolation of the free fraction by dialysis, filtration or column chromatography. The assay can detect as little as 0.15pg/mL.

Total T was measured by the ARCHITECT assay (Abbott Laboratories). This is a one-step immunoassay to determine the presence of T in human serum and plasma using Chemiluminescent Microparticle Immunoassay (CMIA) technology. Sample, anti-testosterone (mouse, monoclonal) coated paramagnetic microparticles, T, acridinium-labeled conjugate, and assay diluent, were combined to create the reaction mixture. T present in the sample competes with the T acridinium-labeled conjugate for binding with anti-testosterone (mouse, monoclonal) coated particles to form an antigen-antibody complex. An inverse relationship exists between the amount of T in the sample and the RLUs detected by the ARCHITECT optical system.

The total levels of PSA were measured by the ARCHITECT Total PSA assay (Abbott Laboratories). This is a two-step immunoassay to determine the presence of total PSA (both free PSA and PSA complexed to alpha-1-antichymotrypsin) in human serum.

In the first step, sample and anti-PSA coated paramagnetic microparticles are combined. PSA present in the sample binds to the anti-PSA coated microparticles. After washing, anti-PSA acridinium-labeled conjugate is added in the second step. The "Trigger" solutions are then added to the reaction mixture. The resulting

chemiluminescent reaction is measured RLUs. A direct relationship exists between the amount of total PSA in the sample and the RLUs detected by the ARCHITECT optical system.

Free PSA was measured also by the Chemiluminescent Microparticle Immunoassay ARCHITECT assay (Abbott Laboratories). In the first step, sample and anti-free PSA coated paramagnetic microparticles are combined. Free PSA present in the sample binds to the anti-free PSA coated microparticles. After washing, anti-PSA acridinium-labelled conjugate is added After the addition of the "Triger" solutions, the resulting chemiluminescent reaction is measured as RLUs. A direct relationship exists between the amount of free PSA in the sample and the RLUs detected by the optical system. The sensitivity of the ARCHITECT free PSA assay was calculated to be better than 0.008ng/mL. This sensitivity is defined as the concentration at two standard deviations above the mean RLU at level 0.

Levels of SHBG were measured with the IMMULITE SHBG assay (Diagnosis Products). The assay tube is coated with a monoclonal antibody specific for SHBG, which is coupled to the enzyme alkaline phosphatase. The chemiluminescent substrate, a phosphate ester of adamantyl phosphate, undergoes hydrolysis in the presence of alkaline phosphatase. The emission of light is proportional to the concentration of SHBG in the sample. The sensitivity of the assay was 0.2nmol/L. For all the tests mentioned above, the inter-and intra-assay coefficients of variation were less than 15%.

The number of CAG repeats in the AR gene was calculated by GeneScan analysis after amplifying by PCR the region of interest following the same procedure as described in the materials and methods section 3 of chapter 1 (page 24)

3. Statistical analysis

The mean values of the various parameters studied were calculated and compared between the infertile and fertile group using a two independent samples t-test.

The correlations among the different variables were analyzed using Pearson's correlation coefficients (r) and when appropriate, adjustments for potentially confounding variables were made by multiple regression analysis. P values less than 0.05 were considered statistically significant.

Statistical analyses were performed using the SPSS Version 9.01 (SPSS) computer software. Statistical significance was defined as a two-sided p-value of less than 0.05 and data was reported as mean±SE.

RESULTS

1. Differences between fertile and subfertile men

Endocrine parameters important for androgen action and metabolism were measured in 91 subjects with proven fertility, and 112 subfertile men with varying degrees (azoospermic, 29.4%; severe oligospermic, 54.1%; oligospermic, 16.5%) of spermatogenic defects. The mean ages for the fertile and subfertile men were 35.57±0.58 and 33.98±0.50 years respectively. The racial composition of both groups was similar, being predominantly of Chinese ethnic origin.

The mean values of PSA (free and total), SHBG, AR CAG repeat lengths, and T (free, total, and corrected for SHBG) between the fertile and subfertile males were compared using the two sided *t*-test. Results are summarised in Table1.

The main observation when the mean values were compared between the two groups was that T levels were lower in subfertile males compared to fertile males. The majority of T is bound to carrier proteins and normally less than 1% is free T, which is considered the active form. One of the main protein carriers of T is SHBG. An indirect method to estimate the levels of free T is to calculate the ratio T/SHBG. Total T, T adjusted for SHBG, and free T were about 17-20% lower in subfertile men compared to their fertile counterparts. Differences in free T and T/SHBG were highly significant at the $P<0.001$ level. The frequency distribution of T levels in subfertile subjects was shifted to the left with respect to the fertile group, indicating that low T levels were a general feature of the subfertile group, and that data were not skewed by a few severely hypogondal patients (Fig. 1).

Table 1. Descriptive statistics of the various parameters measured in the study.

Column 1 lists the various parameters studied, and column two and three indicate the values of the mean, SE of the mean, and the range of the fertile and subfertile group respectively.

To check weather the values of the subfertile and fertile groups were the same, t-tests (2-tailed) were performed and the corresponding p-values are given in column 4. Significant differences were obtained when the levels of total T, free T and T/SHBG were compared between the fertile and subfertile groups.

PARAMETER	FERTILE±SE (Range)	SUBFERTILE±SE (Range)	P-Value
Free PSA(ng/mL)	0.245±0.02 (0.066-0.639)	0.239±0.01 (0.038-0.609)	0.734
Total PSA(ng/mL)	0.754±0.01 (0.212-5.96)	0.828±0.06 (0.094-5.033)	0.435
Free T(pmol/L)	77.90±0.02 (27.1-158.8)	64.37±2.34 (16.47-143.83)	**0.000**
Total T(nmol/L)	14.83±0.57 (4.44-31.24)	12.31±0.53 (2.78-31.77)	**0.002**
SHBG(nmol/L)	21.48±0.80 (8.80-42.20)	22.49±0.87 (6.70-45.70)	0.422
T/SHBG	0.74±0.01 (0.17-1.72)	0.59±0.02 (0.13-1.67)	**0.000**
Number CAG Repeats	22.50±0.31 (11-29)	22.98±0.87 (14-33)	0.253

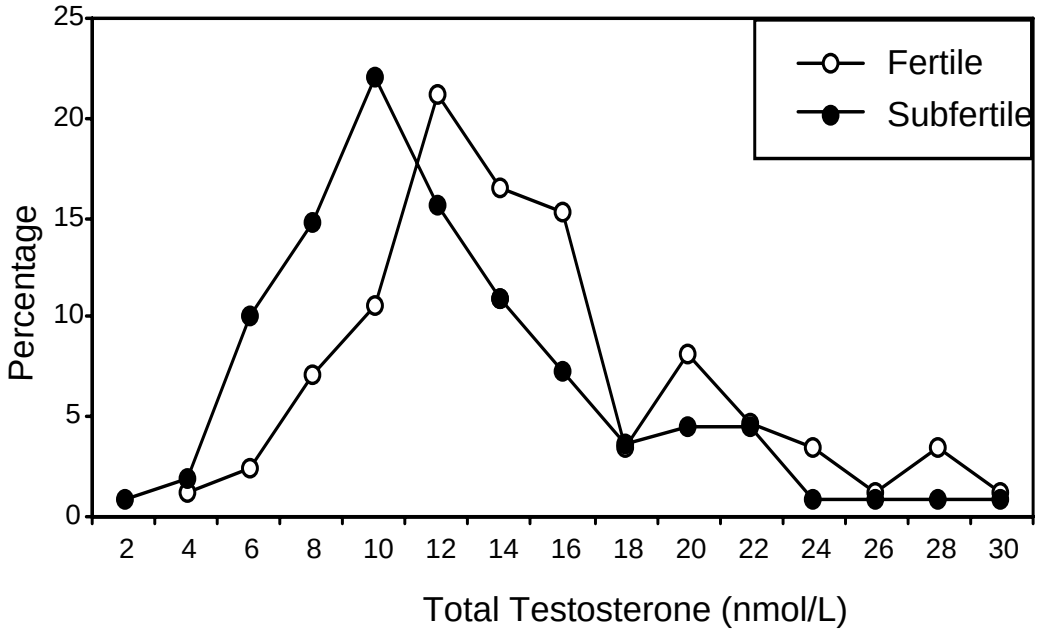


Figure 1. Frequency distribution of testosterone levels in fertile and subfertile subjects. Serum total testosterone in each individual was measured and grouped to the closest even value. Fertile, n=91; Subfertile n=112.

This relative hypogonadism in subfertile men was not age-related since, on average, subfertile subjects were younger than fertile controls.

There was a tendency, not reaching statistical significance, for the AR CAG length to be longer in the subfertile group. No significant differences in SHBG were observed between subfertile patients and fertile controls.

No differences in the free and total PSA levels were encountered between the subfertile and fertile group.

2. Bivariate correlations

All the possible correlations between the different parameters studied, PSA, T, SHBG and the number of CAGs, were performed for the infertile and fertile group independently. Results are summarised in Tables 2 (for fertile subjects), and 3 (for infertile subjects). The most interesting correlations are described below.

2.1 Sex hormone binding-globulin is highly correlated with testosterone

The most significant correlation in this study was observed between SHBG and T (Tables 2 and 3), in both the fertile and subfertile group. SHBG levels were directly correlated with T levels (Fig. 2). This was the case when T was expressed as free or total. The correlation between total T and SHBG for fertile and subfertile men grouped together was of $r=0.546$, $P<0.0019$ (data not shown in the tables). This finding was surprising as since androgens, unlike estrogens, are considered to have an inhibitory effect on SHBG.

Table 2. Results of the bivariate correlations established in the fertile group. Values of Pearson's correlation coefficient (r) are indicated at the top of each cell. The second value in the cell corresponds to the p-value (p) of each correlation. The numbers in bold indicate significant results.

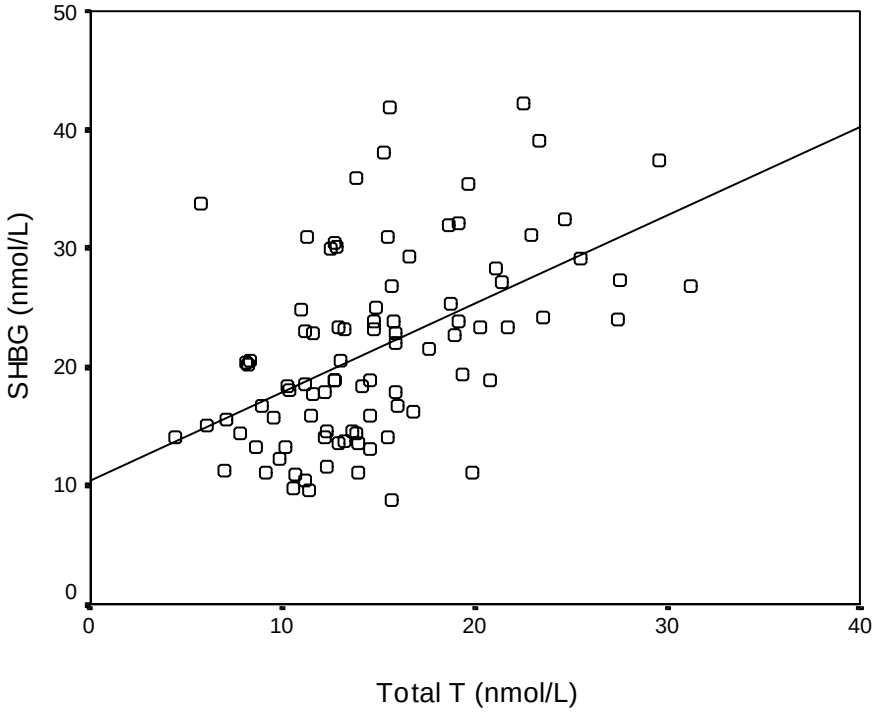
		Free PSA	***Log Total PSA***	***Total PSA***	**Free T**	***Total T***	***SHBG***
Free T	r	-0.006	0.061	0.093			
	p	0.957	0.567	0.383	*****	*****	*****
Total T	r	-0.027	0.069	0.098	**0.903**		
	p	0.798	0.517	0.353	**0.000**	*****	*****
SHBG	r	-0.114	0.022	0.069	**0.462**	**0.502**	
	p	0.284	0.837	0.513	**0.000**	**0.000**	*****
T.Test/ SHBG	r	0.111	0.062	0.019			
	p	0.296	0.559	0.855	*****	*****	*****
No.of CAGs	r	0.064	0.159	0.123	**0.210**	0.185	**0.216**
	p	0.545	0.134	0.245	**0.046**	0.079	**0.040**

Table 3. Results of the bivariate correlations established in the subfertile group. Values of Pearson's correlation coefficient (r) are indicated at the top of each cell. The second value in the cell corresponds to the p-value (p) of each correlation. The numbers in bold indicate significant results.

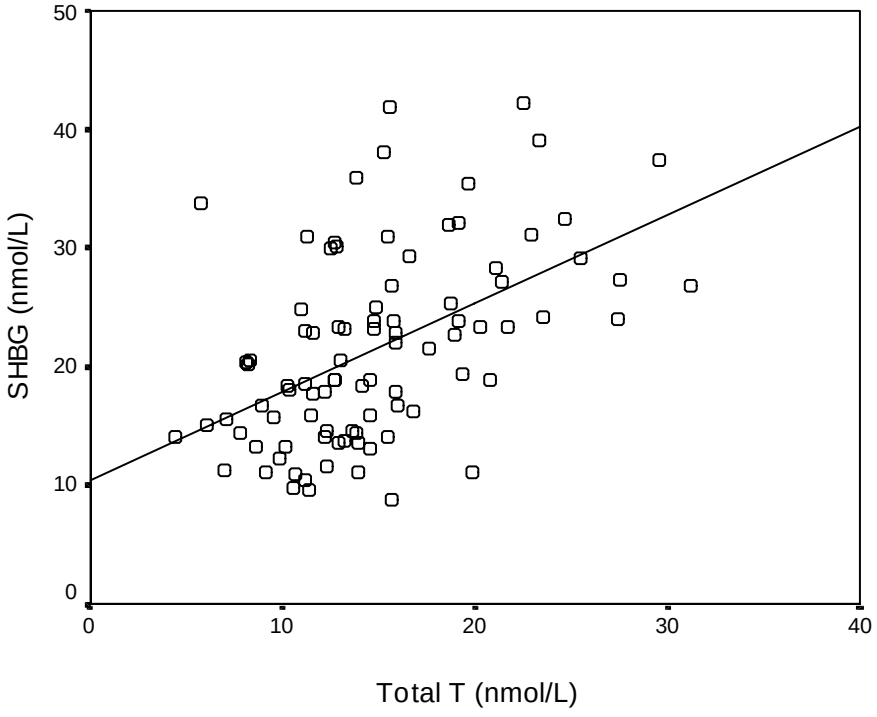
		Free PSA	***Log Total PSA***	***Total PSA***	***Free T***	***Total T***	***SHBG***
Free T	r	**0.328**	**0.330**	**0.271**			
	p	**0.000**	**0.000**	**0.004**	*****	*****	*****
Total T	r	**0.354**	**0.356**	**0.293**	**0.806**		
	p	**0.000**	**0.000**	**0.002**	**0.000**	*****	*****
SHBG	r	**0.242**	0.175	0.159	**0.608**	**0.623**	
	p	**0.008**	0.055	0.083	**0.000**	**0.000**	*****
T.Test/ SHBG	r	0.042	0.154	0.129			
	p	0.661	0.106	0.177	*****	*****	*****
No. of CAGs	r	-0.108	**-0.198**	**-0.229**	-0.139	-0.095	-0.085
	p	0.239	**0.029**	**0.011**	0.145	0.317	0.356

Figure 2. Scatter graphs showing the positive correlation between the levels of SHBG and total T in the fertile group: A), and infertile group: B).

A)


SHBG (nmol/L)
Total T (nmol/L)
0
10
20
30
40
50

B)


SHBG (nmol/L)
Total T (nmol/L)
0
10
20
30
40
50

2.2 Levels of testosterone and prostate specific antigen

T is strongly correlated with PSA levels in subfertile men (Table 3). The relationship between T and PSA in subfertile men remained highly significant ($P<0.001$) when T was measured as total or free T, or when total or free PSA was assayed. Since PSA values in normal populations are highly skewed, we calculated the log total PSA and observed that log total PSA and T remained tightly correlated ($r=0.356$, $p<0.001$). However in fertile controls, no relationship between T and PSA was observed. (Fig. 3).

2.3 Number of CAG repeats and prostate specific antigen

CAG repeats were inversely correlated with total PSA (or log total PSA) only in the subfertile group ($r= -0.229$, $p= 0.011$), (Fig. 4). This correlation was still present after adjusting for T ($r= -0.216$, $p= 0.023$), (data not shown in the tables) and was thus independent of the effect of T on PSA observed. In contrast no correlation was observed between CAG repeats and PSA in the fertile group.

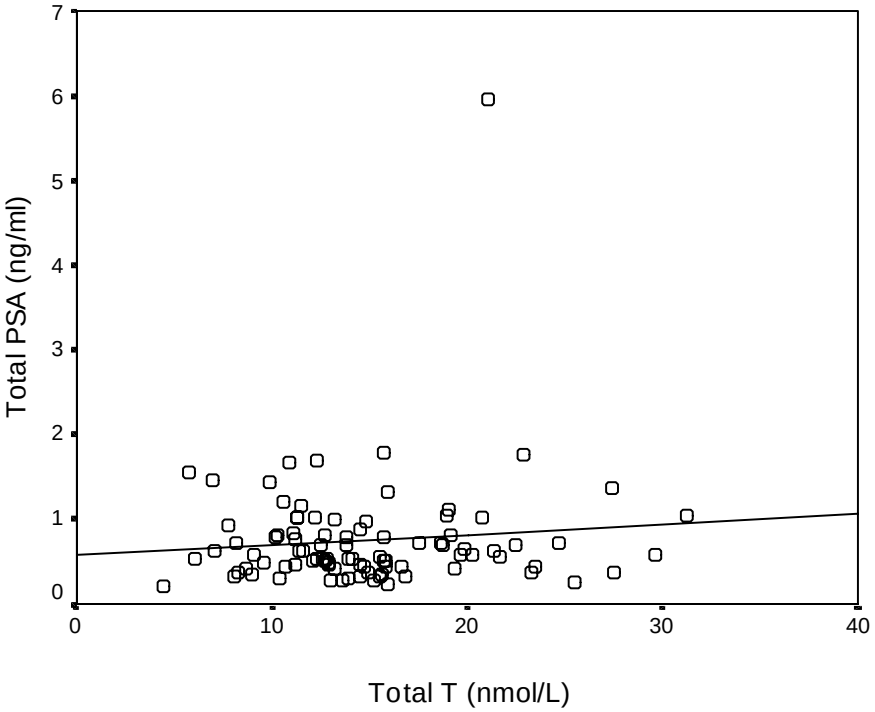
2.4 Number of CAG repeats, testosterone, and sex hormone-binding globulin

The number of CAG repeats was positively correlated to the levels of free T ($r=0.210$, $p=0.046$), (Fig.5), and SHBG ($r=0.216$, $p=0.040$) only in the fertile group. This was not the case for the subfertile population studied.

The relevance of such correlation in the fertile group is unknown.

Figure 3. Scatter graphs showing the correlation between the levels of total PSA and total T. Fig. 3A shows that there was not a correlation between the levels of total T and those PSA in the fertile group. This was not the case in the subfertile group, Fig.3B, where the levels of PSA positively correlated with the levels of total T.

