GENETIC ANALYSIS OF THE ROLE OF ANDROGEN METABOLISM IN THE PATHOGENESIS OF PROSTATE CANCER

ROSHAN HENDRICKS

Thesis presented in partial fullfilment of the requirements

for the degree of Master in Natural Sciences in Medical Sciences

(Medical Biochemistry)

Study Leader: Dr V.M. Hayes

Co-study Leaders: Prof. C.F. Heyns, Dr R. Hillermann

University of Stellenbosch

December 2004

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and has not previously in its entirety or in part been submitted at any university for a degree.

Ms R. Hendricks Date

SUMMARY

Prostate cancer (CaP) has the highest incidence of any malignancy affecting South African males. The aetiology of prostate carcinoma indicate that ethnicity is one of the most important risk factors. The causes of these ethnic differences are unknown but presumably involve both environmental and genetic factors. Carcinoma of the prostate is androgen dependent, and it has been suggested that variations in androgen metabolism and synthesis may affect an individuals' risk. Therefore, genes involved in these pathways are candidates for determining CaP susceptibility.

In this study two candidate genes in the androgen biosynthetic and metabolic pathway were analysed, viz., the androgen receptor gene (*AR*), involved in androgen transport and transcriptional activation, and the cytochrome p450c17α gene (*CYP17*), important for testosterone biosynthesis. Comprehensive mutation detection assays were designed (appropriate for analysis of archival paraffin-embedded material) for almost the entire coding region (excluding polymorphic repeat sequences), and including all splice site junctions of the *AR* gene, as well as the entire coding region of *CYP17*. The aim of this study was thus to determine the type and frequencies of genetic variants of these androgen metabolism genes within the diverse South African population, and to determine if the observed ethnic variation in the incidence and progression of CaP can be explained by ethnic-based genetic differences.

For high sensitivity mutation detection, the most powerful of the pre-screening methods was used, namely denaturing gradient gel electrophoresis (DGGE). 20 CaP and 25 control benign prostatic hyperplasia (BPH) tissue samples were screened in order to identify possible mutations. Blood samples from the same patients were analysed in order to determine whether mutations are germline and therefore present in all cells of the body. Additional blood samples from the Western Province Blood Transfusion Service (WPBTS) (Refer to section 2.1.2, Table) were also analysed in order to determine the frequency of identified polymorphisms within the general population. Certain polymorphisms were further analysed in paraffin-embedded wax material (exclusively from Blacks) to determine the distribution of these polymorphisms in the Black population. Direct sequencing of mutant-containing DNA fragments was performed to determine the exact location and nature of mutation.

Using the *AR*- DGGE assay 4 novel mutations were identified as well as a previously reported codon 211 (E211) polymorphism. With the *CYP17*- DGGE assay, 3 novel single nucleotide polymorphisms (SNPs) were detected. Three base variants occured, in codons 36 (L36), 46 (H46) and 65 (S65), as well as intronic substitutions in intron 4 (IVS+58G4C) and intron 6 (IVS–25C7A). Frequencies of SNPs were measured in the CaP and BPH samples.

In conclusion, the identified polymorphisms could be used as markers in determining CaP susceptibility and may thus facilitate the identification of individuals with a high- or low-risk of developing carcinoma of the prostate.

OPSOMMING

Prostaatkanker vertoon die hoogste voorkoms van enige kwaardaardigheid wat Suid-Afrikaanse mans aantas. Die etiologie van prostaatkarsinoom dui aan dat etnisiteit een van die mees belangrike risikofaktore is. Oorsake van hierdie etniese verskille is onbekend, maar vermoedelik is omgewing en genetiese faktore albei betrokke. Karsinoom van die prostaat is androgeenafhanklik en daar is voorgestel dat variasies in androgeenmetabolisme en androgeensintese 'n persoon se risiko mag affekteer. Gevolglik, is gene betrokke in hierdie paaie kandidate vir die bepaling van prostaatkanker vatbaarheid.

In hierdie studie het ons twee kandidaat gene in die androgeen biosintetiese en metaboliese pad geanaliseer, naamlik, die androgeen reseptor geen (*AR*), betrokke in androgeen vervoer en aktivering van transkripsie, en die sitokroom p450c17α geen (*CYP17*), belangrik vir testosteroon biosintese. Ons het omvattende mutasie-bespeurings-essai-sisteme ontwikkel (ook uitvoerbaar op argivale paraffien-bewaarde materiaal), wat amper vir die hele koderende streek van die *AR* geen gebruik kan word (uitsluitend herhalende polimorfiese reekse) en wat alle splytpunt-aansluitings van die *AR* geen insluit, asook vir die hele koderende streek van *CYP17*. Die doel van hierdie studie was dus om die tipe en frekwensies van genetiese variante van androgeen metabolisme gene in ons diverse Suid-Afrikaanse bevolking te bepaal, en om vas te stel of die waarneembare etniese wisseling in die insidensie en vordering van prostaatkanker verstaan kan word deur etnies gebaseerde genetiese verskille.

Die mees sensitiewe tegniek wat tans beskikbaar is vir vooraf-sifting vir onbekende mutasies is gekies, naamlik denaturerende gradiënt gel elektroforese (DGGE). Om moontlike mutasies op te spoor, het ons 20 prostaatkanker en 25 benijne prostaathiperplasie (BPH) monsters geanaliseer. Analise was gedoen op bloedmonsters van dieselfde pasiënte om vas te stel of kiemlyn mutasies (in alle liggaamselle) teenwoordig is. Bykomstige bloedmonsters (van die Westelike Provinsie Bloedoortappingsdiens) is ook geanaliseer om die frekwensie van bespeurde polimorfismes in die algemene bevolking te bepaal. Argivale paraffien-bewaarde materiaal (eksklusief van Swartes) is ook geanaliseer om die verspreiding van sekere polimorfismes in die Swart bevolking te bepaal. Direkte DNA volgorde bepaling van mutante DNA fragmente is uitgevoer om die ligging en tipe van mutasies te bepaal.

Met die toepassing van ons *AR*-DGGE mutasiesisteem het ons 4 nuwe mutasies ontdek asook 'n kodon 211 (E211) polimorfisme wat voorheen gevind is. Vyf enkel nukleotied polimorfismes is met die *CYP17*-DGGE mutasiesisteem opgespoor. Die polimorfismes sluit in: drie basis veranderinge wat voorkom in kodons 36 (L36), 46 (H46) en 65 (S65), asook introniese substitutisies in intron 4 (IVS+58G4C) en intron 6 (IVS−25C7A). Frekwensies van die polimorfismes was bereken in die prostaatkanker en BPH monsters.

Die resultate aangebied in hierdie tesis dui aan dat die gevonde polimorfismes as merkers gebruik kan word om prostaatkanker vatbaarheid te bepaal en daardeur

individue te identifiseer met 'n hoë of lae risiko vir prostaatkarsinoom ontwikkeling.

ABBREVIATIONS

aa : amino acid

AAG : androstanediol glucuronide

AIS : Androgen insensitivity syndrome

AR : Androgen receptor

ARE : Androgen response element

Bp : base pairs

BPH : Benign prostatic hyperplasia

^{0}C : degrees celsius

cAMP : cyclic adenosine monophosphate

CaP : Prostate cancer

CGH : comparative genome hybridisation

CYP17 : Cytochrome p450c17α

CYP3A4 : Cytochrome p450 3A4

d : deviation

df : degrees of freedom

DGGE : denaturing gradient gel electrophoresis

DHEA : dehydroepiandrosterone

DHT : dihydrotestosterone

DNA : deoxyribonucleic acid

DRE : digital rectal examination

ds : double stranded

E : glutamate

ER : estrogen

ER α : estrogen receptor α

FISH : fluorescence *in situ* hybridisation

GR : glucocorticoid receptor

GRE : glucocorticoid receptor element

GTP : green tea polyphenols

H : histidine

hKLK2 : human glandular kallikrein-1

HRE : hormone response element

HSP : heat shock protein

IL : interleukin

L : leucine

LBD : ligand binding domain

LD : linkage disequilibrium

LH : luteinising hormone

LNCaP : lymph node prostate cancer

LOH : loss of heterozygosity

MAB : maximal androgen blockade

MMTV : mammary tumour virus

MR : mineralocorticoid receptor

mRNA : messenger ribonucleic acid

ORF : open reading frame

p : probability value

PAA : polyacrylamide

PCR : polymerase chain reaction

PR : progesterone receptor

PSA : prostate specific antigen

RARE : retinoic acid response element

RAR γ : retinoic acid receptor gamma

RXR α : 9-*cis*-retinoic acid receptor alpha

S : serine

SBMA : spinal and bulbar muscular dystrophy

scc : side chain cleavage

SHBG : sex-hormone binding globulin

SNP : single nucleotide polymorphism

SSCP : single strand conformation polymorphism

SRD5A2 : steroid 5-α-reductase type II

ss : single stranded

TA : thymine / adenine

TAD : transactivating domain

TIF : transcription intermediary factor

Tm : melting temperature

TNF : tumour necrosis factor

TP53 : p53 gene

TR α : thyroid hormone receptor alpha

TRAMP : transgenomic adenocarcinoma of the mouse prostate

TSC : The SNP consortium

UCSC : University of California at Santa Cruz

UF : urea and formamide

UTR : untranslated region

WPBTS : Western Province blood transfusion service

FIGURES AND TABLES

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The focus of molecular genetic analysis has moved away from single gene disorders which usually have a simple Mendelian pattern of inheritance, to more complex multifactorial diseases like most cancers (Lander and Schork, 1994; Gray *et al*, 2000; Emahazion *et al*, 2001). These complex diseases do not exhibit classic Mendelian recessive or dominant inheritance due to a single causative factor and occur at a much higher frequency than monogenic disorders. Multiple susceptibility genes may be quite common in the population and could each contribute a small effect but collectively increase the risk of developing cancer indirectly. Variant sequences in such genes in a common pathway could result in a "polygenic" aetiology of cancer, facilitating the identification of individuals with a high- or low-risk polygenic profile.

1. PROSTATE CANCER: GENERAL INTRODUCTION

1.1.1 Clinical aspects of prostate cancer

The prostate is a small, male gland which surrounds the urethra[1]. This gland produces semen (prostatic fluid) [2] . Prostate cancer (CaP) develops in the prostate gland and early detection greatly improves the chances of effective treatment. This cancer is generally slow growing and progressive and may not show any symptoms at all. The early warning signs include difficulty in passing urine, a frequent need to urinate (especially at night) and difficulty in starting and stopping the urinary stream, usually caused by the tumour pressing on the urethra and affecting or blocking the flow of urine (http://www.cansa.org.za,2000). These symptoms could also, however,

[1] The tube which carries urine from the bladder to the outside of the body

result from benign prostatic hyperplasia (BPH), a common progressive enlargement of the prostate gland, which obstructs the flow of urine and can eventually, at worst, stop it completely. It occurs in 50-75% of males over the age of 50 in the presence of testes that are functioning normally (Berry *et al*, 1984). The reason for this spontaneous prostate growth is not exactly known, however research has shown that abnormally high levels of the hormone dihydrotestosterone (DHT) is present in the prostates of individuals with BPH (Wilson, 1980 and 1996). Treatment for BPH is normally surgical with removal of the prostate via open prostatectomy or, more usually, transurethal resection.

Most cases of early CaP show no symptoms but can be detected by a blood test which measures the levels of PSA (prostate specific antigen); a protein made by prostate cells. Measurement of serum PSA has proven to be the most sensitive overall marker available for CaP (Catalona *et al*, 1994). The majority of studies have accepted 4.0 ng/ml as the upper limit of normal, while some investigators have suggested a lower cut off level of 3.0 ng/ml (Catalona *et al*, 1991; Labrie *et al*, 1992; Crawford *et al*, 1996; Rietbergen *et al*, 1997; Heyns *et al*, 2001). PSA levels reflect the likelihood of having CaP and generally men with high levels of > 4.0` ng/ml are advised to have a biopsy (a sample of tissue from the gland) to check for the presence of cancerous cells. Early CaP can also be detected by rectal examination. However, the combination of a rectal examination together with the measurement of serum PSA concentration, provides a better method of detecting CaP than PSA alone (Smith and Catalona, 1995).

[2] A white substance which is important for the motility, nourishment and fertility of sperm

If a biopsy indicates the presence of CaP, a pathologist assigns each tissue sample a grade, based on the architectural patterns of cancer cells observed under a light microscope. This indicates how far the cells have travelled along the path from normal to abnormal. The grade offers a good clue to the tumour's behaviour. A tumour with a low grade is likely to be slow growing, while one with a high grade is more likely to grow aggressively or to have spread outside the prostate (metastasised). The most widely used grading method for CaP is known as the Gleason grading system, which assigns a grade to each of the two largest areas of cancer in the tissue samples (http://www.phoenix5.org, 2000). Grades range from 1 to 5, with 1 being the least aggressive and 5 the most aggressive. Grade 3 tumours for example seldom have metastases, but metastases are common with grade 4 or grade 5. The two grades are then added together to produce a Gleason score. A score of 2 to 4 is considered low grade; 5 through 7, intermediate grade; and 8 through 10, high grade. A tumour with a low Gleason score typically grows slow enough that it may not pose a significant threat to the patient in his lifetime.

Advanced symptoms of CaP are blood in the urine, a painful or burning sensation when passing urine, any of the aforementioned symptoms combined with pains in the lower back, upper thighs or pelvic area. If left untreated, the cancerous cells in the prostate could spread to other parts of the body such as bone, the liver, lungs and adrenal glands (http://www.cansa.org.za,2000). CaP treatment involves either surgery to remove the prostate gland (radical prostatectomy) or radiation (radiotherapy).

Carcinoma of the prostate is the most frequently diagnosed malignancy and one of the leading causes of cancer deaths in men in Western countries (Parker *et al*, 1997). It has histologically been diagnosed as the second most common cancer in South African men, with an annual estimated incidence of 19.1 / 100 000 men (Sitas, 1994). CaP incidence varies among race groups and a comparison shows that it is the second most common histologically diagnosed carcinoma in Coloureds and Blacks (comprising 9.12% and 8.00 %, respectively), the third most common cancer in Whites (7.11%), and the fourth most common malignancy in Asian (Indian) (4.56%) men in South Africa (Sitas, 1994). Risk factors are known to influence the incidence of CaP, which will be discussed further in the next section.

1.1.1 Aetiology of prostate cancer

CaP, like all cancers, is a multifactorial disease where environmental stimuli and genetic risk factors, in varying proportions, are important promoters in disease aetiology. Most environmental factors have been linked to hormonal exposures, which will be covered in Chapter 1.2. The only established risk factors are age, race-ethnicity and family history of CaP (Mettlin, 1997).

1.1.2 Age and race-ethnicity

The two most important risk factors for CaP are age and race-ethnicity. CaP is extremely rare before the age of 40 years but the rate of increase with age is greater than for any other cancer; rates increase at approximately the 9th -10th power of age whereas figures for other common epithelial cancers like lung and colon cancer are between 5 and 6 (Cook *et al*, 1969). For CaP the rate of increase is steepest early in the age-specific curve and then gradually dissipates with age. This contrasts with

some epithelial cancers such as colon cancer, for which the relationship between age and incidence is strictly linear in log-log units (Cutler and Young, 1975).

Substantial differences in the prevalence of CaP are observed among populations and several studies have suggested that Black (African) men have a higher incidence of disease. African-American men have been reported to have the highest incidence rates in the world. Even though rates differ slightly among African-American men across the United States (US), average rates are about 50-70% higher in African-Americans than in Caucasians (Ross and Schottenfeld, 1996). The higher incidence of CaP in Black compared to white men has been documented for more than 20 years. Even though incidence rates rose strikingly in both races starting in the early 1990s secondary to PSA-based early detection, the difference between races has remained. The differences in incidence are most dissimilar in the age group younger than 65 years, where Caucasians have a rate of 45.5 / 100 000 and Blacks 81.5 / 100 000 (Brawley *et al*, 1998a and 1998b). However, there has been no sufficient relative data on CaP rates among Blacks in South Africa to substantiate these increased incidence rates. Interestingly, the other end of the racial-ethnic range of CaP incidence shows that Asian populations (Native Japanese and Chinese men) have the lowest CaP rates in the world. Although the difference in incidence between African-Americans and these Asian populations was approximately 50-fold (Ross and Schottenfeld, 1996), it has been demonstrated that different detection strategies for CaP are mainly accountable for this difference and that the difference in incidence is in fact much smaller; in the order of 2- to 3-fold (Shimizu *et al*, 1991). Even so, Japanese- and Chinese-Americans have rates higher than those of men in their respective homelands (about 40-50% higher), however, their rates still remain

much less than US Caucasians and are even lower than those of African-Americans, although the Japanese, in particular, represent a highly acculturated immigrant population in the US (Bernstein and Ross, 1991). Several studies have reported higher proportions of CaP cases with advanced stage and poorer survival rate among Blacks than among Caucasians (Brawn *et al*, 1993; Optenberg *et al* 1995 and Powell *et al*, 1995). Therefore Black men do not only have a higher risk of CaP, they also have a greater risk for the more aggressive forms of the disease (Montie and Pienta, 1999) but as yet there are no published comparable South African data which show similar results.

In the past few years, decreased mortality rates from CaP have been observed, possibly as a result of early detection and treatment. The mortality rate, however, has been greater in Black than in Caucasian men, with the greatest disparity noted in younger men. It has been shown that age-specific mortality rates are twice as high in Blacks between the ages of 40 and 60 years, and slightly less than two times greater in men older than 70 years (Powell, 1998). For men diagnosed after the age of 70 years, mortality rates are similar (Pienta *et al*, 1995).

1.1.2.2. Familial prostate cancer

Many studies have suggested that CaP clusters within male members of a family (familial CaP) (Krain, 1974; Meikle *et al*, 1985; Spitz *et al*, 1991, Eeles *et al*, 1997; Schaid *et al*, 1998). One case-control study (Steinberg *et al*, 1990) was carried out to determine the frequency of CaP in 691 male relatives with CaP and 640 male relatives of their spouses as controls. Results showed that men with an affected father or brother were twice as likely to develop the disease than men with no

affected relatives. It was also found that men with two or three first-degree relatives affected had a 5- and 11-fold increased risk of developing CaP. Inherited risk for CaP, however, represents less than 10% of CaP cases (Carter *et al*, 1992).

Some studies support an X-linked (traits which are determined by genes on the X-chromosome) or recessive model of inheritance (Monroe *et al*, 1995; Narod *et al*, 1995), whereas other data suggest an autosomal dominant mode of inheritance (male to male transmission) for early onset disease (Carter *et al*, 1992; Gronberg *et al*, 1997; Schaid *et al*, 1998). This rare allele (alternative form of a gene) could account for 9% of all CaP cases at $\leq$ 85 years of age and 43% of CaP cases diagnosed at $\leq$ 55 years of age. In addition to the normal allele or alleles, most identified loci also have one or more rare alleles. Many human loci were in fact identified through a clinically significant disorder caused by a rare mutant allele (Thompson *et al*, 1991). Even though the dominant alleles have a low population frequency of 0.36-1.67%, they are highly penetrant (the all-or-none expression of a mutant genotype) so that by the age of 85 years, 63-89% of men carrying a mutation would be likely to be diagnosed with CaP (Carter *et al*, 1992; Gronberg *et al*, 1997; Schaid *et al*, 1998). Inherited susceptibilities [3] seem to play a role in the development of CaP as highly penetrant genes linked to CaP have been identified on the X chromosome (Monroe *et al*, 1995 and Xu *et al*, 1998) and chromosome 1 (Smith *et al*, 1996; Gronberg *et al*, 1997; Eeles *et al*, 1998 and Gibbs *et al*, 1999). Mutations in the Hereditary prostate cancer 1 gene (*HPC1*), located on chromosome 1q24-q25, have been found in patients with hereditary CaP and account for about 1 in 500 cases. Further research into *HPC1* is ongoing.

[3] Hereditary CaP is a subtype of the disease in which Mendelian inheritance of a susceptibility gene is evident

Although there is strong evidence for genetic susceptibility to CaP, most of the genes underlying susceptibility remain to be identified. Recently, the results of eight genome-wide searches were reviewed, based on 1.293 families with multiple cases of CaP (Easton *et al*, 2003). Results from this study showed that even larger family sets would be required to reliably evaluate linkage.

Due to the nature of CaP the search for genetic determinants involved in the disease has been challenging, partly because such a complex disease is not compliant to be characterised by methods used as in the case of diseases with simple Mendelian genetic inheritance. CaP is more likely to result from "inherited susceptibility" — low penetrant genes will be discussed in Chapter 1.3.

1.2 ANDROGENS AND PROSTATE CANCER

Androgens are male hormones (e.g. testosterone and dihydrotestosterone) and are essential in several stages of male development. DHT is responsible for sexual differentiation by sculpturing the human embryo in the uterus to cause development of internal and external male genitalia. During puberty, a flow of testosterone initiates and maintains spermatogenesis (male sex drive and performance) and secondary male characteristics like growth of accessory sex organs, including the prostate gland. Other secondary male characteristics stimulated by testosterone in adults include muscle development, voice deepening, rugation and pigmentation of the scrotum, some phallus growth and penile erections, axillary and pubic hair development, and male psychosexual orientation. DHT is also important in maintaining the structure and function of the mature prostate. Without androgens,

the prostate undergoes apoptosis (programmed cell death) and significant atrophy (Trapman and Brinkmann, 1996).

Androgens therefore play a crucial role in their primary target organs by controlling cell proliferation and an increased cell proliferation appears to be a shared event in disease production or development of many human cancers (Preston-Martin *et al*, 1993). Even though hormonal secretion and metabolism can be environmentally influenced (for example through a change in diet); hormonal-pattern control is directly or indirectly regulated genetically. Environmental and genetic factors which change the hormonal environment and therefore cell division rates in target organs would be expected to alter the rate of malignant transformation in these target tissues as well (Henderson *et al*, 1988a).

1.2.1 Circulating testosterone levels and Prostate cancer risk

Studies of circulating hormone levels and CaP risk are complicated as a result of inter- and intra-assay variability, cyclical and sporadic variations in the target hormone, the requirement for large prospective studies and difficulty in determining tissue-specific exposure. However, one adequate prospective serologic study has been conducted to date in which a rather strong relationship between circulating testosterone levels and CaP development was found (Gann *et al*, 1996). In this study, steroid hormone binding globulin (SHBG) levels, which are firmly linked with testosterone levels, were measured and a strong pattern of increased CaP risk with increasing levels of circulating free testosterone was observed.

1.2.2 Androgens and Prostate cancer risk (Environment and Prostate cancer aetiology)

Results of various epidemiological studies suggest that environmental risk factors like diet and other risk factors such as level of physical activity, vasectomy, obesity and sexual behaviour could be related to a hormonal mechanism.

1.2.2.1 Diet and nutrition

There has been considerable consistency among studies showing that a high intake in fat can increase CaP risk (Berg, 1975; Lew and Garfinkel, 1979; Kolonel *et al*, 1983; Kaul *et al*, 1987; Severson *et al*, 1988 and 1989). Theoretically, high levels of dietary fat could lead to a higher production of sex hormones and high circulating levels of androgens can lead to increased CaP (Nobel, 1959; Coffey, 1979; Hamalainen *et al*, 1984). The exact mechanism is not completely understood but an increased hormonal bioavailability as a result of a high-fat diet has been implied. The concentrations of plasma in fatty acids are responsive to dietary fat and these levels correlate with concentrations of bioavailable sex steroids (Bruning and Bonfrer, 1989). In addition to this, a low intake of dietary fibre and complex carbohydrates and a high intake of protein and energy have been found to be associated with the increased risk for CaP in some studies (Bosland, 1988; Nomura and Kolonel, 1991; Bosland, 1994; Kolonel, 1996).

CaP risk and the associations with individual nutrients or food that have been reported are not very strong. Even though CaP is rare in China and Japan, the studies of migrants from these countries to the US have shown that they have a considerable increased risk (Nomura and Kolonel, 1991; Bosland 1988). These

changes are thought to be the result of differences in the environment such as lifestyle and diet (Nomura and Kolonel, 1991 and Bosland 1988). Therefore one can conceive that the combination of the effects of dietary factors on CaP development and progression are more important than individual effects of any particular dietary factor (Bosland *et al*, 1999). This is supported by lack of effect of dietary fat after CaP was induced in animal models, whereas epidemiological studies (Bosland 1988, Bosland 1994; Bosland *et al*, 1999) show a consistent positive association between an intake of dietary fat and an increased CaP risk (West D *et al*, 1991; Andersson *et al*, 1995). There are a number of promising leads concerning other protective or causative dietary influences on the development of CaP. Evidence from prospective epidemiological studies (Giovannucci *et al*, 1995) and serologic biomarker data (Gann *et al*, 1999) suggest that lycopene, a potent caretenoid found primarily in tomato and tomato products, may decrease CaP risk. Vitamin E, another antioxidant vitamin, was found to be associated with a reduced risk of CaP during a randomised-controlled clinical trial specifically designed to lower the risk of lung cancer in male smokers (Heinonen *et al*, 1998). Selenium, a soil-based mineral, which has both proapoptotic and antioxidant properties significantly reduced CaP incidence as a secondary effect of a prevention trial designed to reduce skin cancer recurrences (Clark *et al*, 1998). Other dietary factors like dietary fibre and phytoestrogens (a plant-based estrogen) have not been investigated thoroughly but are of interest (fibre as a means to enhance faecal excretion and possibly lower the levels of enterohepatic recirculation of androgens, and phytoestrogens by inhibiting androgens in the epithelium of the prostate; interfering with prostate growth) (Ross and Schottenfeld, 1996).

1.2.2.2 Obesity

Present information contradicts whether an increase in body mass index or obesity is a CaP risk factor (Bosland 1994; Denmark-Wahnefried *et al*, 1997; Bosland *et al*, 1999). A significant increase in CaP risk was observed (Severson *et al*, 1988) with an increase in the upper-arm muscle area (but not fat area) and the circumference of the upper-arm. However, a positive association between muscle mass and the risk of CaP could be the result of exposure to exogenous or endogenous hormones or even other anabolic factors (Severson *et al*, 1988; Landry and Primos, 1990).