A)


Total PSA (ng/ml)
0
1
2
3
4
5
6
7
0
10
20
30
40
Total T (nmol/L)

B)

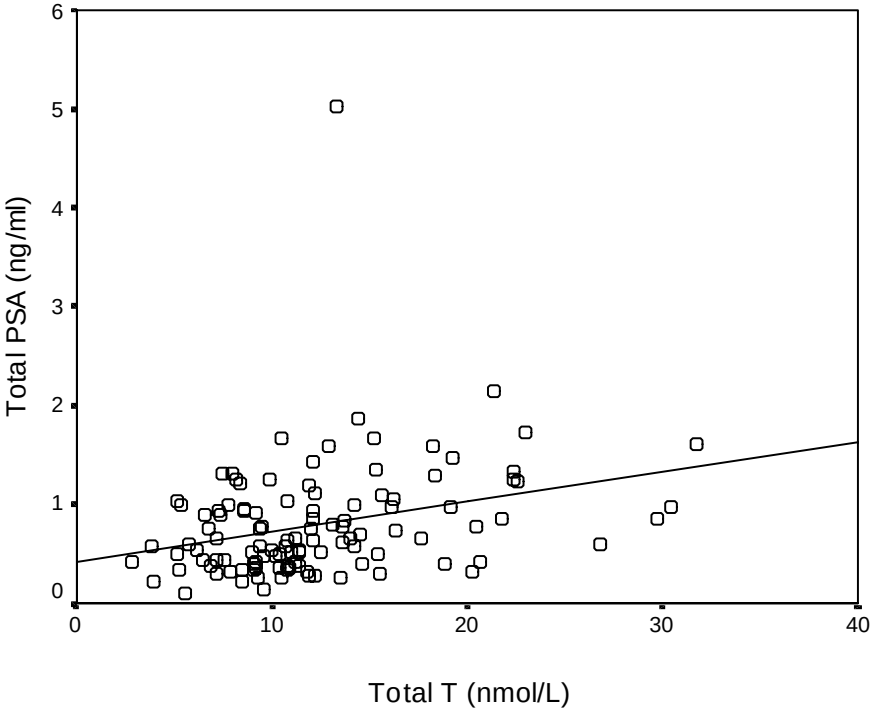
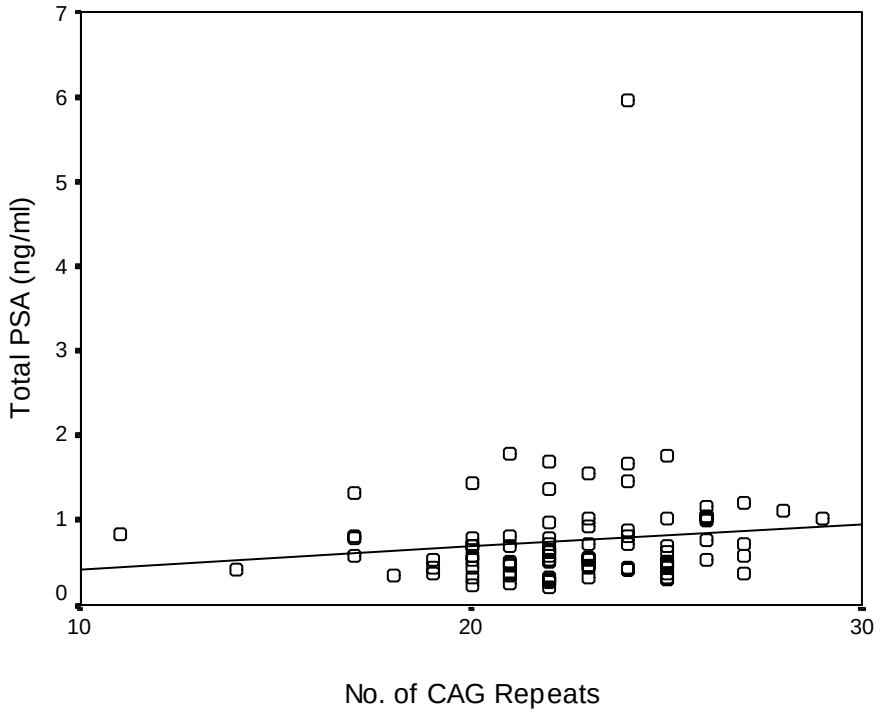

Total PSA (ng/ml)
0
1
2
3
4
5
6
0
10
20
30
40
Total T (nmol/L)

Figure 4. Scatter graphs showing the correlation between the number of CAG repeats in the AR gene and the levels of total PSA. Fig 4A shows that there was not a correlation between the number of CAG repeats and the levels of total PSA in the fertile group. In the subfertile group, Fig.4 B, there was an inverse correlation between the number of CAG repeats and the levels of total PSA.

A)


Total PSA (ng/ml)
0
1
2
3
4
5
6
7
10
20
30
No. of CAG Repeats

B)

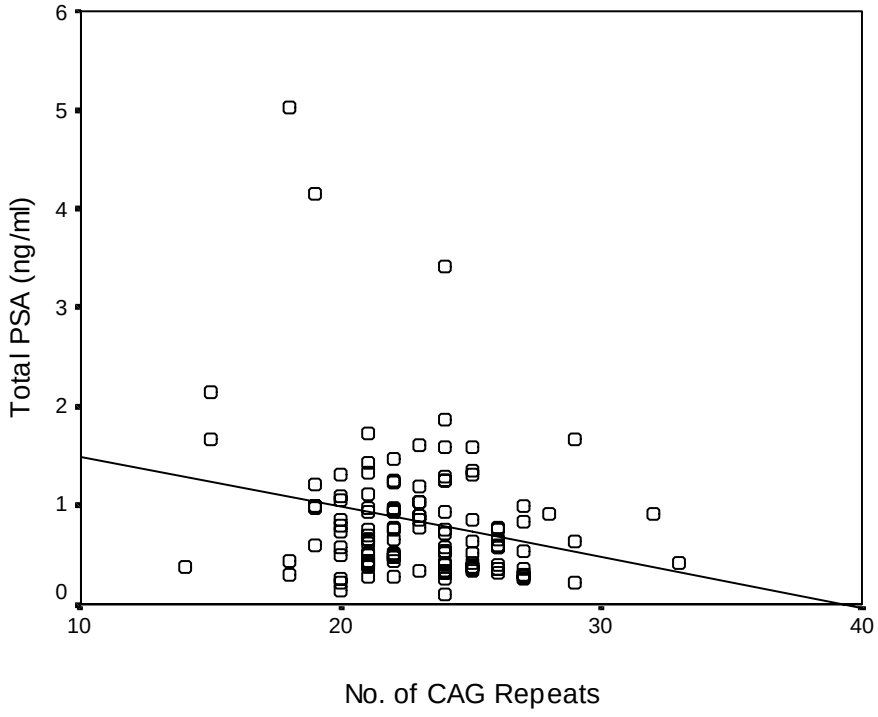
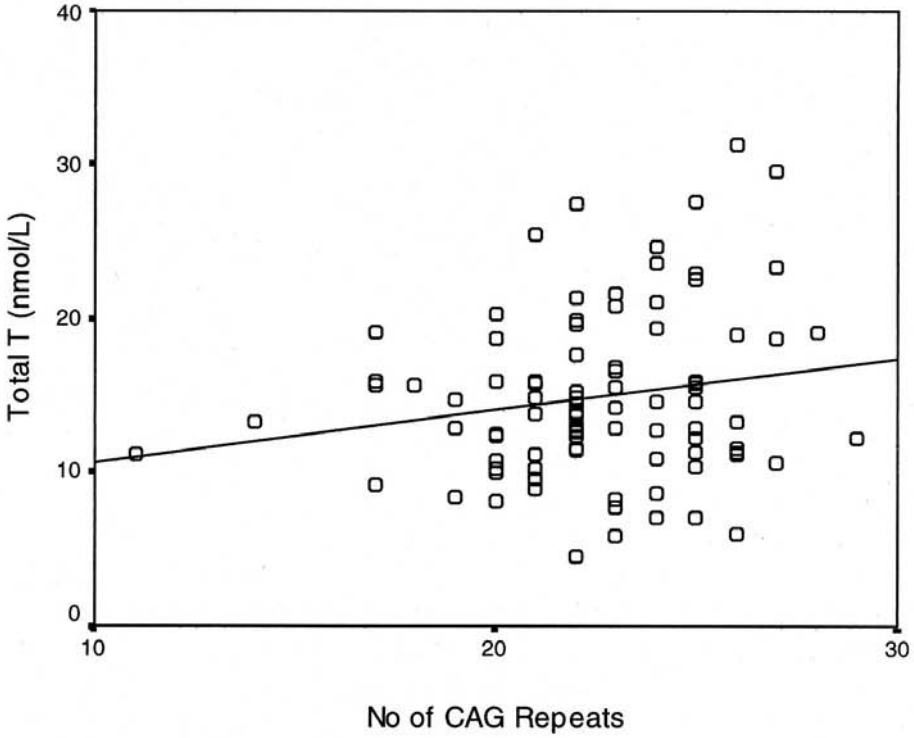
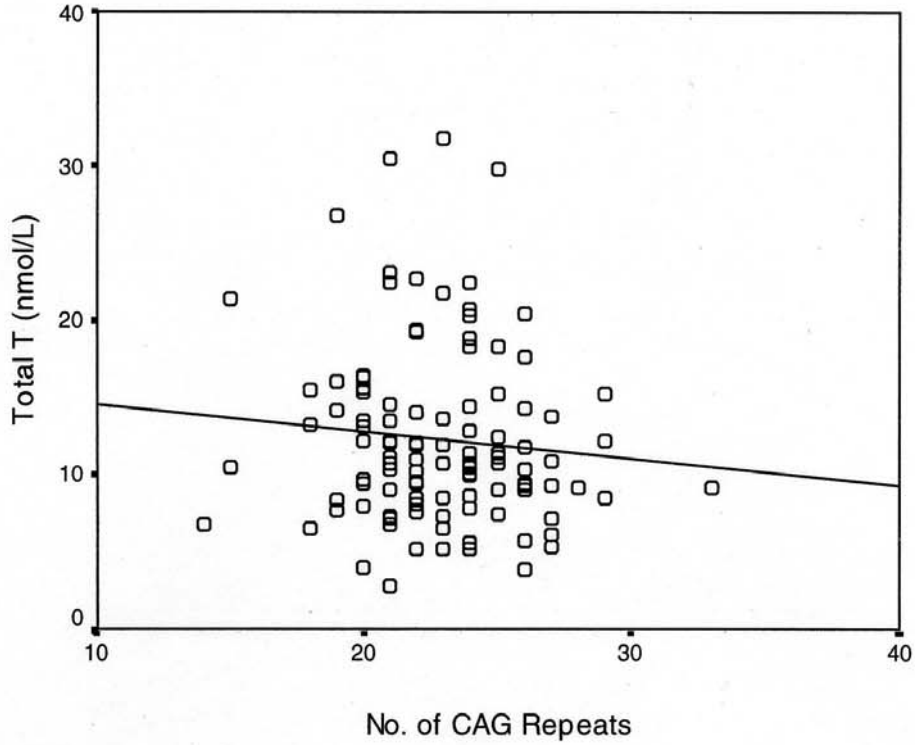

Total PSA (ng/ml)
0
1
2
3
4
5
6
10
20
30
40
No. of CAG Repeats

Figure 5. Scatter graphs showing the correlation between the number of CAG repeats in the AR gene and the levels of total T. Fig 5A shows that there was a positive correlation between the number of CAG repeats and the levels of total T in the fertile group. On the contrary, in the subfertile group, Fig. 5B, this correlation was not present.

A)


40
30
20
10
0
Total T (nmol/L)
10
20
30
No of CAG Repeats

B)


40
30
20
10
0
Total T (nmol/L)
10
20
30
40
No. of CAG Repeats

DISCUSSION

After lung cancer, prostatic adenocarcinoma is the second leading cause of male cancer death in the US (Parker *et al*., 1997). In Singapore prostate cancer ranks sixth. A widely used marker for the diagnosis of prostate cancer is the measurement of the serum level of PSA. PSA is a protein secreted by the prostate gland. In general, PSA levels increase with advancing clinical stages of tumorogenesis and they are proportional to gland and tumour volume (Stamey *et al*., 1987). PSA determinations are commonly used in screening programs for prostate cancer detection and in monitoring the following therapy. However, its use as an early-stage tumor marker has not been standardized, since a wide disparity in PSA levels occurs within any individual clinical stage and not all prostate cancers cause an increase in serum levels PSA. On the other hand, elevated PSA levels may appear in non-cancerous conditions such as BHP and prostatitis (Schellhammer *et al*., 1993). Despite its problems, PSA is the most reliable marker for tumor growth due to the lack of any other standardized tumor marker for prostate cancer.

The mechanism regulating PSA levels remains ambiguous. One theory suggests that this mechanism consists of free T entering the prostatic epithelial cell, converting T into DHT which binds to the AR and increases PSA production through its binding to the androgen responsive elements located upstream from the PSA gene (Andrews *et al.,* 1992). Laboratory studies confirmed that the promoter region of the PSA gene contains two proximal androgen responsive elements (ARE), located at positions -170bp and -400bp respectively, upstream of the start of the transcription point. A third ARE is located at approximately -4Kb, that functions as an enhancer element. *In-vitro*

studies have demonstrated activation of the expression of the PSA gene by the action of the AR. The activation of PSA by androgens is also suggested by the observation that when patients, in early stages of prostate growth are given anti-androgen therapies, as a standard treatment to block the action of the AR, the PSA levels and prostate volume decreases (Ruckle *et al.*, 1993).

Despite the widespread clinical use of PSA as a tumor marker we do not have good understanding about its role in the prostate physiology. The relationship between androgen action and PSA levels in men without prostate cancer or BPH still remains unclear. It was therefore very important to investigate the factors that may regulate the expression of the PSA protein in individuals with no evidence of prostate growth, since this question has not been adequately addressed in the past.

In this study androgenic parameters relevant to the expression of PSA, such as the levels of free and total T, SHBG, the number of the CAG repeats in the AR gene, and the levels of PSA, were measured in two groups of subjects. The first included 91 fertile males and the second 112 subfertile males.

One of the strongest correlations in this study was that serum T levels in subfertile patients were significantly lower than that of fertile controls (Table 1, Fig. 1) section 1 of results (pages 132 and 134 respectively). It has not been generally appreciated that, in terms of T production, the subfertile male population is relatively hypogonadic (Fig. 1), section 1 of results (page 134). Our large study population, totaling more than 200 men, could be divided into two groups; fertile men with normal T levels and subfertile patients whose steady-state T levels were on average 17% to 20% lower. This finding

gave us the opportunity to explore the steady state effect of normal versus low-normal levels of T on the androgen economy.

In addition to the levels of total T, we measured the levels of free T. The measurements of total T by RIA has traditionally been considered a good reliable method of androgen status. However, more than 98% of T is protein bound: approximately 40% to SHBG and 60% to albumin. Certain physiological stages and conditions are associated with changes in the levels of SHBG. This is the reason why some investigators have suggested the use of the free T levels as a more suitable marker for the androgenic status of the individual. To date the procedures to measure free T levels have proved not to be very reliable. The traditional methods for the measurement of free T are equilibrium dialysis and ultra-filtration. In the first method diluted serum is placed inside a bag of semi-permeable membrane and then placed into a protein-free buffer or saline. Small molecules such as free T pass through the membrane from the sample into the surrounding buffer. This method is very tedious, can not process many samples at the same time, and has many technical errors such as the evaporation of water within the flask. The ultra-filtration method is very similar to equilibrium dialysis. The difference is that the filtrate, with the dissolved solutes, is pulled though the membrane by centrifugal forces. The data we obtained from the Count-A-Count for free T levels were very precise. The results were calculated from the interpolation of a standard curve calibrated in free T concentrations, and the variability among the different measurements was below 15% CV. Free T and total T were very strongly correlated ($P<0.001$) in both fertile and subfertile populations, suggesting that use of the more laboratory-friendly total T assay might be sufficient for most purposes in the future.

The place of SHBG in modulating androgen action is controversial. On one hand, it is generally accepted that androgens, unlike estrogens, reduce SHBG production. Thus SHBG levels are lower in males than in females, and syndromes of hyperandrogenization in females (PCOS, hirsutism and acne) are associated with decreased SHBG levels. Administration of T results in a 2-fold lowering of SHBG in normal and hypogonadal men (Plymate *et al*., 1983). On the other hand, other studies concluded that levels of T and SHBG in males appear to be positively correlated (Carlstrom *et al*., 1990; Gann *et al*., 1996). Furthermore, *in-vitro* experiments with hepatic cell lines indicate that androgens increase, rather than decrease, the synthesis and secretion of SHBG (Edmunds *et al*., 1990) In our study, SHBG levels in the serum of both fertile and subfertile men changed in tandem with T. This has important implications for androgen action since about 40% of T is physiologically bound to SHBG, and is therefore not biologically active. The positive correlation of SHBG with T encountered in our population will tend to minimize and moderate the androgenic effects of varying total T in men.

Androgen action is further modulated by our finding that AR CAG length correlates directly with free T in fertile men. There is increasing evidence from cellular (Tut *et al*., 1997; Chamberlain *et al*., 1994) and human studies that the intrinsic androgenicity of the AR is inversely regulated by the length of its CAG tract. Thus long AR CAG tracts, with lower androgenic potential, are associated with male infertility and depressed spermatogenesis in Singapore (Tut *et al*., 1997; Mifsud *et al.,* 2001) and other population as extensively discussed in Chapter 1. On the other hand, short CAG tracts have higher androgenicity and are associated with androgen-dependent prostate cancer (Giovannucci *et al*., 1997), PCOS (Mifsud *et al*., 2000), and male pattern baldness (Sawaya *et al,* 1998). If so, the effect of low T in fertile men would be

balanced by the increased androgenicity of the AR associated with shorter CAG tracts. Strikingly, no association between T and AR CAG tract was observed in subfertile patients. On the contrary, AR CAG tracts tended to be longer in our subfertile men, thereby accentuating the effects of low T, perhaps contributing to defective spermatogenesis.

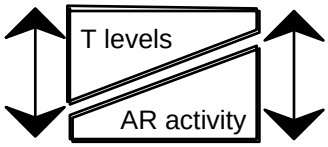
In our study, T strongly correlated with PSA levels only in subfertile men (Table 3), section 2 of results (page 137), suggesting that T drives PSA levels only in conditions of relative low T. Thus T (free T or total T) was highly correlated ($P<0.0001$) with both free PSA and total PSA in these subfertile men, with low-normal T levels. This observation is consistent with previous studies which state that PSA and T are related in conditions where T levels are low, that is at puberty (Kim *et al.,* 1998), in castrated men (Behre *et al.,* 1994), in women (Escobar-Morreale *et al.,* 1998) and in transsexual women where T was administered externally (Goh, 1999). However no correlation between T and PSA was observed in our fertile men with normal T levels. Interestingly, administration of T to healthy young males did not increase the levels of serum PSA (Macindoe *et al*., 1995; Monath *et al*., 1995; Copper *et al*., 1998; Wallace *et al*., 1993). One possible explanation is that in a high androgenic enviroment, the androgen response mechanism is saturated and further increases in T cannot induce additional AR activity or rise in PSA levels.

The relationship between the number of CAG repeats and the levels of serum PSA, was observed only in the infertile group. As in the case of the relationship between the levels of T and PSA, the hypoandrogenaemic status of infertile men could be the reason why the inverse relationship between the number of CAG repeats and PSA is

only observed in infertile men. In these men the AR would not be saturated and therefore it is possible to see that the longer is the number of CAG repeats the lower are the levels of PSA.

Our data suggest that, firstly, PSA correlates with T only in an environment of relatively low androgenicity, such as in subfertile patients. Secondly, in a low androgenic environment, an inverse relationship between the number of CAG repeats and the levels of PSA was found, indicating that short CAG tracts (associated with high AR activity) are related to higher levels of PSA. Fig. 6 summarizes the correlations described above. Overall, our study suggests that interpretation of the clinical significance of PSA levels is best made in conjunction with T levels and AR CAG length. This finding has important implications for screening programs using PSA as the biomarker for prostate cancer in different populations, and may assist in the management of prostate cancer patients before and after androgen blockage.

Figure 6. Factors in the androgen economy in fertile and subfertile subjects.
Although testosterone levels in fertile and subfertile subjects overlap to a degree, differences in AR CAG tract would tend to accentuate the effects of low T in the subfertile group. In this environment of low androgenicity, PSA correlates with both T and AR CAG length in subfertile subjects.

	Fertile Controls	Subfertile patients
Testosterone (T)	High	Low
SHBG*	Correlates with Testosterone in both fertile and subfertile subjects	
AR CAG	Testosterone correlates with AR CAG tract length (Low T compensated by high intrinsic AR activity associated with short CAG tracts, and vice versa)	Tendency to have long AR CAG tracts (Effect of low T increased by low intrinsic AR activity associated with long CAG tracts)
Final effect on androgenicity	 High androgenicity maintained in those with low T	 Low androgenicity accentuated in those with low T
PSA and T	No correlation of PSA with T in environment of high androgenicity	PSA correlates with T in environment of low androgenicity
PSA and AR CAG length	No correlation between PSA and AR CAG tracts in environment of high androgenicity	PSA correlates with AR CAG tracts in environment of low androgenicity

CHAPTER 4: IN VITRO STUDIES OF THE TRANSACTIVATION ABILITY OF THE ANDROGEN RECEPTOR GENE CONTAINING DIFERENT CAG REPEATS

INTRODUCTION

The studies described earlier indicated an effect of the CAG length in male infertility, PCOS, and levels of PSA, strongly suggesting that the CAG length can influence the regulation of androgen driven genes. This chapter will focus on experiments that were designed to illustrate the molecular mechanism(s) whereby CAG repeat length regulates AR transactivation activity. Reports on the correlation between the number of CAG repeats and the activity of the AR to date had been conflicting. Some researchers found an inverse relationship between the number of CAG repeats and the transactivation of the AR, such as Mhatre *et al*., (1993), Chamberlain *et al*. (1994), Jenster *et al*. (1944). Other researchers such as Choong *et al*. (1996) reported a reduction of AR mRNA and AR protein expression with expanded CAG tracts but not a decreased transcriptional activity, when the mouse mammary tumor virus (MMTV) promoter was used as a reporter system in cotransfection studies in CV-1 cells. Similarly, Neuschmid-Kaspar *et al*. (1996) reported no differences in AR activation between expanded CAG tracts containing 40 CAGs (number that is already pathological and causes Kennedy's disease) and values in the normal range.

To date, one of the most commonly used promoters to study androgen action is the (MMTV) promoter. It contains the hormone response element (HRE) consisting of four repeats of the hexanucleotide 5'-TGTTCT-3' to which the steroid receptors bind (Cato *et al*., 1988). There are some commercially available reporter systems that

include in their sequence the MMTV promoter region, such as the pMAM-*neo*-LUC (Clontech). pMAM*neo*-LUC contains the firefly luciferase as a reporter gene to provide a positive control for the expression of the MMTV-LTR promoter of pMAM-*neo.* Luciferase is one of the most sensitive reporter genes available. pMAM*neo*-LUC also expresses ampicilin resistance, which permits positive selection of transformed cells. However, the MMTV promoter is a viral promoter and therefore not endogenous to the human genome. Functional studies of subtle AR mutations using this promoter might not reflect the situation *in-vivo*. Moreover, the hormone response element (HRE) in the MMTV promoter is also the consensus sequence for the glucocorticoid receptor (GR) and progesterone receptor (PR). Therefore, in order to overcome the aforementioned issues and to study the effect of ARs containing different sizes of poly-Gln tracts on their transactivation activity, reporter systems containing human promoter sequences with AREs are necessary.

Although the expression of many genes is thought to be androgen dependent, the number of human genes whose transcription is known to be directly regulated by androgens is limited. Among the best described androgen regulated genes are the aldehyde dehydrogenase 1 (ALDH1) which is expressed among other tissues in the testis and the PSA which is expressed mainly in the prostate,

1. Aldehyde dehydrogenase gene 1

Aldehyde dehydrogenase gene 1 (ALDH1) enzymes are considered as general detoxifying enzymes, which eliminate toxic aldehydes. In the human, ALDH1 has a role in the biosynthesis of retinoic acid, which is a modulator for gene expression and cell differentiation. Northern blot analysis showed that liver tissue, pancreas tissue,

hepatoma cells, and genital skin fibroblast cells express high levels of ALDH 1 (Yohida *et al*., 1993).

The human ALDH1 gene is expressed in the normal genital skin fibroblast cells but not in the cells obtained from X-linked AR negative testicular feminization patients (Yohida *et al*., 1993). These findings suggested that the ALDH1 gene is induced by the AR complex in genital cells. Moreover a putative ARE exist at positions –688/-674 (GTGACAgtgTGTTCC), suggesting the possibility of androgen-dependent regulation of the ALDH1 expression in genital skin (Yuchio *et al*., 1995). However, the molecular mechanism of the androgen-receptor mediated expression of ALDH1 is unknown.

2. Prostate specific antigen gene

Prostate specific antigen gene (PSA), a member of the Human Kallikrein gene family (hKLK), is a serine protease. The mature form of PSA is a single chain glycoprotein of 237 aa and has a molecular weight of approximately 30 KD. Three human kallikrein had been described, hKLK1for human pancreatic/renal kallikrein, hKLK2 for human glandular kallikrein-1, and hKLK3 for PSA which is a well-known prostate-specific tumor marker. The three genes are clustered in an area of 60 Kb on human chromosome 19q13.2-13.4 (Berg *et al*., 1999).

The PSA promoter region at position –170bp to –156bp contains the sequence AGAACAgcaAGTGCT, named androgen response element I (ARE-I), which resembles the consensus sequence GGTACAnnnTGTTCT for the glucocorticoid and progesterone receptor. Riegman *et al*, (1991) demonstrated in cotransfection

experiments that mutations of this motif abolished the transcription activation by the AR. This sequence is thus important for regulation of transcription by their interaction with the AR complex. Upstream this region, at position –394bp, there is another androgen response element (ARE-II) with the sequence (GGATCAgggAGTCTC) that acts in synergy with the ARE-I.

However, the PSA promoter is not specifically activated by androgens alone. Creutjens *et al*., (1997), demonstrated that it can also be activated by glucocorticoids in most of the cell lines tested except in the prostatic cells named LNCaP cells. The specificity of the PSA promoter in LNCaP cells could be due to the absence of other members of the steroid receptor family in this particular cell line, or the presence of certain regulatory proteins involved in PSA expression (Creutjens *et al*., 1997).

In humans, PSA is highly expressed in the luminal epithelial cells of the prostate, an androgen dependent organ, and expressed at very low levels in other tissues. For instance PSA has also be found in breast cancers, salivary gland neoplasms, cells of the male urethra, breast milk, blood and urine. PSA produced in the prostate is secreted into the seminal fluid in high concentrations. A major function of PSA is the proteolytic cleavage of gel-forming proteins in the seminal fluid, resulting in the liquification of the seminal gel and increased sperm mobility.

MATERIALS AND METHODS

Promoter regions of two human genes, PSA and ALDH1, that contain ARE in their sequences were cloned in front of the luciferase gene of a reporter system, and their responsiveness to androgen action tested in cell culture experiments. The commercially available reporter system pMAM-*neo*–LUC (Clontech) uses the MMTV promoter to drive the expression of the luciferase gene. Firstly, the ALDH1, and PSA reporter systems, were constructed by subcloning the promoter regions in front of the luciferase gene of pMAM*neo*-LUC after removing the MMTV promoter fragment.

1. Amplification of the prostate specific antigen promoter regions of 1.6Kb, 630bp and 100bp

The regions of 1.6Kb and 630bp upstream of the start point were selected because they contained the two ARE necessary for the androgen driven activity of the PSA promoter. The longer 1.6 Kb region was selected in order to include some neighbouring sequences that might be the target site for proteins acting as coactivators of the AR. The PSA 1.6Kb promoter region was isolated from total genomic DNA of a healthy individual by PCR. The specific primers designed were the following:

Forward: 5'-ACGGTCCATATG GATCAAGTCAGCTACTCTGG-3'

Reverse: 5' AGCCGTCAGCTGAAGCTTGGGGCTGGGGAGCC-3'

The underlined regions correspond to the restriction sites *Nde I* and *Sal I* in the forward and reverse primer respectively. The total volume of the reaction was 100µl and it contained 200ng of total genomic DNA used as a template, 10x cloned buffer, 0.2µM of each primer, 2.5 mM of deoxyrobonucleotide triphosphates and 5U of cloned Pfu DNA polymerase enzyme (Stratagene). PCR conditions were as follows, a first cycle

at 95°C for 5 minutes followed by 30 cycles of 95°C for 30 sec, 60°C for 30 sec and 72°C for 5 min. The product of the PCR reaction was purified by the use of Wizard PCR preps Rapid Purification System (Promega). The purified PCR product was then double digested with the restriction enzymes *SalI* and *Nde I*. The reaction contained 10 units of each enzyme, 10x *Sal I* buffer and was incubated at 37°C for 4 hours.

The method to isolate the 630bp and 100bp PSA promoter regions was very similar. Specific forward primers containing the restriction site for *NdeI* were designed.

630bp: 5'-ACGGTC CATATG AGACTCATATTTCTCGGCTA-3',

100bp: 5'-ACGGTC CATATGAGGGGGTTGTCCAGCCTCCA-3'

The reverse primer remained the same as for the PSA 1.6 Kb fragment.

Partial digestion of the pMAM-*neo*-LUC Vector.

The aim was to remove the mammalian viral promoter from the reporter system pMAM and to insert instead the human PSA promoter previously isolated. The vector contained more than one restriction site for *SalI* therefore partial digestion on the vector was performed in order to isolate the appropriate band. The enzymes used for digestion were *SalI* and *NdeI*. Firstly, the vector was digested with *NdeI* and subsequently partially digested with *Sal I*.

2. Amplification of the aldehyde dehydrogenese 1 gene promoter region

The ALDH reporter system was similarly constructed. The restriction enzymes chosen did not cut the vector or the ALDH1 promoter region selected. The promoter region of the ALDH1 gene was obtained from total genomic DNA by PCR amplification using the specific primers:

Forward: 5’ ACGGTCCATATGGTAACGTTTGCTAGAGCTAC

Reverse: 5’ ATAGCACGATCGTCTGATTCGGCTCCTGGAAC

The underlined regions correspond to the restriction sites *NdeI* and *NheI* in the forward and reverse primers respectively. The PCR reaction included 200ng of genomic DNA used as a template, 10x cloned Pfu buffer, 0.2µM of each primer, 2.5 mM of deoxynucleotide triphosphate, and 5 U of cloned Pfu DNA polymerase enzyme (Stratagene). The PCR conditions were 30 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 5 minutes with an initial cycle of 95°C for 5 minutes. Both insert and vector were double digested using 10units of the enzymes *NdeI* and *NheI* and then purified with Wizard PCR Preps Rapid Purification System (Promega).

3. Ligation reactions

The PSA and ALDH promoter fragments were ligated to their respective vector by using the enzyme T4 ligase (NE Biolabs). The relative amounts of insert and vector were previously estimated in an agarose gel.

4. Transformation

Epicurian Coli XL1-Blue MRF’ Kan (Stratagene) were used for transformation. Firstly 20µl of cells were thawed on ice and incubated for 10 minutes with occasional swirling. 1 µl of the ligation reaction was added to the cells and then incubated on ice for another 30min. Cells were heated at 42°C for 90 sec, and afterwards placed on ice for another 2 min. 0.25ml of SOC media previously warmed at 37°C was added and the new mixture incubated at 37°C for 1 hour while shaking at the speed of 250rpm. 75µl of the cells were plated in a LB-ampicillin agar plate while the remaining quantity was plated into another plate to avoid overgrowth in a single plate. Plates were

incubated overnight at 37°C. The following morning the positive colonies were selected, and the presence of the expected insert checked. The procedure was the following; colonies were slightly touched with a tip and the tip deposit on 30µl of TE media. The mixture was heated at 95°C for 5min to brake the cells. Then 1µl was used as template for a subsequent 50µl PCR reaction. The PCR conditions were the same as the previously used to amplify the promoter region of the respective genes.

5. Purification of plasmid DNA

The positive colonies were selected and incubated overnight with 5ml of LB-ampicillin media at 37°C with vigorous shaking for 16 hours. 2ml of the bacteria culture were used for small-scale extraction using Qiagen plasmid preparation kit. The remaining media was used to prepare glycerol stock. From the glycerol stock cells were plated on a LB ampicillin agar plate O/N. One single colony was selected and incubated on 30ml of LB-ampicillin media for about 10 hours. Afterwards the 30ml were transferred to 500 ml of LB-ampicillin media. The culture was allowed to grow overnight at 37°C under 250rpm of shaking. The plasmid isolation was performed following the Qiagen Plasmid maxi preparation Kit protocol.

6. Sequencing the PSA-LUC construct

The complete region of the 1.6Kb region was sequenced to confirm the fidelity of the enzymatic manipulations. Internal primers were designed in order to cover the whole region. Each of the sequence reaction contained 1.6 µg of the PSA-LUC plasmid, 3.2 pmol of the primer, 8µl of terminator ready reaction mix (PE) in a total volume of 20µl. The cycling conditions were 5 cycles at 96ºC for 10 seconds, 50ºC for 5 seconds and 60ºC for 4 minutes. In order to purify the sample 2µl of 3M sodium acetate

(pH5.2) and 50µl of 95% ethanol were added to the previous reaction, mixed, and then place on ice for about 10 minutes. After being centrifuged at a maximum speed for 30 minutes the ethanol solution was discarded and the pellet washed with 70% ethanol. The pellet was dried in a vacuum pump. The samples were then re-suspended in 6µl of loading buffer. The loading buffer solution contains formamide and EDTA/Blue dextran in a ratio 5:1. The composition of the sequencing gel was the same that the GeneScan gel, a 4% acrylamide solution. About 2µl of the sequence sample were loaded into the gel. The sequencing run took about 7 hours. The ABI PRISM Dye Terminator Cycle Sequencing kit (Perkin Elmer) was used for sequencing. The automated sequencing was carried out on an ABI PRISM 377 DNA automated sequencer.

7. Transient transfection experiments

Transient co-transfection experiments were optimized using various cell lines such as CV-1, COS-7, LNCaP, and He-La. The reporter system transfected and the amount of plasmid employed varied depending on the purpose of the experiment but the protocol was very similar. The method to introduce the DNA into the cells was lipofectamine (Gigco BRL), a liposomal solution. Since the size of tissue culture plates varied throughout the experiments, the relative amounts of DNA were adjusted for this factor. When the plates used for transfection were of 60mm of diameter, called p60, the procedure was as follow: 200 µl of the solution A, containing the plasmid DNA and 200µl of solution B, containing the lipofectamine solution were carefully mixed. The mixture was incubated for 45 minutes to allow for the formation of the liposomes. Cells to be transfected were washed twice in serum free media. The mixture (400µl) was added to the cells and supplemented with 2ml of serum free media. Cells were

then incubated for 5 hours at 37°C and 5%CO2. During this period the liposomes merged with the cellular membrane allowing the DNA to enter into the cell. It was important to transfect the cells when they were actively dividing that is in their exponential growth phase. After 5 hours of liposomal exposure the media on the cells was replaced by the growth media which was supplemented with 10% charcoal treated FBS and the hormone used for induction. Cells were then placed in the incubator for another 48 hours.

8. Harvesting the cells

After the 48 hours period of incubation the cells were harvested from the culture plates. Firstly, the cells were washed twice with PBS solution buffer. For the p60 plates 400µl of the commercially available Reporter Lysis buffer (Promega) was added to the cells and with the help of a syringe plunger the cells were scrapped. The lysate was deposited in a tube and placed on ice for the subsequent measurement of luciferase activity.

9. Meassurement of luciferase activity

Cell lysates were centrifuged at 13000rpm to separate the cellular debris from the soluble fraction where the luciferase protein was present. The substrate employed was the commercially available Luciferase Assay Substrate Lyophilazed (Promega). 100µl of the substrate were aliquoted into the each tube. Then 20 µl of the cell extract where the luciferase enzyme is present was added. After a quick mix, the tube was placed in the luminometer cuvette where the emission of light was measured. The luminometer used was TD 20/20 Turner Designs Luminometer (Promega). Luciferase activity was measured in RLU.

10. Construction of the new PSA reporter system

The previous reporter systems were not very effective because the intrinsic base line activity, in the absence of hormonal treatment, was very high. In order to solve this problem another reporter plasmid was selected, the pGL3-basic luciferase vector (Promega). The promoter region selected was the same, the 1.6 Kb region upstream the start of the PSA transcription point. It was isolated from the previously engineered pM-PSA-LUC reporter system after double digestion with *Hind III* and *KpnI*. The 20µl reaction contained 10 units of each enzyme, 4µg of the pM-PSA-LUC, and 10x buffer, and was incubated at 37C for 2 hours. The fragment of interest was isolated from a TAE buffer agarose gel and subsequently gel purified with gene clean kit II (Promega). The reporter system, pGL3 was double digested with the same enzymes, and then purified with gene clean kit II (Promega). The following steps include ligation, transformation, selection of the transformant bacteria containing the insert of interest, and large-scale preparation of the new construct. The new reporter system was called pGL3- PSA-LUC.

11. Mammalian cell two-hybrid system experiments

The TAD region of the AR gene containing 14, 20 and 28 CAG repeats in the exon 1 was fused to the VP16 activation domain (pVP16AD-ARTAD), and the DBD-LDB region of the AR gene was fused to the pGAL4 domain (pGAL4 DBD-LBD). The reporter system used was the GAL4-TATA-LUC that contains multiple Gal 4 binding sites. (Lim *et al*., 2000). Tranfections experiments were performed in HeLA cells using the lipofection technique. The transfection experimental procedure was the same as described previously. Three independent transfections were performed. The hormone used for induction was MB.

RESULTS

Reporter systems containing human promoters sequences of the PSA and ALDH1 genes were constructed. Initially the regions of interest were cloned into the commercially available pMAM*neo*-LUC by first removing the mouse mammary virus promoter region that contains the hormone responsive element (HRE) and inserting the selected PSA and ALDH promoter regions. The reporter systems were tested in cell culture experiments.

1. Construction of the ALDH1 reporter system

The promoter region of the ALDH gene spanning from –977 to +12 was amplified using the following flanking primers:

Forward primer: 5'ACGGTC CATATG GTAACGTTTGCTAGAGCTAC 3'

Reverse primer: 5' CAAGGTCCTCGGCTTAGTCT GCTAGC ACGATA 3'

The underlined regions correspond to the restriction sites for *NdeI* in the forward primer and the *Nhe I* in the reverse one. The PCR reaction yielded a unique fragment of the expected size (989 bp) as shown in Fig. 1A. The fragment was purified with the geneclean kit II and then double digested with the corresponding restriction enzymes. In order to remove the MMTV promoter region and incorporate the ALDH region, the reporter plasmid pMAM was digested using the same enzymes *Nde I* and *NheI* and subsequently the products of the digestion resolved in a 0.8% agarose gel. The band corresponding to the vector without the MMTV promoter was isolated and gel purified with the geneclean II Kit (Fig. 1B). The Insert and vector are shown in Fig. 1 C. E. Coli supercompetent cells (Epicurian Coli XL1-Blue MRF'Kan) were transformed and after an incubation period of 16 hours at 37°C, a few colonies were

Figure 1. Construction of the ALDH reporter system.

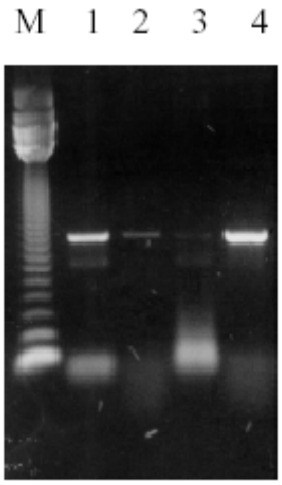
A) Amplified region of the ALDH promoter region of 989bp after the PCR reaction.

Each reaction (lanes 1-4) contained different amounts of template (total genomic DNA). The marker used was the 100bp ladder (M).

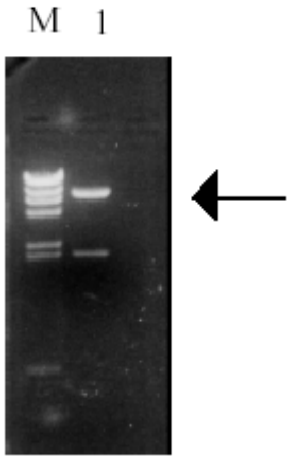
B) Product of the pMAM plasmid digestion with the enzymes *NdeI* and *NheI*. The first band in lane 1 (arrow) corresponds to the pMAM reporter system without the MMTV promoter region (8Kb). The marker used to estimate the sizes was Hind III-digested λ-DNA (M).

C) Agarose gel to estimate the relative amounts of the vector (V) and insert (I) before proceeding with the ligation reaction. The markers used were 100bp marker (M1) and Hind III-digested λ-DNA (M2).

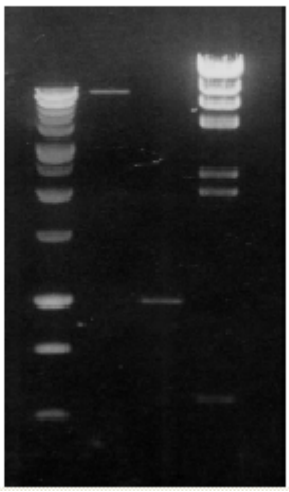
A)



B)



C)



screened for detection of the presence of the plasmid with the insert. One of the selected colonies resulted in positive. The positive colony was placed on 5ml of LB-ampiciline media and incubated overnight at 37°C. Large-scale preparation of the plasmid (Quiagen kit) yielded enough DNA to perform the subsequent transfection experiments.