1.2.2.3 Physical activity

There is contradictory evidence that the level of physical activity could be a possible risk factor for CaP (Bosland, 1994; Giovannucci *et al*, 1998 and Bosland *et al*, 1999). Exercise could increase, decrease or have no effect on circulating androgen concentrations. This depends on the level of exercise, the time the blood sample was drawn in relation to the exercise and the training program followed (Adlercreutz *et al*, 1986 and Keizer *et al*, 1989). Thus it is a possibility that the nature and extent of physical activity could influence circulating androgen concentrations and even CaP risk.

1.2.2.4 Vasectomy

Many sources of literature have identified vasectomy as a possible risk factor for CaP in cohort studies (Giovannuci *et al*, 1992 and 1993) as well as case-control studies (Honda *et al*, 1988; Mettlin *et al*, 1990; Rosenberg *et al*, 1990; Perlman *et al*, 1991; Spitz *et al*, 1991; Peterson *et al*, 1992; Hayes *et al*, 1993). Three mechanisms have been proposed by which vasectomy could possibly enhance CaP risk namely, by

increasing levels of circulating androgens, reducing of seminal fluid production (Honda *et al*, 1988; Mettlin *et al*, 1990; Rosenberg *et al*, 1990; Howards and Peterson 1993; Giovannuci *et al*, 1993; John *et al*, 1995; Mo *et al*, 1995) and immunologic mechanisms involving antisperm antibodies. Some studies (Kobrinsky *et al*, 1976; Purvis *et al*, 1976; Smith *et al*, 1979; Whitby *et al*, 1979; Alexander *et al*, 1980; John *et al*, 1995; Mo *et al*, 1995) have found slight, yet significant changes in the circulating levels of certain hormones. In one study, elevated levels of 5α-dihydrotestosterone, the metabolite of testosterone, was found in the prostate of men who had undergone vasectomies (Mo *et al*, 1995). This could be an important factor associated with CaP risk.

1.2.2.5 Sexual behaviour

The possibility that sexual factors could play a role in the aetiology of CaP, was investigated in several case-control studies (Mandel and Schuman, 1987; Ross *et al*, 1987; Bosland, 1988; Honda *et al*, 1988; Fincham *et al*, 1990; Yoshida *et al*, 1990; Hayes *et al*, 1992). A significant positive association was reported between the level of sexual activity (masturbation and intercourse) and the circulation of total testosterone levels in men between 69 and 70 years of age (Tsitouras *et al*, 1982). No decrease in testosterone levels with aging in men who are sexually active was found as well. This suggests that a hormonal mechanism could underlie a possible association between sexual activity and the risk of CaP.

1.2.2.6 Requirement of androgens in experimental prostate carcinogenesis

Men and dogs are the only mammalian species that spontaneously develop CaP with significant frequency (Coffey, 1993). If castrated before puberty, they do not develop

CaP. It has proven to be very difficult to produce CaP experimentally which imitates the human disease histologically. The lack of naturally occurring and experimental models of disease has slowed down the progress in understanding the pathogenesis of CaP. In 1977 the first experimental model of prostate adenocarcinoma (histologically equivalent to the human disease) was produced (Nobel, 1977) by the subcutaneous administration of testosterone in rats. Additional experimental rodent models have since been developed; all of which require androgens for tumour induction and progression.

1.2.2.7 Hormonal patterns among racial-ethnic groups

Since racial-ethnic variation in incidence is such a powerful characteristic of CaP epidemiology, investigations over the past decade have explored whether underlying differences exist in the androgen environment of populations at the extremes and in the middle of CaP incidence.

A comparison of the hormonal status of small groups (n = 11-20) of Black South African, White-American and African-American men with ages ranging from 40 to 55 years was made by Hill and his co-workers (Hill *et al*, 1979; Hill *et al*, 1980). In a separate study, Black South African, White-American and African-American boys of age 12 to 15 years; and young Black South African and African-American men in the 18 to 21 year old category were compared (Hill *et al*, 1984). In the older men (40-55 years), the plasma levels of dehydro-epiandrostenedione (DHEA), the testosterone precursor, were significantly lower in the two groups of Black men than in the White men. The plasma levels of androstenedione, also a testosterone precursor, and estradiol (another sex steroid) were significantly higher in the African men than in the

two American groups and no differences were observed in testosterone levels among these groups. Results for DHEA and testosterone were similar in the other study with the younger men and 12 to 15 year old boys. Androstenedione levels however, were much lower in the African than in the American subjects and estradiol was lower in young Black boys of age 12 to 14 years than in White boys but higher in older Black boys and young Black men than in White boys and men. These results suggest an interaction between environmental differences and ethnic background that changes over the years as an individual becomes more sexually mature. In the studies of Hill and colleagues among South African Black men, the men in the 18-21 year old group were different from the 40-50 year olds with respect to DHEA and androstenedione, suggesting the importance of separately analysing the hormonal profiles of older and younger men.

Hormonal differences among racial-ethnic groups could be present even in the *in utero* period, as a study shows which compares the circulating hormone levels in African-American and White-American women in the first trimester of their pregnancies (Henderson *et al*, 1988a). A 47% higher serum testosterone level was found in Black than in White women and estradiol levels were also 37% higher; suggesting that African-American males are exposed to higher levels of androgen concentrations than White-American males even before birth.

In another study, testosterone levels of healthy young adult African-American (at a very high risk for CaP) and White men (at half the risk for CaP than Black men) were compared (Ross *et al*, 1986). Even though CaP is a disease of older men, the focus of this study was on young adult men in order to avoid confusion as a result of effects

of other chronic diseases associated with ageing. It was also recognised that the differences in CaP incidence relating to racial-ethnicity is at its maximum during the early stages at which CaP appears, implying that the hormonal milieu responsible occurs quite early in life. In the study, it was found that the total circulating testosterone was 19% higher in the Black subjects than in the group of White subjects and was even 15% higher after adjusting for known correlates of testosterone, which include smoking, alcohol use, body mass index and time of sampling. The free testosterone was 21% higher in the African-Americans than in the White-Americans. The authors have estimated that differences of such a magnitude in circulating testosterone is large enough to explain the substantial difference in CaP incidence (Ross *et al*, 1986). This study suggests an association between CaP risk and high concentrations of circulating androgens and is based on the fact that CaP incidence rates increase as a power function of age and as tissue age is a function of the rate of cell division in target tissue, a direct relationship exists between hormonal changes and prostate tissue age (Pike *et al*, 1983).

The study of hormone patterns in African-American and White men (Ross *et al*, 1986) were extended to include young Japanese men that were born and raised in rural Japan and have a low expected lifetime risk of CaP compared to their previously contrasted U.S. Black and White counterparts (Ross *et al*, 1992). These Japanese men had testosterone levels that were intermediate between those of African-Americans and Whites and did not differ significantly from these two groups. In this study circulating levels of androstanediol glucuronide, a reliable index of in vivo 5α-reductase activity, was measured in all three groups of men. The levels of this hormone in Japanese men were 25-35% lower than the African-American and

White men (Ross *et al*, 1992) and these two U.S. groups had significantly higher levels of this androgen metabolite, with African-Americans having elevated levels compared to whites (Kelsey and Bernstein, 1996). Chinese men in Hong Kong were also shown to have similarly low levels of androstanediol glucuronide as the Japanese men (Lookingbill *et al*, 1991). These findings suggest that, in comparison with the high-risk U.S. groups, the low-risk Japanese population has a lower testosterone metabolism, probably due to a low activity of 5α-reductase that converts testosterone to DHT and the testosterone precessor androstenedione to androsterone. The higher levels of androsterone glucuronide in U.S. men, however, could indicate a higher production of testosterone when compared with Japanese men. Therefore the differences in CaP incidence among African-Americans, White Americans and Asians (Chinese and Japanese men) could be due to differences in testosterone biosynthesis and metabolism.

1.2.3 Androgens and the treatment of Prostate cancer

Androgen stimulation of the prostate over a prolonged period has been identified as a risk factor that promotes prostate carcinogenesis. In 1941 Charles Huggins and his colleague reported one of the most important developments in the treatment of CaP. They documented the clinical benefits of bilateral orchiectomy in CaP patients with advanced disease (Huggins and Hodges, 1941). Bilateral orchiectomy is the primary androgen ablation therapy and deprives tumour cells of the proliferative influence of testosterone. Even decades later androgen ablation therapy is still the mainstay of treatment for metastatic CaP and is associated with stabilisation or regression of disease in up to 80% of hormone therapy naive patients, i.e. before treatment with androgen ablation (The Leuprolide Study Group, 1984; Peeling, 1989). Conversely,

cytotoxic chemotherapy has not been shown to offer a survival advantage and has only been administered in patients after all attempts at hormonal therapy have failed. The mechanical basis of androgen ablation can be observed in the majority of normal and malignant prostate cells that undergo apoptosis when deprived of androgens. The AR mediates the maintenance of androgen growth. Antiandrogens are competitive inhibitors that prevent the natural ligands of the AR from binding to the receptor, and are used therapeutically in CaP patients. Prostate tumours that have not previously been exposed to androgen ablation consist of a combination of cells that differ in their requirements of antiandrogen for growth. Most of the tumour cells need androgen for growth while a few can grow in the absence of androgen. When prostate cells are now exposed to hormone therapy, the group of cells that grow independently of androgens becomes more prominent and the disease progresses despite androgen ablation. So even though there is an observation of an initial response in patients with advanced disease, androgen-independent tumours eventually grow, leaving hormone therapy ineffective. The AR has been identified as a component in this hormone-resistant tumour growth (Trapman and Cleutjens, 1997). However, resistance to androgen ablation could also develop as a consequence of a deregulated androgen signalling axis, resulting from *AR* gene mutation or amplification, altered interactions of coregulatory molecules during transcription, or non-steroidal activation of the AR by growth factors and cytokines (Buchanan *et al*, 2001).

1.2.4 Androgen biosynthesis

Since the prostate is an androgen-regulated organ that requires the influence of androgens for growth and development, one can agree that these hormones are the

main stimulus for cell proliferation in the prostate (Coffey, 1979). Because cell division and malignant transformation are inextricably linked (Preston-Martin *et al*, 1993), the role of androgens in prostate carcinogenesis generates an intense interest.

Testosterone and DHT are the main biologically active androgens. Less than 10% of circulating androgens are secreted by the adrenal gland (Canovatchel and Imperato-McGinley, 1992). Testosterone is synthesised within the Leydig cells of the testis and is produced from cholesterol following a series of enzymatic reactions involving several of the cytochrome P450 enzymes (Berkovitz *et al*, 1983; Waterman and Keeney, 1992; Geller *et al*, 1997; Miller *et al*, 1997). Its rate of synthesis is controlled by luteinizing hormone (LH), which is secreted by the pituitary gland. The release of LH by the pituitary gland is controlled, in turn, by luteinizing hormone-releasing hormone (Singh *et al*, 2000). The major metabolic androgen pathways relating to the prostate and the sites of activity of the gene products that pertain to the prostate are shown in Figure 1. The first hormonally regulated and rate-limiting step in the biosynthesis of all steroid hormones is the conversion of cholesterol to pregnenolone. Steroid secretion is directly related to steroid synthesis, as steroidogenic cells do not store significant amounts of steroid hormones. Regulation of steroid synthesis occurs both acutely and chronically. The acute stimulation of steroidogenesis happens within minutes and is mediated by a steroidogenic acute regulatory protein, which facilitates the movement of cholesterol into the mitochondria for its conversion to pregnenolone (Stocco and Clark, 1996; Miller, 1997). Cholesterol is converted to pregnenolone by the mitochondrial cholesterol side-chain cleavage system, catalysed by a mitochondrial cytochrome P450 enzyme; P450scc (scc refers to **s**ide

chain **c**leavage). The human adrenal cortex may now direct this pregnenolone toward one of three different pathways. When P450c17 is absent in steroidogenic cells such as the adrenal zona glomerulosa, pregnenolone is converted to progesterone and further converted to other C-21,17-deoxysteroids, leading down the mineralocorticoid pathway toward aldosterone. In the human adrenal and gonad, when the 17α-hydroxylase activity of P450c17 is present, pregnenolone is converted to 17OH-pregnenolone as indicated by reaction I in Figure 1; and to a lesser extent, progesterone to 17OH-progesterone as shown in Figure 1, reaction II. These steroids can be converted to other C-21,17-hydroxysteroids and proceed to the principal glucocorticoid, cortisol. In the gonad and post-adrenarchal adrenal (adrenarche refers to the prepubertal rise in the secretion of adrenal androgens which occurs only in human beings and higher primates; Cutler *et al*, 1978), the 17,20-lyase activity of P450c17 converts 17-OH pregnenolone to the C19 steroid DHEA (as depicted by reaction III in Figure 1) which is the precursor of androgen biosynthesis. The conversion of 17OH-progesterone to androstenedione (reaction IV Figure 1) by the 17,20-lyase activity of human P450c17 does not occur very often (Lin *et al*, 1991; Lin *et al*, 1993), even though the catalysis of this reaction by P450c17 occurs readily in rodents and cattle.

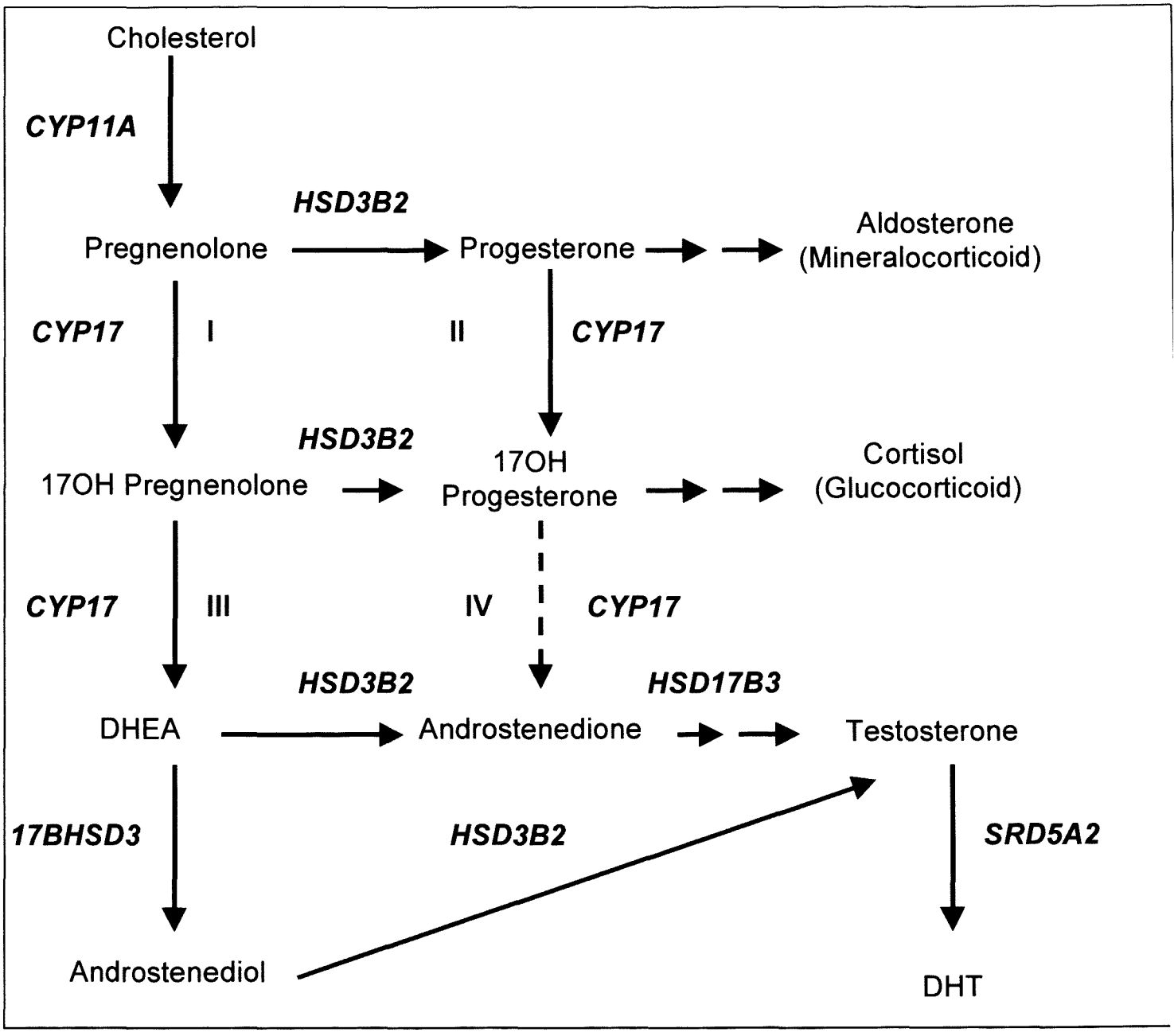


Fig. 1. The pathways of steroid hormone biosynthesis. Initiation of synthesis of all steroid hormones occurs by the conversion of cholesterol to pregnenolone by the cholesterol side-chain cleavage enzyme, mitochondrial cytochrome P450scc. Pregnenolone may be converted to 17OH-pregnenolone by 17α-hydroxylation. The adrenal and gonad may cleave the steroid C17,20 bond and convert 17OH-pregnenolone to DHEA which is the precursor of all sex steroids.

1.2.5 Molecular mechanism of androgens

In the bloodstream, testosterone circulates either free or is bound to specific binding proteins such as SHBG[4], or even more loosely bound to albumin (Randall, 1994). Testosterone enters prostate cells freely, down an activity gradient by a passive diffusion mechanism (Wilson JD, 1992) and is rapidly converted to its metabolically more active form, DHT, via cytoplasmic 5α-reductase. These androgens diffuse into the nucleus where DHT and to a lesser extent, testosterone, bind to the androgen receptor (AR) as depicted in Figure 2. AR ligand interactions also occur in the cytoplasm of target cells, leading to dimerisation and subsequent translocation to the nucleus. Although testosterone and DHT bind to the same androgen receptor protein, DHT has a much higher affinity for the receptor (Grino *et al*, 1990) based on the decreased rate of dissociation of the DHT-receptor complex resulting in an increased efficiency in the stability of the binding. In contrast with DHT, testosterone has an accelerated off-time and at equimolar concentrations is only 33% as active as DHT (Grover and Odell, 1975).

Only a small amount of DHT re-enters the plasma while a larger proportion is converted to androsterone or androstenediol that circulate as glucuronide conjugates (Lookingbill *et al*, 1991). The AR has a relatively low affinity for adrenal androgens such as DHEA and androstenedione (Culig *et al*, 1993) and for non-androgenic steroids like progesterone and estradiol (Wilson and French, 1976).

[4] A serum protein which demonstrates a high affinity for androgenic steroid hormones

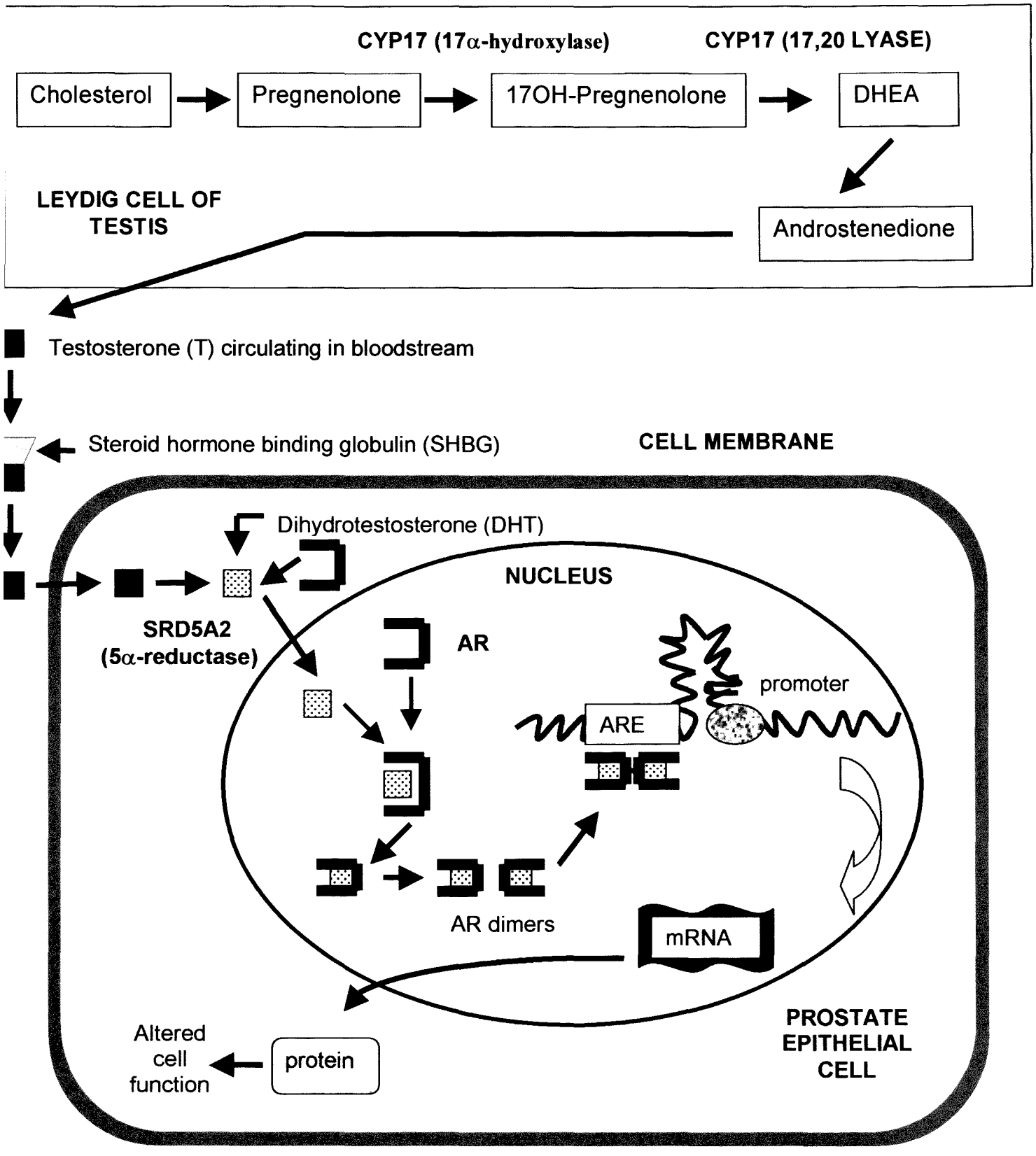


Fig.2. Diagrammatic representation of the model for the mechanism of androgen action. T or its metabolite, DHT, binds to specific androgen receptors (ARs) in cell nuclei. The activated AR-DHT complex binds to androgen response elements (AREs) of target genes, triggering the appropriate alteration in mRNA and protein synthesis.

Binding of androgen-ligands to the AR, activates the receptor and induces an allosteric conformational change of the protein which results in hormone-dependant, DNA-independent phosphorylation of the receptor at several sites and dissociation from receptor-associated heatshock proteins (HSPs) (Carson-Jurica *et al*, 1990; Housley *et al*, 1990; Smith and Toft, 1993). The mechanisms of receptor activation are quite complex and could be influenced by local growth factors, neurotransmitters or vitamins (O'Malley and Tsai, 1992). Removal of HSPs allows for AR dimerization with another bound AR, which in turn unmasks the highly conserved region of the DNA-binding domain that is concealed in the inactive receptor. The binding domain takes the form of zinc fingers which are able to interdigitate with the DNA helix, binding to specific palindromic[5] androgen response elements (AREs) on androgen-regulated target genes and their flanking DNA (Freedman, 1992; Wong *et al*, 1993). Binding of AR homodimers to DNA is associated with hyperphosphorylation as a result of activation of hormone-dependant, DNA-dependant phosphokinase. The hyperphosphorylated AR-ARE complex interacts with the preinitiation complex in the promoter region of a regulated gene. The preinitiation complex consists of several converging *cis* and *trans*-factors and enzymes, including RNA polymerase II, which is involved in gene transcription. It is thought that the upstream AR-ARE complex associates with the preinitiation complex via DNA loops thus stabilizing the promoter and increasing the rate of gene transcription. The appropriate genes in the cell are switched on or off to alter the mRNA and therefore the proteins produced by the cell, as shown in Figure 2 (O'Malley *et al*, 1991). Accessory factors are recruited into this complex by steroid receptors and include coactivator and corepressor proteins which function as signalling intermediates between receptors and the general transcription

[5] Palindromes are nucleotide sequences in which the 5' to 3' sequence of one strand of a segment of DNA is the same as that of its complementary strand

machinery. Some transcription factors such as SRC-120 and TIF-1 (Le Douarin *et al*, 1995) seem to interact with most steroid receptors, while others are thought to be specific for certain steroid receptors. ARA70 has been the first identified ligand-dependant-associated protein for the AR and shows preference for the AR by functioning as a specific coactivator for enhancing receptor-dependent transcriptional activity (Yeh and Chang, 1996).

1.2.5.1 Genes regulated by androgens

Androgen-regulated genes are thought to have hormone response elements (HREs) in their first exon or 5' flanking region which interact with AR-dimers (Adler *et al*, 1992; Roche *et al*, 1992). If functional HREs are not palindromic, they exist as repeats or inverted repeats which contain variable spacer base pairs between repeat sequences that are important for determining specificity. There are two groups of HREs; those that aid thyroid, retinoic acid and estrogen responses and those which regulate the responses of glucocorticoids, progesterone and androgens. The aforementioned class of HREs are called glucocorticoid receptor elements (GRE); TGTTCT, as glucocorticoid receptors transduce the greatest response in tyrosine aminotransferase (*TAT*) genes and in the mouse mammary tumour virus (*MMTV*). A consensus sequence described as a 15-bp element of 2 imperfect indirect repeats with a 3-bp spacer (Roche *et al*, 1992), has been derived for the AR and is quite similar to the GRE. Two AREs were identified in the 5' flanking region of the androgen-regulated probasin gene and it has been shown that they have a high AR specificity (Rennie *et al*, 1993). Even though the first ARE was similar to the GRE, the second ARE was distinct. Together, these AREs mediate the androgen response. It was further shown that tissue specific accessory factors or so called

trans-acting factors, could be important in determining the binding specificity to HREs/ AREs (Rennie *et al*, 1993). Specific upstream *cis*-acting elements have been identified on androgen-responsive genes which regulate androgen responsiveness (Adler *et al*, 1992; Supakar *et al*, 1993). This includes one which interacts with a nonreceptor *trans*-acting protein (Adler *et al*, 1992).