2. Construction of the PSA reporter systems containing the -1.6kb, -630bp and -100bp promoter regions

The region comprising the section (–1600, +12) of the PSA promoter gene was amplified from total genomic DNA using specific primers with restriction sites for the enzymes *Sal I* and *NdeI* respectively. As shown in Fig. 2A, the PCR yielded a single product of 1.6Kb, the expected size. The product of the PCR was then purified. The PCR product was then double digested with the restriction enzymes *NdeI* and *SalI*. The pMAM plasmid was firstly digested with *Nde I* followed by partial digestion with *SalI*. The partial digestion resulted in five fragments of the predicted size Subsequently the band of interest (approximately 8kb) was selected and gel-purified (Fig.2B). The purified plasmid and the PSA promoter region were subsequently ligated. After the ligation reaction, E. Coli competent cells (Epicurian Coli XL1-Blue MRF'Kan) were used for transformation. After an incubation period of 16 hours at 37°C some colonies were selected for screening.

One of them contained the plasmid with the newly cloned PSA promoter region (Fig. 3). The positive colony was allowed to grow for 16 hours at 37°C in LB-ampicilin media and then a small-scale plasmid preparation performed. The newly inserted fragment was sequenced to confirm its integrity and to exclude the presence of new

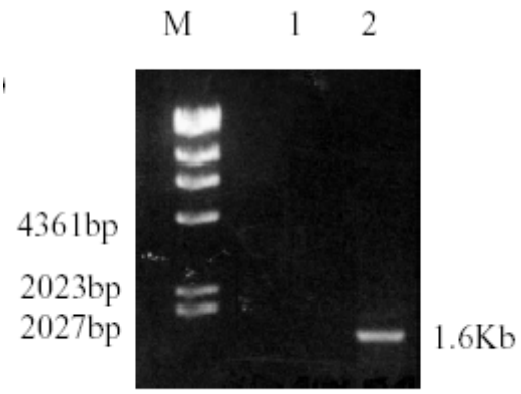
Figure 2. Cloning the 1.6Kb region of the PSA promoter gene into the pMAM*neo*-LUC reporter plasmid.

A) Isolation of the 1.6kb fragment from genomic DNA by PCR. 10μl of the PCR was analyzed on a 1% agarose gel previously stained with ethidium bromide. The reaction resulted in a unique product (lane 2). Lane 1 corresponds to the negative control (same reaction components with no template DNA). Hind III-digested λ-DNA was used as a size standard marker (M)

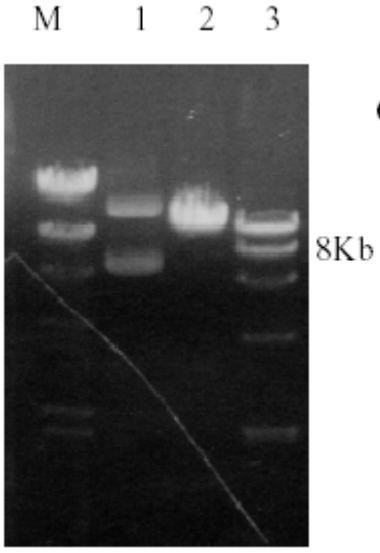
B) pMAM plasmid after partial digestion with *SalI*. Lane 1 corresponds to the uncut pMAM vector, lane 2 corresponds to the pMAM vector after single digestion with *NdeI* and lane 3 to the pMAM plasmid after partial digestion with *SalI*. The second band of 8Kb corresponded to the band of interest thus is the pMAM plasmid after the removal of the MMTV promoter sequence. Hind III-digested λ-DNA was used as a size standard marker (M).

C) Schematic representation of pMAMneo-LUC indicating the position of the restriction sites for *SalI*.

A)


M
1
2
4361bp
2023bp
2027bp
1.6Kb

B)


M
1
2
3
8Kb

C)

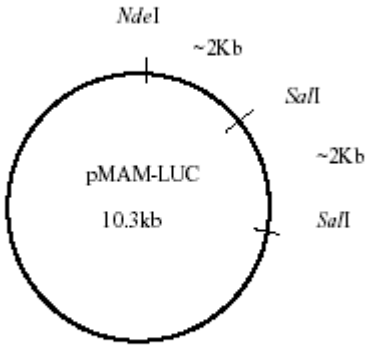
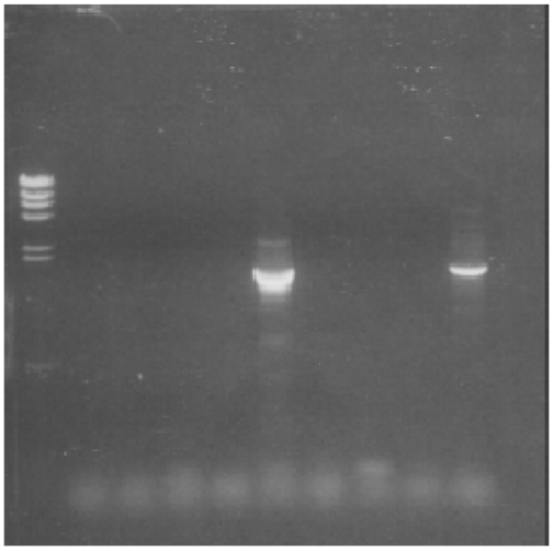

NdeI
~2Kb
SalI
~2Kb
pMAM-LUC
10.3kb
SalI

Figure 3. Screening for positive colonies after transformation.

Each lane in the 1% agarose gel corresponds to each of the colonies selected for screening. Only one of the colonies selected (lane 5) contained the plasmid with the newly engineered insert. The last lane (C) corresponds to the positive control, where the source of DNA was total genomic DNA. Hind III-digested λ-DNA was used as a size standard marker (M).

M 1 2 3 4 5 6 7 8 C



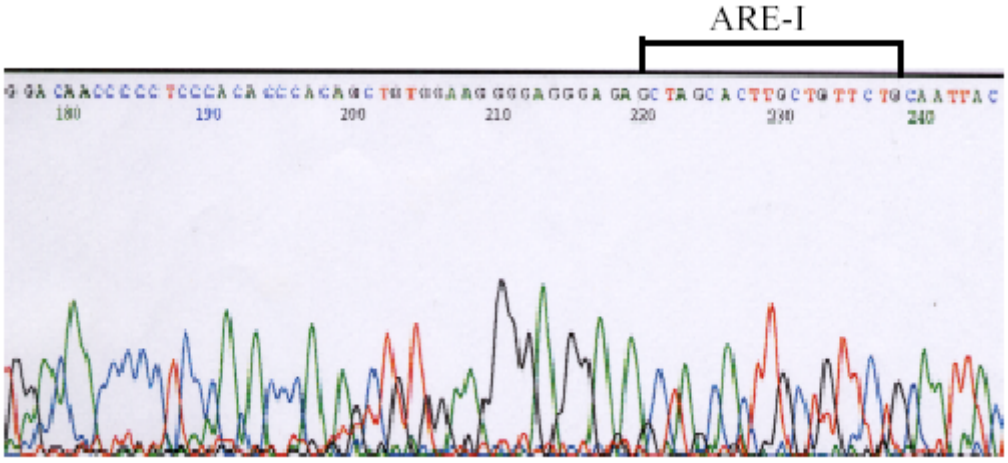
incorporated mutations. Fig. 4 A and B show the sequence of the PSA promoter region containing the two ARE. Sequence analysis of this promoter revealed the presence of a first androgen response element, (ARE-I) at position –170, and (ARE-II) at position –400.

To determine which segments of the PSA promoter are necessary and sufficient for androgen action, two other reporter systems comprising the regions -630bp and –100pb upstream of the start of the translation point were similarly constructed. The 100bp PSA promoter fragment contains no ARE but has a TATA box. This reporter system served as negative control for androgen action.

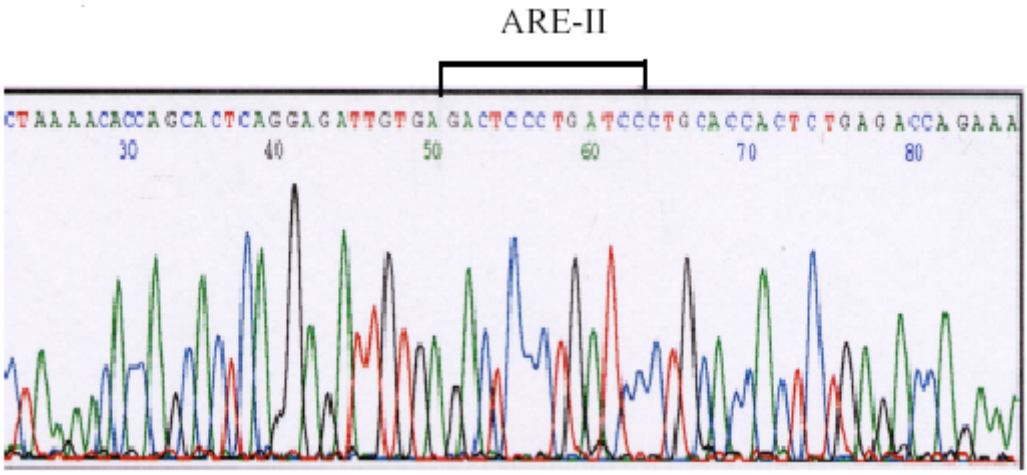
Figure 4. Sequence analysis of the promoter region of the PSA gene.

Sequence analysis revealed the presence of two ARE. A) ARE-I located at position -170 and B) ARE-II at position -400 of the start of the transcription point respectively.

A)



B)



3. Transient co-transfection experiments

Firstly, the capability of the AR protein to bind to the ARE at the promoter regions of the ALDH and PSA genes and activate the luciferase gene was evaluated. The next step was to determine if the size of the poly-Gln stretch in the AR played a role in the transactivation capability of the AR. The lengths of the poly-Gln selected in the AR were those coded from expression plasmids containing 14, 20 and 28 CAG repeats respectively. They were previously constructed in our laboratory (Tut *et al*., 1997). Cells were co-transfected with the AR expression plasmid and one of newly engineered reporter systems either pM-PSA-LUC or pM-ALDH-LUC.

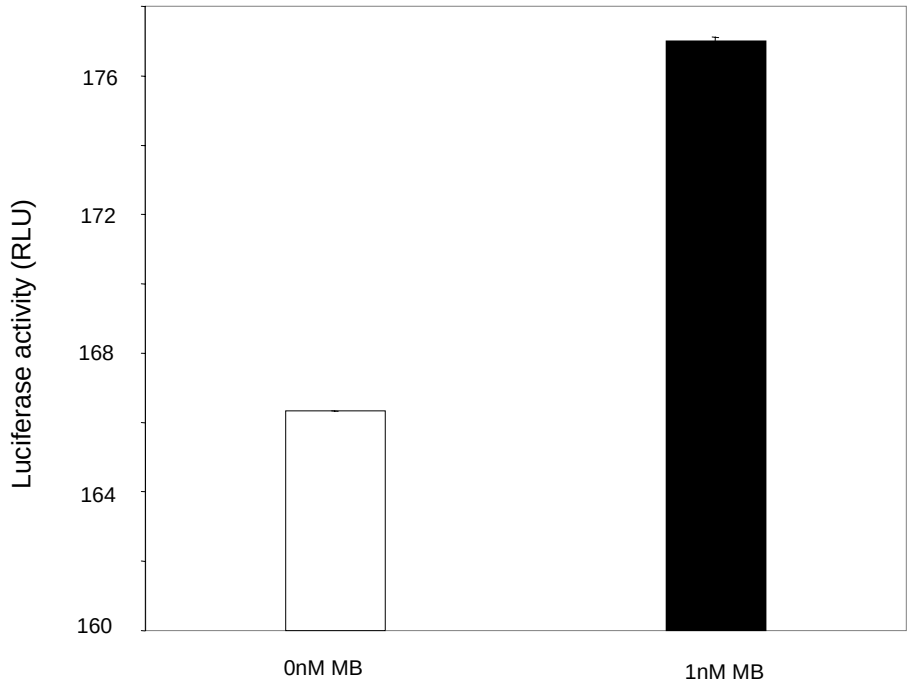
3.1 Transient co-transfection experiments to evaluate the responsiveness of the pM-ALDH-LUC reporter system to androgen action

The newly constructed pM-ALDH-LUC reporter system was tested in cell culture experiments in CV-1 and COS–7 cell lines. They were selected for transfection experiments because they do not express the AR protein. Cells were co-transfected with the AR expression plasmid (pSVhARo) containing 20 CAG repeats and the reporter system pM-ALDH- LUC. The AR was induced with 1nM of MB. As we can see from Fig. 5A and B, the base line activity of the reporter system in the absence of hormonal treatment was very high because of high base levels, and the fold induction after the hormonal exposure very low (folds induction of 1.06 and 1.30 in C V-1 and COS-7 cells respectively), indicating that this ALDH-LUC was not a useful reporter system of androgen action.

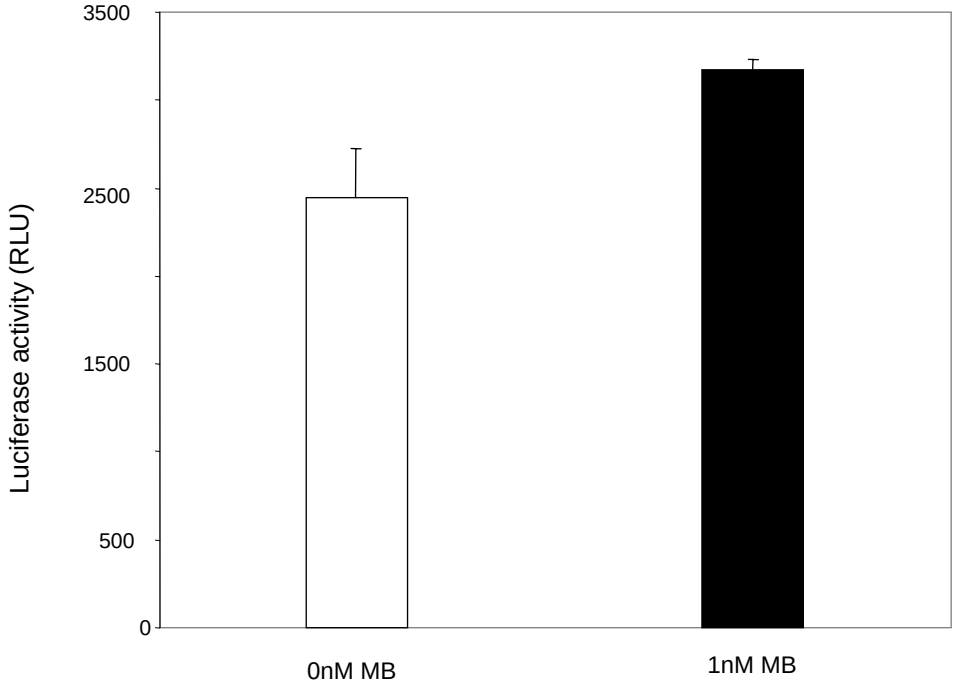
Figure 5　Androgen regulation of the ALDH promoter in CV-1 and COS-7 cells.

CV-1 and COS-7 cells were transiently cotransfected with the ALDH-LUC reporter system and the AR expression plasmid. Cells were incubated for 48 h in the absence (white bars) and presence (black bars) of MB (1nM). The Y axis indicates the absolute values of luciferase activity (RLU).

A) ALDH Reporter system in CV-1 cells



B) ALDH Reporter system in COS-7 cells



3.2. Transient co-transfection experiments using the pM-PSA-LUC reporter systems

Co-transfection studies were performed in CV-1 cells using one of the reporter systems containing the selected regions of the PSA promoter gene (1.6Kb, 630bp or 100bp) along with the AR expression plasmid. As shown in Fig. 6, the reporter system containing the 1.6 Kb fragment induced in a higher manner the expression of the luciferase gene compared to the reporter systems containing the 630bp and 100bp fragments. The 1.6 kb fragment was therefore selected for further studies.

3.3 Inducibility of the PSA 1.6Kb fragment of the PSA promoter by the androgens DHT and MB

CV-1 cells were cotransfected with either the pM-PSA-LUC or pMAM reporter systems, and the AR expression plasmid (PSVhARo). Cells were induced with the natural ligand DHT or the more powerful synthetic androgen MB (Fig. 7). The pM-PSA-LUC reporter system induced 4 fold the expression of the luciferase gene at doses of 10nM of DHT. Similar results were obtained with MB whereby the maximum fold induction was 5 fold at doses of 1nM. On the contrary, cotransfection with the commercial reporter system pMAM resulted in a higher fold induction (approximately 20). This was due to its lower base line activity.

3.4 The activity of the 1.6 Kb PSA promoter was evaluated in other cell lines

In order to elucidate if the high base line of PSA-LUC reporter system was cell type specific, co-transfection experiments were done in different cell lines such as CHO, COS-7 and LNCaP. The hormone used for induction was MB at doses of 1nM (Fig. 8). With CHO cells, a 3-fold induction was achieved after hormomal exposure.

Figure 6 Androgen regulation of the PSA promoter in CV-1 cells.

CV-1 cells were co-transfected with one of the three PSA-LUC reporter systems containing the PSA promoter fragments of different sizes (100bp, 630bp or 1.6Kb) and the AR expression plasmid (PSVhARo). The Y axis indicates the fold induction values after exposure to 1nM MB. The maximum fold induction was achieved with the 1.6Kb fragment of the PSA promoter gene.

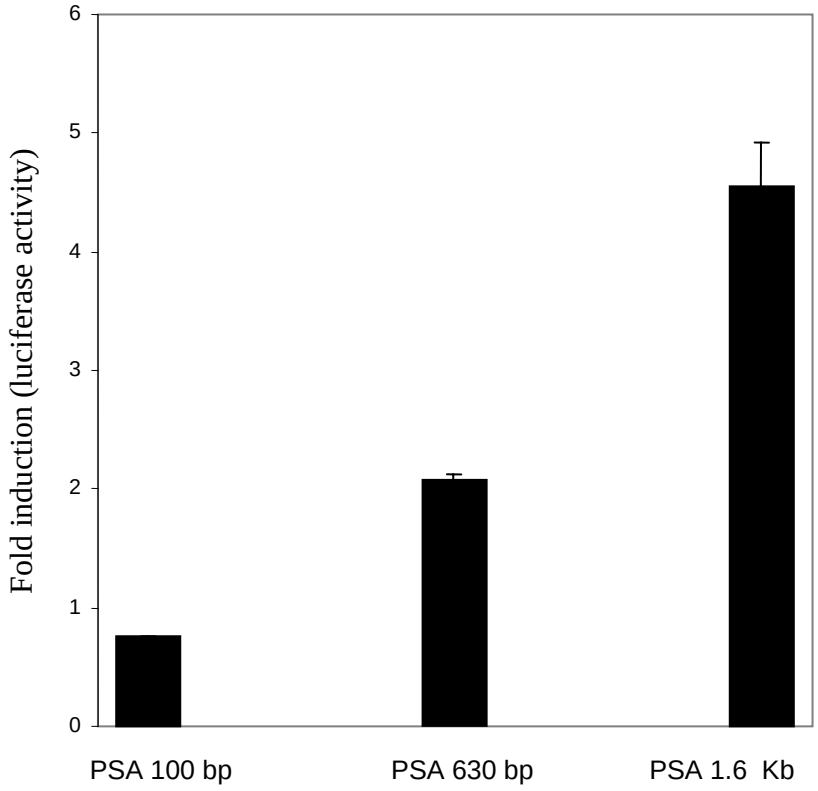
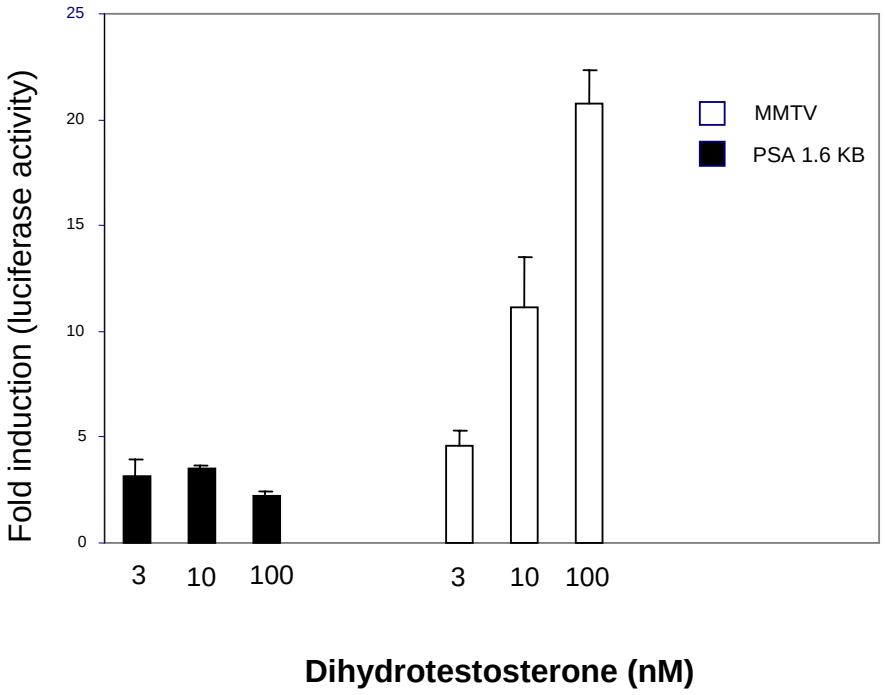

Fold induction (luciferase activity)
6
5
4
3
2
1
0
PSA 100 bp
PSA 630 bp
PSA 1.6 Kb

Figure 7 Dose response activities of the PSA and MMTV promoter in CV-1 cells.

Cells were cotransfected with either pMAM (solid bars) or the PSA-LUC,1.6Kb reporter system (blank bars) and the AR expression plasmid (PSVhARo). After transfection cells were induced with three different doses of A) DHT and B) MB.Y axis indicates the fold induction values

A)


Fold induction (luciferase activity)
25
20
15
10
5
0
MMTV
PSA 1.6 KB
3
10
100
3
10
100
Dihydrotestosterone (nM)

B)

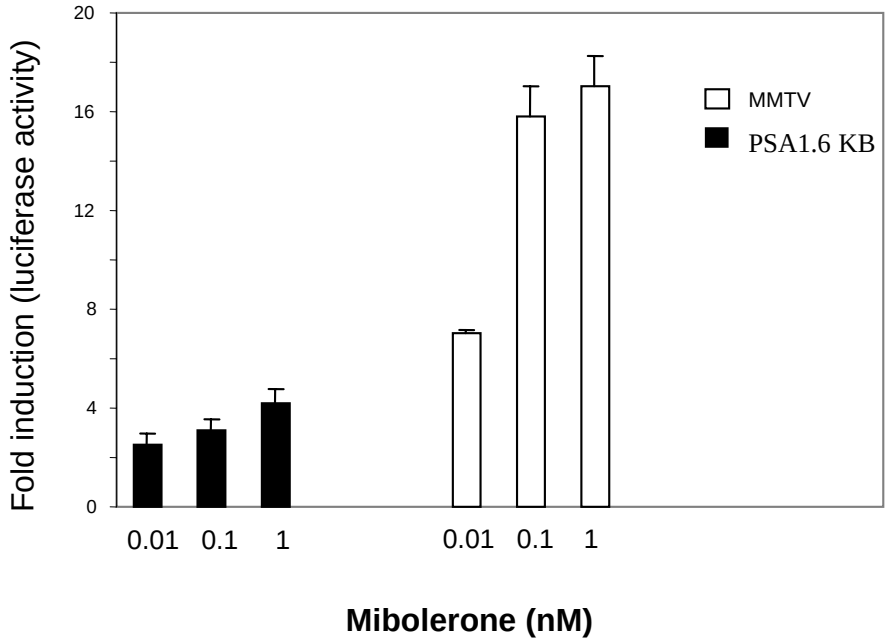
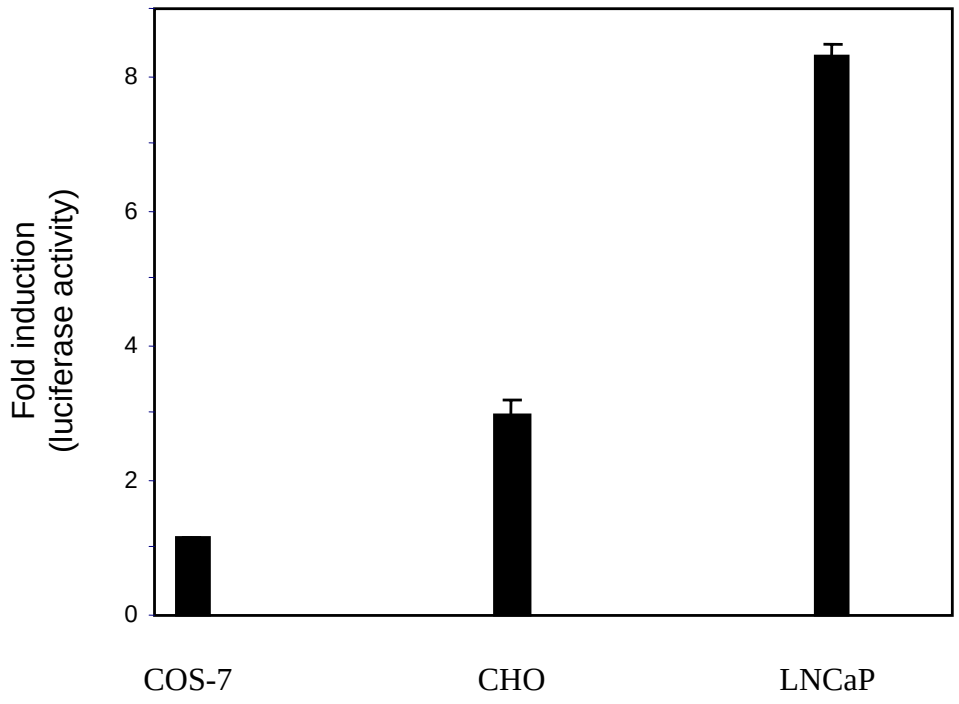

Fold induction (luciferase activity)
20
16
12
8
4
0
MMTV
PSA1.6 KB
0.01
0.1
1
0.01
0.1
1
Mibolerone (nM)

Figure 8 Co-transfection experiments with PSA reporter system in different cell type lines.

Co-trasfection experiments were performed in the COS-7, CHO and LNCaP cell lines. The highest fold induction of more than 8 was achieved when LNCaP cells were transfected.


Fold induction
(luciferase activity)
0
2
4
6
8
COS-7
CHO
LNCaP

When the cells co-transfected were COS-7 cells the fold induction was of only 1.13. The highest fold induction was observed in LNCaP cells whereby a 8.3 fold increased was achieved when cells where exposed to doses of 1nM MB. The reason for the higher fold induction when LNCaP cells were co-transfected may be due to the presence of cell specific coactivators than interact with the AR and modulate the activation of the PSA promoter.

In summary, the pM-PSA-LUC reporter system did not prove to be as useful as expected. The reason for the low level of induction after androgen exposure was the high base line activity of the promoter in the absence of hormone.

3.5 pGL3-PSA- LUC reporter system

The subsequent steps were to design another reporter system using a different vector. The vector chosen was pGL3 basic-luciferase reporter vector (Promega). The promoter region selected remained the same, the 1.6Kb region of PSA promoter gene.

The 1.6Kb promoter region of the PSA gene was obtained from the previously engineered reporter system. After double digesting them with *Hind III* and *KpnI,* the band of interest was isolated from a 1% TAE agarose gel and then gel-purified with genescan Kit II. The pGL3-basic vector was double digested with the same enzymes and after the ligation reaction, E.Coli supercompetent cells were transformed. The positive colonies were screened for the presence of the insert. The newly engineered reporter system was called pGL3-PSA-LUC.

3.6 Cotransfection studies with the reporter system PGL3-PSA-LUC

The newly engineered reporter system, PGL3-PSA-LUC, was tested in CV-1 cells and results compared to those of the commercially available pMAM. The newly engineered pGL3-PSA-LUC proved to be quite responsive to the action of androgens, more so than pMAM. In the first experiment, before optimising the system, the fold induction after hormonal stimulation was of 27. The main difference with the previously engineered pM-PSA-LUC was that the base line activity, in the absence of hormonal treatment, was very low.

3.7 Optimisation of the relative amounts of AR and PGL3-PSA-LUC .

Various experiments were designed to optimise the best relative amounts of PSVhARo and PGL3-PSA-LUC when cells where co-transfected. During the optimisation procedure transfection experiments were carried out in 6 well culture plates. Firstly, the amount of PGL3-PSA-LUC was kept constant to 333ng/well and the amount of PSVhARo added was varied from 40, 80, 166 to 320 ng/well respectively. Results indicated that the highest induction was achieved when the amount of PSVhARo used was of approximately 100ng/well. Then, the amount of PSVhARo was kept constant, and that of pGL3-PSA-LUC was varied from 75, 150, 333 to 666ng/well respectively. Results indicated than the amount of PSA-LUC of 666ng/well gave the highest fold induction (30 fold) after hormonal treatment with 1nM of MB.

3.8 Differences in the CAG length in the AR may transactivate the PGL3-PSA-LUC reporter system differently

The aim of the construction of reporter systems with human promoters containing ARE was ultimately to determine if there were differences in the way AR with different length of poly-Gln would activate such human promoters. To address this question, co-transfection experiments with the reporter system pGL3-PSA-LUC and one of expression plasmids containing 14, 20 or 28 CAG repeats in their AR gene were performed. The hormones used for induction were the synthetic MB and the naturally occurring DHT and T. A correlation between the number of CAG repeats in the AR and its the transactivation capability was not found regardless of the hormone and hormonal dose used for induction. The results obtained from the various co-transfection experiments were quite inconclusive. Perhaps the methodology employed was not sensitive enough to detect the hypothesised subtle differences that may exists, in spite of repeated experiments.

4. TAD and LBD interaction using the mammalian cell two-hybrid system

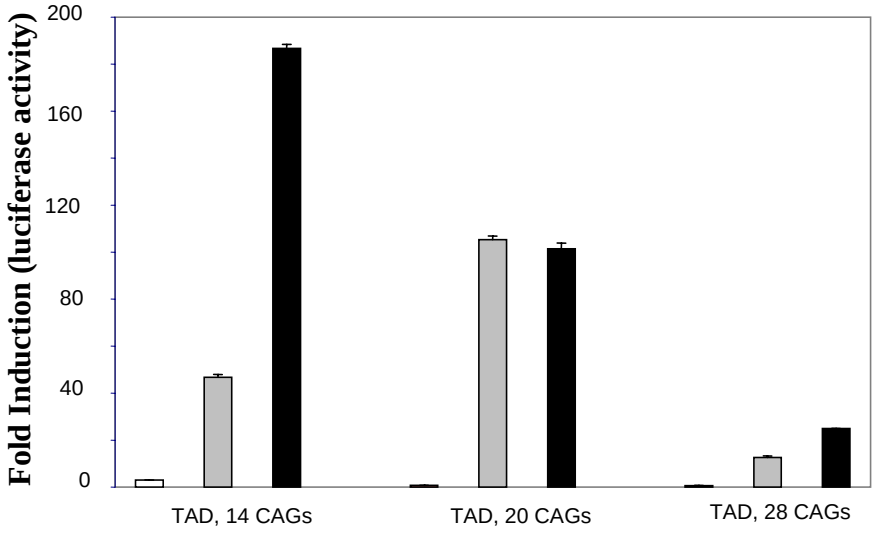
To further understand the mechanism by which the length of the poly-Gln tract in the AR may modulate the transactivation capability of the AR, interactions between the TAD and LBD domain of the AR gene were studied by using the mammalian two-hybrid system.

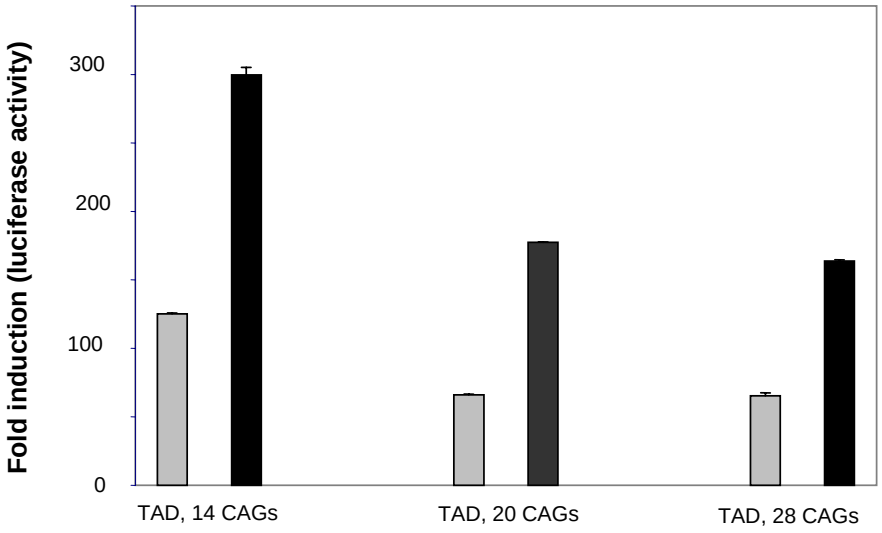
The TAD domain of the AR containing different CAG repeats length tracts (14, 20 and 28) were fused to the yeast VP16 activation domain, and the AR-DBD-LBD domain to the yeast GAL4 domain. The fusion proteins were co-expressed in HeLa cells, and protein-protein interactions were measured with a reporter vector containing multiple GAL4 DNA-binding sites, named GAL4-TATA-LUC.

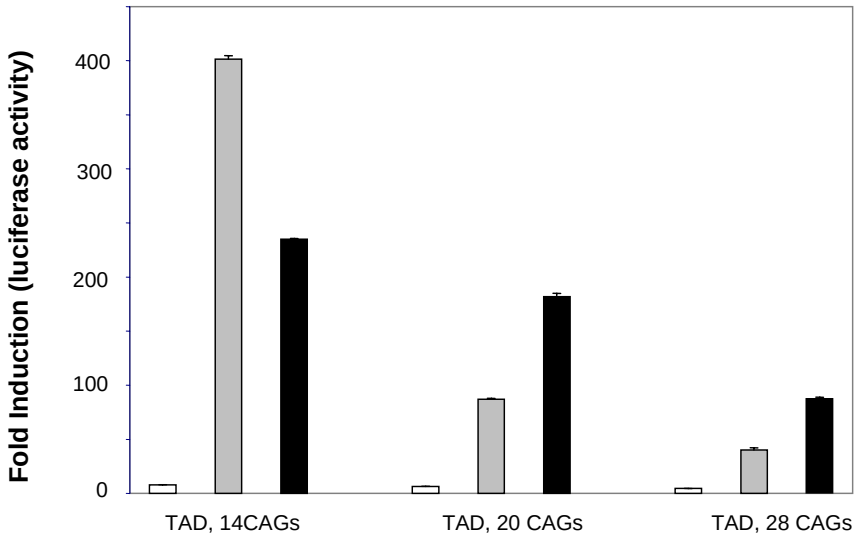
Results from three independent transfections provided evidence that the lower the number of CAG repeats in the TAD domain of the AR, the stronger was the interaction with the LBD domain after hormonal induction with different doses of MB (Fig. 9). Because of the high levels of sensitivity of the mammalian two hybrid system it was possible to detect the subtle differences in AR activity and explain the molecular mechanism by which the length of the CAG tract would affect the intrinsic activity of the AR

Fig. 9 Interaction between the NTD and LBD in the mammalian two-hybrid system.

The fusion proteins TAD-VP16 and the LBD-GAL4 were co-expressed and protein-protein interactions measured with the GAL-4-TATA-LUC reporter system. For each experiment three co-transfections were being compared, whereby the lengths of the poly-Gln tract in the TAD were 14, 20 and 28 Gln respectively. The shorter the poly-Gln tract the strongest was the interaction of the TAD-VP16 with the LBD-GAL4. The hormone used for induction was MD at doses of 0.01nM (white bars), 0.1nM (gray bars) and 0.5nM (black bars).


Fold Induction (luciferase activity)
200
160
120
80
40
0
TAD, 14 CAGs
TAD, 20 CAGs
TAD, 28 CAGs


Fold induction (luciferase activity)
300
200
100
0
TAD, 14 CAGs
TAD, 20 CAGs
TAD, 28 CAGs


Fold Induction (luciferase activity)
400
300
200
100
0
TAD, 14CAGs
TAD, 20 CAGs
TAD, 28 CAGs

DISCUSSION

Although the expression of many genes is thought to be androgen dependent, the number of human genes whose transcription is documented to be directly regulated by androgens is limited. A well-known example of a non-human promoter with ARE is that of the mouse mammary tumour virus (MMTV). This promoter is regulated by glucocorticoids, progestins, and androgens although less efficiently. Nevertheless, one of the concerns when using the MMTV promoter it that it is not endogenous to the human genome, and therefore studies using this promoter may not reflect the situation *in vivo.* In order to overcome these issues two reporter systems with human promoters containing ARE were constructed, the ALDH1 and PSA

ALDH1 is a cytosolic enzyme ubiquitously distributed in various tissues such as liver, stomach, brain and red blood cells. The ALDH1 promoter was chosen because this gene is expressed in the normal genital skin fibroblast cells but not in the cells obtained from AR negative testicular feminization patients (Yohida *et al*., 1993), suggesting that the ALDH1 gene is up-regulated by androgen action. Moreover, two putative AREs exist in the promoter sequence indicating the importance of the AR machinery in the expression of the gene. Co-tranfection experiments of the AR expression plasmid and the newly constructed ALDH-LUC reporter system indicated that the activation of the ALDH1 promoter region by androgens was very low and moreover, the base line activity of the reporter in the absence of androgen action very high. The reason for the high base line activity could be that the ALDH 1 protein acts as a modulator for cell differentiation and is expressed in many tissues; therefore, other proteins besides androgens may activate the expression of the ALDH1 gene. This made the ALDH1 promoter an unsuitable reporter for androgen action.

The other promoter selected was the PSA promoter. The PSA protein is a member of the human kallikrein family whose expression is regulated by androgens (Riegman *et al*., 1991) and it is highly expressed in the luminal epithelial cells of the prostate and at low levels in other tissues. The PSA promoter contains two ARE located at positions –170bp and -400bp respectively at which the AR binds as homodimer (Kitty *et al*., 1995). Thus, the regions selected to construct the reporter system were the –630bp and -1.6 Kb upstream of the start of the translation point. The region of -1.6 Kb was chosen because adjacent sequences to the ARE may play an important role if the activation function of the PSA promoter. As previously described by Riegman *et al*. (1991), when the PSA -1.6 Kb promoter fragment was coupled to a CAT reporter gene it induced more potently the expression of the CAT gene than the -630bp fragment.

Initially, when the PSA promoter region was inserted into the commercially available pMAM reporter system by removing the MMTV promoter region and introducing the PSA promoter region, the activation by androgens was not very high. This was due to the high base line activity of the reporter system in the absence of hormonal induction. Perhaps an enhancer element was created whereby the new promoter would drive the expression of the luciferase gene, even in the absence of androgens.

Thus, it was decided to construct a new reporter system. When the plasmid pGL3 was used as a reporter system, despite inserting the same PSA promoter sequence, the results were more satisfactory. The base line luciferase activity, in the absence of androgens, was very low. Correspondingly, the fold induction in the presence of androgens was high being about 20- to 30-fold higher than vehicle. This makes the

pGL3-PSA-LUV reporter system a viable one for testing androgen response. This reporter system was used subsequently in several publications from our laboratory (Loy *et al.*, 2003, Lim *et al*., 2000), demonstrating its utility as a measure of androgen action. The next step was to analyse if the ARs containing different CAG length would activate differently the expression of this PSA-LUC reporter system. The AR-CAG lengths selected for the transfection experiments were 14, 20 and 28 respectively. These values are included into what is considered a healthy range (above 40 CAGs the person develops a neurological disease called spinal bulbar muscular atrophy or Kennedy's disease). HeLa and CV-1 cells were selected for co-transfection experiments because they express the AR protein at very low levels.

Finding differences in the way the ARs containing different lengths would differently activate the PSA promoter prove to be a difficult task because of many factors, which more than likely led to the inconsistency of the results achieved. One of the difficult aspects of this experimental procedure was that 3 different co-transfections were being compared. The liposomal structure enclosing the DNA is not a stable chemical structure and brusque manipulation could result in a reduction in the number of liposomes. Extreme care was taken to overcome this problem. Another important aspect to consider is the tertiary structure of the plasmid DNA used for transfection. It appears that the higher the amount of supercoiled DNA structure in the plasmid, the higher is the transfection efficiency. In time the supercoiled structure is converted into a circular DNA. Taking this fact into consideration, the large-scale preparation of the three AR plasmids used for transfection was performed at the same time. Another issue was that the amount of DNA used for the three transfections that had to be

exactly the same. The amount of DNA to be transfected was calculated by measuring the absorbance of the plasmid DNA at 260nm.