The human prostate-specific kallikreins, human glandular kallikrein-1 (*hKLK2*) and prostate-specific antigen[6], members of a subgroup of serine proteases (Watt *et al*, 1986; Schedlich *et al*, 1987; Evans *et al*, 1988; Lundwall 1989), have been shown to be regulated by androgens on interaction with the AR-complex. There is about 80% homology between PSA and *hKLK2* in the 5' flanking region from -300 to -1 with respect to the cap site (site of initiation of transcription). The promoters of both genes contain AREs in their 5' regions, and it is believed that androgens directly regulate their transcription rates (Wolf *et al*, 1992; Klee *et al*, 1999). The PSA protein is well characterised and is now widely used as a serum marker for human CaP (Wang *et al*, 1981). New studies also demonstrate the utility of human kallikrein as an independent CaP marker (Parker *et al*, 1988).

1.2.6 Genes in the Androgen Metabolism Pathway

Due to the androgen dependence of prostate carcinoma (Huggins and Hodges, 1941) and the assumption that variations in androgen metabolism and synthesis could affect a person's risk, genes that encode enzymes involved in the androgen pathway are considered ideal candidates for CaP susceptibility. It has also been

[6] PSA has been used classically in clinical studies, so the term is used interchangeably with *hKLK3*

suggested that a cascade of genetic alterations, modulated by epigenetic events[7], cause the initiation and progressive growth of prostate tumours.

The genetic control of androgen biosynthesis, transport and metabolism is a very complex process, which directly or indirectly involves hundreds of genes. A few such genes in the androgen metabolic pathway have been explored in this study and have been selected based on criteria suggested by Ross and colleagues (Ross *et al*, 1998): The products of the genes involved should possibly play a critical role in androgen stimulation of the prostatic epithelium, polymorphic genetic markers must be known and it must be evident that the various polymorphic alleles have functional consequences or are linked to loci with functional relevance. Since race-ethnicity is such a powerful feature in CaP, this group also suggested that interracial allelic variation of the polymorphic markers among the populations at different underlying risk of CaP should occur in a way predicted by the functional studies of the polymorphisms. Multiple functional polymorphisms could also be present in the same gene, so an individual could carry high-risk and low-risk markers and have an overall risk level that is the same as the general population.

The genes of interest are listed in Table 1 together with the main function of their protein products.

[7] Refers to any factor that can affect the phenotype without change in the genotype

Table 1. Target genes possibly involved in prostate carcinogenesis in the androgen metabolism pathway

Target gene	Function
Cytochrome p450c17α (*CYP17*)	Testosterone biosynthesis
Cytochrome p450 3A4 (*CYP3A4*)	Testosterone metabolism
Steroid 5-α-reductase type II (*SRD5A2*)	Testosterone bioactivation
Androgen receptor (*AR*)	Androgen transport and transcriptional activation

The major metabolic androgen pathway relating to the prostate and the relevant sites of activity of the gene products of interest is illustrated in Figure 2.

The cytochrome p450c17α (*CYP17*) gene encodes the enzyme that regulates critical steps in testosterone biosynthesis, the cytochrome p450 3A4 (*CYP3A4)* gene is involved in the biotransformation of testosterone to 2β-, 6β-, or 15β-hydroxytestosterone. The steroid 5α-reductase type II gene (*SRD5A2*) encodes the enzyme responsible for converting testosterone to the metabolically more active DHT, and the androgen receptor (*AR*) gene encodes the androgen receptor allowing for androgen binding and transport, DNA binding and transactivation of androgen-regulated target genes.

1.2.6.1 The androgen receptor: structure and function

After the discovery was made that uterine cells preferably retain radiolabeled estrogen over other steroid molecules (Toft and Gorsky, 1966; Jensen *et al*, 1968), the concept of intracellular steroid-specific binding proteins or receptors was introduced. Shortly afterwards, other binding proteins for steroid ligands were found, which include those for progesterones, glucocorticoids and androgens (Baxter and

Tomkins, 1971; Belsham *et al*, 1995). The AR was first described in 1969 (Fang *et al*, 1969). The gene was subsequently cloned by several independent groups (Lubahn *et al*, 1988; Trapman *et al*, 1988; Tilley *et al*, 1989) and has since been studied extensively in attempts to elucidate its role in the development and progression of CaP.

Southern blot analysis revealed that the human *AR* gene is a single-copy gene, localised to the centromeric region on the X-chromosome at band q11-12 (Brown *et al*, 1989). The entire gene spans a total length of 75-90kb and produces a messenger ribonucleic acid (mRNA) transcript of approximately 11kb. Northern analysis of human and rodent tissues has demonstrated two AR mRNA species of 8.5 and 11kb, respectively, in several cell lines resulting from differential processing of precursor mRNA (Lubahn *et al*, 1988; Faber *et al*, 1991; Wolf *et al*, 1993). The different sizes of the mRNA species and the translational open reading frame[8] is due to the presence of variable untranslated 5' and 3' sequences (Refer to Section 1.2.6.1, describing the NH_2-terminal domain). In the human prostate and in genital skin fibroblasts, the 11kb size mRNA is predominantly expressed. The *AR* gene is divided into 8 exons, separated by introns of up to 26kb in size (Kuiper *et al*, 1989). The coding region is spread over an ORF of almost 2760 nucleotides. Exon 1 is a relatively large 1,613 bp and contains two polymorphic trinucleotide repeat segments. The structural organisation of the AR gene is essentially identical to those of genes encoding for other steroid hormone receptors as the exon/intron boundaries are highly conserved (Kuiper *et al*, 1989; Lubahn *et al*, 1989; Zong *et al*, 1990; Keaveney *et al*, 1991). The structure of the *AR* gene is illustrated in Figure 3. Two transcription

[8] ORF: the interval between the start and stop codon that encodes amino acids for insertion into polypeptide chains; codons: nucleotide triplet encoding for a single specific amino acid

initiation sites are found in a 13-bp region in the promoter area (Tilley *et al*, 1990; Faber *et al*, 1991). Homologous down-regulation of *AR* gene expression occurs in the prostate but an ARE has not been identified in the promoter region. The promoter does not contain the classic TATA or CCAT boxes but rather has a GC box that could play a role in initiation of transcription (Faber *et al*, 1993). The presence and functional involvement of an SP1 site has, however, been reported.

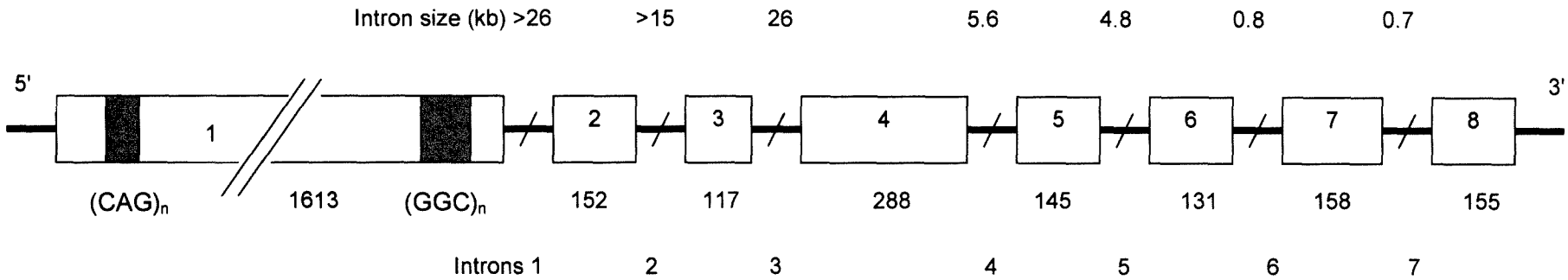


Fig.3. The *AR* gene spans 75-90kb of genomic DNA. The eight exons (numbered within boxes) with their sizes in base pairs (drawn to a scale of 10mm = 100bp) shown directly below the boxes, are separated by introns (numbered below the diagram) with sizes ranging from 0.7 kb to more than 26 kb. Exon 1, which is relatively large, contains two polymorphic trinucleotide repeats, viz. CAG and GGC tracts that vary in length within and across racial/ethnic populations.

A 50-bp purine-rich region situated 70 bp upstream from the transcription start site has been described as the most important *cis*-acting element for *AR* gene transcription (Mizokami *et al*, 1994). In addition, a cyclic adenosine monophosphate (cAMP) responsive element, AR/CRE, has been identified upstream, as well as several potential *cis*-acting elements, including retinoic acid response element (RARE) and AP-1; implying that the AR gene expression could be regulated by various mechanisms.

The *AR* gene codes for a ligand-activated nuclear transcription factor that mediates the action of steroid hormones, namely the androgen receptor (AR) (Beato, 1989). It belongs to a large group of proteins with regions of highly conserved structural homology and includes specific receptors for oestrogen, progesterone, glucocorticoid, mineralocorticoids, vitamin D, thyroid hormone, retinoic acid as well as orphan receptors[9] (Trapman and Brinkmann, 1996; Trapman and Cleutjens, 1997; Prins, 2000). The *AR* ORF codes for a 910-919 amino acid peptide with a calculated molecular mass of 110 kD (Trapman *et al*, 1988; Faber *et al*, 1989) (The variable length of the protein is explained further in the text when the domains of the AR are described). This primary structure of the AR is depicted in Figure 4. The protein is nearly ubiquitous, as only bone marrow and the spleen show complete absence of AR expression (Lindzey *et al*, 1994; Taplin *et al*, 1995).

[9] The hormone or ligand required for functional activation has not been identified for some receptors; thus the term orphan receptors

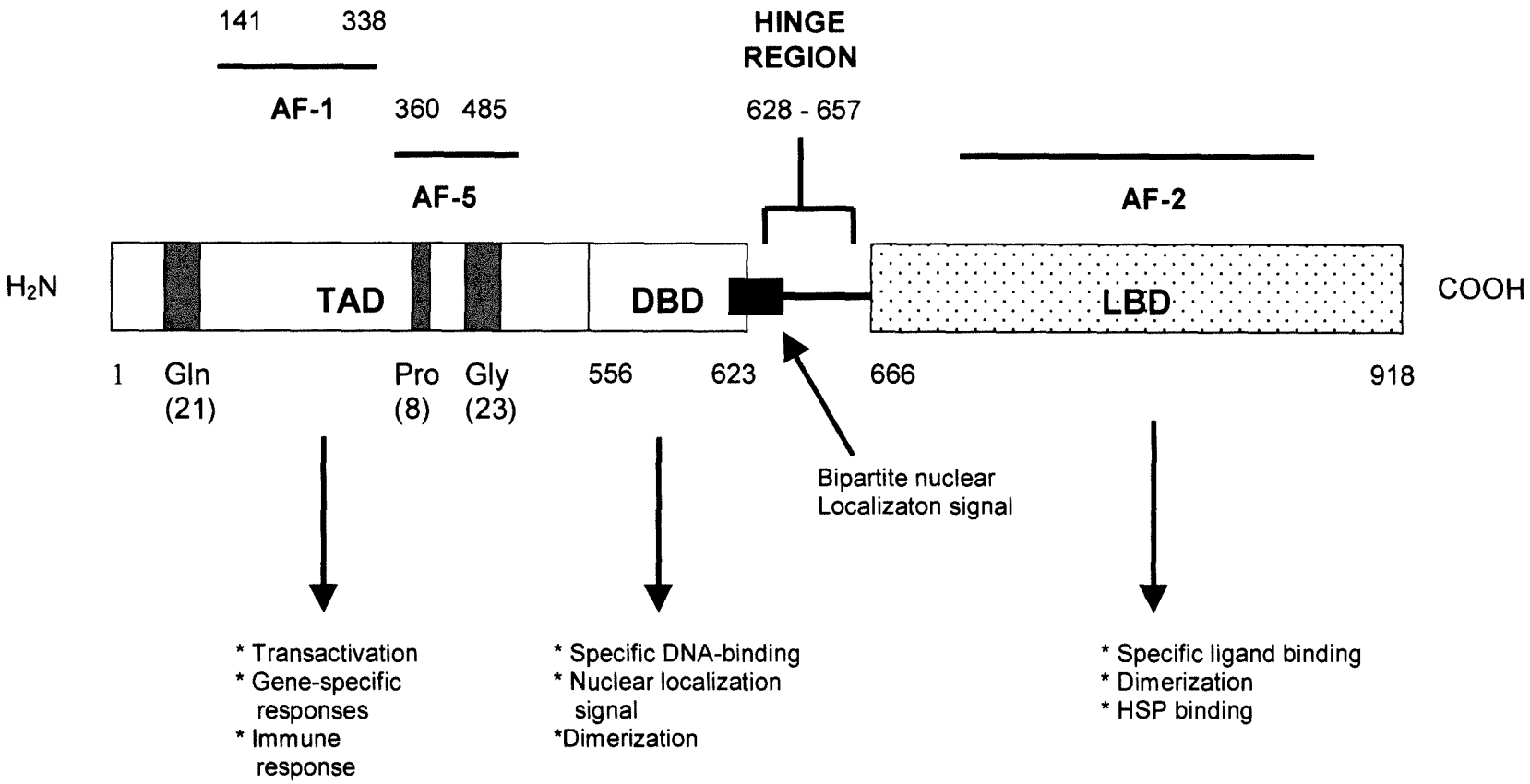


Fig.4. Schematical representation of the primary structure of the human AR with its different functional domains. Between the DBD and the LBD is the hinge region, an area with low sequence homology between the AR and other steroid receptors, which contains the major part of the AR nuclear targeting signal. Grey blocks with the corresponding residue and number of amino acids below indicate the polyamino acid repeats in the TAD. Regions representing the transactivating capacity of the AR in the NH_2 -terminal and COOH-terminal domains are shown by black lines. The rest of the numbers indicate the amino acid positions of the various functional and structural domains.

The AR was first characterised based on its physicochemical properties such as the affinity for binding testosterone or DHT (expressed as Kd), specificity for binding various androgens, sedimentation coefficient, thermolability and DNA binding. Activation of receptors by androgens changes the structural characteristics of the receptor from a heteromeric structure[10] that forms complexes with other proteins and does not bind to DNA, to a homodimer[11] with a capacity to bind to specific DNA sequences (Carson-Jurica *et al*, 1990; Brinkmann *et al*, 1992; Truss and Beato, 1993). Based on these functional studies of the AR protein in cell extracts, researchers have found that the AR is structurally divided into four domains (as shown in Figure 4): the amino (NH_2)-terminal transcription activation (transactivation) domain (TAD) encoded by a single exon (exon 1), the central DNA-binding domain (DBD) encoded by exons 2 and 3, a hinge region encoded by the 5' portion of exon 4 and the carboxy (COOH)-terminal ligand-binding domain (LBD) encoded by sequences in exons 4 to 8 (Faber *et al*, 1989; Kuiper *et al*, 1989; Lubahn *et al*, 1989; O'Malley, 1990; Cude *et al*, 1999). The domains are independently folded structural modules linked by extended polypeptides. Three-dimensional structures for the DBDs or LBDs for a number of these receptors in the superfamily were determined by X-ray crystallography (Luisi *et al*, 1991) or nuclear magnetic resonance (Hard *et al*, 1990), and reveal a folding pattern that is common across the family as predicted by the conserved amino acid sequences. The ligands for these steroid hormones are also closely related in structure, however, individual receptors distinguish ligands with stereochemically specific recognition and accurately activate specific genes in response to the binding of selected ligands.

[10] Composed of different molecules
[11] 2 identical molecules joined together

Unbound AR can be found distributed within the cytoplasm and nucleus of prostate cells (Thompson and Chung, 1984; Brinkmann and Trapman, 1992; Zhuang *et al*, 1992; Paris *et al*, 1994; Bourguet *et al*, 1995; Trapman and Brinkmann, 1996). Androgens bind to the COOH-terminus in the steroid binding pocket of the AR. After ligand binding, the DBD and a lysine-rich nuclear localisation region, located between the DBD and LBD, are exposed (Veldscholte *et al*, 1992; Ikonen *et al*, 1994; Zhou *et al*, 1994; Beato *et al*, 1996; Trapman and Brinkmann, 1996; Culig *et al*, 1998). Activated AR's then dimerise and translocate to the nucleus (Beato *et al*, 1996). The DNA binding domains of AR dimers include two "zinc fingers" that recognise and bind to specific sequences known as AREs in the promoter region of target genes; the first zinc finger stabilizes the receptor-DNA interaction (Huggins and Hodges, 1941). At the promoter site, the NH_2-terminal and steroid binding domains of the receptors interact with other transcription factors and cause either an increase or decrease in the transcription rate of the gene (Fronsdal *et al*, 1994; Gao *et al*, 1996; Yeh and Chang, 1996). This process of androgen receptor activation leading to the eventual alteration of transcriptional activity of target genes has previously been described in more detail in Chapter 1.2 (Molecular mechanism of androgen).

The large TAD comprises more than half of the AR protein and has a length of over 500 amino acids. Of the three major domains shared by nuclear receptors, this region exhibits the least conserved sequences and as result represents the most immunogenic part of the protein (Wilson *et al*, 1992). The NH_2-terminus contains the sequence information to optimise the receptor's transactivation capability (Jenster *et al*, 1991; Simental *et al*, 1991) and to specify gene recognition in the regulation of transcription (Jenster *et al*, 1992). This TAD, has been assumed to interact directly

or indirectly (mediated by co-activators) with other specific and general transcription factors. It has also been supposed that an interaction takes place between the NH_2-terminal domain and protein components of the chromatin complex. The experimental evidence for these hypotheses is however, still lacking. Deletion mapping of the NH_2-terminal domain has been performed to identify the regions necessary for transactivation function. Almost the whole NH_2-terminal region was found to be essential for maximum transactivating activity (Jenster *et al*, 1991; Jenster *et al*, 1995; Simental *et al*, 1991). The boundaries of the NH_2-terminal transactivation domain have not exactly been specified but a core segment (AF-1 in Figure 4), with over 50% activity, has been defined in the region between amino acids 141 to 338 in the full length AR. In the absence of ligand, the AR and therefore this domain, is inactive. This is not due to the cytoplasmic location of the receptor because a truncated AR without the complete LBD, which is constitutively active, also has almost no transactivating activity. In this artificial construct, the region from amino acid 360 to 485 (designated AF-5 in Figure 4) is responsible for most of the transactivational activity in this constitutively active AR mutant, lacking the LBD (Jenster *et al*, 1995). The finding of two separate transactivating domains and the observation that their functionality depends on the presence or absence of the LBD suggests an interaction between the two domains. In fact, recent *in-vivo* protein-protein interaction studies with separate NH_2-terminal and ligand-binding domains demonstrated that such a ligand-dependent, functional interaction is possible (Wong *et al*, 1993; Langley *et al*, 1995; Doesburg *et al*, 1997). It is however, unknown whether these interactions can take place in the full length receptor, whether interactions are direct or indirect (via one or more bridging factors), and whether these proposed interactions take place between different parts of one receptor

protein (intramolecular) or between two receptor proteins (intermolecular) in the full length AR. The former model is possible if the proteins' structural flexibility allows such an interaction. It also has to be established which parts of the domain form the protein-protein interphase. In the latter case, the interaction will assist in the formation of AR homodimers. It is not clear whether interactions can take place between two LBDs during AR dimerisation, as has been confirmed in the case of the estrogen receptor (Fawell *et al*, 1990). The NH_2-terminus may also contribute to the three-dimensional structure of the receptor molecule through interactions with other regions of the protein (Evans, 1988).

The NH_2-terminal domain of the AR receptor is characterised by a relatively high amount of acidic amino acids between residues 100 and 325 which results in a nett negative charge at this site that could be important in transcription regulation (Faber *et al*, 1989), as acidic domains in yeast factors have been found to be involved in the regulation of transcription (Ptashne, 1988). This domain of the receptor is unique among the nuclear receptor superfamily as it contains several homopolymeric amino acid stretches. These sequences include a polyglutamine tract, a polyglycine repeat and a polyproline tract (Lubahn *et al*, 1988b). The most NH_2-terminal of these is the polyglutamine stretch (~amino acids 58-79) encoded by a polymorphic CAG triplet repeat region (Sleddens *et al*, 1992), containing an average of 21 $\pm$ 2 glutamine residues (La Spada *et al*, 1991). A similar polyglutamine tract is also present in the glucocorticoid receptor and is found in the rat AR (Choong *et al*, 1998). Further downstream is a shorter stretch of eight proline residues, situated approximately at amino acids 372-379, that does not vary in size. The polymorphic polyglycine region of about 23 residues (~amino acids 449-472) is closer to the DBD and is encoded by

GGC repeats. Due to the polymorphic CAG and GGC repeats, the length of the AR may vary between 900 and 930 amino acids, leading to confusion concerning the numbering of the individual amino acid residues. Therefore the AR has been described as having amino acid positions which correspond to a receptor of 918 amino acid residues as indicated in Figure 4. The biological functions of these polymorphic motifs of the transactivation domain is not known as yet, but evidence from studies of other transcription factors show that particularly the polyproline and polyglutamine tracts could play an important role in regulation of transcription via protein-protein interactions with other transcription factors. Most proteins in which similar repeats were identified are involved in regulation of gene expression and/or developmental control (Haynes *et al*, 1987; Burglin, 1988). The CAG repeat is especially extremely polymorphic and offers the possibility of linkage studies in families with supposed defects in AR function. Variation in the length of this repeat can also be used as a marker for the *AR* gene as well as an X-chromosome marker (Boehmer *et al*, 1997; Li *et al*, 1998). Based on *in vitro* experiments after transient transfection of AR cDNA's which contain different lengths of CAG repeats (Jenster *et al*, 1994; Kazemi-Esfarjani *et al*, 1995), differences in repeat length (8-35 glutamine residues) were observed in the normal population and have been implicated to be associated with a mild modulation of AR activity (Nance, 1997; Buchanan *et al*, 2001)).

DNA binding domain

The DBD of the AR reveals approximately 80% homology with the corresponding domains in the mineralocorticoid receptor (MR) and progesterone receptor (PR). The 66 amino acid DBD is rich in cysteine, lysine and arginine, and was predicted to

associate with DNA based on its similarity to regions within known DNA binding proteins e.g. TFIIIa (Weinberger *et al*, 1985). Detailed crystallographic information is also available of the DBD of the estrogen and glucocorticoid receptors complexed with DNA (Luisi *et al*, 1991). In the glucocorticoid receptor (GR) and estrogen receptor (ER) the DBD region is arranged as a pair of loop structures folded to form a single structural unit made up of two zinc-binding motifs (Freedman, 1992; Schwabe *et al*, 1993). Central in the structure of the DBD are four cysteine residues, invariably present in all steroid receptors, which co-ordinately bind a zinc ion, tetrahedrically, in each of the two zinc clusters (commonly referred to as zinc fingers) (Evans, 1988). These structures are highly conserved among all members of this gene superfamily. By analogy with other steroid receptors, it is likely that an α-helix structure, if formed by the last part of the NH_2-terminal zinc finger, directly contacts the bases in the ARE in the DNA (Green *et al*, 1988; Berg, 1989; Freedman, 1992; Truss and Beato, 1993). Nucleotides at the distal end of the first zinc finger encode three amino acids (glycine577, serine578, valine581) referred to as the "P box", which is supposed to be important in specific recognition of the ARE. These groups of amino acids are conserved between the AR, PR, GR and MR. Therefore the homodimers of the four receptors recognise the same, imperfect, transcriptional enhancer palindromic sequence (GGA/TACAnnnTGTTCT) in or near target genes, in the 5' flanking region. This consensus sequence is a high affinity binding site for these receptors (Nordeen *et al*, 1990; Roche *et al*, 1992; Lieberman *et al*, 1993; Kallio *et al*, 1994). In the regulatory regions of natural target genes, however, the sequence of the binding site might differ considerably from the consensus sequence. Nucleotides at the proximal end of the second zinc-finger, which is highly basic, encodes a region known as the "D box" that facilitates co-operative binding of a receptor homodimer to an ARE by

protein-protein interactions with the DNA phosphate backbone, determining the spacing of the two ARE half-sites and also mediating dimerisation (Green *et al*, 1988; Berg, 1989; Luisi *et al*, 1991; Wong *et al*, 1993). At the carboxy-terminal end of this region of the AR is an arginine/lysine pair that forms part of the nuclear targeting sequence as described (Zhou *et al*, 1994) under the heading of "The hinge region".

The hinge region

In the absence of its ligand, the AR is distributed over the cytoplasm and the nucleus and addition of androgen results in a rapid migration from the cytoplasm to its site of action in the nucleus (Jenster *et al*, 1991; Zhou *et al*, 1994). This nuclear translocation depends on the presence of an intact bipartite nuclear localisation signal situated in the region between the DBD and the LBD. A low sequence homology of this region known as the hinge exists between the AR and other steroid receptors. The main part of the AR nuclear targeting signal consists of a cluster of basic residues at positions 628-657.

The ligand binding domain (carboxy terminus)

The COOH-terminal third of the AR is predominantly hydrophobic and encompasses approximately 250 amino acid residues, containing the receptor's ligand binding function. This region displays about 50% identity with the corresponding residues in PR, MR and GR. There are, however, short stretches that have between 65% and 100% similarity. The integrity of the entire domain is important for the specific, high integrity binding of androgens as most deletions or point mutations in this region suppress proper ligand binding (Jenster *et al*, 1991; Simental *et al*, 1991; Quigley *et*

al, 1995). A part of the LBD is also important for receptor dimerisation (Fawell *et al*, 1990; Doesberg *et al*, 1997).