All the issues mentioned above were taken into consideration. Despite that, the disparity of results obtained from the different co-transfection experiments made the task of drawing conclusion from the results a problematical one. A possible explanation as why the results from the different transfection experiments were inconsistent is that the PSA is a human promoter and therefore other transcription factors besides the AR can activate its expression. Secondly, the AR acts in conjunction with other proteins named coactivators. Various coactivators interact in a direct or indirect way with the poly-Gln region and modulate the activity of the AR such as the Ras-related nuclear protein ARA 24 (Hsiao *et al*., 1999). AR-CAG/ARA24 interaction becomes stronger as the number of glutamines decreases, increasing the coactivation capability. So a possibility as why the differences in transactivation were not encountered in cotransfections experiments using HeLa and CV-1 cells could be due to the absence of the coactivators essential for the poly-Gln interaction. In a recent study by Beiling *et al*. (2000) the researchers reported that AR constructs containing 15, 24 and 31 CAGs respectively, all in the normal range, activated differently the human probasin promoter when the cells used for cotransfection were COS-1 and LNCaP. On the contrary, when the co-transfected cells were PC3 and MCF-7, no differences were observed. The cell specific nature of this correlation suggests that the presence of coactivators, besides the AR protein is involved in the transactivation process. Thirdly, the differences in AR transactivation conferred by the CAG tract length could be promoter specific. In a recent report by Anuradha *et al*. (2002) the investigators confirmed that not statistically significant or

consistent differences were noted when AR plasmids with two different CAG repeats length (9 or 29) were co-transfected in PC-3 cells with a reporter system containing the PSA promoter. On the contrary, Irvine *et al*. (2000) performed co-trasfection studies in PC-3 cells using AR expression plasmids containing 9, 21, 29, 42 and 50 CAG repeats respectively and the MMTV promoter as reporter system. The authors reported the existence of a statistically significant poly-Gln size effect on AR transactivation activity. Therefore it is probable that the differences in the transactivation capability of the AR containing different CAG length are promoter specific.

In order to pursue our aim to understand the molecular mechanism by which the CAG length plays a role in the intrinsic activity of the AR, the interaction between the TAD with different CAG lengths, and the LBD of the AR were tested in the mammalian two-hybrid system. The physical and functional interactions between the LBD and the TAD in the AR have been reported, and it has been postulated that this interaction is essential for the dimerization of the AR (Langley *et al*., 1995). As a result of this close interaction many mutations in the LBD affect this contact, resulting in a decrease of the activity of the AR (Lim *et al*., 2000). Some investigators have postulated that this interaction is the result of a direct AR LBD-TAD contact (He *et al*., 1999), and the importance of coactivators in facilitating this interaction has been extensively reported in the literature (He *et al*., 2002). Mutations in the AR TAD have been shown to affect the recognition motifs sequences for coactivator binding, thereby severely disrupting the action of the AR (Alen *et al*., 1999). In addition, it has been demonstrated that a change in CAG length can affect the interaction of the AR with certain coactivators, such as ARA24 (Hsiao *et al*., 1999) and GRIP1 (Irvine et al., 2000), becoming

stronger as the number of CAGs decreases. With the above in mind, we investigated the role of the poly-Gln tract in the TAD-LBD interaction using the mammalian two-hybrid system. The high sensitivity of this method makes it suitable for the detection of small interactions between the proteins studied. Our results from three different experiments indicated that the shorter the poly-Gln tract in the TAD, the stronger was the binding to the LBD. This interaction was androgen-dependent, and preliminary experiments indicated that it might be hormone dose dependent. The TAD contains a ligand-independent, constitutive activation function generally referred to as AF-1 that is repressed by the unliganded LBD in the wild type AR (Jenster *et al.,* 1991). The AR LBD also contains an activation function (AF-2) but unlike AF-1, AF-2 is ligand dependent. Importantly, targeted transactivation of AR responsive genes requires de coordinated activities of both AF-1 and AF-2, although their relative contributions to this process can vary depending on the promoter and cellular context (Snoek *et al*., 1998).

The results from the experiments reported in this thesis suggest a new molecular mechanism to explain the implications of the CAG tract length in the AR activity. The interaction between the LBD and TAD could be modulated by the length of the poly-Gln tract in the TAD, being stronger as the number of CAG decreases, due to perhaps specific coactivator binding and ultimately facilitating the AR dimerization process.

CONCLUSIONS

Androgens are critical for sexual differentiation and maintenance of secondary sexual characteristics in man. All androgens act in conjunction with the AR protein which is encoded by a single copy gene in the X-chromosome. Since the cloning and structure-elucidation of the AR gene (Lubahn *et al*., 1988), numerous AR gene mutations have been described that result in complete androgen insensitivity and sexual reversal in otherwise healthy 46 XY individuals. To date more than 300 such mutations are documented in a mutational database (www.mcgill.ca/androgendb). Recently, it is increasingly apparent that AR mutations can result in a spectrum of androgen insensitivity phenotypes, ranging from the complete to the minimal forms of androgen insensitivity. The latter is the least well characterized of all the androgen insensitivity phenotypes. In this thesis, we asked what phenotypes may result from minimal changes in AR function. In this regard, we were intrigued by reports that patients with pathologically expanded AR CAG repeats in SBMA have gynecomastia, low virilization, oligospermia or azoospermia and testicular atrophy (La Spada *et al*., 1991). Furthermore preliminary studies have suggested that the length of the CAG repeats might be related to changes in AR transactivation and function (Chamberlain *et al*., 1994; Tut *et al*., 1997). Therefore, the purpose of this work was to determine the role of the CAG length in the AR gene activity and its effects in AR-regulated phenotypes in men and women. The work presented in this thesis aimed to examine the mildest form of changes in androgen action in those whose phenotypes are generally regarded as normal.

The first question was to identify a proper phenotype for study. We decided to study firstly male subjects with defective spermatogenesis and infertility, secondly the regulation of serum PSA in men, and thirdly PCOS in women. When this study was initiated only one publication had reported an association between the CAG repeat length and male infertility in a population of predominantly Chinese ethnic background (Tut *et al*., 1997). Our studies involved (Chapter 1; CAG repeat polymorphism in the androgen receptor gene and male infertility) two populations of different ethnic backgrounds; one from Singapore consisting mainly of ethnic Chinese subjects, and the other from the US composed of Caucasian subjects. The technology employed to measure the number of CAGs (GeneScan) was precise with resolution to the base pair level. To our gratification, our results showed that azoospermic patients have statistically significant longer CAG repeat tracts compared to fertile controls, and this difference was observed in both Singapore and US populations. This indicated to us that the length of the CAG tract might have crucial role in the spermatogenic process.

Since the first studies from our group (Tut *et al*., 1997), more than 20 follow-up publications have reported the relationship between the length of AR CAG tract and male infertility. Consistent with the inverse relationship between CAG repeat length and AR function, this study and those of other groups have shown that longer CAG repeats are found in the AR of patients with defective spermatogenesis. In a Singaporean population of mixed Asian origin, patients with a CAG tract of more than 28 repeats in their AR had more than a 4-fold increased risk of reduced spermatogenesis (Tut *et al.*, 1997). This trend was also observed in Australian (Dowsing *et al.*, 1999), Japanese (Yoshida *et al.*, 1999) American (Mifsud *et al.*,

2001; Patrizio *et al.*, 2001) and French (Wallerand *et al.*, 2001) populations wherein subfertile men with idiopathic azoospermia or oligozoospermia had significantly longer CAG repeat tracts than controls. Remarkably, longer AR CAG repeats are also associated with moderate to severe undermasculinization in male infants (Lim *et al.*, 2000). Together, the data support the hypothesis that longer CAG tracts have reduced intrinsic AR activity, leading to depressed spermatogenesis and male infertility. However, these findings could not be corroborated in studies involving Swedish (Giwercman *et al.*, 1998) or German populations (Dadze *et al*., 2000; von Eckardstein *et al.*, 2001), Dutch (Van Golde *et al.*, 2002) or Danish (Rajpert-De Meyts *et al.*, 2002) studies wherein no statistically significant relationship was found between the size of CAG repeats and idiopathic defective spermatogenesis. The causes of the disparity from different studies remain unclear and it has been suggested to be related to ethnicity differences (Rajpert-De Meyts *et al.*, 2002). In my opinion, one key contributory factor may be the selection criteria for patients. The patients presented in this thesis were karyotyped and genetic tests for male infertility, such as Y-chromosome microdeletions and AR point mutations, were performed. The patients who tested positive for the above mention genetic causes of male infertility were not included in our study. In addition to the genetic analysis, all patients underwent a physical examination and those with obstructive genital syndromes were excluded from the study. Using these selection criteria, our data indicate that there is an association between azoospermia and CAG length. Azoospermic patients have statistically longer CAG tracts compared to the fertile controls, and this difference was independent of the ethnic background since it was present in both populations. In addition, for the US population each unit increase in CAG length was associated with a 20% increase in the odds of being azoospermic. Interestingly long CAG repeat length

can influence the effectiveness of hormonal male contraception by rendering those with incomplete GnRH suppression azoospermic (Eckardsterin *et al*, 2002). In an Israeli infertile population, the mean length of the AR-CAG was also significantly longer than in fertile men (Madgar *et al*., 2002). Our data indicate that having more than 26 CAG repeats, the probabilities of being azoospermic increase. This number may be of value as a reference point in other populations since it was significant for both Singapore (Tut *et al*., 1997) and US patients (Mifsud et *al.*, 2000). Our studies provide the basis for an additional test in the assessment of male infertility. This is of special importance for those patients involved in assisted reproductive programs, specifically for those with an unclear cause of infertility classified as "idiopathic". Recent advances in assisted reproductive technologies have made it possible for couples to conceive, even in the absence of spermatozoa in the ejaculate through testicular biopsy and epididymal sperm aspiration (Chan *et al*., 2000). Screening and genetic counselling would be recommended for these patients. Another useful clinical observation was that most azoospermic patients with $\geq$ 26 CAGs had spermatozoa present in their seminiferous tubules. It is plausible that boosting intratesticular levels of androgens might improve spermatogenesis in these patients. This therapeutic modality would be explored in future studies.

The second aim of the project was to test the hypothesis that the length of the CAG stretch in the AR gene could play a role in the development of PCOS, a disease with a significant genetic component. A number of genes have been implicated in the development of PCOS, mainly those relating to the androgen synthetic pathway (Carey *et al*., 1994; Gharani *et al*., 1997). The present study included 91 PCOS patients and 112 healthy controls of a population of predominantly Chinese ethnic background.

Women have two copies of the AR gene and these two alleles were examined for each subject. The data showed that when the shorter of the two alleles were compared, women with PCOS had a lower number of CAG repeats compare to the controls, although this difference was not statistically significant. When the PCOS patients were divided into two groups, those with low and high levels of testosterone, PCOS patients with low testosterone levels had significant shorter CAG tracts, as well as higher frequency of the short allele compared to those patients with high levels of testosterone. Thus we have identified a subset of PCOS patients whose testosterone levels might not be increased, but have short CAG tracts at least in one of the alleles. The higher activity conferred by the short AR allele might enhance the intrinsic androgenicity of the AR leading to PCOS in this subset of patients. Women have two X chromosomes, in most cases have two different AR alleles, one of them being inactivated through methylation (Lyon *et al*., 1989). This indicates the possibility that in the subset of PCOS patients with low testosterone levels, the AR allele with the longer number of CAG repeats might be preferentially inactivated. More studies are needed to confirm the skewed pattern of X-chromosomal inactivation in PCOS women. Thus, from the results of our study we can introduce the novel concept that hyperandrogenism in the subset of PCOS patients with normal androgen levels could be conferred by short AR CAG tracts. This finding can explain why 37% of Singaporean subjects have clinical and ultrasound evidence for PCOS without raised serum androgens. Since the publication of our report (Mifsud *et al*., 2000), several studies have indicated the validity of our findings. Women with relatively few CAG repeats in the AR gene, resulting in higher transcriptional activity of the receptor, displayed higher levels of serum androgens, but lower levels of LH than women with longer CAG repeat sequences in a Swedish population (Westberg *et al*., 2001). Short

CAG lengths also increases the risk of precocious pubarche (pubic hair <8 yrs) and ovarian hyperandrogenism in a Spanish population (Ibanez *et al*., 2003). Among these menarchal girls with precocious pubarche, shorter CAG biallelic mean were associated with higher 17-hydroxy-progesterone levels post leuprolide administration, higher testosterone levels, acne, hirsutism scores, and more menstrual irregularities. Collectively, all these studies highlight the role of androgen sensitivity in the patho-physiology of PCOS. Our studies will help resolve the problem of definition that bedevils the studies in PCOS. By defining specific subsets of PCOS, entry criteria can be further refined, hopefully leading to definitive studies of this very common and distressing condition. Clearer appreciation of how different defects in the androgen signalling pathways cause PCOS, will lead to novel therapeutic solutions. The challenge for the future is to ascertain whether anti-androgen therapy may still be useful in these subjects with clinical evidence for PCOS but whose androgens are within normal limits.

Having gained the knowledge that the CAG repeat length plays a role in the activity of the AR, the third objective of the study was to determine the androgen-related factors, including the levels of T and SHBG, as well as the CAG repeat number in the AR gene, that may be involved in the regulation of the concentrations of the tumor marker PSA. Our study population was composed of two different groups of individuals, a subfertile and a fertile male population. The first observation was that the levels of T were significantly lower for the subfertile group when compared to the fertile one. Different correlations between the different androgenic parameters were investigated and the main conclusions were that in the fertile group, there was a positive correlation between the levels of T and the CAG repeat number and this correlation was not found

in the subfertile group. In subfertile subjects, T strongly correlated with PSA concentrations, and independent of T, total PSA negatively correlated with the AR CAG length. Overall our data suggest that, firstly, PSA correlates with T only in an environment of relatively low androgenicity, such as in subfertile men. Secondly, in such a low androgenic environment, short CAG tracts (associated with high AR activity) correlate positively with PSA concentrations. These results suggest that interpretation of PSA is best made in conjunction with T concentrations and the AR CAG length. These findings are relevant for the screening of prostate cancer, which to a large extent is dependent on serum PSA. One of the current problems is to distinguish between a "suspicious cancer-related PSA increase" or "cancer-related PSA increase". Cut-off points of PSA levels are controversial especially at the lower end of the spectrum (Ito *et al*., 2003). There is a debate whether increase from cero to 4.0 ng/mL is significant and whether diagnostic/therapeutic modalities like more intensive monitoring and biopsies should be undertaken in these cases. Our data suggest that for those individuals with low testosterone levels, a lower cut-off point of PSA may be appropriate as these may be due to an increased AR androgencity, and not prostate cancer. More clinical studies are required to prove the validity of this approach and whether current rates of over-detection (Draisma *et al*., 2003) of prostate cancer may be improved.

The last part of the project aimed to study *in vitro* the effect of the CAG repeat length on the AR activity. Various in-vitro co-transfection experiments were performed. Two human promoters (PSA and ALDH1) that were known to be androgen regulated were constructed. The purpose of these studies was to determine if AR genes of different CAG length would activate the luciferase reporter system with different

intensities and consequently, to find a correlation between the number of repeats and the intrinsic activity of the AR. No consistent correlation was observed while using the PSA promoter. Perhaps these differences in AR activities were very subtle and difficult to detect under the experimental conditions employed or are promoter specific. To overcome this problem, a more sensitive method was designed, this time using the mammalian two-hybrid system technology. The AR LBD-TAD interactions were studied using TAD of different CAG repeat lengths. The results indicated that the shorter the CAG tract the stronger is the interaction between the two domains. A recent study (Callewaert *et al*., 2003) also reported that a deletion of the repeat positively affected the interactions of the ligand-binding domain with the amino-terminal domain. The authors concluded that this was dependent on the recruitment of the p160 coactivator SRC-1e to the amino-terminal domain of the AR as reflected by an enhanced coactivation of the AR by SRC-1e. More studies are needed to validate this mode of action, and to further comprehend the underlying molecular mechanism(s).

Glutamine-rich regions of proteins occur widely in both prokaryotes and eukaryotes. They are frequently found as polyglutamine tracts, encoded by CAG repeats, often of considerable length. Despite this wide distribution, the functions of CAG repeats in healthy subjects are often unclear. The work presented in this thesis aimed to further clarify the role of the AR-CAG repeat length in diseases/conditions where the CAG repeat length still falls into the healthy range. We concluded that the CAG tracts were longer in those subfertiles/infertile patients with azoospermia in two ethnically different populations such as Caucasian and Chinese, and shorter in one of the alleles in a subset of female PCOS patients whose testosterone levels were not raised.

Furthermore, PSA levels were negatively correlated with AR CAG length in those individuals with low testosterone levels. These three clinical studies associated with the in-vitro studies suggest an effect of the CAG length on the AR TAD-LBD interactions and validate the model that shorter CAG tracts confer higher intrinsic activity to the AR and leading to diseases characterized by high androgenic levels such as PCOS. In contrast, longer CAG tracts will reduce AR activity predisposing the individuals to diseases were the androgen action is impaired such as defective spermatogenesis.

Since the publication of our studies in 2000, there has been increasing evidence from numerous studies in a variety of patient populations and diverse disciplines, that AR CAG tracts have a subtle but critical role in diseases and processes regulated by androgens. There has been growing interests in the topic by many authors and a number of publications have followed our studies. A summary of these publications is presented in Table 1. Several reports have appeared indicating that CAG tracts regulate the AR function in ageing men and women. Harkonen *et al*. reported that men with long CAG repeats show significantly more depressed mood, anxiety, wish to be dead and decreased beard growth (Harkonen *et al*., 2003). According to the studies by Zitzmann *et al*., high number of repeats are related to the increased age-dependent bone loss in men (Zitzmann *et al*., 2001) and increased risk of osteoporosis in women (Chen *et al*., 2003). The prostate volume and growth in testosterone-substituted hypogonadal men was also found dependent on the CAG repeat polymorphism of the androgen receptor gene (Zitzmann *et al*., 2003). On the other hand, short alleles appear to be associated with early onset of Alzheimer's disease in men (Lehmann *et al*., 2003). Short alleles are also independently correlated with cardiovascular risk by

increasing low body fat mass and plasma insulin while at the same time, decreasing the level of high density lipoprotein concentrations in older men (Zitzmann *et al*., 2003) and obesity indices in older women (Gustafson *et al*., 2003).

Epidemiological studies are emerging that suggest associations between CAG repeats and hormone dependent malignancies. Sasaki et al. reported that endometrial cancer patients have significantly longer AR (Sasaki *et al*., 2003). Short CAG repeats were associated with an increased prostate cancer risk in Hispanic men (Balic *et al*., 2003). Long AR alleles have been also associated with increased risk of early onset of breast cancer (Suter *et al*., 2003), increased breast cancer risk among women with a first-degree of family history of breast cancer (Haiman *et al*., 2002) and ovarian cancer (Santarosa *et al*., 2002). Although not uniformly unanimous, these studies cumulatively provide strong evidence that the AR CAG length within the normal polymorphic range can indeed regulate AR function in cases of health or disease.