Elucidation of the 3D crystallographic structure of the LBD of different retinoic acid receptors (9-*cis*-retinoic acid receptor alpha — RXRα and all-*trans* retinoic acid receptor gamma — RARγ), thyroid hormone receptor alpha (TRα), estrogen receptor alpha (ERα) and PR reveals that a common tertiary structure with a variable number (10-12, depending on the type of receptor) of α-helices and an antiparallel β sheet arranged in a helical sandwich are involved in the formation of an interaction surface for transcriptional coactivators, the transcriptional intermediary factors (TIFs) (Bourguet *et al*, 1995; Renaud *et al*, 1995; Wagner *et al*, 1995; Brzozowski *et al*, 1997; Williams and Sigler, 1998; Feng *et al*, 1998). Despite differences in structure and the number of α-helices, a high degree of homology is observed in the 3D structure of the LBDs of the TRα, ERα and PR. Predominantly involved in the formation of the hydrophobic binding pocket are the β-turn and helices 3, 5, 7, 11 and 12. Amino acid residues in these helices are in contact with the ligand. An interaction surface is formed with ligand binding, allowing interactions with other proteins (e.g., TIFs).

Various nuclear receptors contain a ligand-dependent transactivation function (AF2) in their LBDs (Danielian *et al*, 1992; Barettino *et al*, 1994; Gronemeyer and Laudet, 1995). In this AF2 region, an independent activating domain, AF-2 AD, which is conserved among many nuclear receptors, is found in the COOH-terminal part of the LBD (Wurtz *et al*, 1996). A core region in the AF-2 AD, situated in helix 12, seemed to be important for the hormone-independent interaction with TIFs and transcriptional

activity (Danielian *et al*, 1992; Barettino *et al*, 1994; Lanz and Rusconi, 1994, Montano *et al*, 1996; Berrevoets *et al*, 1998; Feng *et al*, 1998). The transcriptional activity of a broad range of nuclear receptors can be modulated by these TIFs (Cavailles *et al*, 1995; LeDouarin *et al*, 1995; Onate *et al*, 1995; Vom Baur *et al*, 1996; Voegel *et al*, 1996; Yeh and Chang, 1996; Berrevoets *et al*, 1998). The *in vitro* association of the receptor with these TIFs can be destroyed by mutations in the AF2 AD core region. It has recently been found that TIFs display acetyltransferase activity and this has provided further understanding about the molecular events which occur at the chromatin level during the activation of transcription. A general mechanism for nuclear receptor activation was proposed by Wurtz *et al* (1996), in which the AF-2 AD core, present in helix 12, plays a central role in the formation of an interaction surface, allowing binding of TIFs to the LBD. The amino acid residues involved in this surface have been investigated and identified in the TR LBD, and it appears that the interaction surface contains a hydrophobic cleft (Feng *et al*, 1998). Residues located in helices 3, 4/5, 6 and 12 form the surface cleft. Formation of this surface occurs with ligand binding when the COOH-terminal alpha helix 12 is folded against the scaffold of the three other helices (Feng *et al*, 1998).

No information is available for the AR LBD with respect to 3D structure but based on the high homology in the various LBDs of the various nuclear receptors, it can be predicted to resemble the PR to a large extent (Trapman *et al*, 1988; Tanenbaum *et al*, 1998). A common 'mouse trap-like' mechanism of ligand binding has been proposed for the AR. Helix 12, which contains the AF-2 core sequences, folds back to the core region of the LBD, then comes into close contact with the ligand and seals the ligand pocket. The conformational change induced by hormone binding could

create an interaction surface that allows binding of transcriptional co-activators, like TIF-2 or its mouse homologue GRIP-1 and the AR-specific co-activator ARA70 (Hong *et al*, 1996; Voegel *et al*, 1996; Yeh and Chang, 1996).

A transcription activation function (AF2) has been identified in the LBD of the human AR even though it is very weak when compared with that found in other steroid receptors such as ERs and GRs (Hong *et al*, 1996; Voegel *et al*, 1996; Berrevoets *et al*, 1998). The AF2 domain can be activated in a hormone-dependent way and is strongly enhanced in a promoter dependent way by the co-activators TIF2 and GRIP1 (Hong *et al*, 1996; Voegel *et al*, 1996; Berrevoets *et al*, 1998). The boundaries of the AF2 domain in the AR LBD have not yet been determined but contain the core region as defined in the LBDs of several members of the ligand-dependent nuclear receptor family (as shown in Figure 4). This AF2 activation domain core region contains the conserved sequence 884-Glu-Met-Met-Ala-Glu-888. Mutations in this region can result in a decrease in activation capability without affecting the ligand-binding function, indicative that the amino acid residues of the AF2-AD core region are not directly involved in ligand binding, but are part of or determine the interaction surface. This presumption has been confirmed by studies on mutations in this region and the interaction of coactivators (Feng *et al*, 1998; Berrevoets *et al*, 1998).

Evidence has been presented for a possible interaction between the LBD and the AF functions in the NH_2-terminal domain as explained under the subheading, 'The trasactivational domain'. Investigation of this interaction in more detail revealed that only certain regions in the NH_2-terminal domain are involved in this interaction

(Langley *et al*, 1995, 1997; Doesberg *et al*, 1997; Berrevoets *et al*, 1998). The AF1 core region is not involved in this interaction, but about 33 amino acid residues before this transactivating region as well as the AF5 region are necessary for an effective functional interaction. By the substitution of an essential amino acid residue (Glu 888) for a glutamine residue, it has been found that amino acid residues: 884Glu-Met-Met-Ala-Glu888 in the AF2-AD core region of the COOH-terminal is involved in the interaction. The functional interaction of the AR LBD with TIF2 is affected by a similar sequence variant, which implies that TIF2 and the NH_2-terminal domain recognise the same interaction surface of the LBD with hormone binding (Berrevoets *et al*, 1998).

A large multi-protein complex consisting of heat shock proteins is associated with the unliganded LBD of the AR and maintains the receptor in a transcriptionally inactive conformation (Housley *et al*, 1990; Smith and Toft, 1993). Binding of hormones results in the dissociation of chaperone proteins from the receptor and initiates conformational changes in the LBD that are important for receptor dimerisation, DNA binding and interaction with transcription mediators. Diffusion of the receptor-ligand complex of the AR and androgens to the nucleus requires the nuclear localisation signal located in the hinge region between the DBD and the LBD.

Androgen binding to the AR results in two conformational changes of the receptor molecule (Kuil and Mulder, 1994; Kuil *et al*, 1995). Initially, a fragment of 35-kDa, which spans the complete LBD, and part of the hinge region, is protected by the ligand, but after prolonged incubation times a second conformational change occurs resulting in protection of a smaller fragment of 29-kDa. Antagonists inhibit the

biological effects of androgens and compete with agonists for binding to the receptor. In the presence of various antiandrogens such as cyproterone acetate, hydroxyflutamide and bicalutamide, only the 35-kDa fragment is protected and no smaller fragments are detectable upon longer incubations. The 35-kDa fragment is associated with an inactive conformation, whereas the second conformational change, which is only inducible by agonists and considered as the necessary step for transcription activation, is lacking upon binding of antiandrogens. Further analyses with specific antibodies against different epitopes in the 35- and 29-kDa fragments have shown that the most COOH-terminal end of the AR protein is represented by the 29-kDa fragment (Kuil *et al*, 1995). Other studies also reveal that antagonists induce a different conformational change upon binding to the LBD (Zeng *et al*, 1994; Kuil *et al*, 1995). This aberrant conformation leads to a receptor with a reduced or no transcriptional activity. Different antagonists could affect various aspects of receptor functioning like dissociation of heat shock proteins, dimerisation, DNA binding and interaction with general or specific transcription factors, with coactivators, and with the NH_2-terminal domain.

1.2.6.2 Cytochrome P450c17α

The human 17α-hydroxylase gene has been localised to a specific band, q24.3, (Matteson *et al*, 1986; Fan *et al*, 1992) on chromosome 10, spans 6569 bases and is divided into eight exons by seven introns (Picado-Leonard and Miller, 1987). This gene is the sole member of a unique gene family (P450XVII) within the P-450 supergene family (Picardo-Leonard and Miller, 1987; Kagimoto *et al*, 1988; Nelson *et al*, 1993).

Promoter function is driven by the SF1 protein and the 5' flanking region of the gene contains 3 functional SF1 elements that collectively mediate 25-fold or greater induction of promoter activity by SF1 (Hanley *et al*, 2001). Lin *et al.* (2001) studied the transcriptional regulation of *CYP17* and found that the most regulatory activity was confined to the first 227bp. In addition to this a TATA box[12], a steroidogenic factor-1 site, and 3 previously uncharacterised sites at -107/-85, at -178/-152 and at -220/-185 were identified. Further experimentation suggested that the -107/-85 site and the -178/-152 site bind members of the nuclear factor-1 (NF1) family of transcription factors. Mutations of both these sites reduced basal transcription by half. Assays also showed that the ubiquitous proteins Sp1 and Sp3 both bind to the -227/-184 region, and that mutation of their binding sites reduced transcription by 75%. Mutation of the Sp1/Sp3 site and the 2 NF1 sites eliminated almost all detectable transcription. The authors concluded that Sp1, Sp3 and NF1C proteins binding to their previously mentioned sites are crucial for adrenal transcription of the human gene for P450c17 (Lin *et al*, 2001).

CYP17 encodes a single protein of 508 amino acids, expressed in the adrenal cortex and gonads (Chung *et al*, 1987). This microsomal polypeptide, cytochrome P450 steroid 17α-hydroxylase (P450c17), catalyses both 17α-hydroxylase and 17,20 lyase activities. The dual function of the enzyme in the steroidogenic pathway has been described in more detail under the section — 1.2.4 Androgen biosynthesis.

[12] An 8-base-pair sequence conserved in most or all genes studied. It is located about 25 to 30 bases upstream from the site of transcription initiation and is composed only of thymine = adenine pairs. RNA polymerase ultimately binds here as part of the initiation complex of transcription

1.2.6.3 Steroid 5α-reductase type 2

SRD5A2 is located on chromosome 2 at 2p23 (Thigpen *et al*, 1993) and spans over 40kb of genomic DNA, with 5 exons and 4 introns (Labrie *et al*, 1992b; Russell *et al*, 1994). The gene translates into a protein of 254 amino acids (Labrie *et al*, 1992b). The steroid 5α-reductase type II enzyme product is transiently expressed in new-born skin and scalp, as well as in the liver. It is also found in foetal genital skin, the male accessory glands, and the prostate, including both BPH and adenocarcinoma of the prostate (Thigpen *et al*, 1993; Luu-The *et al*, 1994). The best–known role of this enzyme is the transformation of testosterone to the metabolically more active DHT, which in turn transactivates a number of genes involved in prostate development and growth (Coffey, 1988; Coffey, 1993). Increased DHT synthesis may be involved in the development of both benign and neoplastic growth of the prostate in elderly men (Labrie *et al*, 1992a) (See section under 1.2.4 Androgen metabolism for further explanation). It has been suggested that modulation of steroid 5α-reductase activity could be responsible for some variations in CaP risk among US ethnic groups (Ross *et al*, 1992).

1.3 GENETICS OF PROSTATE CANCER

1.3.1 Genetic progression of Prostate cancer

The molecular mechanisms fundamental to the development and progression of CaP are poorly understood. Transformation of a cell from a normal to a malignant phenotype and cancer progression depends on a series of genetic changes that results in activation of proto-oncogenes[13] and inactivation of tumour-suppressor

[13] Normal genes involved in cell division or proliferation; if activated by some mutational event they become involved in unregulated cell growth and proliferation responsible for tumour development

genes[14] normal genes involved in the regulation of cell growth. It has been estimated that 5-10 such genetic alterations which affect the expression and function of these critical genes, are required before a normal cell transforms to a fully malignant one. The majority of these events result from mechanisms like somatic mutations[15] and can also occur due to translocations[16] the transfer of a segment from one chromosome to another or amplification[17].

Development of comparative genomic hybridisation (CGH); allowing the screening of tumours for DNA sequence copy number changes, and loss of heterozygosity (LOH) studies, have identified several recurrent chromosomal alterations in CaP. CGH studies have shown that the loss of genetic material is an event which occurs early in CaP and is five times more common than DNA copy number gains or amplifications (Visakorpi *et al*, 1995b). This suggests that the inactivation of tumour suppressor genes is very important in the early stages of CaP. Hormone-independent CaPs often contain gains and amplifications, which implies that the late progression of CaP is also characterised by activation of oncogenes (Visakorpi *et al*, 1995a). Commonly occurring chromosomal aberrations include losses at chromosome regions 6q, 8p, 9p, 10q, 13q, 16q, 17p, 18p and 18q and gains at 7p, 7q, 8q and Xq (Cher *et al*, 1994; Joos *et al*, 1995; Visakorpi *et al*, 1995; Cher *et al*, 1996 and 1998; Nupponen *et al*, 1998; Sattler *et al*, 1999; Alers *et al*, 2000). Only a few target genes have, however, been associated with these alterations. Besides the *AR*, a number of other mutated genes have been found in CaP and includes *p53*, *Rb*, *CDKN2* and *ras* to mention a few. These genes and their genetic mechanisms are shown in Table 2.

---

[14] Normal genes involved in the regulation of cell growth
[15] Mutation occurring in a somatic cell
[16] The transfer of a segment from one chromosome to another
[17] The production of multiple copies of a sequence of DNA

1.3.1.1 *AR* gene amplification

With the aid of fluorescence *in situ* hybridisation (FISH), about 30% of hormone refractory prostate carcinomas were shown to contain *AR* gene amplification. About 20% of hormone refractory tumours contain additional copies of the X chromosome and the *AR* gene (Visakorpi *et al*, 1995; Koivisto *et al*, 1997). None of the tumours analysed prior to endocrine therapy for *AR* gene-copy number shows high-level amplification of the gene (Visakorpi *et al*, 1999; Koivisto *et al*, 1997; Bubendorf *et al*, 1999). So it seems that *AR* amplification is exclusively associated with regrowth of the tumour during endocrine therapy, when the supply of androgens is rate limiting and the selection of this genetic event is likely to provide an advantage for cell growth. It has been reported that similar examples of gene amplification lead to chemotherapy resistance. *AR* gene amplification is, however, the first example of such a mechanism that is related to hormone therapy, implying that castration could act as an equally strong selection force as chemotherapy.

It is believed that gene amplification can lead to an overexpression of the target gene of amplification (Alitalo and Schwab, 1986). Therefore, it has been suggested that amplification of the *AR* gene could result in overexpression, allowing the continued growth of androgen-dependent cancer cells even in the presence of low androgen concentrations left in serum after castration (Visakorpi *et al*, 1995; Kallionemi and Visakorpi, 1996). By using mRNA *in situ* hybridisation, it was in fact shown that *AR* amplification is associated with increased expression of the gene in hormone refractory prostate tumours (Koivisto *et al*, 1997). Even the refractory tumours, which do not contain amplification of the AR gene, seem to express *AR* quite strongly, suggesting that the AR signalling pathway could also be active in the nonamplified

cases. Uniform expression of the *AR* displays a clear difference between breast and prostate cancer, as hormone refractory breast cancer typically loses the expression of the estrogen and progesterone receptors (Kuukasjarvi *et al*, 1996). Little direct evidence is available in favour of a functional *AR* which is amplified and overexpressed in hormone refractory prostate tumours. Data exist which show that tumours with an amplified *AR* gene initially respond better to castration than tumours that do not acquire amplification, suggesting that these tumours are androgen sensitive (Koivisto *et al*, 1997). Also, preliminary studies show that there is a statistically significant improvement in progression-free and overall survival in patients with *AR* gene amplification who received the so-called maximal androgen blockade (MAB); which abolishes the effects of adrenal androgens, than patients without the amplification. This is suggestive of a functional AR signalling pathway (Palmberg *et al* 1997 and 2000).

Finding of this gene amplification illustrates how the genome-wide surveys by CGH provide a powerful tool to identify specific genetic alterations with an important role in cancer progression. Together with studies in classical cytogenetics, loss of heterozygosity (LOH) measurements and other techniques, these results are beginning to shape a model of the somatic genetic evolution of CaP (Figure 5). The precise sequence of genetic events in this process still remains to be determined and only a few of the genes in the affected chromosomal regions have been cloned.

1.3.1.2 *AR* mutations

In addition to *AR* gene amplification, the gene can also be altered by mutation. More than 300 different mutations of the *AR* gene have been reported so far, therefore making it the most often mutated transcription factor known (McPhaul *et al*, 1991, Quigley *et al*, 1995, Choong and Wilson, 1998; Gottlieb *et al*, 1998). The majority of these mutations are spread throughout the LBD, with some clustering in exons 5 and 7. This variability could represent as much as two-thirds of AR coding region mutations (Quigley *et al*, 1995). Most of these mutations have been identified in patients with varying degrees of Androgen Insensitivity Syndrome (AIS) (McPhaul *et al*, 1991; Quigley *et al*, 1995) but presently more than 60 different mutations have been reported in CaP (Gottlieb *et al*, 1998). Twenty-eight percent of these CaP mutations occur in a "hotspot" between codons 726 and 763, however, only 11% (2 of 17) of the mutations in this hotspot are at a CpG site and only 4 of the 17 mutations have been shown to be pathogenic (Culig *et al*, 1993; Elo *et al*, 1995; Taplin *et al*, 1995; James *et al*, 1997). The AR mutations present in patients with AIS are germline mutations (Quigley *et al*, 1995), (present in every cell in the body) and they are loss-of-function mutations as they inactivate receptor function. Mutations in the AR of CaPs are more often somatic mutations (present only in the tumour) (Gottlieb *et al*, 1998).

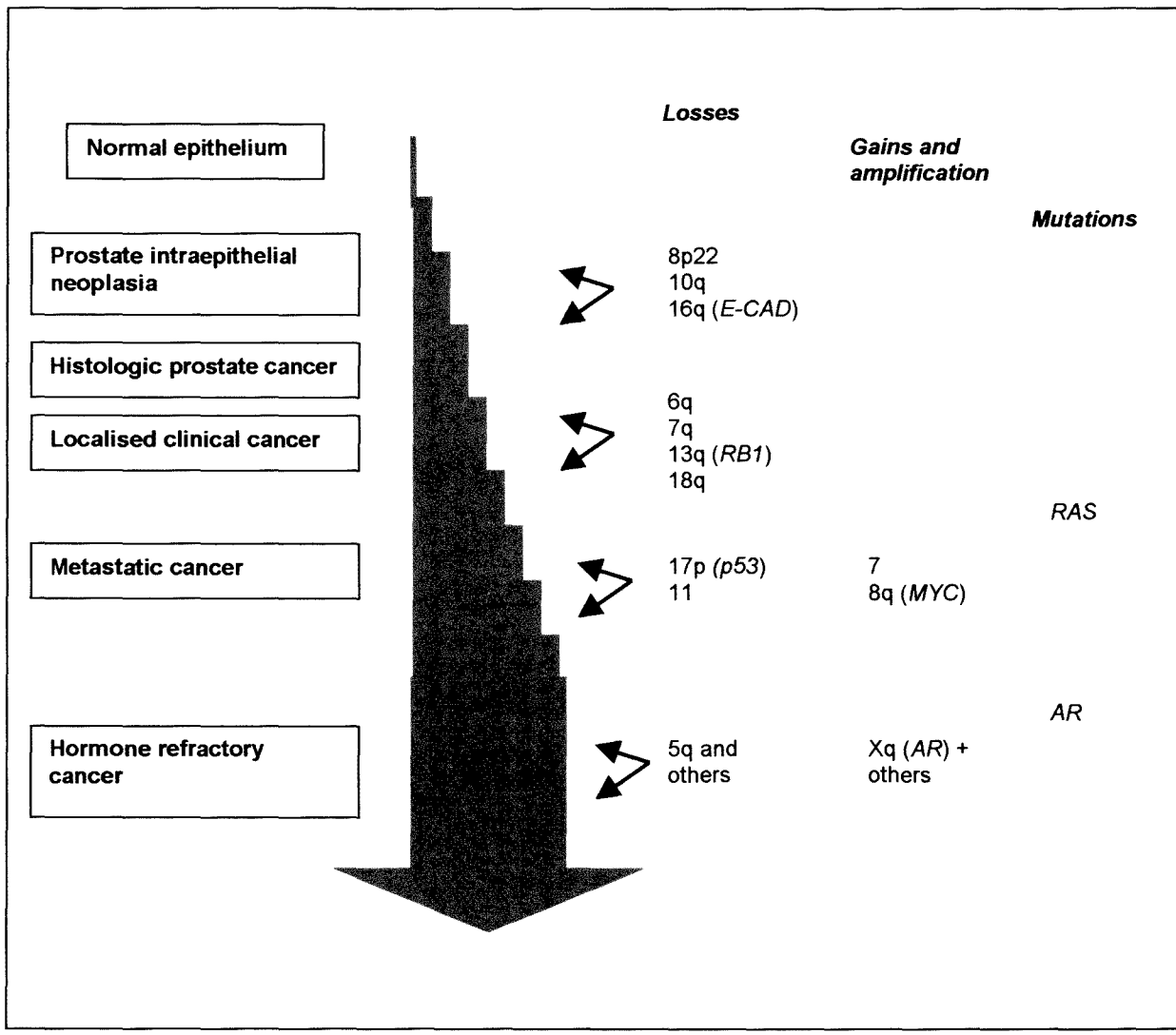


Fig.5. A schematic representation of somatic changes which are likely to underlie the step-wise progression of CaP. (Adapted from Visakorpi *et al*, 1995). The sketch is based on information from classical cytogenetics (Sandberg, 1992), CGH analyses (Cher *et al*, 1994 and Visakorpi *et al*, 1995b), allelic loss (Isaacs *et al*, 1994), chromosome transfection (Rinker-Schaeffer *et al*, 1994) and other molecular genetic studies (Macoska *et al*, 1993 and Isaacs *et al*, 1994). Affected chromosomal regions are shown followed by known candidate genes; in parenthesis. Each chromosomal region could also contain other genes that play an important role in tumour progression. The exact sequence of genetic events in this process still remains to be determined and only a small fraction of the genes in the affected chromosomal regions have been cloned. The thickness of the arrow in the middle of the diagram illustrates the accumulative number of vital genetic changes at each step of tumour progression, e.g. hormone refractory tumours on average contain twice as many genetic changes than untreated primary tumours (Visakorpi *et al*, 1995).

The first AR gene mutation found in CaP was in a well-characterised human prostate carcinoma cell line, lymph node prostate cancer (LNCaP). This single base mutation in the HBD (exon 8) converts a threonine to alanine and its position was reported by Harris *et al* (1990) to be codon 877 (ACT→GCT) of a 919 codon AR cDNA (Lubahn *et al*, 1988b and 1989; Harris *et al*, 1990), while Veldscholte and colleagues (1990) found the same mutation in codon 868 of a 910 codon AR cDNA (Jenster *et al*, 1991; Veldscholte *et al*, 1990); these are equivalent positions. Unlike AR gene mutations in AIS, the LNCaP mutation does not abolish receptor function. The LNCaP mutant binds androgen with an affinity which is similar to that of wild-type AR, and, like wild-type AR, is transcriptionally activated by androgen (Veldscholte *et al*, 1990). However, this mutant AR has altered steroid specificity compared to wild-type AR and can also be transcriptionally activated by estrogen, progesterone, or the antiandrogens cyproterone acetate or hydroxyflutamide (the active metabolite of flutamide) (Veldscholte *et al*, 1990). The same codon 877 mutation was found in 6/24 (25%) patients with advanced hormone-resistant prostate cancer (Gaddipati *et al*, 1994) and another study described the mutation in metastatic specimens of one patient (Suzuki *et al*, 1993), providing a strong support for the *in vivo* significance of this *AR* mutation.

The first somatic mutation in the *AR* gene in clinical CaP, occurring in the tumour before hormonal therapy, was reported in 1/26 specimens as a single base change found in codon 730, replacing valine with methionine (Newmark *et al*, 1992). This Val (GTG) → Met (ATG) mutation results in an altered binding specificity of the AR, similar to that seen in LNCaP cells (Culig *et al*, 1993).

Numerous mutations have been found in the HBD of advanced CaP (Culig *et al*, 1993; Gaddipati *et al*, 1994; Taplin *et al*, 1995; Bentley and Tilley, 1996). Culig and associates (1993) found a mutation in one hormone refractory tumour at codon 715 (Val $\rightarrow$ Met) that increased the efficacy with which progesterone and adrenal steroids like DHEA and androstenedione could activate transcriptional activity of the AR, allowing tumour cell proliferation at low levels of these circulating androgens. These findings suggest that the mutant receptor leads to a gain of function, rather than a loss.

Some studies report a relatively high frequency of AR mutations, while other investigations show a much lower frequency of mutation in clinical specimens; with the frequency of AR mutations varying between 0% and 50% in different reports (Newmark *et al*, 1992; Culig *et al*, 1993; Suzuki *et al*, 1993; Gaddipati *et al*, 1994; Ruizeveld *et al*, 1994; Schoenberg *et al*, 1994; Elo *et al*, 1995; Takahashi *et al*, 1995; Taplin *et al*, 1995; Evans *et al*, 1996; Suzuki *et al*, 1996; Paz *et al*, 1997). Discrepancies in methodology and patient sampling could account for this variation.

Marcelli and colleagues examined 137 samples of CaP from different stages in disease development, for the presence of AR mutations (Marcelli *et al*, 2000). Ninety-nine of these were specimens from localised cancer and 34 from local nodal metastases. Eighty-four of the 99 localised disease samples were microdissected to enrich for tumour cells. Analyses revealed no mutations in any of the samples of primary CaP, irrespective if microdissection was performed or not. In samples of locally metastatic cancer, however, 8/38 samples contained AR mutations. These analyses were performed on samples of genomic DNA from local nodal metastases,

before any therapy. The mutations were only found in samples that were enriched for tumour DNA by microdissection. Although functional analyses of these mutations have not been reported yet, many of the mutations are predicted to be inactive in assays of AR function, based on analyses of similar mutations that were identified and analysed from patients with various forms of androgen insensitivity.

In a study by Tilley and associates, all 8 exons of the *AR* gene were analysed from 25 CaP patients prior to hormonal therapy (Tilley *et al*, 1996). A high frequency of point mutations (44%), half of which were located in exon 1, were detected in the cancer samples. This could partly explain the differences in reported AR mutation frequencies, as most studies have not analysed exon 1, which comprises 58% of the coding sequence and encodes the entire transactivating domain of the receptor.