Table 1. Role of the AR-CAG length in diseases and processes regulated by androgens

DISEASE	OUTCOME	ETHNICITY	REFERENCE
Hormone Related Cancers			
Ovarian Cancer	Women with ≥22 greater risk than women < 22 CAGs (OR=2.17, 95% CI 1.10-3.20)	Caucasian	Santarosa et al., 2002
Prostate volume	Prostate growth under substitute testosterone levels were dependent on (CAG)n with lower effect on longer CAGs	Cauasian	Zitzmann et al., 2003
Prostate cancer	Men with ≤ 18 CAGs have 3-fold increased risk of prostate cancer than men with <18 CAGs	Caucasian (Hispanic)	Balic et al., 2003
Endometrial cancer	Endometrial cancer patients had longer alleles than controls	Japanese	Sasaki et al., 2003)
Breast cancer	Women with one or two alleles with ≥ 22 CAGs and first degree family history of breast cancer have greater risk of breast cancer (OR=1.70, 95%CI 1.20-2.40)	Caucasian	Haiman et al., 2002
Breast cancer	Women with alleles cumulative size of ≥ 43 CAGs have increased risk (OR=1.3, 95% CI 1.0-1.7)	Caucasian	Suter et al., 2003

DISEASE	OUTCOME	ETHNICITY	REFERENCE
Ageing			
Age dependent bone density in healthy males	Men with CAGs of 22-31 have increased age dependent bone loss compared to men with CAGs of 14-21	Caucasian	Zitzmann et al., 2001
Osteosporosis in postmenopausal women	Women with > 20 CAGs greater risk than women <20 (OR 4.2, 95% CI 1.0-17.2)	Chinesse	Chen et al., 2003
Central obesity in women	Women with $\leq$ 25 CAGs had higher waist and hip circumferences	Caucasian	Gustafson et al., 2003
Andropausal symptoms in ageing men	CAG repeat number correlates with depression, depressed mood, anxiety, decreased beard growth.	Caucasian	Harkonen et al., 2003
Cardiovascular risk	Positive correlation of CAG repeat number with serum levels of leptin, insulin and body fat mass	Caucasian	Zitzmann et al., 2003
Others			
Ovarian hyperandrogenism in girls.	Biallelec mean of $\leq$20 CAGs was associated with ovarian hyperandrogenism	Caucasian	Ibanez et al., 2003
Alzheimer's disease (AD) in men	$\leq$ 20 CAGs were associated with AD OR=2.5, 95% CI:1.2-5.0 in men	Caucasian	Lehmann et al., 2003

The occurrence of the more than 32 AR alleles (CAG length from 8 to 39) in the normal healthy population implies corresponding differences in androgen sensitivity among healthy subjects. What are the advantages to the species of such variation in AR sensitivity? Evolutionary studies of a number of genes involved in the human triplet expansion diseases have indicated that the repeats in these genes have arisen by gradual expansion of the tandem repeat, apparently resulting from replication slippage. Studies in *Escherichia coli* (Rosenberg *et al.,* 1994) have shown that nutritional factors or other type of stress might increase the rate of mutation by down-regulating the mismatch DNA repair system. Thus environmental challenges would increase the rate of variability in the genome, enabling species to adapt to new challenges. Repeats are generally absent in rodent homologue genes and comparative studies indicate an increase in repeat length during primate evolution, with humans generally having the longest repeats (Rubinsztein *et al.,* 1994). Repeat number variation of these polymorphic regions would also provide a ready and inexhaustible supply of new quantitative genetic variation for rapid evolutionary adaptation of species to novel ecological challenges, or rapid divergence of an isolated population. This rate of mutational variation can enable a population faced with continuing environmental challenges to adapt smoothly and rapidly, thereby contributing to survival of the species.

REFERENCES

Achard C, Thiers J. 1921. Le virilisme pilaire et son association a l'insuffisance glycolytique (diabete des femme a barbe.) Bull Acad Natl Med 86:51-83

Adams J, Polson DW, Franks S. 1986. Prevalence of polycystic ovaries in women with anovulation and idiopathic hirsutism. Brit Med J. 293: 355-359.

Aiman J, Griffin JE. 1982. The frequency of androgen receptor deficiency in infertile men. J Clin Endocrinol Metab. 54:725-32.

Aitken RJ, Gordon E, Harkiss D, et al. 1998. Relative impact of oxidative stress on the functional competence and genomic integrity of human spermatozoa. Biol Reprod. 59:1037–1046.

Allen SE, Potter HD, Azziz R. 1997. Prevalence of hyperandrogenemia among nonhirsute oligo-ovulatory women. Fertil Steril. 67:569–572

Amato AA, Prior TW, Barohn RJ, Snyder P, Papp A, Mendell JR. 1993. Kennedy's disease: a clinicopathologic correlation with mutations in the androgen receptor gene. Neurology. 43:791-794.

Andrews PE, Young CY, Montgomery BT, Tindall DJ. 1992. Tumor-promoting phorbol ester down-regulates the androgen induction of prostate-specific antigen in a human prostatic adenocarcinoma cell line Cancer Res. 52:1525-9

Arbizu T, Santamaria J, Gomex JM, Quilez A, Serra JP. 1983. A family with adult spinal and bulbar muscular atrophy, X-linked inheritance and associated testicular failure. J Neurol Sci. 59:371-382.

Aus G, Bergdahl S, Hugosson J, Lodding P, Norlen L. 1994. Related Influence of benign prostatic hyperplasia, testosterone and age on serum levels of prostate specific antigen. Scand J Urol Nephrol. 28: 379-84.

Azziz R, Kashar-Miller MD. 2000. Family history as a risk factor for the polycystic ovary syndrome J Pediatr Endocrinol Metab. 13:1303-6

Balen AH, Conway GS, Kaltsas G, Techatrasak K, Manning PJ, West C, Jacobs HS. 1995. Polycystic ovary syndrome: the spectrum of the disorder in 1741 patients Hum Reprod. 10:2107-11.

Basso D, Fogar P, Piva MG, Navaglia F, Mazza S, et al. 2000. Total PSA, free PSA/total PSA ratio, and molecular PSA detection in prostate cancer: which is clinically effective and when? Urology. 555:710-5

Balic I, Graham ST, Troyer DA, Higgins BA, Pollock BH, et al. 2003. Androgen receptor length polymorphism associated with prostate cancer risk in Hispanic men. J Urol. 168:2245-8

Behre HM, Bohmeyer J, Nieschlag E. 1994. Prostate volume in testosterone-treated and untreated hypogonadal men in comparison to age-matched normal controls Clin Endocrinol. 40(3):341-9

Beiling J, Ball EM, Favaloro JM, Zajac JD. 2000. Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. J Mol Endocrinol. 251:85-96

Berg W, Linder C, Eschholz G, Link S, Schubert J. 1999. Possibility of improving the acceptance rateof early detection testing for prostate cancerwith a one-step test for prostate-specific antigen in whole blood Urol Int. 63:102-6

Bergh C, Olsson JH, Selleskog U, Hillensjo T. 1993. Steroid production in cultured theca cells obtained from human ovarian follicles. Hum Reprod. 8:519-24.

Boring CC, Squires TS, Tong T, Montgomery S. 1994. Cancer statistics, 1994 CA Cancer J Clin. 44:7-26.

Brooks BP, Paulson HL, Merry DE, Salazar-Grueso EF, Brinkmann AO, Wilson EM, Brownell JE, Allis CD. 1995. An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. Proc Natl Acad Sci U S A. 3;92:6364-8

Brooks BP, Paulson HL, Merry DE, Salazar-Grueso EF, Brinkmann AO, Wilson EM, Fischbeck KH. 1997. Characterization of an expanded glutamine repeat androgen receptor in a neuronal cell culture system. Neurobiol Dis. 3:313-23

Callewaert L, Christiaens V, Haelens A, Verrijdt G, Verhoeven G, Claessens F. 2003. Implications of a polyglutamine tract in the function of the human androgen receptor. Biochem Biophys Res Commun.20;306:46-52

Carey AH, Chan KL, Short F, White D, Williamson R, Franks S. 1993. Evidence for a single gene effect causing polycystic ovaries and male pattern baldness. Clin Endocrinol. 38:653-8.

Carey AH, Waterworth D, Patel K, White D, Little J, Novelli P, Franks S Williamson R. 1994. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet. 3:1873-6.

Carlstrom K, Eriksson A, Stege R, Rannevik G. 1990. Relationship between serum testosterone and sex hormone-binding globulin in adult men with intact or absent gonadal function. Int J Androl. 13: 67-73

Cato AC, Stock G, Fuhrmann U. 1998. Steroid hormone action 1998: novel challenges for drug design. J Mol Med. 1998 76:459-60.

Casella R, Maduro MR, Misfud A, Lipshultz LI, Yong EL, Lamb DJ. 2003. Androgen receptor gene polyglutamine length is associated with testicular histology in infertile patients J Urol. 2003 169:224-7

Chamberlain NL, Driver ED, Miesfeld RL. 1994. The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res. 22:3181-3186.

Chen HY, Chen WC, Wu MC, Tsai FJ, Tsai CH. 2003. Androgen receptor (AR) gene microsatellite polymorphism in postmenopausal women: correlation to bone mineral density and susceptibility to osteoporosis. Eur J Obstet Gynecol Reprod Biol. 26;107:52-6

Chia KS, Lee HP, Seow A, Shanmugaratnam K. 1996. Trends in cancer incidence in Singapore 1968-1992. Singapore Cancer Registry 106.

Choong CS, Kemppainen JA, Wilson EM. 1998. Evolution of the primate androgen receptor: a structural basis for disease. J Mol Evol. 47:334-42.

Choong CS, Kemppainen JA, Zhou ZX, Wilson EM. 1996. Reduced androgen receptor gene expression with first exon CAG repeat expansion. Mol Endocrinol. 10:1527-35.

Cleutjens CB, Steketee K, van Eekelen CC, van der Korput JA, Brinkmann AO, Trapman J. 1997. Both androgen receptor and glucocorticoid receptor are able to induce prostate-specific antigen expression, but differ in their growth-stimulating properties of LNCaP cells. Endocrinology.138:5293-300

Conway GS, Honour JW, Jacobs HS**.** 1989. Heterogeneity of the polycystic ovary syndrome: clinical, endocrine and ultrasound features in 556 patients. Clin Endocrinol. 30:459–470

Conway GS, Avey C, Rumsby G. 1994. The tyrosine kinase domain of the insulin receptor gene is normal in women with hyperinsulinaemia and polycystic ovary syndrome. Hum Reprod. 9:1681-3.

Cooper CS, Perry PJ, Sparks AE, MacIndoe JH, Yates WR, Williams RD. 1998. Effect of exogenous testosterone on prostate volume, serum and semen prostate specific antigen levels in healthy young men. J Urol. 159: 441-3.

Cunha GR, Donjacour AA, Cooke PS, Mee S, Bigsby RM, Higgins SJ, Sugimura Y. 1987. The endocrinology and developmental biology of the prostate. Endocr Rev. 8: 338-62

Cunningham SK, Loughlin T, Culliton M, McKenna TJ. 1985. The relationship between sex steroids and sex-hormone-binding globulin in plasma in physiological and pathological conditions. Ann Clin Biochem. 22: 489-97.

Dadze S, Wieland C, Jakubiczka S, Funke K, Schroder E, Royer-Pokora B, Willers R,Wieacker PF. 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 DM, Baker HW. 1999. Infertility in men: recent advances and continuing controversies. J Clin Endocrinol Metab. 84: 3443-3450.

De Kretser DM. 1997. Male infertility. Lancet.15;349: 787-90.

Diamandis EP, Yu H. 1997. Nonprostatic sources of prostate-specific antigen. Urol Clin North Am. 24: 275-82.

Diamandis EP, Yu H. 1995. Prostate-specific antigen and lack of specificity for prostate cells. Lancet. 6: 345:1186.

Douglas TH, Connelly RR, McLeod DG, Erickson SJ, Barren R, Murphy GP. 1995. Effect of exogenous testosterone replacement on prostate-specific antigen and prostate-specific membrane antigen levels in hypogonadal men. J Surg Oncol. 59:246-50.

Dowsing AT, Yong EL, McLachlan RI, de Kretser D, Trounson A. 1999. Linkage between male infertility and trinucleotide expansion in the androgen receptor gene. Lancet. 354: 640-643.

Draisma G, Boer R, Otto SJ, van der Cruijsen IW, Damhuis RA, Schroder FH, de Koning HJ. 2003. Lead times and overdetection due to prostate-specific antigen screening: estimates from the European Randomized Study of Screening for Prostate Cancer. J Natl Cancer Inst. 18:868-78

Dunaif A, Finegood DT. 1996. Beta-cell dysfunction independent of obesity and glucose intolerance in the polycystic ovary syndrome J Clin Endocrinol Metab. 81: 942-7.

Dunaif A. 1993. Insulin resistance in polycystic ovarian syndrome. Ann N Y Acad Sci. 28;687: 60-4.

Eckardstein SV, Schmidt A, Kamischke A, Simoni M, Gromoll J, Nieschlag E. 2002. CAG repeat length in the androgen receptor gene and gonadotrophin suppression influence the effectiveness of hormonal male contraception. Clin Endocrinol. 57:647-55

Edmunds SE, Stubbs AP, Santos AA, Wilkinson ML. 1990. Estrogen and androgen regulation of sex hormone binding globulin secretion by a human liver cell line. J Steroid Biochem Mol Biol. 37: 733-9.

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics. 12: 241-253.

Ehrmann DA, Sturis J, Byrne MM, Karrison T, Rosenfield RL, Polonsky KS. 1995. Insulin secretory defects in polycystic ovary syndrome. Relationship to insulin sensitivity and family history of non-insulin-dependent diabetes mellitus. J Clin Invest. 96:520-7.

Escobar-Morreale HF, Serrano-Gotarredona J, Avila S, Villar-Palasi J, Varela C, Sancho J. 1998. The increased circulating prostate-specific antigen concentrations in women with hirsutism do not respond to acute changes in adrenal or ovarian function. J Clin Endocrinol Metab. 83: 2580-4.

Famiglietti G. 2002. Varicocele and cryptorchism: diagnosis and follow up Arch Ital Urol Androl. 74:157-65

Franks S, Gharani N, Waterworth D, Batty S, White D, Williamson R, McCarthy M. 1998. Genetics of polycystic ovary syndrome. Mol Cell Endocrinol. 25;145:123-8.

Franks S, Gharani N, Waterworth D, Batty S, White D, Williamson R, McCarthy M. 1997. The genetic basis of polycystic ovary syndrome. Hum Reprod. 12:2641-8.

Franks S, White D, Gilling-Smith C, Carey A, Waterworth D, Williamson R. 1996. Hypersecretion of androgens by polycystic ovaries: the role of genetic factors in the regulation of cytochrome P450c17 alpha. Baillieres Clin Endocrinol Metab. 10:193-203.

Franks S. 1995. Polycystic ovary syndrome. N Engl J Med. 333:853-6

Fulghesu AM, Cucinelli F, Pavone V, Murgia F, Guido M, Caruso A, Mancuso, Lanzone A. 1999. Changes in luteinizing hormone and insulin secretion in polycystic ovarian syndrome. Hum Reprod. 14:611-7.

Gann PH, Hennekens CH, Ma J, Longcope C, Stampfer MJ. 1996. Prospective study of sex hormone levels and risk of prostate cancer. J Natl Cancer Inst. 21; 88:1118-26.

Ghadessy FJ, Lim J, Abdullah AAR, Panet-Raymond V, Choo CK, Lumbroso R, et al. 1999. Oligospermic infertility associated with an androgen receptor mutation that disrupts interdomain and coactivator (TIF2) interactions. J Clin Invest. 103:1517-1525.

Gharani N, Waterworth DM, Batty S, White D, Gilling-Smith C, Conway GS, McCarthy M, Franks S, Williamson R. 1997. Association of the steroid synthesis gene CYP11a with polycystic ovary syndrome and hyperandrogenism. Hum Mol Genet. 6:397-402.

Gharani N, Waterworth DM, Williamson R, Franks S. 1996. 5' polymorphism of the CYP17 gene is not associated with serum testosterone levels in women with polycystic ovaries. J Clin Endocrinol Metab. 81:4174.

Gilling-Smith C, Willis DS, Beard RW, Franks S. 1994. Hypersecretion of androstenedione by isolated thecal cells from polycystic ovaries. J Clin Endocrinol Metab. 79:1158-65

Giovannucci E, Stampfer MJ, Krithivas K, et al. 1997. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci, USA. 94: 3320-3323.

Giwercman YL, 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. Clin Genet. 54: 435-6

Goh VH. 1999. Breast tissues in transsexual women--a nonprostatic source of androgen up-regulated production of prostate-specific antigen. J Clin Endocrinol Metab. 84: 3313-5.

Grulet H, Hecart AC, Delemer B, Gross A, Sulmont V, Leutenegger M, Caron J. 1993. Roles of LH and insulin resistance in lean and obese polycystic ovary syndrome. Clin Endocrinol. 38: 621-6.

Guo Q, Kumar TR, Woodruff T, Hadsell LA, DeMayo FJ, Matzuk MM. 1998. Overexpression of mouse follistatin causes reproductive defects in transgenic mice. Mol Endocrinol. 12:96-106.

Gustafson DR, Wen MJ, Koppanati BM. 2003. Androgen receptor gene repeats and indices of obesity in older adults. Int J Obes Relat Metab Disord. 27:75-81

Haiman CA, Brown M, Hankinson SE, Spiegelman D, Colditz GA, et al. 2002. The androgen receptor CAG repeat polymorphism and risk of breast cancer in the Nurses' Health Study. Cancer Res. 15:1045-9

Hamilton-Fairley D, Kiddy D, Watson H, Sagle M, Franks S. 1991. Low-dose gonadotrophin therapy for induction of ovulation in 100 women with polycystic ovary syndrome. Hum Reprod. 6:1095-9.

Hardy DO, Scher HI, Bogenreider T, Sabbatini P, Zhang ZF, Nanus DM, Catterall JF. 1996. Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. J Clin Endocrinol Metab. 81: 4400-4405.

Harkonen K, Huhtaniemi I, Makinen J, Hubler D, Irjala K, et al. 2003. The polymorphic androgen receptor gene CAG repeat, pituitary-testicular function and andropausal symptoms in ageing men. Int J Androl. 26:187-94.

He B, Kemppainen JA, Voegel JJ, Gronemeyer H, Wilson EM. 1999. Activation function 2 in the human androgen receptor ligand binding domain mediates interdomain communication with the NH(2)-terminal domain. J Biol Chem. 274:37219-25

He B, Minges JT, Lee LW, Wilson EM. 2002. The FXXLF motif mediates androgen receptor-specific interactions with coregulators. J Biol Chem. 22:10226-35.

Hiort, O., Holterhus, P.M., Horter, T., Schulze, W., Kremke, B., Bals-Pratsch, M., Sinnecker, G.H. and Kruse, K. 2000. Significance of mutations in the androgen receptor gene in males with idiopathic infertility. J. Clin. Endocrinol. Metab. 85: 2810-2815.

Holte J, Bergh T, Berne C, Berglund L, Lithell H. 1994. Enhanced early insulin response to glucose in relation to insulin resistance in women with polycystic ovary syndrome and normal glucose tolerance. J Clin Endocrinol Metab. 78: 1052-8.

Holte J. 1996. Disturbances in insulin secretion and sensitivity in women with the polycystic ovary syndrome. Baillieres Clin Endocrinol Metab. 10: 221-47.

Hsiao PW, Chang C. 1999. Isolation and characterization of ARA160 as the first androgen receptor N-terminal-associated coactivator in human prostate cells. J Biol Chem. 274:22373-9

Hsiao PW, Lin DL, Nakao R, Chang C. 1999. The linkage of Kennedy's neuron disease to ARA24, the first identified androgen receptor polyglutamine region-associated coactivator. J Biol Chem. 274: 20229-20234.

Hsing AW, Gao YT, Wu G, Wang X, Deng J, Chen YL, Sesterhenn IA, Mostofi FK, Benichou J, Chang C. 2000 Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res. 15;60: 5111-6.

Ibanez L, Ong KK, Mongan N, Jaaskelainen J, Marcos MV, et al. 2003. Androgen receptor gene CAG repeat polymorphism in the development of ovarian hyperandrogenism. J Clin Endocrinol Metab. 88:3333-8.

Irvine RA, Yu MC, Ross RK, Coetzee GA. 1995. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res. 54: 2861-2864.

Irvine, RA., Ma, H., Yu, MC., Ross, R.K., Stallcup, MR. and Coetzee, GA. 2000. Inhibition of p160-mediated coactivation with increasing androgen receptor polyglutamine length. Hum. Mol. Genet. 9: 267-274.

Ito K, Yamamoto T, Ohi M, Kubota Y, Fukabori Y, et al. 2002. Usefulness of prostate-specific antigen velocity in screening for prostate cancer. Int J Urol. 9:316-21.

Jahanfar S, Eden JA, Warren P, Seppala M, Nguyen TV. 1995. A twin study of polycystic ovary syndrome. Fertil Steril. 1995 63:478-86

Jenster G, van der Korput HA, van Vroonhoven C, van der Kwast TH, Trapman J, Brinkmann AO. 1991. Domains of the human androgen receptor involved in steroid binding, transcriptional activation, and subcellular localization. Mol Endocrinol. 5:1396-404.

Jequier, AM., Ansell, ID, Bullimore, NJ. 1985. Congenital absence of the vasa deferentia presenting with infertility. J. Androl. 6: 15–19

Kashar-Miller M, Azziz R. 1999. Heritability and the risk of developing androgen excess. J Steroid Biochem Mol Biol. 69: 261-8.

Kenkel S, Rolf C, Nieschlag E. 2001. Occupational risks for male fertility: an analysis of patients attending a tertiary referral center Int J Androl. 24:318-26

Kim MR, Gupta MK, Travers SH, Rogers DG, Van Lente F, Faiman C. 1999. Serum prostate specific antigen, sex hormone binding globulin and free androgen index as markers of pubertal development in boys. Clin Endocrinol. 50: 203-10.

Koh JS, Cheng CW, Foo KT. 2001. Spectrum of prostate cancer in the Singapore General Hospital (1980 to 1985). Ann Acad Med Singapore. 30:513-5.

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. Mol Hum Reprod. 5:14-6.

Kukuvitis A, Georgiou I, Bouba I, Tsirka A. 2002. Association of oestrogen receptor alpha polymorphisms and androgen receptor CAG trinucleotide repeats with male infertility: a study in 109 Greek infertile men. Int J Androl. 25:149-52.

Kupfer SR, Marschke KB, Wilson EM, French FS. 1993. Receptor accessory factor enhances specific DNA binding of androgen and glucocorticoid receptors. J Biol Chem. 15;268:17519-27.

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature. 352: 77-9.

Langley E, Zhou ZX, Wilson EM. 1995. Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. J Biol Chem. 15:29983-90.

Lanzone A, Caruso A, Di Simone N, De Carolis S, Fulghesu AM, Mancuso S. 1995. Polycystic ovary disease. A risk factor for gestational diabetes? J Reprod Med. 40: 312-6.

Legius E, Vanderschueren D, Spiessens C, D'Hooghe T, Matthijs G. 1999. Association between CAG repeat number in the androgen receptor and male infertility in a Belgian study Clin Genet. 56:166-7

Lehmann DJ, Butler HT, Warden DR, Combrinck M, King E, et al. 2003. Association of the androgen receptor CAG repeat polymorphism with Alzheimer's disease in men. Neurosci Lett. 10:87-90

Legro RS, Driscoll D, Strauss JF 3rd, Fox J, Dunaif A. 1998. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. Proc Natl Acad Sci U S A. 95: 14956-60.

Li H, Gomes PJ, Chen JD. 1997. Nucleotide RAC3, a steroid/nuclear receptor-associated coactivator that is related to SRC-1 and TIF2. Proc Natl Acad Sci U S A. 5;94: 8479-84.

Lieberman AP and Fischbeck KH. 2000. Triplet repeat expansion in neuromuscular disease. Muscle Nerve 23: 843-850.

Lim J, Ghadessy FJ, Yong EL. 1997. A novel splice site mutation in the androgen receptor gene results in exon skipping and a non-functional truncated protein. Mol Cell Endocrinol. 8;131: 205-10.

Lim J, Ghadessy FJ, Abdullah AA, Pinsky L, Trifiro M, Yong EL. 2000. Human androgen receptor mutation disrupts ternary interactions between ligand, receptor domains, and the coactivator TIF2 (transcription intermediary factor 2). Mol Endocrinol. 14:1187-97.

Lim, HN., Chen, H., McBride, S., Dunning, AM, Nixon, RM, Hughes, IA, Hawkins JR. 2000. Longer polyglutamine tracts in the androgen receptor are associated with moderate to severe undermasculinized genitalia in XY males. Hum. Mol. Genet. 9: 829-834.

Lin, HK, Wang, L, Hu, Y.C., Altuwaijri, S, Chang C. 2002. Phosphorylation-dependent ubiquitylation and degradation of androgen receptor by Akt require Mdm2 E3 ligase. EMBO. 21: 4037-4048.

Liow SL, Ghadessy FJ, Ng SC, Yong EL. 1998. Y chromosome microdeletions, in azoospermic or near-azoospermic subjects, are located in the AZFc (DAZ) subregion. Mol Hum Reprod. 4: 763-8.

Loy CJ, Sim KS, Yong EL. 2003. Filamin-A fragment localizes to the nucleus to regulate androgen receptor and coactivator functions. Proc Natl Acad Sci U S A. 15;100:4562-7.

Lubahn DB, Brown TR, Simental JA, Higgs HN, Migeon CJ, Wilson EM, French FS. 1981. Sequence of the intron/exon junctions of the coding region of the human androgen receptor gene and identification of a point mutation in a family with complete androgen insensitivity. Proc Natl Acad Sci U SA. 86:9534-9538.

Lubahn DB, Joseph DR, Sullivan PM, Willard HF, French FS, Wilson EM. 1988. Cloning of human androgen receptor complementary DNA and localization to the X chromosome. Science. 15;240(4850): 327-30.

Lunde O, Magnus P, Sandvik L, Hoglo S. 1989. Familial clustering in the polycystic ovarian syndrome. Gynecol Obstet Invest. 28: 23-30.

Lyon MF. 1999. Imprinting and X-chromosome inactivation. **:** Results Probl Cell Differ. 25:73-90

MacIndoe JH, Cooper CS, Perry PJ, et al.1995. The effect of exogenous testosterone on prostate specific antigen levels in healthy men. J. Urolo. 153:397

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

Mais V, Kazer RR, Cetel NS, Rivier J, Vale W, Yen SS. 1986. The dependency of folliculogenesis and corpus luteum function on pulsatile gonadotropin secretion in cycling women using a gonadotropin-releasing hormone antagonist as a probe. J Clin Endocrinol Metab. 62: 1250-5.

Marshall JC, Eagleson CA. 1999. Neuroendocrine aspects of polycystic ovary syndrome. Endocrinol Metab Clin North Am. 28: 295-324.

Matias PM, Donner P, Coelho R, et al. 2000. Structural evidence for ligand specificity in the binding domain of the human androgen receptor: Implications for pathogenic gene mutations. J Biol Chem. 275: 26164-26171

McCormack RT, Rittenhouse HG, Finlay JA, Sokoloff RL, Wang TJ, Wolfert RL, Lilja H, Oesterling JE. 1995. Molecular forms of prostate-specific antigen and the human kallikrein gene family: a new era. Urology. 45: 729-44.

McKenna NJ, Xu J, Nawaz Z, Tsai SY, Tsai MJ, O'Malley BW. 1999. Nuclear receptor coactivators: multiple enzymes, multiple complexes, multiple functions. J Steroid Biochem Mol Biol. 69: 3-12

Mengual L, Oriola J, Ascaso C, Ballesca JL, Oliva R. 2003. An increased CAG repeat length in the androgen receptor gene in azoospermic ICSI candidates. J Androl. 24:279-84.

Mhatre AN, Trifiro MA, Kaufman M, et al. 1993. Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nat Genet. 5:184-188.

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

Mifsud A, Sim CK, Boettger-Tong H, Moreira S, Lamb DJ, Lipshultz LI, Yong EL. 2001. Trinucleotide (CAG) repeat polymorphism in the androgen receptor gene: Molecular markers of male infertility risk. Fertil Steril. 75:275-81

Moller DE, Flier JS 1988. Detection of an alteration in the insulin-receptor gene in a patient with insulin resistance, acanthosis nigricans, and the polycystic ovary syndrome (type A insulin resistance). N Engl J Med. 8;319:1526-9

Monath JR, McCullough DL, Hart LJ, Jarow JP. 1995. Physiologic variations of serum testosterone within the normal range do not affect serum prostate-specific antigen. Urology. 46:58-61.

Neuschmid-Kaspar F, Gast A, Peterziel H, Schneikert J, Muigg A, Ransmayr G, Klocker H, Bartsch G, Cato AC. 1996. CAG-repeat expansion in androgen receptor in Kennedy's disease is not a loss of function mutation. Mol Cell Endocrinol. 25:149-56.

O'Donnell WT and Warren ST. 2002. A decade of molecular studies of fragile X syndrome. Annu. Rev. Neurosci. 25, 315–338

Pan H, Li YY, Li TC, Tsai WT, Li SY, Hsiao KM. 2002. Increased (CTG/CAG)(n) lengths in myotonic dystrophy type 1 and Machado-Joseph disease genes in idiopathic azoospermia patients Hum Reprod. 17:1578-83

Parker ME. 1997. PSA thresholds for prostate cancer detection. JAMA. 278:699-700

Patrizio P, Leonard DG. 2001. Expansion of the CAG trinucleotide repeats in the androgen receptor gene and male infertility: a controversial association. J Androl. 22: 444-448.

Perutz MF, Johnson T, Suzuki M, Finch JT. 1994. Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. Proc Natl Acad Sci U S A. 7;91: 5355-8

Plymate SR, Leonard JM, Paulsen CA, Fariss BL, Karpas AE. 1983. Sex hormone-binding globulin changes with androgen replacement. J Clin Endocrinol Metab. 57: 645-8.

Rajpert-De Meyts, E, Leffers, H, Petersen, JH, Andersen, AG, Carlsen, E, Jorgensen, N, Skakkebaek, NE. 2002. CAG repeat length in androgen-receptor gene and reproductive variables in fertile and infertile men. Lancet. 359: 44-46.

Rebbeck TR, Kantoff PW, Krithivas K, et al. 1999. Modification of BRCA1-Associated Breast Cancer Risk by the Polymorphic Androgen Receptor CAG Repeat. Am J Hum Genet. 64:1371-1377.

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

Riegman PH, Vlietstra RJ, van der Korput JA, Brinkmann AO, Trapman J. 1991. The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. Mol Endocrinol. 5:1921-30.

Rodin DA, Bano G, Bland JM, Taylor K, Nussey SS. 1998. Polycystic ovaries and associated metabolic abnormalities in Indian subcontinent Asian women. Clin Endocrinol. 49:91–99

Rosenberg SM, Longerich S, Gee P, Harris RS. 1994. Adaptive mutation by deletions in small mononucleotide repeats. Science.15;265:405-7.

Rosner W, Hryb DJ, Khan MS, Nakhla AM, Romas NA. 1999. Androgen and estrogen signalling at the cell membrane via G-proteins and cyclic adenosine monophosphate Steroids 64:100-6

Rubinsztein DC, Leggo J, Amos W, Barton DE, Ferguson-Smith MA. 1994. Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphism in human and primate populations. Hum Mol Genet. 3:2031-5.

Ruckle HC, Oesterling JE. 1993. Prostate-specific antigen and androgen deprivation therapy. World J Urol. 11:227-32.

Sack JS, Kish KF, Wang C, Attar RM, Kiefer SE, An Y, Wu GY. 2001. Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. Proc Natl Acad Sci U S A. 24;98:4904-9.

Santarosa M, Bidoli E, Gallo A, Steffan A, Boiocchi M, Viel A. 2002. Polymorphic CAG repeat length within the androgen receptor gene: identification of a subgroup of patients with increased risk of ovarian cancer. Oncol Rep. 9:639-44.

Sar M, Lubahn DB, French FS, Wilson EM. 1990. Immunohistochemical localization of the androgen receptor in rat and human tissues. Endocrinology. 127:3180-6.

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. J Androl 22:804-8

Sasaki M, Sakuragi N, Dahiya R. 2003. The CAG repeats in exon 1 of the androgen receptor gene are significantly longer in endometrial cancer patients. Biochem Biophys Res Commun. 305:1105-8.

Sawaya ME, Shalita AR. 1998. Androgen receptor polymorphisms (CAG repeat lengths) in androgenetic alopecia, hirsutism, and acne. J Cutan Med Surg. 3:9-15

Schellhammer PF, Wright GL Jr. 1993. Biomolecular and clinical characteristics of PSA and other candidate prostate tumor markers. Urol Clin North Am. 20:597-606.

Schoor RA, Elhanbly SM, Niederberger C. 2001. The pathophysiology of varicocele-associated male infertility. Curr Urol Rep. 26:432-6

Schulman IG, Chakravarti D, Juguilon H, Romo A, Evans RM. 1995. Interactions between the retinoid X receptor and a conserved region of the TATA-binding protein mediate hormone-dependent transactivation. Natl Acad Sci U S A. 29;92:8288-92

Shibata H, Kanzaki M, Takeuchi T, Miyazaki J, Kojima I. 1996. Two distinct signalling pathways activated by activin A in glucose-responsive pancreatic beta-cell lines. J Mol Endocrinol. 16: 249-58.

Shimada N, Sobue G, Doyu M, Yamamoto K, Yasuda T, Mukai E, Kachi T, Mitsuma T. 1995. X-linked recessive bulbospinal neuronopathy: clinical phenotypes and CAG repeat size in androgen receptor gene. Muscle Nerve. 18:1378-84.

Simoni M, Gromoll J, Dworniczak B, Rolf C, Abshagen K, Kamischke A, Carani C, Meschede D, Behre HM, Horst J, Nieschlag E. 1997. Screening for deletions of the Y chromosome involving the DAZ (Deleted in Azoospermia) gene in azoospermia and severe oligozoospermia. Fertil Steril. 67: 542-7.

Skakkebaek NE, Giwercman A, de Kretser D. 1994. Pathogenesis and management of male infertility. Lancet. 343:1473-9

Smith DF. 1993. Dynamics of heat shock protein 90-progesterone receptor binding and the disactivation loop model for steroid receptor complexes. Mol Endocrinol. 7:1418-29.

Snoek R, Bruchovsky N, Kasper S, Matusik RJ, Gleave M, et al. 1998. Differential transactivation by the androgen receptor in prostate cancer cells. Prostate. 36:256-63.

Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS, Redwine E. 1987. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. N Engl J Med. 317:909-16

Stanford JL, Just JJ, Gibbs M, et al. 1997. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res. 57 :1194-1198.

Stephan C, Jung K, Lein M, Sinha P, Schnorr D, Loening SA. 2000. Molecular forms of prostate-specific antigen and human kallikrein 2 as promising tools for early diagnosis of prostate cancer. Cancer Epidemiol Biomarkers Prev. 9:1133-47

Suter NM, Malone KE, Daling JR, Doody DR, Ostrander EA. 2003. Androgen receptor (CAG)n and (GGC)n polymorphisms and breast cancer risk in a population-based case-control study of young women. Cancer Epidemiol Biomarkers Prev. 12:127-35

Takahashi K, Eda Y, Okada S, Abu-Musa A, Yoshino K, Kitao M. 1993. Morphological assessment of polycystic ovary using transvaginal ultrasound. Hum Reprod. 8:844-9.

Taketani Y. 1990. Pathophysiology of polycystic ovary syndrome. Overview. Horm Res. 33 Suppl 2:3-4.

Talbot JA, Bicknell EJ, Rajkhowa M, Krook A, O'Rahilly S, Clayton RN. 1996. Molecular scanning of the insulin receptor gene in women with polycystic ovarian syndrome. J Clin Endocrinol Metab. 81:1979-83.

Techatraisak K, Conway GS, Rumsby G. 1997. Frequency of a polymorphism in the regulatory region of the 17 alpha-hydroxylase-17,20-lyase (CYP17) gene in hyperandrogenic states. Clin Endocrinol. 46:131-4.

Timchenko LT, Caskey. Trinucleotide 1996. Repeat disorders in humans: discussions of mechanisms and medical issues. CT. FASEB J. 10:1589-97

Tropeano G, Lucisano A, Liberale I, Barini A, Vuolo IP, Martino G, Menini E, Dell'Acqua S. 1994. Insulin, C-peptide, androgens, and beta-endorphin response to oral glucose in patient with polycystic ovary syndrome J Clin Endocrinol Metab.78:305-9.

Tsukada T, Inoue M, Tachibana S, Nakai Y, Takebe H. 1994. An androgen receptor mutation causing androgen resistance in the undervirilised male syndrome. J Clin Endocrinol Metab.79:1202-1207.

Tut TG, Ghadessy F, Trifiro MA, Pinsky L, Yong EL. 1997. Long polyglutamine tracts in the androgen receptor are associated with reduced transactivation, defective sperm production and male infertility. J Clin Endocrinol Metab. 82: 3777-3782

Usdin K. 1998. NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases Nucleic Acids Res. 1;26:4078-85

Urbanek M, Legro RS, Driscoll DA, Azziz R, Ehrmann DA, Norman RJ, Strauss JF 3rd, Spielman RS, Dunaif A. 1999. Thirty-seven candidate genes for polycystic ovary syndrome: strongest evidence for linkage is with follistatin. Proc Natl Acad Sci U S A. 20;96: 8573-8

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

von Eckardstein S, Syska A, Gromoll J, Kamischke A, Simoni M, Nieschlag E. 2001. Inverse correlation between sperm concentration and number of androgen receptor CAG repeats in normal men. J Clin Endocrinol Metab. 86:2585-90.

Vottero A, Stratakis CA, Ghizzoni L, Longui CA, Karl M, Chrousos GP. 1999. Androgen receptor-mediated hypersensitivity to androgens in women with

nonhyperandrogenic hirsutism: skewing of X-chromosome inactivation. J Clin Endocrinol Metab. 84:1091-5

Wallace EM, Pye SD, Wild SR, Wu FC. 1993. Prostate-specific antigen and prostate gland size in men receiving exogenous testosterone for male contraception. Int J Androl. 161:35-40

Wallerand H, Remy-Martin A, Chabannes E, Bermont L, Adessi GL, Bittard H. 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, Ghadessy FJ, Yong EL. 1998. Analysis of the transactivation domain of the androgen receptor in patients with defective spermatogenesis. Clin Genet. 54:185-192.

Warhol MJ, Longtine JA. 1985. The ultrastructural localization of prostatic specific antigen and prostatic acid phosphatase in hyperplastic and neoplastic human prostates. J Urol.134: 607-13.

Waterworth DM, Bennett ST, Gharani N, McCarthy MI, Hague S, Batty S, Conway GS, White D, Todd JA, Franks S, Williamson R. 1997. Linkage and association of insulin gene VNTR regulatory polymorphism with polycystic ovary syndrome. Lancet. 5;349: 986-90.

Weber JP, Oesterling JE, Peters CA, Partin AW, Chan DW, Walsh PC. 1989. The influence of reversible androgen deprivation on serum prostate-specific antigen levels in men with benign prostatic hyperplasia. J Urol.141: 987-92.

Westberg L, Baghaei F, Rosmond R, Hellstrand M, Landen M, et al. 2001. Polymorphisms of the androgen receptor gene and the estrogen receptor beta gene are associated with androgen levels in women. J Clin Endocrinol Metab. 86:2562-8.

World Health Organisation. 1999. Laboratory manual for the examination of human semen and sperm-cervical mucus interaction 4th Edition. Cambridge, Cambridge University Press.

Yeh S, Chang C. 1996. Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells.Proc Natl Acad Sci U S A. 93:5517-21.

Yong EL, Baird DT, Yates R, Reichert LE, Hillier SG. 1992. Hormonal regulation of the growth and steroidogenic function of human granulosa cells. J Clin Endocrinol Metab. 74:842–849

Yong EL, Chua KL, Yang M, Roy A, Ratnam SS. 1994. Complete androgen insensitivity due to a splice-site mutation in the androgen receptor gene and genetic screening with single stranded conformation polymorphism (SSCP). Fertil Steril. 61: 856-852.

Yong EL, Ng SC, Roy AC, Yun G, Ratnam SS. 1994. Pregnancy after hormonal correction of severe spermatogenic defect due to mutation in androgen receptor gene. Lancet. 344: 826-827.

Yong EL, Tut TG, Ghadessy FJ, Prins G, Ratnam SS. 1998. Partial androgen insensitivity and correlations with the predicted three dimensional structure of the androgen receptor ligand-binding domain. Mol Cell Endocrinol. 137:41-50.

Yoshida KI, 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: 1078-81.

Yoshida A, Rzhetsky A, Hsu L, Chang C. 1993. Human aldehyde dehydrogenase gene family Eur. J. Biochem. 1;251:549-57

Young CY, Montgomery BT, Andrews PE, Qui SD, Bilhartz DL, Tindall DJ. 1991. Hormonal regulation of prostate-specific antigen messenger RNA in human prostatic adenocarcinoma cell line LNCaP Cancer Res. 51:3748-52.

Yuchio Y, Chen J, Hsu L, Yoshida A. 1995. The transcriptional activation of Human Aldehide dehydrogenase gene J. Biolog. Chemistry. 29:17521-27

Zhou ZX, Sar M, Simental JA, Lane MV, Wilson EM. 1994. A ligand-dependent bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH2-terminal and carboxyl-terminal sequences. J Biol Chem. 6;269: 13115-23

Zitzmann M, Gromoll J, Von Eckardstein A, Nieschlag E. 2003. The CAG repeat polymorphism in the androgen receptor gene modulates body fat mass and serum concentrations of leptin and insulin in men. Diabetologia. 46:31-9.

Zitzmann M, Brune M, Kornmann B, Gromoll J, Junker R, Nieschlag E. 2001. The CAG repeat polymorphism in the androgen receptor gene affects bone density and bone metabolism in healthy males. Clin Endocrinol . 55:649-57.

Zitzmann M, Depenbusch M, Gromoll J, Nieschlag E. 2003. Prostate volume and growth in testosterone-substituted hypogonadal men are dependent on the CAG repeat polymorphism of the androgen receptor gene: a longitudinal pharmacogenetic study J Clin Endocrinol Metab. 88:2049-54.

Zirkin BR, Santulli R, Awoniyi CA, Ewing LL. 1989. Maintenance of advanced spermatogenic cells in the adult rat testis: quantitative relationship to testosterone concentration within the testis. Endocrinology. 124: 3043-9.