A smaller study was performed in Dallas to investigate the frequency of AR mutations in samples from patients with stage D cancer, in which tumour growth was androgen-independent. Results of these experiments, however, showed a very low frequency of AR mutations of less than 10%, even in these samples (Avila and McPhaul, unpublished observations). The outcomes of the majority of these studies suggest that the frequency of AR mutations is low in primary CaPs.

AR mutations have also been found in latent CaP, i.e., histological evidence of CaP. One such study found 18 of 74 (25%) inactivating mutations in the *AR* gene in the latent cancers of Japanese men, whereas no mutations were seen in 43 latent tumours from American men (Takahashi *et al*, 1995). The presence of these mutations causes the ablation of AR activity which precludes latent tumours from

progressing to clinically evident disease. Because the inactivating mutations occur so frequently, they could help to account for the lower rates of clinically evident CaP in Asian men than in other races.

In contrast to the relatively low frequency of AR mutations found in localised CaP, various other studies have suggested that a much higher incidence occurs in metastatic tumours. These studies have focused on the analyses of bone marrow samples from patients with stage D CaP.

A preliminary investigation by Taplin and co-workers found a high frequency of AR mutations in the metastatic deposits of patients (5/10) that have been treated with different forms of androgen ablation (Taplin *et al*, 1995). The cDNAs in this group of patients were analysed and shown to encode a number of different mutant ARs, ranging in frequency from 20-100%. In a later study, many of these mutations were shown to induce novel activation patterns on the mutant receptors (Fenton *et al*, 1997). Taplin and colleagues further investigated the frequency and nature of mutations in metastatic CaP samples from patients treated with AR antagonists and from patients not treated with AR antagonists (Taplin *et al*, 1999). An increased proportion of these samples were found to express mutated ARs after treatment with AR antagonists (1/17 untreated as opposed to 5/16 treated). Interestingly, the mutations identified in the AR antagonist-treated groups were predominantly composed of substitution mutations at amino acid 877. This mutation together with a position 874 sequence variant, corresponds to mutations found in cell lines from patients with CaP (Veldscholte *et al*, 1990; Tan *et al*, 1997; Zhao *et al*, 2000).

From these studies it seems that the frequency of AR mutations is low in localised forms of CaP, even in androgen-independent tumours. Also, the frequency of AR mutations increases as the tumours become metastatic to local lymph nodes and even rises further when distant metastatic deposits are analysed. These observations suggest that AR mutations confer some sort of selective advantage that is most effective as the tumours metastasise.

1.3.2 Genetic susceptibility to Prostate Cancer

The familial nature of complex diseases such as diabetes, heart disease, hypertension, osteoporosis, psychiatric conditions, infection and cancer, clearly indicates a significant genetic component. Unlike monogenic disorders, the genetic contribution consists of multiple disease genes of modest effect. The genetic complexity could be partly due to heterogeneity where different combinations of gene mutations may result in similar phenotypes, such as when genes are needed for a common biochemical pathway or cellular structure. Also, some individuals who inherit a susceptibility allele may not display the disease (incomplete penetrance), whereas others who do not inherit susceptibility alleles may still get the disease as a result of environmental or random causes (phenocopy). Therefore the genotype at a given locus could affect the probability of disease, but not fully determine the outcome.

1.3.2.1 Polymorphic variants and disease association

Linkage analysis and association studies are methods with different approaches to analyse the genetic aetiology of complex disease. Linkage refers to the physical

relationship between two or more genetic markers[18] used to identify the position of a disease gene. Linkage analysis uses family materials and involves constructing a transmission model[19] to compare the inheritance pattern of genetic markers to a disease outcome. Sequences that are closely linked are less likely to be separated by the reshuffling of deoxyribonucleic acid (DNA) that occurs during meiotic recombination than more distantly spaced sequences. Therefore markers that are co-inherited with the disease pinpoint the approximate chromosomal region of an underlying disease gene. The advantage of pedigree analysis is that disease inheritance can be compared to patterns of linkage over large genomic areas in order to map gene mutations. Pedigree studies have been very successful in mapping hundreds of highly penetrant disease loci for Mendelian diseases due to mutations in a single gene. It is, however, very difficult to recruit affected families of sufficient sizes or numbers in order to effectively apply this method. For complex diseases, the number of families required is not practical and the results obtained for the families are not additive. Therefore, linkage studies could have limited power for identifying the moderate gene effects hypothesised to contribute to these diseases that depend on the interaction of several or many genes and the environment (Risch and Merikangas, 1996).

Alternatively, association studies based on single nucleotide polymorphisms (SNPs) have become the prevalent way to approach the genetics of complex diseases (Kwok *et al*, 1996; Brookes, 1999). A SNP is a stable, single base substitution that occurs at a frequency of more than 1% within the general population and about 90% of genetic variations in humans are attributable to these allelic variants (Collins *et al*,

[18] Predefined polymorphic sequences with known sublocations on chromosomes predefined polymorphic sequences with known sublocations on chromosomes

1998; Brookes, 1999). The rest of the sequence variation results from rearrangements, deletions or insertions of one or more bases and repeat length polymorphisms.

Techniques for SNP identification on a large scale within the human population are resulting in an exponential increase in the number of SNPs (Wang *et al*, 1998). Within the last two years the search and discovery of SNPs have been so successful that 1.69 million have been reported since June, 2001, as part of the publication of the draft human genome sequence (The International SNP Map Working Group, 2001). The NIH SNP database (http://www.ncbi.nlm.nih.gov/SNP) already contains over 2.8 million cases. There are several other databases including ENSEMBL (http://www.ensembl.org), UCSC (University of California at Santa Cruz; http://genome.ucsc.edu) and TSC (The SNP Consortium; http://snp.cshl.org). After coverage of SNPs on 24,953 exons, it was also reported by The International SNP Map Working Group, that approximately 85% of all exons were within 5kb of the nearest SNP.

About 5% of SNPs in the public database have been discovered by gene-based studies (The International SNP Map Working Group, 2001). Efforts to re-sequence a gene or set of genes that belong to a common biological pathway have proceeded slowly. Progress is underway to identify and validate SNPs present in genes from critical pathways, like the Th1 and Th2 cytokines, complement cascade factors and mediators of apoptosis. These in turn can be applied to specific studies based on a reasonable hypothesis. Using this approach enables one to examine a biological

[19] Multi-generation family pedigrees

process as a whole and the choice of candidate SNPs can be extended to include known genes that take part in the active process.

Even though SNP databases are an important resource, it is not the end result needed to map complex diseases. It is therefore useful to refer to these SNPs as candidate SNPs because nothing is known about the frequency in populations for most of them. Various groups have started characterising the candidate SNPs and one pilot study in particular was undertaken to see how effective these SNPs in the public database would be if they were to be developed into genetic markers (Marth *et al*, 2001). Results from this study show that 17% of the SNPs were monomorphic, i.e., only one of the two predicted alleles was found in all 3 population samples of African-Americans, Asians and Caucasians in the US and no detectable difference was observed in any of these 3 major US population groups. Thus these candidate SNPs would not be useful for genetic studies. The majority of the candidates represent real variants but with no significant population frequency. Approximately 6% were uncommon SNPs and had rare allele frequencies below 20% in any population, 52% of candidates were common with minor allele frequencies in 2 or 3 populations and 27% of SNP candidates were common in all 3 populations. The outcome of this research estimates that if one uses these publicly available candidate SNPs, there is only a 66-70% chance that they have an appreciable minor allele frequency and a 50-50 chance that the SNPs are common in a specific population.

According to a study by Miller and Kwok, 2001, the "life-cycle" of a SNP can be divided into 4 stages (lettered a – c):

(a) occurrence of a new variant allele by nucleotide mutation

In spite of effective DNA repair mechanisms (Wood *et al*, 2001), mutations cause disease alleles and normal variants in humans. A common SNP is found more or less every 1000 bp. All combinations of substitution polymorphisms are observed in humans, with A/G substitution and the reverse complement T/C SNPs as the most prevalent (Taillon-Miller *et al*, 1999; Miller *et al*, 2001).

Methyl CpG to TpG transitions are the most frequently found mutations in humans and represent 25% of all mutations (Wang *et al*, 1998a; Taillon-Miller *et al*, 1998). This is probably a main cause of deficiency of CG dinucleotides observed in the genomic sequence, as many will eventually become TG. New CpG sites, however, will be created by other less common mutations. It is a challenge in understanding how and when mutations cause disease. SNPs in coding (cSNPs) and regulatory regions are the most likely to affect gene function (Collins *et al*, 1997; Chakravarti, 1998; Syvanen *et al*, 1999). SNPs, which occur in the promoter regions of genes that co-ordinate the immune response, for example tumour necrosis factor (*TNF*), interleukin 4 (*IL4*), *IL6* and *IL10*, have been associated with various infectious and autoimmune disorders (Taylor *et al*, 2001). This suggests that small differences in the regulation of key genes can strengthen or lessen the effect of a biological pathway. Since only a small proportion of the genome contains coding regions (Lander *et al*, 2001), the majority of SNPs are likely to have little functional consequence. Approximately 50% of cSNPs cause missense mutations in the corresponding proteins, while the other half are silent (Cargill *et al*, 1999; Halushka *et*

al, 1999). Missense SNPs could be neutral and not detectably alter the phenotype, or they could cause differences in individual characteristics, different responses to drugs or varied susceptibility to disease. It has been proposed that a combination of SNPs confer susceptibility in many complex diseases but it is not certain whether rare or common variants will be the responsible determinants (Collins *et al*, 1997; Pritchard, 2001).

(b) survival against odds of the allele through early generations

The second stage of the SNP life cycle is referred to as "survival against odds" as the majority of new mutations are expected to be lost in early generations due to sampling. As an example, if a person is heterozygous for a new, selectively neutral, autosomal mutation and has two children, there is a probability of 0.75 that the mutation will be found in at least one of the children. Providing the mutant allele copies are carried by heterozygote individuals in the population who each have two children, the probability that the new mutation will be lost is $1- (0.75)^{g,}$ where g = generations, with the mutation occurring in g = 0.

(c) an increase to significant frequency as well as survival through population changes:

If a mutation survives early generations and increases in frequency to a point where it becomes homozygous in some people, the third stage is entered and the risk of loss is reduced. Due to random variables and genetic hitchhiking[20], the frequency of a new allele is expected to show variation, but hindrances in population size will result in survival of the alleles with the highest frequencies and the loss of rare alleles.

[20] Occurs when linked neutral variants are dragged along with advantageous mutations

(d) fixation

During the fourth stage of the lifecycle, a few SNPs will cease to exist because the "new" allele will reach 100% in the population and become a difference between humans and a closely related species. Based on the divergence and variability within humans, Miller and Kwok, 2001, predicted that the lifetime in which a SNP becomes fixed for a new allele is 284 000 years.

1.3.2.2 Association studies

One of the main motivations for the discovery of SNPs has been the potential search for complex disease loci using association studies. In these studies SNPs can be used as markers for linkage disequilibrium (LD). LD can be described as a measure of the degree of association (co-segregation) of two closely linked genetic markers and is therefore used to identify chromosomal regions that are associated with disease in a population (Morton *et al*, 2001). Surrogate markers are used to identify the gene regions, which show association with the disease and usually indicate the location of the adjacent causative gene. LD can be generated by selection, mutation or admixture of populations with different allele frequencies and generally depends on the population size, generations and distance between genetic markers. Usually with greater distance between markers, the quicker disequilibrium will decay. For highly polymorphic markers such as microsatellites, the high mutation rate significantly contributes to random association of alleles.

Theoretical approaches using computer simulations have predicted that LD would be limited to ~3kb in the human population, implying that 500 000 SNPs would be required for a whole-genome association study (Kruglyak, 1999). However,

predictions from the literature based on empirical data, suggest that 30 000 SNPs would be sufficient (Collins *et al*, 1999).

One study, representing the largest genotyping effort, was performed to evaluate variation and the extent of LD between the *AR* trinucleotide repeat loci (CAG and GGC tracts) in human populations; specifically those of African descent such as African Americans (Kittles *et al*, 2001b). This would be an important requirement to determine if there are subpopulations of disease chromosomes segregating in high-risk groups like African Americans. Results showed that the populations of African descent possess significantly shorter alleles than the non-African groups at a *P*-value less than 0.0001. The allelic diversity was higher for both markers among African Americans[21] than other populations. Analysis of molecular variance revealed that almost 20% of CAG and GGC repeat variance could be ascribed to differences between populations. All non-African populations had the same common haplotype while the three populations of African descent possessed three varied common haplotypes, where haplotypes are the particular combination of alleles at different sites on a single chromosome. Significant LD was observed in the sample of healthy African Americans and could be as a result of recent migration of African Americans from various rural communities following urbanisation, admixture with European Americans and recurrent gene flow from diverse West African populations. The assessment of LD in the African American population is quite significant because the high level of stratification could be confusing in disease association studies if the substructure is not controlled for. Also, the identification of high-risk haplotypes may be more powerful in disease studies than single locus analyses. The research team

[21] Including indigenous Africans from Sierra Leone and Nigeria ncluding indigenous Africans from Sierra Leone and Nigeria

of Kittles and co-workers intend to increase the resolution in identification of these possible high-risk haplotypes for CaP by typing SNPs within the gene and performing haplotype analysis. Finally these studies will provide a clearer understanding of the role AR variation plays in CaP aetiology.

One can also search for disease association by genotyping one or a few SNPs from specific candidate disease genes with hypothesised functional importance. Association studies of this nature do not concern familial inheritance patterns at all. They are case-control studies based on a comparison of unrelated affected and unaffected individuals (controls). An association could be inferred if susceptibility factor at a gene of interest occurs at a significantly higher frequency among affected compared with control individuals. A variant might not directly result in a disease phenotype, but could be a genetic marker in linkage with a nearby locus. Association studies could be quite useful in identifying genetic markers that modify the risk for the outcome of disease. Even though association can be performed for any randomly occurring polymorphism, it is most meaningful when it uses genetic markers assumed to have a biological relation to the trait.

There are also limitations when using case-control studies, as a positive association can occur as a result of population admixture. In a mixed population, any characteristic that is present at a higher frequency in an ethnic group will show a positive association with any allele that happens to be more common in that group. Limitations in genotyping technology make it impractical to test large numbers of SNPs in large numbers of individuals. As a result, published studies have considered a small number of mutations in candidate disease genes, based on

previous research. However, only a few of the disease associations appear to be reproducible when they are retested with independent clinical materials. This may lead to increased doubt concerning the usefulness of genetic association analysis (Lernmark and Ott, 1998; Editorial, 1999), or scepticism about whether results can be generalised to other groups with various genetic and environmental profiles (Terwilliger and Goring, 2000).

Studies using SNPs to investigate the genetic basis of human disease may provide insight into the susceptibility to a disease. The study of SNP distribution, especially in different populations can also be useful for investigating molecular events that underlie evolution such as genetic drift, selection and recombination (Nachman, 2001). SNPs are a product of the past and could provide clues about human history like tracing the origin of populations and their migrations.

1.3.2.3 *AR* gene susceptibility alleles

The highly variable trinucleotide microsatellite of CAG-repeats in exon 1 of the *AR* gene has been the target of unprecedented interest in recent years and is implicated in various disease processes. Previous studies show that this CAG repeat varies in length between 11 and 31 repeats in healthy men (Edwards *et al*, 1992) and variations of the CAG repeat length have been linked to alterations in AR function. Individuals with a marked expansion of CAG repeats (between 40 and 52 repeats) suffer from Kennedy syndrome/ spinal and bulbar muscular atrophy (SBMA), a rare neuromuscular disorder of adult onset resulting in severe depletion of lower motornuclei in the spinal cord and brainstem (Kennedy *et al*, 1968; La Spada *et al*, 1992; Nance, 1997). SBMA is characterised by progressive muscle weakness and

atrophy (Kennedy *et al*, 1968) and partial androgen insensitivity[22]. Patients with SBMA often exhibit endocrinological abnormalities which include testicular atrophy, infertility, gynaecomastia, oligo- or azoospermia[23], impotence and elevated serum gonadotropin levels (Arbizu *et al*, 1983; Amato *et al*, 1993). An earlier onset, severeness of muscle weakness (Igarashi *et al*, 1992; La Spada *et al*, 1992), slight androgen resistance (Shimada *et al*, 1995) and a diminished transcriptional activity of the AR (Choong and Wilson, 1998) are associated with extended CAG repeats. A study by Tut and colleagues found that patients with 28 or more CAG repeats in their AR have a four-fold increased risk for impaired spermatogenesis which is an exquisitely androgen-dependant process (Tut *et al*, 1997). This defective sperm production could be related to the partial androgen insensitivity, which is seen in individuals with SBMA. In another study, breast cancer diagnosis was made earlier among women with mutations in their *BRCA1* genes who also have very long CAG repeats (Rebbeck *et al*, 1999). On the other hand, one study showed that men with shorter CAG tracts of $\leq$ 19 repeats have a substantially greater risk for developing BPH than those with CAG repeats of $\geq$ 25 in length (Giovannucci *et al*, 1999). Short CAG repeats have also been associated with androgen-mediated skin disorders in men and women (Sawaya and Shalita, 1998) where men with androgenic alopecia [24] had 19 $\pm$ 3 repeats, whereas women with this condition had 17 $\pm$ 3 repeats as opposed to the controls where a range of 12-29 repeats was found. In rheumatoid arthritis in males, shorter CAG repeats of the AR are related to a younger age of onset (Kawasaki *et al*, 1999). These rheumatoid arthritis patients had significantly

[22] Androgen insensitivity is an X-linked disorder associated with abnormalities in male sexual morphogenesis; genetic males with severe or mild defects in the development of the male phenotype are partially androgen insensitive
[23] Reduced or no sperm production
[24] Partial or complete loss of hair

shorter CAG repeat lengths than the younger male controls (21.8 vs. 23.2) or the older-onset male patients (21.8 vs. 23.2).

Differences in AR CAG repeat length have also been linked to CaP risk in some but not all studies (Hardy *et al*, 1996; Giovannucci *et al*, 1997; Ingles *et al*, 1997; Stanford *et al*, 1997). Several population-based studies have shown that African-American men have a higher incidence of CaP, they are younger at diagnosis and often present with a more advanced form of disease, when compared with other ethnic groups (Mebane *et al*, 1990; Boring *et al*, 1992; Morton, 1994). The average number of CAG repeats in African-Americans, Caucasians and Asians is 18, 21 and 22, respectively (Coetzee and Ross, 1994). Shorter CAG repeat lengths result in an increased transcriptional activity of the AR (Chamberlain *et al*, 1994). Based on the differences of CAG repeats in the aforementioned ethnic groups in combination with the inverse relationship between the length of CAG tracts and AR transactivation function, it has been hypothesised (Coetzee and Ross, 1994) that the variation in race in CaP risk could partly be due to differences in AR CAG repeat length. There has since been several groups that have investigated the relationship of CAG repeat length to different epidemiological and clinical variables in CaP. The results from some of these studies are found in Table 3. Depending on the study, shorter CAG repeats fall in the range between < 17 to < 23 and have been correlated with a younger age at diagnosis of CaP (Hardy *et al*, 1996; Bratt *et al*, 1999), an increased overall CaP risk (Giovannucci *et al*, 1997; Hakimi *et al*, 1997; Ingles *et al*, 1997), lymph node-positive disease (Hakimi *et al*, 1997) and extraprostatic extension (Giovannucci *et al*, 1997; Ingles *et al*, 1997). Various other studies, however, have shown that the CAG tract is not associated with CaP risk (Irvine *et al*, 1995; Bratt *et*

al, 1999; Correa-Cerro *et al*, 1999; Edwards *et al*, 1999), earlier age of cancer onset (Hakimi *et al*, 1997; Correa Cerro *et al*, 1999), stage of disease at diagnosis (Hardy *et al*, 1996; Stanford *et al*, 1997), histological grade of disease at diagnosis (Stanford *et al*, 1997; Correa-Cerro *et al*, 1999) or different clinical variables, including response to treatment and time to progression after hormonal therapy (Hardy *et al*, 1996), disease-free and overall survival (Edwards *et al*, 1999). It is difficult to compare the results of many of these studies as they have different designs, variable study populations and reference CAG lengths.

When looking at the case-control studies (Irvine *et al*, 1995; Giovannucci *et al*, 1997; Hakimi *et al*, 1997; Ingles *et al*, 1997; Stanford *et al*, 1997; Correa-Cerro *et al*, 1999), one cannot be certain that individuals of the control groups did not have undiagnosed CaP or would even develop the disease, therefore affecting the results.

Table 3. Summary of results from 10 groups which analysed the link between CAG/GGN repeats or both and CaP association. Compilation of patient populations (Cases/controls) and CAG/GGN results are shown. CAG analysis and GGN analysis entailed comparison of various repeat lengths used in statistical analysis in the different studies

Author	**Irvine et al (1995)**	**Giovannucci et al (1997)**	**Hakimi et al (1997)**	**Hardy et al (1997)**	**Ingles et al (1997)**
Cases/controls	68/123; 57 cases Caucasian; controls: 45 African-American, 39 Caucasian, 39 Asian	587/588; study cohort predominantly Caucasian (>95%)	34 stage B, 25 stage D/ 370; entire study group Caucasian	101 cases, 95% Caucasian, 82% with stage C or D disease	57/169; Caucasian patients and control subjects
CAG analysis	< 22 vs. ≥ 22	≤ 18 vs. ≥ 26	< 17 vs. ≥ 17	≤ 22 vs. >22	< 20 vs. ≥ 20
CAG results	No statistical significance between CAG length and CaP risk	1.5 times higher CaP risk for men with CAG < 22 than men with long CAGs; shorter CAG repeats associated with extraprostatic extension, distant metastases, high grade and disease mortality	Cases have 4-times greater likelihood of having short CAG repeats; men with lymph-node positive disease have 8-times higher probability of having CAG repeats of < 17	Significant association between short CAG and younger age at diagnosis; no correlation with CAG length and stage of disease at diagnosis, baseline PSA, response to hormonal therapy, or disease progression	Short CAGs associated with a two-fold increase in CaP risk and also linked to extraprostatic disease
GGN analysis	16 vs. ≠ 16	NA	≤ 14 vs. >14	NA	NA
GGN results	A two-fold, non-significant elevated risk among patients with CAG < 22 combined with GGN ≠ 16 where CAG and GGN tracts are in linkage disequilibrium in CaP patients	NA	4-fold higher risk of short GGN repeats in men with CaP; GGN not linked to age of diagnosis or aggressive disease; CAG and GGN alleles not in linkage disequilibrium	NA	NA

Author	Stanford et al (1997)	Platz et al (1998)	Bratt et al (1999)	Correa-Cerro et al (1999)	Edwards et al (1999)
Cases/controls	301/277; study cohort consists of Caucasians of ages 40-64	592/588 - CAG analysis, for GGN analysis 794 controls; 95% of population Caucasian	190/ 186; control alleles from women. Entire group Swedish*	132/105; French German subjects for all cases and controls	178/390; X-chromosome female controls; British Caucasian population+
CAG analysis	< 22 vs. ≥ 22	< 21, 21-23, >23	15-20, 21-23, 24-31	< 22 vs. ≥ 22	≤ 21 vs. > 21
CAG results	Small, non-significant increased CaP risk in men with CAG repeats of < 22; 3% decrease in risk of CaP with each additional CAG repeat; no association between CAG and disease stage/grade at diagnosis	CaP risk highest in men with CAG < 21 and GGN length = 23 when compared with CAG >23 and GGN ≠ 23	No significant correlation between CAG repeat length and CaP risk; shorter CAG repeats linked to younger age at diagnosis; patients with longer CAGs (mean 22.7) have good responses to endocrine therapy	No association between CAG length and CaP susceptibility, age at diagnosis, or histological grade of disease	CAGs not significantly linked to risk for CaP, disease free and overall survival; inclination towards correlation between CAG repeats of > 21 and metastatic disease
GGN analysis	≤ 16 vs. >16	23 vs. ≠ 23	NA	NA	≤ 16 vs. >16
GGN results	Men with GGN repeats of ≤ 16 had a 1.6-fold elevated CaP risk and increased risk for aggressive disease; those with CAG < 22 and GGN ≤ 16 had a 2-fold increased risk of CaP development; CAG and GGN are not in linkage disequilibrium	CaP risk decreases for each repeat change in GGN length of < or > 23; no significant association between GGN and age at diagnosis; CAG and GGN in linkage disequilibrium in population	NA	No correlation between GGN repeat length and CaP risk	GGN repeats not associated with CaP risk; long GGN repeat lengths linked to elevated risk for disease relapse and mortality; CAG and GGN alleles not in linkage disequilibrium

* CAG alleles of female controls and male patient CAGs were used to determine the correlation between CAG length and CaP risk.
+ Female GAG alleles in the control group in addition to male case CAGs were used to assess the ethnic distribution of GAG repeats.

The trinucleotide GGN polymorphism (where N=C, G or T) in the NH_2-terminal domain of the AR has also been associated with CaP. The length of this polyglycine tract varies from 10-30 repeats in normal individuals (Platz *et al*, 1998) and is less polymorphic than the CAG repeat (Edwards *et al*, 1999). As for the CAG repeat, results are conflicting when comparing the GGN repeat data and CaP variables. The results from a few of these studies are shown in Table 3. Certain studies have found short GGN repeats (Stanford *et al*, 1997; Hakimi *et al*, 1997) of $\leq$ 16 and $\leq$ 14, respectively, to be associated with an elevated CaP risk, whereas other groups (Correa-Cerro *et al*, 1999; Edwards *et al*, 1999) reported no link between the GGN tracts and risk for CaP. Some studies have shown that the GGN repeat is not associated with disease aggressiveness (Hakimi *et al*, 1997) or age at diagnosis (Hakimi *et al*, 1997; Platz *et al*, 1998), though longer GGN alleles of $>$ 16 are associated with an increased risk for disease relapse and mortality (Edwards *et al*, 1999). It has even been reported that there is linkage disequilbrium between the CAG and GGN alleles and therefore certain combinations of alleles are co-inherited and together could be associated with an increased CaP susceptibility (Irvine *et al*, 1995; Platz *et al*, 1998). Other studies, however, have found that the CAG and GGN alleles are not in linkage disequilibrium (Hakimi *et al*, 1997; Stanford *et al*, 1997; Edwards *et al*, 1999).

Due to the great variation in the above findings, it is difficult to see exactly how the CAG and GGN repeats affect the development and progression of CaP. The genetic influences of these polymorphisms, together with environmental factors, could possibly contribute to a person's overall risk for CaP development as well as

assisting to determine disease characteristics like histological grade, tumour aggressiveness and the response to therapy.

1.3.2.4 *CYP17* susceptibility alleles

The 5' untranslated region of *CYP17* contains a common single nucleotide substitution (T→C) located 34 base pairs upstream from the translation initiation site. This polymorphism creates a potential Sp1-like (CCACC box) binding site, that may lead to an increased transcription and enhanced androgen biosynthesis (Carey *et al*, 1994).

To date, only a few studies have shown an association of the *CYP17* gene and CaP risk, however there have been contradictions in terms of which allele is associated. In a North Carolinian Caucasian population, Lunn and co-workers (1999) observed a borderline significant association between the T/C genotype and CaP risk. A smaller study by Gsur *et al.* (2000) also reported increased risk for Caucasian men homozygous for the C allele. In contrast, among Japanese men (n = 252 cases and n = 131 controls), Habuchi and colleagues (2000) observed men with the T/T genotype to have increased risk of CaP and BPH. Similarly in a Swedish study, (cases, n = 178; controls, n = 160), Wadelius *et al.* (1999) also reported a significant elevation in risk for the T/T genotype. The opposing findings of previous studies could be attributable to inaccurate genotype frequency estimates resulting from small sample sizes (all studies have less than ~ 250 cases and ~ 160 controls). Linkage disequilibrium between the T and C alleles of *CYP17* with different functionally relevant polymorphisms in specific ethnic groups may explain the contradictory

findings observed in studies conducted in different white or nonwhite ethnic populations.

There seems to be little difference between *CYP17* allele and genotype frequencies in African Americans and European Americans (Feigelson *et al*, 1997; Weston *et al*, 1998; Lunn *et al*, 1999). This is unlike several other candidate genes for CaP which display striking allele frequency differences that parallel differences in CaP incidence (Rebbeck *et al*, 1998; Ross *et al*, 1998; Paris *et al*, 1999). To date, no allele and genotype frequency data exist on clinically evaluated indigenous Africans and African-American CaP patients. To further investigate the involvement of the *CYP17* variant with CaP, Kittles *et al*. (2001a) typed the polymorphism in three different populations and evaluated its association with CaP and clinical presentation in African Americans. This group genotyped the *CYP17* polymorphism in Nigerian (n = 56), European-American (n = 74), and African American (n = 111) healthy male volunteers, along with African American men with CaP (n = 71). They found that the common *CYP17* variant was associated with increased risk of CaP in African American men. Comparison of genotypes revealed a significantly higher risk among individuals homozygous for the C allele for developing high grade/stage CaP. African American patients with the CC genotype were seven times more likely to present with more aggressive disease. As sample sizes were moderate for the African American samples, results should be interpreted with caution until larger studies further evaluate the polymorphism.

1.3.2.5 Other susceptibility alleles and Prostate cancer: *SRD5A2, CYP3A4*

SRD5A2

Two polymorphisms in the coding region have been identified in the *SRD5A2* gene. The one, a germline missense substitution, results in an alanine residue at codon 49 being replaced with threonine (A49T) (Reichardt *et al*, 1997). This variant has been linked to pathological characteristics of CaP and increases the levels of steroid 5-α-reductase activity. It has been reported to be more common in African-American and Latino men with CaP, as compared with healthy African-American and Latino controls, and was found to be most common in African-American and Latino men with advanced disease (Reichardt *et al*, 1997). More recently, Makridakis *et al.* (1999) provided additional support for the hypothesis that the A49T variant was associated with CaP risk in African-American and Hispanic men and also reported that the variant enzyme had a higher activity *in vitro* than the normal enzyme. The effects of this mutation in other races have not been previously reported.

The other polymorphism, V89L, reported in *SRD5A2* is caused by a G to C transversion that results in the substitution of a leucine for a valine amino acid in position 89 (denoted the L allele) and was associated with lower levels of androstanediol glucouronide (AAG). It was found that the distribution of this mutation appears to parallel the frequency of CaP within different ethnic groups, with Caucasian and African-American men having a low prevalence of the L allele (24% and 22%, respectively), compared with a prevalence of 46% among Asian men. The L/L genotype was associated with a significantly lower mean serum AAG concentration within the Asian men, suggesting that the L allele may reduce 5-α-reductase activity (Makridakis *et al*, 1997). This polymorphism, however, has not

been found to be significantly associated with Caucasian men (Febbo *et al*, 1999; Lunn *et al*, 1999). If small differences in serum AAG levels reflect larger differences in intraprostatic androgen activity then these genetic differences may be of biological relevance to CaP risk.

CYP3A4

Genetic variations of *CYP3A4* have recently been reported. A mutation in the 5' - upstream region termed *CYP3A4*1B* (A290G) results in an alteration in a transcriptional regulatory element that may be required for *CYP3A4* gene expression. This variant was observed in 52% of African-Americans and 9.6% of Caucasians, but has not been identified in Asians (Ball *et al*, 1999; Rebbeck, 2000; Sata *et al*, 2000). It was suggested to be associated with advanced stage CaP in men and the association between *CYP3A4* genotype and tumour stage was most pronounced in men diagnosed at a relatively old age who reported no family history of CaP (Rebbeck *et al*, 1998). The identification of the *CYP3A4* genotype as a biomarker associated with CaP has potential implications for the treatment and prevention of CaP and warrants further investigation.

1.4 AIM OF THE STUDY

CaP has the highest incidence of any malignancy affecting South African males (National Cancer Registry of South Africa, 1992). The incidence of this cancer type shows strong age, race and geographical dependence. It is a disease of older men, with its highest incidence and mortality rates having been reported in African-Americans. The differences in incidence and mortality rates of CaP observed among different ethnic groups suggest that environmental and genetic factors may be

important risk factors for the development of this cancer. A model for the malignant progression of CaP remains incomplete. Various genes, however, have been identified as possible candidates involved in this malignant progression, either as a result of "gain of function" or "loss of function".

The aim of this study is therefore the identification of genes involved in the predisposition to and / or progression of CaP within the diverse South African population, specifically:

1) to determine the type and frequencies of genetic variants of genes involved in androgen metabolism (*AR*, *CYP17*), and
2) to determine if the observed ethnic variation in the incidence and progression of CaP can be understood by these ethnic-based genetic differences (i.e. differences in the frequencies of genetic variation).

Our understanding of the genetic changes involved in the progression from normal epithelium to malignancy and thus, ultimately the knowledge gained from such a study, should yield information important for the early diagnosis and prognosis of patients, as well as providing mechanisms for the development of therapies.

CHAPTER 2

MATERIALS AND METHODS

2.1 HUMAN POPULATIONS

The human species is divided into many distinct subpopulations, the largest of which are commonly called races. Races are defined as major population groups whose gene pools differ from each other (Thompson *et al*, 1991). There are three major racial divisions in South Africa, Whites, Blacks, and Coloureds, each of which has numerous genetically different subgroups. Although the human chromosomes and the loci they contain are identical in all members of the species, allele frequencies at many loci vary widely among population groups. Some variants are virtually restricted to members of a single group, although they are not necessarily present in all members of the group. More often, variant alleles have different frequencies in different populations. Within each population there is extensive variation, greater on the average than the mean differences between groups.

The basis of the genetic differences among races and among their subpopulations is mutation. The selection of favourable mutations in response to environmental conditions or the chance survival of specific neutral or even harmful mutations, together with a degree of reproductive isolation between the groups, allowed genetic differences between population groups to be established. Subsequently, each group is subdivided further into numerous distinct subpopulations, often called ethnic groups, with their own characteristic set of gene frequencies. There are often dramatic differences in allele frequencies between population groups, both for alleles

that cause genetic disease and for presumably selectively neutral genetic markers, such as certain blood group and protein polymorphisms and some DNA polymorphisms. Although the former are highly significant for determining recurrence risks for genetic diseases in specific population groups, the latter are important as markers of recent human evolution.

2.1.1 The South African Population (Historical background)

Portuguese sailor, Bartholomeu Dias, was the first recorded European to traverse the South African coast (in 1488) in his search for a sea route from Europe to the riches of the East. It was only, however, until the middle of the 17th century that Western European settlers established colonies on the southern tip of the continent. Today this White population in South Africa are mainly Dutch, French, German and British immigrants of European descent (Mead, 1997).

In 1652, the Dutch East India Company established a small white settlement in Table Bay (now known as Cape Town) under the command of Jan Van Riebeeck (Grütter and Van Zyl, 1981). The Black population, the nomadic Hottentots, came into contact with the European settlers by trading their cattle. Soon after the Europeans arrived at the Cape, they made contact with the earliest indigenous inhabitants, the San. As the colonists intruded further upon the land and water sources, and stepped up their demands for livestock and labour, the San were forced further inland and today the only living members of these people are found in the Kalahari Desert of Namibia. The San were not enslaved and therefore foreign slaves were brought in from Angola and Benin and thereafter they were mostly imported from Madagascar and Mozambique. Most of these slaves were men and some of them took San

women as their wives, so there was a natural increase in the slave population. For much of the eighteenth century slaves were imported at the rate of two or three hundred a year. By the 1790s the slave population had risen to 25,000 compared with 21,000 white colonists.

From the 1770s, colonists also came into contact and conflict with Bantu-speaking chiefdoms some 700km east of Cape Town. Five main groups of Bantu are distinguished today (Government Communication and Information System, 2001). The first, the Nguni, spread through Mpumalanga[25] down the east coast. Several Nguni nations can be distinguished. Of these, the Xhosa live in the old Transkei and Ciskei regions, the Zulu in Zululand and KwaZulu Natal, the Swazi in Swaziland and the Transvaal[26] and the Ndebele in Zimbabwe. A second group of Bantu, the Sotho, spread westward about 500 years ago, the North Sotho (Pedi) settling in the Northern Province and Mpumalanga, the West Sotho (Tswana) furthest west in Botswana and adjacent areas of South Africa and the South Sotho (Basotho) in Lesotho and the Free State (formerly Orange Free State). A third group, the Ovambo-Herero, went even further west, settling in what is today known as Namibia. About 300 years ago, a fourth group, the Venda made their appearance in the Northern Province. Even later, in the course of the 19th century, a fifth group, the Tsonga, migrated from the east coast to the Transvaal. Maps of South Africa with its former divisions and the provinces as they are today are depicted in Figures 6A and 6B, respectively. (South African maps and boundaries data with the kind complements of Antonio Martins; http://www.fotw.ca, 2000).

[25] Formerly eastern Transvaal formerly eastern Transvaal

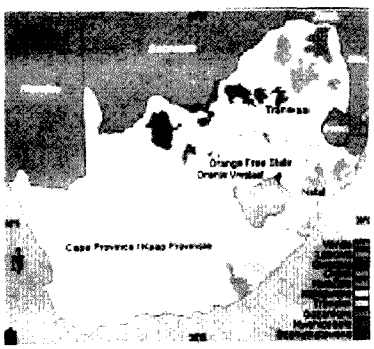
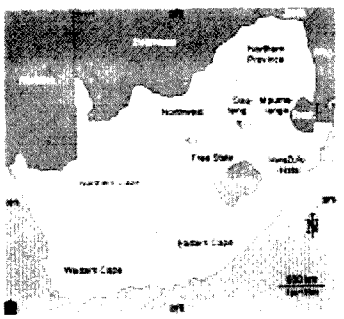


FIG.6A. Shows a map of South Africa with it's former divisions; viz, the Cape Province, Natal, Orange Free State and Transvaal.

FIG.6B. Since 27 April 1994, South Africa has 9 provinces: The Eastern Cape, Free State, Gauteng, Kwazulu-Natal, Mpumalanga, Northern Province, Northern Cape, North-West Province and the Western Cape.



The origins of the 'Coloureds', or people of mixed race go back to the Cape region around the 17th century when many of the Hottentots mixed with and married the early Dutch sailors. At the same time, Malay women were brought to the Cape as wives for Dutch sailors and settlers. Nine tenths of the Cape Coloured population are Christian and 90,000 are represented by a subgroup, the Cape Malays, who practice Islam as their faith and are geographically limited to the Cape Peninsula (Shillington, 1987).

The so-called 'coloured' population of the Cape was the result of interbreeding between Europeans and Khoikhoi (local Hottentot inhabitants), and Khoikhoi and slaves (Grütter and Van Zyl, 1981). Many of the early white settlers at the Cape were single men and a number took Khoikhoi women as their unofficial wives. On the farms there was also interbreeding between male slaves and female Khoikhoi servants. Gradually there developed a 'coloured' servile population, part slave and

[26] Now divided into Mpumalanga, Limpopo (formerly Northern Transvaal) and the North-West Province (formerly Western Transvaal)

part 'free'. They adopted the Dutch language of the colonists and began to lose their African or Asian cultural identities. During the nineteenth century whites began to use the term Cape Coloured to refer to the whole of this Dutch (Afrikaans)-speaking servile population of slaves, mixed race and Khoisan. These groups have all interbred and blended so that there are no recognisable separate descendants from any of them, except for the Cape Malays (Shillington, 1987).

2.1.2 Patient cohort and specimen collection

In this study, CaP tissue and peripheral blood samples from the diverse South African population was used. The study population included patients of a Black, Coloured and Caucasian descent who have undergone a prostatectomy / prostate needle biopsy at the Tygerberg Hospital. Peripheral blood was drawn from the same patients after fresh tissue was obtained. Personal and clinical data (including tumour Grade 1 to 3 classification, metastasis, hormone dependence and PSA levels where possible) were obtained from the medical records at Tygerberg Hospital and all samples were obtained with informed consent (extra samples were obtained from the Western Province Blood Transfusion Service (WPBTS) in order to calculate population frequencies. Our samples included 20 blood and tissue CaP's, 25 control (BPH) blood and tissue samples (either with or without prostatitis) and 6 blood and tissue samples of which the clinical data are outstanding. The remaining samples include 30 archival paraffin embedded CaP biopsies (5x10μm slices) from Blacks. The number of samples of CaP and BPH for each race group in addition to the numbers of controls from the WPBTS[27] is summarised in Table 4. DNA was isolated from the nuclei of peripheral blood, fresh frozen prostate tissue and paraffin samples

[27] Differs across markers as the same patients were not tested for all markers

using standard procedures (Qiagen Kit) as described in the QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook (blood and body fluid spin protocol: pg 22-24; tissue protocol: pg 28-31; genomic DNA isolation from paraffin-embedded tissue: pg 40-41). This study forms part of a project that was ethically approved by the Research Committee of the University of Stellenbosch (Ethical approval number: 2001/C101).

Table 4. The study cohort consisting of Black, Coloured and White population groups

	Blacks	Coloureds	Whites	TOTAL
Blood/tissue CaP	2	15	3	**20**
Blood/tissue BPH	1	16	8	25
Blood/tissue unknown *	2	4	0	6
Paraffin-embedded material	30	—	—	**30**
WPBTS controls°: *AR* E211	50	57	52	159
CYP17 H46	23	20	—	43
CYP17 S65	24	20	—	44
CYP17 IVS4+58G>C	50	52	2	104
CYP17 IVS6-25C>A	33	30	1	64

* Clinical diagnosis inconclusive

° Controls consisting of samples from males and females were tested for identified markers (no controls were looked at for the *CYP17* L36 marker at the time of writing up)

2.2 MUTATION DETECTION

2.2.1 DNA Amplification

One of the prerequisites for many recombinant DNA techniques is the availability of large amounts of a specific DNA segment. Polymerase chain reaction (PCR) allows a direct, dramatic amplification of specific DNA segments and it can be used on fragments of DNA that are initially present in infinitesimally small quantities.

PCR reactions were performed for the *AR* and *CYP17* amplicons using the DGGE primer sets shown in Table 5. Each reaction mixture consisted of 100ng / 200ng denatured genomic DNA[28], 0.1mmol/l dNTPs, 20pmol of each primer, 2.5mmol/l 10X Mg^{2+} reaction buffer, and 0.5 units *Taq* polymerase (Roche Diagnostics, Mannheim,

Germany) in a total volume of 50µl. The automated amplification cycle was carried out on a thermocycler as follows: an initial denaturation step at 96^{0}C for 3min, followed by 32 cycles of denaturation at 96^{0}C for 45s, annealing for 1min (various temperatures shown in Table 5), and elongation at 72^{0}C for 1min20s. An additional extension step of 72^{0}C for 7min followed the last cycle. Before DGGE analysis of samples, an additional heteroduplexing step was performed. This involved denaturation at 96^{0}C for 10min, followed by renaturation at the various annealing temperatures of PCR amplification. PCR products were checked by electrophoresis of a mixture of 5µl (10%) of each amplified sample with 5µl loading buffer on a 2% agarose gel. The gels were run in an agarose electrophoresis system containing 0,5 X TBE buffer with ethidium bromide (10mg/ml) (10µl of the stain per litre of buffer), at 150V for about 30 minutes and viewed under UV light to verify amplification.

Nested PCRs were carried out for the analysis of DNA extracted from paraffin-embedded material for exon 1D of the *AR* gene and exon 1A, 4 and 7 of the *CYP17* gene. The PCRs with the external primer sets (Table 6) were performed using 1µl of DNA isolated from paraffin material in a 30µl reaction mixture as described before, with a variation in primer concentrations as listed in Table 6. PCR reactions consisted of 25 cycles as specified previously with annealing temperatures as shown in Table 6. Re-amplification of 1µl of the external amplified PCR products was carried out using the internal DGGE primer sets in Table 5 (AR1D, CYP17.1A, 4 and 7) and the same PCR conditions as initially described. Direct PCR amplification, however, is possible, but fixation of tumour-rich material in paraffin degrades the DNA and greatly reduces its amplification rate due to the low concentrations of DNA.

---

[28] 200ng of template was required for amplifications of DNA isolated from tissue samples

2.2.2 Denaturing gradient gel electrophoresis

In this study denaturing gradient gel electrophoresis (DGGE), a polymerase chain reaction (PCR)-based mutation detection method, was used to analyse almost the entire coding region of the *AR* gene and the polymorphic coding regions (exons 4 and 7) of the *CYP17* gene.

Application of the DGGE technique, developed by Fischer and Lerman in 1983, facilitates the separation of double-stranded (ds) DNA molecules based on their high sequence dependant melting behaviour. The order of bases in a sequence determines the melting behaviour of that fragment. The principle of DGGE is shown in Figure 7 as a time-travel gel (Myers *et al*, 1985). As ds DNA is electrophoresed through a gradient of increasing concentration of denaturants (urea and formamide [UF]), it will melt at specific UF concentrations. A high melting temperature (T_m) of a fragment requires a high [UF] to melt the fragment. Partial denaturation of a DNA molecule causes a conformational change which results in reduced electrophoretic mobility of the fragment. Using DGGE, sequence variants have been identified that differ by only a single base. When a fragment becomes completely single-stranded (ss), its mobility continues so that it will run off the gel and variants could be missed. To prevent this, a GC-rich sequence (GC-clamp) is introduced during fragment amplification to the 5' end of one of the primers (Myers *et al*, 1985).

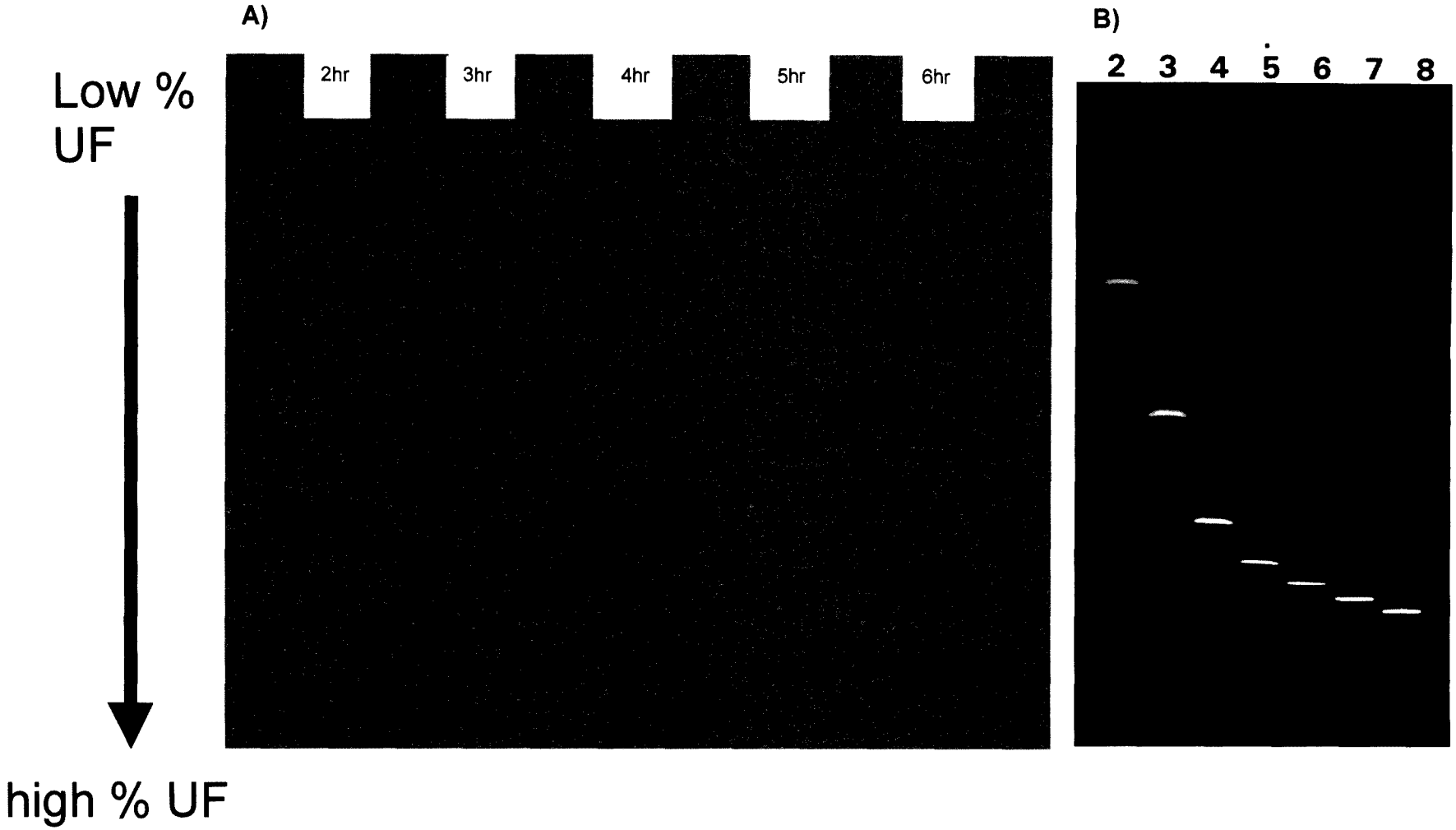


Fig.7 A) Diagrammatic representation of the principles of DGGE (Figure obtained complement of V.M.Hayes). During fragment amplification, a GC-rich clamp is attached to one end of an amplicon, creating a high stability domain. The DGGE gel is a PAA gel with an increasing denaturing gradient of urea and formamide. The gel is run at a constant gel temperature of 60^0C. As the ds DNA passes through the gel, its mobility decreases. The fragment melts out and becomes ss, yet held together by its GC-clamp. This is known as its optimal melting temperature as depicted on the gel. **B)** Time-travel gel (V.M.Hayes).

The GC-clamp alters the melting characteristics of the clamped fragment and increases the percentage of single nucleotide changes detectable by DGGE (Sheffield *et al*, 1989; Abrams *et al*, 1990). The sensitivity of DGGE is also enhanced by heteroduplex analysis which results in the formation of heteroduplexes during the denaturation and re-annealing of target DNA with two different alleles. These heteroduplexes have a lower T_m than the more stable homoduplexes and will therefore melt earlier. Thus small deletions or insertions, where homoduplexes have a similar melting profile would otherwise be undetected without a heteroduplexing step.

Addition of a GC-clamp during PCR-amplification, together with heteroduplexing, has been estimated to increase the number of detectable sequence variants from 40% to virtually 100% (Myers *et al*, 1985; Sheffield *et al*, 1989; Abrams *et al*, 1990). Due to this high sensitivity, DGGE is believed to be the most powerful of the pre-screening methods currently available (Sheffield *et al*, 1989; Guldberg *et al*, 1993).

The success of DGGE requires optimal experimental conditions. Appropriate selection of PCR fragments and primers is very important. A DNA fragment needs an optimal melting profile, i.e. one with a single melting domain for the fragment to be analysed and a higher melting domain that represents the GC-clamp (Wu *et al*, 1998). Theoretical melting profiles can be predicted by using special computer programs designed to analyse the melting behaviour of a DNA fragment. This facilitates the selection of PCR primers that will give rise to fragments that are suitable for DGGE (Lerman and Silverstein, 1987). From the T_m , the range of

denaturant concentration needed for optimal melting of the fragment can be predicted.

Ways have been found to adapt the melting profiles of DNA fragments by modification of primers and their positions, thereby increasing the effectiveness of DGGE. Analysing the natural melting curve (without the GC-clamp), one could understand why fragments that seem suitable for DGGE did not pick up mutations. Modified primers include addition of G/C or T/A tails and result in the detection of mutations previously missed (Wu *et al*, 1998).

Other experimental conditions such as the gel composition and electrophoretic parameters also contribute to the success of DGGE. It has been discovered that the use of 9% polyacrylamide (PAA) gels, a denaturing gradient with a difference of 30% to 50% between the lowest and the highest concentration of denaturant, and electrophoresis in 0.5 X TAE-buffer at a voltage greater than 100 and lower than 200 volts (optimal at 110V), ensures optimal mutation detection for broad-range DGGE analysis (Hayes *et al*, 1999). DGGE is highly sensitive for detection of somatic mutations in tumour material even when surrounded by large amounts of normal tissue. Even for archival paraffin-embedded material, where DNA fragments of a maximal length of 300bp are amplifiable, DGGE is well-suited (Hayes *et al*, 1999). If DGGE assays are designed correctly one can achieve a mutation detection rate of 100% (Wu *et al.*, 1998; Hayes *et al.*, 1999).

See Appendix I for solutions used for DGGE gels.

2.2.3 Primer design

DGGE primers (oligonucleotides that hybridise to opposite strands of the sequence to be amplified) were designed for almost the entire coding region (excluding polymorphic repeat sequences), and including all splice site junctions of the *AR* gene, as well as the promoter region (amplicon 1A) and polymorphic coding regions (amplicons 4 and 7) of the *CYP17* gene. The DGGE primers are listed in Table 5. To ensure appropriate selection of DGGE fragments and primers the MELT 87 computer program was used (Lerman and Silverstein, 1987), keeping in mind the conditions as set out in Wu *et al.* (1998).

The coding sequence of the *AR* gene was divided into 14 amplicons for analysis. Due to its large size, exon 1 was separated into 6 overlapping fragments and exon 7 was divided into 2 overlapping fragments according to the various melting domains within this exon. To prevent fragments from becoming completely single-stranded, a GC-rich sequence (GC-clamp) was attached to the 5' end of an amplicon in each primer set during amplification. In this way the mobility of fragments is decreased and they will not run off the gel, thereby enhancing mutation detection. To ensure optimal mutation detection, primers for exons 1B, 1F and 6 were modified by the insertion of TA (thymine, adenine) nucleotides between the GC-clamp and primer.

TABLE 5. DGGE-PCR primers and experimental conditions for the AR gene as well as the promoter region (exon1A) and exons 4 and 7 of the CYP17 gene

Amplicon	Amplimers, 5'-3'	Size (bp)	Temperature (^{0}C) Melting	Annealing
1B	AR.1BF: [40GC][6AT]GCAGCAGCAGCAAGAGACTAGC AR.1BR: AGGCTCTGGGACGCAACCT	224	79	65
1C	AR.1CF: CAACCTTCACAGCCGCAGTC AR.1CR: [40GC]TGAAGGAGTTGCATGGTGCT	282	81	60
1D	AR.IDF: [7GC]ACCTTAAAGACATCCTGAGCGA AR.1DR: [40GC]TGGACACCGACACTGCCTTA	248	77	60
1E	AR.1EF: [40GC]TTCTGACAACGCCAAGGAGT AR.1ER: TCTTCAGTGCTCTTGCCTGC	259	78	58
1F	AR.1FF: [40GC][6AT]AGGTTCTCTGCTAGACGACA AR.1FR: AGTGCTCCGGACTTGTAGA	232	76	55
1G	AR.1GF: [40GC]AGCTCCGGGACACTTGAACTGC AR.1GR: CAGGCGCTGCCGTAGTCCA	234	80	65
2	AR.2F: [40GC]CCTGAGACTTCACTTGCCTA AR.2R: GAGAAGTGCATGTGCAAGAC	289	73	58
3	AR.3F: TCAGGTCTATCAACTCTTGTA AR.3R: [40GC]GAAGGAGGAGGAAGAGAA	224	70	52
4	AR.4F: [40GC]ATTCAAGTCTCTCTTCCTTCC AR.4R: TATCTCATGCTCCCACTTC	386	74	55
5	AR.5F: CACTGCCTCTGCCTCTTCT AR.5R: [40GC]TCTGGCCAAGCTGCTGTAT	262	75	58
6	AR.6F: CCTCTGTGTATCTCCTTCC AR.6R: [40GC][7AT]CCTCTCTGAATCTCTGTGC	255	74	50
7A	AR.7AF: GTCTAATGCTCCTTCGTGG AR.7AR: [40GC]AGCGTCTTGAGCAGGATGT	219	71	58
7B	AR.7BF: [40GC]TCAAGGAACTCGATCGTAT AR.7BR: CTCTATCAGGCTGTTCTCC	190	75	52
8	AR.8F: [40GC]GGCCACCTCCTTGTCAAC AR.8R: CAAGGCACTGCAGAGGAGT	341	73	58
1A	CYP17.1AF: [6GC]CCACAGCTCTTCTACTCCA CYP17.1AR: [40GC]GCTTGAAGAAGTTGTTATGC	266	77	52
4	CYP17.4F: [40GC]TGGAAGAAGGGTGGATT CYP17.4R: CAGTGTGTAGAATGGACTCC	245	67	58
7	CYP17.7F: [40GC]CATGAGGCTGAGCAAGGAAG CYP17.7R: AGAGTCCAGGCTCGCTGTGT	261	77	58

bp, base pair.
GC-stretch (Wu *et al*, 1998) and GC-clamp used:
[7GC]CGCCCGC;[6GC]CGCCCG;[40GC]CGCCCGCCGCGCCCCGCGCCCGGCCCGCCGCCCCCG CCCG.
AT-stretches used: [6AT]TATAAT; [7AT]TATAATA.

A comprehensive DGGE assay was designed for the entire coding region of *CYP17*. Amplicon 1A covers the region from the CTG nucleotide in the 5' untranslated region (UTR) to base pair 364. Amplicon 4 includes base pair 186 in intron 3 to base pair 353 in intron 4, and base pair 177 in intron 6 to base pair 357 in intron 7 is the region which covers amplicon 7. For the partial analysis of exon 1 of the *CYP17* gene, the

forward primer was modified by the addition of a short GC-clamp of 6 nucleotides to its 5' end, to ensure optimal mutation detection.

For the amplification of small amounts of mutant DNA extracted from paraffin-embedded material, external nested primers were designed for the polymorphic regions of the *AR* and *CYP17* genes. Amplicons of less than 300bp in length were chosen for the purpose of a nested PCR. Direct PCR was possible but the success rate would have been reduced due to the low concentrations of DNA. Nested PCR was therefore the preferred means of amplification to obtain sufficient product for DGGE analysis. The primers used in the nested PCR are listed in Table 6. All nested amplifications were confirmed by repeated PCR to reduce the error rate created by *Taq* polymerase.

TABLE 6: External primer sets for Nested AR and CYP17- PCRs with their annealing temperatures and primer concentrations

Exon	Amplimers, 5'-3'	Size (bp)	Annealing Temperature (^{0}C)	Primer concentration (pmol)
1D	AR.1DextF: TCGTCCCTGTTGCACCTACC AR.1DextR: GATGCTCCAACGCCTCCACA	289	60	10
1	CYP17.1AextF: GCCTCCTTGTGCCCTAGAGT CYP17.1AextR: TTGGTGCCCATACGAACAGA	288	53	15
4	CYP17.4extF: GAGAGACTCTGGCAGCTGGA CYP17.4extR: ATGTGCCAGGTTCTCTGCTT	270	58	15
7	CYP17.7extF: GTGAGTGGGAATGAGGGAGTAA CYP17.7nR: CGCAGGACAGGACAGACTCA	248	60	10

bp, base pair.

Optimised DGGE conditions were used for broad-range analysis as laid out in Hayes *et al*, 1999. All amplicons of the *AR* gene were resolved by electrophoresis after multi-layering of the samples, in a 9% polyacrylamide gel containing a 35-75% urea and formamide (UF) denaturing gradient for amplicons 2-8 and a 50-90% UF

gradient for amplicons 1B-1G. The *CYP17* gene was also analysed using two different gel conditions, i.e. in a 9% PAA gel containing 50-90% (amplicon 1A and 7) and 10-50% (amplicon 4) UF denaturing gradients parallel to the direction of electrophoresis. 100% UF = 7mol/l urea per 40% deionised formamide. Electrophoresis was performed overnight at 110V at a constant gel temperature of 59^0C. The gels were stained with ethidium bromide and photographed under an ultraviolet transilluminator.

2.2.4 DNA Sequencing

The ultimate characterisation of a DNA segment is the determination of its nucleotide sequence. The ability to sequence DNA has added immensely to our understanding of gene structure and the mechanisms of gene regulation. One of the fluorescence-based methods for DNA sequencing uses fluorescently labelled dideoxynucleotides to terminate the extension of the sequencing primer and is known as the dye terminator method (David and Menotti-Raymond, 1998). On an Applied Biosystems sequencer, each different dideoxynucleotide is labelled with a different fluorophore (fluorescent dye), so that the sequencing reaction can be performed in a single tube containing an unlabelled primer and all of the four different dideoxynucleotides. Generally, DNA fragments of the same size generated in the sequencing reaction are labelled at the 3' end with the same dye. After performing the sequencing reactions, additional sample processing is generally required before loading the samples on the gel. The dye-labelled dideoxynucleotides must be removed (accomplished by ethanol precipitation or using a spin column). The samples are resuspended in a loading buffer and denatured prior to sample loading. The electrophoresis produces

a coloured pattern of peaks, representing the four bases that can be read to provide the sequence.

PCR products which showed mobility shifts on DGGE gels were purified using the High Pure PCR Product Purification Kit (Roche Molecular Biochemicals). Direct sequencing of these samples was performed on an ABI sequencer using a non-GC-clamped primer and the Big Dye Terminator V 3.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems). Variable banding patterns were confirmed by sequencing bidirectionally.

2.3 ASSOCIATION STUDIES

From a genetic point of view, a population is a local group of individuals who belong to a single species and in which mating is actually or potentially occurring. The set of genetic information carried by reproductive members of a population is referred to as the gene pool. Genes are the fundamental physical units of heredity and occupy specific positions on a chromosome known as a locus. Alleles are the alternative forms of the same gene that contain modified genetic information and often specifies an altered gene product. For a given locus, the gene pool includes all the alleles of that specific gene present in the population. One approach used to study a population's genetic structure is to measure the frequency of a given allele. This allele frequency equals the frequency of homozygote alleles plus half the proportion of heterozygotes within a population (Klug and Cummings, 1997a).

In an ideal population, a dominant allele does not automatically increase in frequency until it replaces its recessive counterpart. Such an infinitely large population, in which

random mating occurs in the absence of factors like mutation and selection, and, where genotype distribution is proportional to the individual allele frequencies and remains constant from one generation to the next, is the basic tenet of the cornerstone of the Hardy-Weinberg law (HWL). The HWL can be demonstrated as follows: when one thinks of a locus as having two alleles, A and a, and the frequency of allele A is represented by p, and the frequency of a is q, then p+q =1 because the sum of p and q represents 100 percent of the alleles for a particular gene in the population. A population that demonstrates the basic features of the HWL is said to be in Hardy-Weinberg equilibrium, which describes the expected relationship between allele frequencies and genotype frequencies of individuals in the same population group. In the state of genetic equilibrium the genotype frequencies remain constant from one generation to the next, at frequencies of p^2, 2pq and q^2. Unlike these hypothetical alleles, not all alleles in a population are in equilibrium. Populations in nature are dynamic, so changes in structure or size are part of their life cycles. The ideal situation, however, allows one to perform statistical goodness-of-fit tests where results obtained from a given sample are compared to a prespecified distribution (Thompson *et al*, 1991).

Statistical hypothesis tests are most commonly used in analysing and reporting the results of scientific studies. The Fisher exact test is a nonparametric test that is used when the independent or predictor variable is dichotomous, i.e. this variable is made up of two mutually exclusive groups or independent samples and there are no repeated measures. This test is used when both the independent and dependent variables are at the nominal level of measurement and is especially useful when sample sizes are small (McNemar, 1969). The main purpose of such a test is to

examine whether two populations differ from each other in the proportion of subjects who fall into one of two classifications. To undertake the Fisher exact test, the dichotomous data to be analysed are first cast into 2 x 2 contingency tables in which entries represent scores. The probability associated with all possible 2 x 2 tables that have the same row and column totals as the observed data are now evaluated, making the assumption that the null hypothesis is true. The null hypothesis here is that the row and column variables are unrelated. These probabilities are then used to calculate the overall probability of getting the observed data, or a less likely result, when the null hypothesis is true. The mathematical formula used to calculate each probability is rather complicated, so the calculation is much better done with the use of computer programs. A more common procedure has been described by Cox and Hinkley in 1974, where t=t(x) is a function of the observations and T=t(X) the corresponding random variable. T is called a test statistic for testing H_0 if certain conditions are satisfied. These criteria are that the distribution of t when H_0 is true is known at least approximately and the larger the value of t the stronger the evidence of departure from H_0 of the type it is required to test. For given observations x, t_{obs} = t(x) and the level of significance p_{obs} can be calculated by $p_{obs} = pr\ (T \geq t_{obs}\ ;\ H_0)$. The result is a number, $p_{obs,}$ called the level of significance, or p-value and the procedure is called a significance test. An essential component of significance tests is the concept of strength of evidence. A p-value is supposed to indicate "the strength of evidence against the hypothesis" (Fisher, 1958), with conventional interpretations as described by Burdette and Gehan in 1970, as shown in Table 7.

Table 7. Reasonable interpretations of the results of significance tests are as follows:

Significance Level of Data	*Interpretation*
Less than 0,01	Very strong evidence against null hypothesis
0,01 to 0,05	Moderate evidence against null hypothesis
More than 0,05 and less than 0,1	Suggestive evidence against null hypothesis
0,1 or more	Little or no real evidence against null hypothesis

In genetics it is important to be able to evaluate observed deviation. When we assume that data will fit a given ratio such as 1:1, 3:1, or 9:3:3:1, we establish the null hypothesis. The hypothesis assumes that there is no real difference between the measured ratios and the predicted ratios. The apparent difference can be attributed purely to chance. Evaluation of the null hypothesis is accomplished by statistical analysis. Based on this, the null hypothesis may either be rejected or fail to be rejected. If it is rejected, the observed deviation from the expected is not attributed to chance alone. If the null hypothesis fails to be rejected, any observed deviations can be attributed to chance. Therefore statistical analysis provides a mathematical basis for examining how well observed data fit or differ from predicted or expected events, testing the goodness of fit.

One of the simplest statistical tests devised to assess the goodness of fit of the null hypothesis is Chi-square analysis (χ^2) (Klug and Cummings, 1997a). This test takes into account the observed value of each component of an expected ratio as well as the sample size and reduces them to a single numerical value. This value (χ^2) is then used to estimate how frequently the observed deviation can be expected to occur strictly as a result of chance. The formula used in chi-square analysis is $\chi^2 = \Sigma \frac{(o - e)^2}{e}$, where o is the observed value for a given category and e is the expected value for that category. Σ (sigma) represents the "sum" of the calculated

values for each category of the ratio. As (o - e) is the deviation (d) in each case, the equation can be reduced to $\chi^2 = \Sigma \frac{d^2}{e}$

The final step in the chi-square analysis is to interpret the χ^2 value. To do this one must determine the value of the degrees of freedom (df), which is equal to n -1 where n is the number of different categories into which each datum point may fall. Degrees of freedom must be taken into account because the greater the number of categories, the more deviation is expected as a result of chance. The χ^2 value can now be interpreted in terms of a corresponding probability value (p). This calculation is complex and p values are usually located on a table or graph. Statistical programs are available, making chi-square analysis much easier.

A probability value of 0.05 is the relative standard set to serve as the basis for either rejecting or failing to reject the hypothesis. When applied to chi-square analysis, a p value < 0.05 means the probability is only 5% or less that the deviation in the set of results could be obtained by chance alone. Such a p value indicates that the difference between the observed and predicted results is substantial and thus serves as a basis for rejecting the null hypothesis. P values of 0.05 or > (1 – 0.05) indicate that the probability of the observed deviation being due to chance is 5% or more. The conclusion would therefore be not to reject the null hypothesis.

2.3.1 Data Comparisons and Statistical Analysis

(a) The occurrence and frequency of somatic mutations (if any) occurring in the CaP samples will be compared according to grade classification[29] so as to determine

[29] including any relevant information on metastasis or hormone resistance

possible involvement in tumourigenesis and if that involvement is either an early or late event in the progression of the cancer.

(b) The blood samples of patients who had mutations in their CaP samples, will be analysed genetically, for possible occurrence of germ-line mutations to clarify whether the genetic variation in the tumour sample was somatic.

(c) Genetic comparisons of tumour and blood material of the CaP patients and that obtained from the BPH (normal tissue), as well as blood samples from the general population will be performed, to determine the allele frequencies (by manual allele counting) of polymorphic sites. Chi-square and Fisher exact tests will be performed to compare genotype distribution and allele frequencies among patient[30] and control groups[31] . The statistical heterogeneity as assessed by Fisher exact and Chi-square tests were determined with the aid of computer programs, viz., Instat version 3.1 and JXK Chi-squared 2002 programs, respectively.

Data comparisons and Statistical analysis for the identified *AR* and *CYP17* SNPs are summarised in Appendix II. Raw data for the *CYP17* L36 SNP was not included since no extra controls were analysed at the time of writing-up and the polymorphism was found in only one individual from the patient cohort.

[30] Tumour and blood from grade 1,2 or 3 prostate carcinoma, with or without metastasis and with or without hormone resistance

[31] Patients with BPH, with or without prostatitis

2.4 RESULTS

AR-DGGE assay

A comprehensive mutation detection assay was developed for almost the entire coding region (excluding polymorphic repeat sequences) and including all splice site junctions of the *AR* gene. This coding region was divided into 14 amplicons for analysis and due to its large size, exon 1 was divided into six overlapping fragments (1B-1G) and exon 7 was divided into two overlapping fragments (7A and 7B). Examples of the normal banding patterns are shown below in Figure 8.

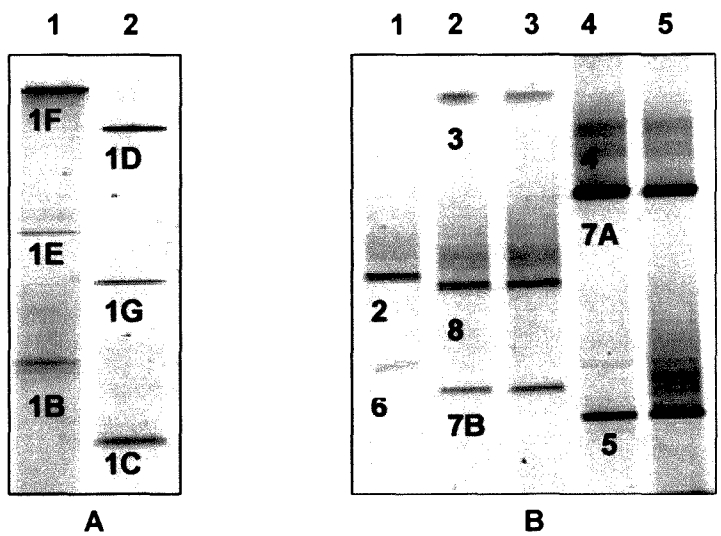


FIG.8. The denaturing gradient gel electrophoresis patterns of 14 pooled PCR amplicons of the *AR* gene. The amplicons 1B-1G (**A;** lanes 1 and 2) and 2-8 (**B;** lanes 1-5), were analysed on single 9% PAA gels with 50-90% and 35-75% UF gradients, respectively. **A.** Amplicons 1F, 1E, 1B (lane 1) and amplicons 1D, 1G and 1C (lane 2) can be layered. **B.** The layering of pooled amplicons 2 and 6 (lane 1); 3, 8 and 7B (lanes 2 and 3); 4, 7A and 5 (lanes 4 and 5).

The assay allowed for the complete analysis of 13 patients on a single gel for amplicons 1B-1G and 9 patients per gel for amplicons 2-8. Since the AR gene is

single-copy, mutant PCR products were mixed with standard controls before heteroduplexing for easier visualisation of aberrant bands.

AR mutations

Using the *AR*-DGGE assay and the primer sets in Table 5 for the analysis of 20 CaP, 25 BPH and 6 tissue samples of unknown clinical diagnosis, five germline mutations were identified, including a previously reported polymorphism. Direct sequencing of the aberrant DGGE banding patterns confirmed these variant sequences which are listed in Table 24 according to the intron or amplicon in which they occur.

TABLE 24. List of identified *AR* mutations ordered according to the intron or amplicon in which they occurred

Int/Amplicon	bp / codon	Mutation	Amino acid change	Previously reported
Int 1	-38 bp	g-a	—	Novel
1D	211	GAG-GAA (E211)	Silent	Yes
1E	242	TCG-TCT (S242)	Silent	Novel
1F	Nd			
1G	341	TCT-TCC (S341)	Silent	Novel

Int, intron; -bp, number of base pairs downstream from acceptor splice site.
E211: silent codon 211 change coding for glutamate (E), S242: silent codon 242 change coding for serine (S), S341: silent codon 341 change coding for serine (S).
Novel indicates new mutations reported in this study; Yes indicates previously reported mutation.
Nd, not determined.

One mutation was located in intron1 (IVS1-38G>A) and the E211, S242 (amplicon 1E), amplicon 1F and S341 (amplicon 1G) variants were all found in exon 1. No mutations were found in the remaining exons 2 through to 8. The

single base E211 substitution occured at a frequency of more than 1% within the various populations (Table 26) and was therefore classified as a single nucleotide polymorphism (SNP). Examples of these DGGE banding patterns are shown in Figure 9. The E211 SNP in lanes 2 and 3 (Figure 9B) and the mutations in lane 1 (Figure 9A, C, D and E), appear heterozygous on the gels, due to mixing of normal control DNA to the patient DNA prior to heteroduplexing.

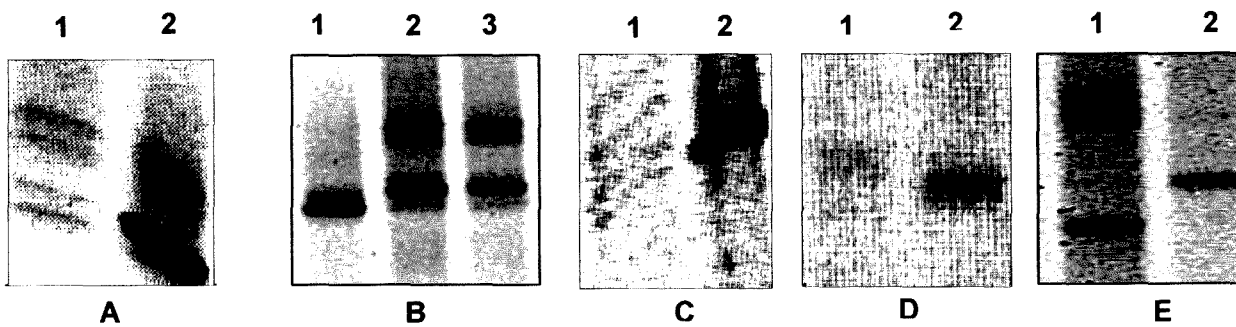


FIG.9. Aberrant denaturing gradient gel electrophoresis banding patterns found in the *AR* gene. All variants are heterozygous on the gels and three of these variant banding patterns are represented as four bands; two upper heteroduplexes and two lower homoduplexes (A, B and C). **(A)** The IVS1-38G>A variant is shown in lane 1 with the normal control in lane 2. **(B)** The E211 mutation (lanes 2 and 3) as compared to a normal control in lane1. **(C)** S242 variant (lane 1), normal control DNA (lane 2). **(D)** Amplicon 1F mutation (lane 1) and **(E)** the S341 mutation (lane 1) are seen as three bands due to the presence of more mutant than normal cells resulting in loss or weaker intensity of the homozygous normal band. Normal controls for amplicon 1F mutation in lane 2 (D) and the normal control for the S341 variant in lane 2 (E).

The variation that was observed in amplicon 1F could not be detected in the sequencing data. For sequencing of this variant the reverse primer without the GC-clamp was used. Since the forward primer was designed with a GC-clamp, it could not be used for sequencing. One could redesign the forward primer without a clamp and use it for sequencing.

All these mutations occur in the NH_2-terminal region of the AR, except for the IVS1-38G>A variant. A splice site usually occurs at the position 38bp downstream from this intron 1 variant and therefore one cannot exclude the possibility that it could affect the splicing mechanism. The IVS1-38G>A change and the S242 mutation co-existed in a cancer sample (Gleason grade 2) of a 75-year-old patient of Coloured descent. The S341 mutation occurred in 2 BPH samples, one from a 64 year old Black patient and the other from a patient of Coloured origin of age 74. These same BPH patients also have the amplicon 1F mutation, indicating that the S341 variant and the amplicon IF change could be linked polymorphisms that may be in linkage disequilibrium i.e. associated with the disease locus. Both the S242 and S341 mutations resulted in no amino acid change and are therefore silent. The average age of these patients with sequence variants is 71 years.

The previously reported *AR*-E211 SNP was found at a relatively high allelic frequency in the Black and Coloured populations, at 44/121 (36%) and 42/120 (35%), respectively, and at a lower frequency of 16/88 (18%) within the White population group (combining frequencies of males and females for all three population groups). The allelic frequencies in the different population groups were obtained by analysing the E211 polymorphism in the patient cohort in addition to samples from the Western Province Blood Transfusion Service. For the *AR* E211 SNP, there were 41 Black females and therefore 41x2 alleles = 82. There were 39 Black males and therefore 39 alleles since males only have one

copy of the *AR* gene, which is located on the X chromosome. The combined allele number for males and female Blacks was therefore 39+82 =121 (Table 26). The total number of alleles was calculated the same way for Coloureds and Whites. Although the frequency of this allele in CaP versus BPH was higher, at values of 17/52 (32%) to 7/26 (26%), we found a *P*-value of 0.1876, considered not statistically significant. Further statistical analysis yielded a Chi-square value of 1.529 and a resulting probability of 0.465. This single nucleotide polymorphism has previously been reported in individuals with varying degrees of androgen insensitivity syndrome (Batch *et al*, 1992; Gottlieb *et al*, 1999) and male infertility (Wang *et al*, 1998b). The research team of Ross and his colleagues (1998) have already tested the SNP as a potential marker for CaP susceptibility. Results from their study show the SNP to be statistically significant with a near 3-fold increased risk of CaP among men under the age of 65. This, however, does not correlate with findings from the patient cohort in this study where no statistical significance with disease association was found.

Mutations in paraffin-embedded prostatic carcinomas

In order to analyse the distribution of the *AR*-E211 SNP within the Black population specifically, our *AR*-DGGE assay was further tested on 30 paraffin-embedded CaP samples in a nested PCR-amplification, using the external nested primers as shown in Table 6. Two sequence variations were found in amplicon 1D of the *AR* gene and these variant banding patterns are shown in Figure 10.

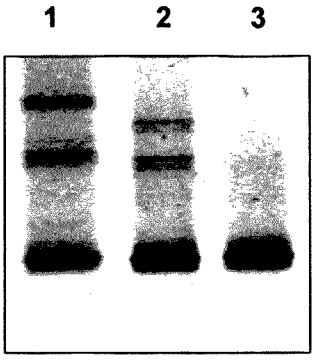


FIG. 10. The denaturing gradient gel electrophoresis (DGGE) gel shows mutations that occur in amplicon 1D of *AR* after nested amplification of DNA extracted from paraffin-embedded prostate tumours. Lanes 1 and 2 represent DGGE variations, showing three bands. The lower homoduplex bands appear as a single, more prominent band. The control DNA was run in lane 3.

From lanes 1 and 2 on the gel, it is clear that one cannot distinguish between the normal and mutant homozygous alleles. This could be because mutant alleles are exhausted by competition during formation of the heteroduplex hybrid bands. The resulting three bands on the gel could more likely be from the homoduplex bands that are lying on top of each other, due to the band intensities. This sequencing data was incomplete at the stage of writing-up.

CYP17 DGGE assay

A comprehensive DGGE assay was designed for the entire coding region of *CYP17*. Using this assay, three novel *CYP17* SNPs were identified (Hayes and de Beer, Communication). The three amplicons 1A, 4 and 7, where these SNPs were located in the CaP, BPH and control population, were screened to determine the allele frequencies, as well as search for possible associations with CaP risk . The novel L36 SNP, and the H46 and S65 variants were found in exon 1. Examples of these aberrant *CYP17* DGGE banding patterns are shown in Figure 11.

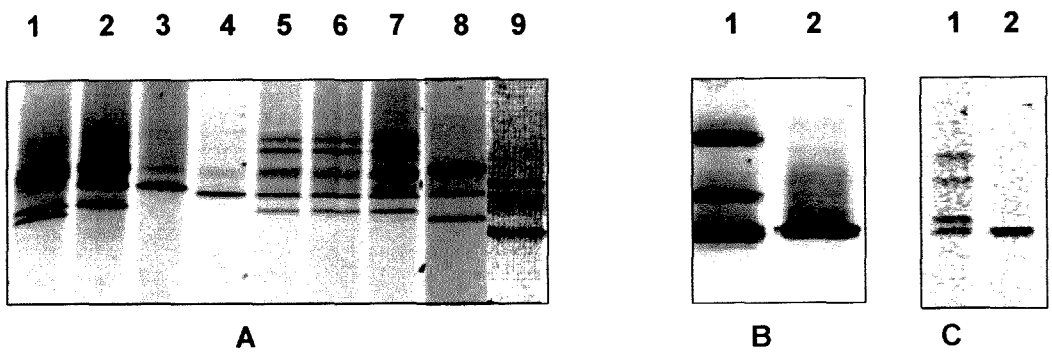


FIG.11. The DGGE banding patterns of *CYP17* SNPs. **(A)** Amplicon 1A was analysed on a 9% PAA gel containing a 50-90% UF gradient. The heterozygous S65 variant can be seen in lane 2 with the homozygous form in lane 3. The homozygous H46 SNP can be visualised in lane 4 with the heterozygous form of the mutation in lanes 5, 6 and 7. Combinations of SNPs include the heterozygous H46 and –34T>C variants in lane 1 and the homozygous S65 and heterozygous H46 variants in lane 8. The heterozygous L36 variant can be seen in lane 9. *CYP17* amplicon 4 **(B)** and 7 **(C)** DGGE analysis was performed on 9% PAA gels with 10-50% and 50-90% UF gradients, respectively. **(B)** Lane1 shows the IVS4+58G>C variant and is represented by three bands viz., 2 heteroduplex bands and 2 homoduplex bands which almost have the same melting behaviour and mobilities and therefore appear as a single slightly more diffuse band. Lane 2 consists of normal control DNA. **(C)** Lane 1 contains the IVS6-25C>A substitution and is represented by 2 upper heteroduplex bands and 2 lower homoduplex bands. Control DNA is seen in lane 2.

Direct sequencing of the *CYP17* exon 1 variants showed that they all represent silent mutations as listed below in Table 25.

TABLE 25. *CYP17* mutations ordered according to the region or exon in which they occurred

Intron/Exon	bp / codon	Base change	Amino acid change	Previously reported
4	+58 bp	T-C	—	Novel
6	-25 bp	T-A	—	Novel
1	36	CTG - CTC (L36)	Silent	Novel
1	46	CAT - CAC (H46)	Silent	Yes
1	65	TCT - TCG (S65)	Silent	Yes

+bp: number of base pairs downstream from exon 4, -bp: number of base pairs upstream from exon 7.

L36: silent codon 36 change coding for Leucine (L), H46: silent codon 46 change coding for Histidine (H), S65: silent codon 65 change coding for Serine (S). Novel indicates new mutations reported in this study.

From the sequencing data, a previously reported SNP was detected in the 5' untranslated region of the *CYP17* gene. This SNP, a T/A -C/G transition, is found 34bp upstream from the initiation of translation and 27 bp downstream from the transcription start site and creates an additional Sp-1 type (CCACC box) promoter site. Results from a study by Kittles, *et al* (2001a), suggest that the C allele of the *CYP17* SNP is significantly associated with an increased CaP risk and a clinically advanced form of the disease in African Americans.

Allele frequencies of single base substitutions

The allele frequencies of the identified polymorphisms within the patient cohort and additional samples from the WPBTS (Refer to Table 1; Section 2.1.2), was analysed. This was used to obtain the allele frequencies within the different population groups. These results are shown in Table 26.

The allelic variants found in the *CYP17* gene are all substitutions of a single base and from the data in Table 26 it can be seen that they occur at frequencies of 1% or more within the various populations. For allelic frequency data and comparative analysis of CaP versus BPH for identified SNPs (Fisher exact and Chi-square analysis), refer to the raw data in Appendix II.

Three silent mutations are located within exon 1 of the *CYP17* gene. A heterozygous *CYP17* L36 substitution was found in a patient of mixed ancestry diagnosed with carcinoma of the prostate assigned a Gleason Grade of 2. This rare SNP occurs at an allelic frequency of 1% (1/70). The H46 SNP in *CYP17*,

was identified in individuals from all three ethnic backgrounds, at allelic frequencies of 40% in the Africans (24/56) and Coloureds (42/110) and at a slightly elevated frequency in Whites at 50% (11/22). The H46 SNP occurred more frequently in the cancer samples, at 50% (21/40), compared to 38% (21/56) in BPH samples. Further statistical analysis showed no significance (Fisher exact $P = 0.1408$) and yielded a $\chi^2 = 0.0057$ and p (probability) of 0.997.

The third silent polymorphism was located in the *CYP17* gene in codon 65 (S65) and was found exclusively in the Coloured population at an allelic frequency of 10% (15/110). This SNP was only present in BPH patients of Coloured ethnicity at 13% (7/56) and absent in all individuals of Coloured origin with CaP. The comparison of the frequency of S65 among CaP and BPH patients was performed to see if there was a possible association of this SNP with CaP risk. The result of the comparison was a *P*-value for statistical significance of 0.0174 (considered to be significant) when the numbers for individuals of all race groups were included ($\chi^2 = 2.53$, $p = 0.282$). By excluding the Blacks and the Whites from the equation, we applied the Fisher exact test again and found a *P*-value of 0.0053 (considered to be very significant), indicating a possible association with a reduced CaP risk in Coloureds.

TABLE 26. Allelic frequencies of Single Base Substitutions identified in the *AR* and *CYP17* genes within our patient cohort and including additional samples from the WPBTS (controls), with their possible association to Prostate cancer susceptibility.

	*Number of Individuals			Allelic frequencies			Possible association with CaP risk		
Sequence Variants	Blacks	Coloureds	Whites	Blacks	Coloureds	Whites	CaP	BPH	*P*-value
AR E211 Females	41	28	25	33/82	19/56	11/50	—	—	—
				0.40	0.34	0.22			
AR E211 Males	39	64	38	11/39	23/64	5/38	17/52	7/26	
				0.28	0.36	0.13	0.32	0.26	0.1876
CYP17 L36	5	35	11	0/10	1/70	0/22	1/40	0/56	—
				0	0.01	0	0.02	0	
CYP17 H46	28	55	11	24/56	42/110	11/22	21/40	21/56	
				0.4	0.4	0.5	0.5	0.38	0.1408
CYP17 S65	29	55	11	0/58	15/110	0/22	0/40	7/56	0.0174
				0	0.1	0	0	0.13	Significant
CYP17 IVS4+58G>C	85	87	13	8/170	7/174	0/26	3/98	3/52	0.3530
				0.047	0.04	0	0.037	0.058	
CYP17 IVS6-25C>A	68	65	12	4/136	11/130	2/24	5/98	4/52	0.2545
				0.027	0.084	0.083	0.05	0.07	

Sample numbers vary between the different population groups and sample types. Analysis was performed as samples were collected. The patient samples consisted of prostatic carcinoma and benign prostatic hyperplasia tissues, as well as blood from the same patients. Controls from the WPBTS was DNA isolated from blood samples. * Number of individuals = patients + controls

Two intronic polymorphisms were also found in *CYP17*. The intron 4 SNP (IVS4+58G>C) was identified exclusively in an African based population at similar allele frequencies of 4.7% (8/170) in Blacks and 4% (7/174) in Coloureds. An elevated frequency of 5.8% in BPH versus 3.7% in CaP samples was observed even though statistical analysis showed no significance (Fisher exact *P* = 0.353). A χ^2 value of 2.45 was obtained with p = 0.293. An intron 6 transversion (IVS6 –25C>A), where a pyrimidine (C) was substituted for a purine (A), was found in all three population groups at high allele frequencies of 8.2% (11/130) in Coloureds and 8.3% (2/24) in Whites and at a lower frequency of 2.9% (4/136) in Blacks. The frequency of the A allele of this SNP was higher in BPH (7%) than in cancer (5%), even though sample numbers were so low. Further statistical analysis showed a χ^2 value of 2.65 with a corresponding probability of 0.265.

2.5 DISCUSSION

CaP is the second most common histologically-diagnosed cancer in South African males (Sitas, 1994). Since the prostate is an androgen-regulated organ, there has been a tremendous interest in the role that androgens play in prostate carcinogenesis. The two most important apparent risk factors for CaP are age and race or ethnicity. CaP is very rare before the age of forty but the subsequent rate of increase with age is greater than for any other cancer. African-American men have been reported to have the highest rates of CaP worldwide. There has been no appropriate comparative data on CaP rates among African Blacks, however, there are findings which suggest high rates of CaP in some African populations (Kehinde, 1995). Various ethnic and cultural groups also exhibit different environmental factors (diet, lifestyle, effects of migration, etc), levels of genetic variation and patterns of gene-environment interactions. A combination of all these factors could play a key role in determining CaP predisposition.

Two types of germline variants predispose to cancer (Ross and Coetzee, 1996). Some germline mutations directly predispose to the disease. These single inherited genetic traits are uncommon in the population and therefore carry low population attributable risks. Genes containing these variants are, however, highly penetrant and individuals with these genetic traits could have a high overall risk for disease development.

Susceptibility genes are much more common and have a low overall risk but a possible strong population influence, since the higher risk genotype could be relatively common and maintained in the population. Androgen metabolism genes like these, in combination with one another are highly penetrant and could indirectly affect the risk of cancer.

Part of this study was to design a comprehensive mutation detection assay for almost the entire coding region (excluding polymorphic repeat sequences), and including all splice site junctions of the *AR* gene in order to identify possible SNPs in our population. Secondly, the identified *CYP17* SNPs were analysed within our South African cohort (the assay was already up and running), to investigate possible associations with CaP risk. (DGGE was our method of choice for SNP detection because it is virtually 100% sensitive and also suitable for the analysis of archival-embedded paraffin material).

AR gene: type and frequency of variants

Five mutations were identified in the *AR* gene, including a previously reported E211 SNP. No additional SNPs were found. All the sequence variants represented germline events. Except for the intronic base substitution (g-a) 38bp downstream from the acceptor splice site in intron 2, the remaining four amino acid substitutions all occur within the coding region of exon 1. Three of these four variations represent silent changes.

A previous study by Coetzee and Ross (1994) have involved the analysis of the polymorphic CAG tract in exon 1, since AR activity seems to be negatively associated with CAG length. Shorter CAG alleles may be associated with CaP (Hardy *et al*, 1996; Ingles *et al*, 1997; Stanford *et al*, 1997). Additional studies even identified phenotypic subgroups among CaP cases based on CAG repeats (Giovannucci *et al*, 1997; Hakimi *et al*, 1997). Another study analysed the CAG and GGC repeats in addition to the previously identified E211 SNP, located roughly between the two polymorphic markers (Ross *et al*, 1998 and 1999). The SNP was found to be associated with a statistically significant almost 3-fold increase in CaP risk among African-American men under the age of 65 (n = 208, healthy men without CaP were used as controls). In this study the *AR*-E211 SNP was found to be more common in the Coloured (n=92) and African (n=80) than in the White population group (n=63). Although sample numbers were small, when comparing the allele frequency in CaP (n=52) and BPH (n=26) for this SNP, our observation conflicts with the study of Ross and Coetzee (1998) in that no statistically significant association was found. The low frequency of this allele in our prostate carcinoma patients is a discrepancy that may be explained by our small sample size, which creates a major analytical limitation.

The silent S242 variant and the intronic base substitution, IVS2-38G>A, were found in the same 75 year old patient of mixed ancestry. After two biopsies were taken from this patient a Gleason grade 2 prostate carcinoma was diagnosed. Tumours of this grade are well-differentiated, rarely have metastases, are

typically slow growing and is not life threatening to the patient. The CaP samples in our cohort (excluding paraffin material) have been assigned with low Gleason grades of 1 to 3 which are indicative of generally slow growing cancers. Our findings of such few mutations within our cancer samples therefore correlates with various other studies which found that *AR* mutations are very rare in localised disease and occur more frequently in metastatic tissue.

Age and functioning testis are the only known etiological risk factors for BPH (Lee *et al*, 1997a). This condition develops when mean testosterone levels are falling. The role of androgens in BPH is therefore likely to be facilitatory rather than causative. In BPH epithelium the AR is mostly expressed in luminal cells rather than basal cells, while AR expression in the stroma has been detected in smooth muscle and fibroblasts (Cooke *et al*, 1991). There are almost no studies reporting AR mutations in BPH. In one study, the genetic and immunodetection of p53 gene[32] mutations were compared in malignant and benign prostate tissues. Two reference CaP cell lines, 26 patient specimens (including 8 BPHs, 16 CaPs and 2 lymph node metastases), 1 prostate and 9 kidney cell lines were used for p53 analysis. The *TP53* status of the samples were characterised using single-strand conformational polymorphism (SSCP) analysis. This approach identified 14 of 15 (93%) cell lines and patient samples having *TP53* missense mutations in the exons 5 to 8 region (Wertz IE *et al*, 1996). More recently, another group evaluated the correlation between specific genotypes in androgen and oestrogen-regulating genes (*AR*, *SRD5A2*, *CYP17* and *CYP19*), and age-

related prostatic changes. The genetic susceptibility to morphological and pathological criteria in 195 French Caucasians were tested, using allelic variants for candidate genes involved in androgen/oestrogen prostatic activity: androgen receptor (CAG repeats), 5alpha-reductase type 2 (TA repeats, V89L and A49T mutations), A2 variant of the 17alpha-hydroxylase (*CYP17*) and the simple tandem repeat polymorphism (*STRP*) and the aromatase *(CYP19)* polymorphisms. Results from this study suggested that common variants of the *CYP17* gene are associated with prostate enlargement and therefore may increase the risk of BPH development in this population, while infrequent variants of the *CYP19* gene could be of a protective nature (Azzouzi AR *et al*, 2002). Our results support the findings of other studies which have used BPH as a factor of comparison and have found either no mutations or detected silent mutations (Elo *et al*, 1995; Evans *et al*, 1996). Currently, the focus of comparison between CaP and BPH has been in protein expression studies. One study compared serum biological androgen activity between men with newly diagnosed CaP and age-matched men with BPH (Raivio T *et al*, 2003), while another group determined the luteinizing hormone-releasing receptor gene mRNA expression in CaP and BPH tissue (Straub B *et al*, 2003). In general, caution should be taken when comparing CaP to BPH to determine risk of developing prostate cancer, since the relationship between BPH and CaP is not well established. There are a number of similarities that make it difficult to disprove a relationship as both are common in ageing men, both diseases show androgen dependence and may be under comparable genetic control. BPH has been linked to CaP diagnosis in some

---

[32] *TP53*, the most commonly mutated gene in human cancers

studies while other association studies have failed to find an association between BPH and CaP (Brawley *et al*, 1998b). It is very possible that BPH does not increase the risk of developing clinically significant CaP but increases the risk of being diagnosed with CaP. Here we report an unsequenced variant in exon 1 and a silent S341 mutation, both occurring in clinical BPH specimens from a Black patient (aged 64) as well as another 74 year old patient of Coloured ethnicity. BPH values[33] of 35 and 20 were assigned to these patients, respectively, after digital rectal examinations (DRE). Histopathological examination of the samples verified the diagnosis of BPH and both individuals still displayed symptoms following transurethral prostatic resection. The occurrence of the two sequence variants in both patients suggests that they could be linked polymorphisms but confirmation is limited by the small sample size. One would have to increase the number of patients in the study cohort to determine the frequency of this variant. Of a total number of eleven White patients, eight were diagnosed with BPH and none had the unsequenced variant or the S341 mutation. Both these mutations were identified in one of two Blacks with BPH and in one of fifteen Coloureds with the condition and therefore appear to be restricted to an African-based population.

The frequency of *AR* somatic mutations in CaP varies between studies and has ranged from very rare to as high as 44% in clinical specimens of advanced tumours (Tilley *et al*, 1996). A general finding has been that the frequency of AR somatic mutations increases with stage, being low in organ confined disease

[33] This is the estimated mass of the prostate in grams, estimated by the doctor on DRE

(Tilley *et al*, 1994; Takahashi *et al*, 1995; Taplin *et al*, 1995) and higher in metastatic disease (Culig *et al*, 1993; Taplin *et al*, 1995; Tilley *et al*, 1996). Many investigators analysed only exons 2 through 8. Tilley and colleagues (1996), however, have shown that 50% of their amino acid substitutions identified in the AR were within exon 1. This partly explains the discrepancies in the reported mutation frequencies since exon 1 accounts for more than half of the *AR* gene and encodes the entire transactivation domain of the receptor. Similarly, we identified four germline sequence variants that were all located within exon 1.

Due to the few *AR* mutations we identified, different sections of the same CaP samples were sent to pathologists for further analysis to confirm tumour-containing status of the material. The sensitivity of mutation detection was considered by using DGGE, which allows for optimal detection of mutations by increasing the percentage of single base changes detectable to virtually 100%. Also, by DGGE, as little as 5% of mutant DNA (Hovig *et al*, 1992) can be detected in "tumour material", while some studies have even detected as little as 0.1% of mutant DNA (Keohavong *et al*, 1997). *AR* SNPs, however, are very rare and the previously reported E211 SNP is currently the only SNP that has been detected in the coding region of the gene.

Failure to use tumour-enriched DNA could account for our findings and those that some authors have published in the past. Reports from the clinicians did not specify whether or not specimens were enriched for tumour before analysis. The

presence of a substantial proportion of nontumour tissue may very likely hamper attempts to detect a mutation. It is probable that more sensitive techniques, such as laser microdissection, could have improved the tissue integrity for mutation detection. There were, however, additional paraffin-embedded CaP tissue samples from Blacks to screen. The fixation process of these samples damages the DNA and a nested-PCR is required in order to amplify sufficient amounts of DNA. This experiment is quite time-consuming and nested-PCR is prone to contamination. One would only be able to detect contamination after two rounds of PCR and testing of the PCR products on an agarose gel. Taking this into consideration, we decided to screen only those samples with polymorphism positive status. It is therefore possible that a number of base changes within the other regions, could have been missed.

The important roles of androgen in the differentiation of prostate tissue and the development and progression of CaP make the AR a link in the multistep pathway to oncogenesis. As with other multifactorial cancers, CaP risk may not be as a result of a single (or small number) of yet unknown predisposing genes, but rather due to a large number of common but weaker genes that collectively lead to a patients' "risk profile".

CYP17 gene: type and frequency of variants

Using the *CYP17* DGGE-assay four SNPs viz. H46, S65, intronic substitutions IVS4+58G>C and IVS6-25C>A, as well as the L36 mutation were identified. The

heterozygous L36 mutation, identified in one Coloured patient with CaP, is very rare and its frequency will probably decrease when screening larger numbers. The H46 SNP is very common and occurs at a relatively high allelic frequency of 40% in Coloureds (42/110) and Blacks (24/56), and at a slightly higher frequency of 50% in Whites (11/22), though sample numbers for whites were very low. H46 is present in both the heterozygous and homozygous forms and from sequencing data, it was found that the heterozygous H46 SNP occurs in combination with the previously reported 5'UTR-34T>A variant as well as the homozygous S65 SNP. When testing for an association with CaP risk, the H46 SNP was present at a slightly higher proportion in the CaP group compared to BPH samples, though not at statistical significance level. The intronic IVS4+58G>C SNP was found at allelic frequencies of 4.7% in Blacks (8/170), 4% in Coloureds (7/174), and absent in Whites and therefore appears restricted to African-based populations. The other intronic SNP, IVS6-25C>A, is present in all three population groups at similar higher allelic frequencies of 8.2% in Coloureds (11/130) and 8.3% in Whites (2/24) when compared to 2.9% in Blacks (4/136). Numbers, however, for Whites were quite low. Both the IVS4+58G>C and IVS6-25C>A SNPs were present at higher frequencies in BPH when compared to CaP and no association with CaP was found. Small sample size is definitely a limiting factor when analysing these SNPs. Therefore possible associations should be further evaluated in a larger sample group.

Our results reflect that there is a low risk of CaP development within our South African cohort. This, however, may be due to sampling bias, i.e. the small number of cases in the study. Difficulties in involving the various ethnicities or heterogeneous reproducibility of phenotype assignment related to anatomical and clinical definition in cases and controls could explain the results. The extent of this problem could be so great that the frequency of any SNP contributing to a disease phenotype will only be slightly elevated in the disease group compared with unaffected controls. Association studies with a large sample size, where cases of disease are compared to controls from the same population, are likely to give a greater chance of detecting small effects. Prostate cancer is selected in the first step on PSA, an androgen dependant parameter that could cause a bias in the relation observed between prostate carcinoma risk and genes involved in androgen pathways. On the other hand, it is possible that for a single gene there could be situations in which an individual carries both a high- and a low-risk marker at different loci resulting in no overall difference in risk. We did, however, detect a silent S65 SNP that was significantly different in controls and cases, as revealed by statistical analysis. The S65 allele was only found in the mixed ancestry group and also exclusive to the BPH controls. Chi-square analysis yielded a value of 2.53 ($p = 0.28$), indicating that there is a 28% probability that the observed deviation is due to chance, giving us no reason for rejecting the null hypothesis. As individuals with BPH have never been shown to develop CaP, this *CYP17* genotype (S65) may be associated with a reduced risk of prostate carcinogenesis in Coloureds.

Complex multifactorial diseases such as CaP may not be amenable to characterisation by linkage analysis and positional cloning, as is the case for monogenic diseases such as cystic fibrosis, sickle cell disease, etc. The results presented in this dissertation therefore form a basis for further investigation as there is a new emphasis on using genome-wide scans to look for single nucleotide polymorphisms, including short deletions, insertions and multinucleotide changes. In light of the difficulties surrounding the genetics of CaP, technologies such as microarrays (DeRisi *et al*, 1996) provide possibilities for the identification of new target genes involved in CaP. With improvements in technology and the collection of family data, novel methods for the statistical modelling of CaP will be key to the understanding of the multifactorial nature of the condition. CaP still remains one of the major health problems in the world and therefore the understanding of the genetics of this disease will ultimately lead to better diagnostic and therapeutic strategies.

The greatest limitation of this study was the lack of patients and tissue and blood samples that were studied. To perform association studies of any statistical relevance would require much larger numbers of cases (CaP versus BPH) and controls within each of the population groups to be studied. Statistical programs like Epi Info[34] and Statistica Power Analysis[35] can be used to find the ideal sample size to enrich research with a variety of tools for estimating confidence intervals and conducting comprehensive power analysis to avoid Type II errors.

---

[34] Version 6, Nov 1993. © 1999, Henry Ford Health System

[35] © Statsoft, Inc 1984-2004

In conclusion, comprehensive DGGE assays were designed for the analysis of the entire coding region of the *AR* and *CYP17* genes. Using these assays, novel sequence variants were identified, including 5 *AR* and *CYP17* SNPs. One *CYP17* SNP appears to be exclusive in an African-based population and another possibly reduces CaP risk in Coloureds. Using a patient cohort of adequate size, these SNPs can be used as markers in determining CaP risk.

2.6 FUTURE PROSPECTS

Polymorphic trinucleotide repeat analysis

Since the length of the AR trinucleotide CAG and GGC repeats in the South African population remains unknown, future analysis of these polymorphic tracts should be performed since case-control studies have reported associations between CaP risk and one or both the trinucleotide repeat lengths.

Possible candidate genes and multigene interactions

Other genes involved in androgen metabolism have also been implicated in the aetiology of CaP. Polymorphisms in *CYP3A4*, involved in the oxidative deactivation of testosterone and possibly the 3α-, 3β-, and 17β-hydroxysteroid dehydrogenases, involved in DHT inactivation and testosterone biosynthesis, may influence prostate carcinogenesis (Reichardt *et al*, 1995; Devgan *et al*, 1997; Ingles *et al*, 1997) and require further investigation.

Possible associations may therefore be explained by the fact that relevant interactions among multiple genes and exposures have not been studied. Thus complex interactions among these genes in the androgen metabolism pathway may be required to understand their effects in CaP aetiology. As a result, associations found may not only be due to linkage disequilibrium alone, but could also be attributed to some combination of linkage disequilibrium and multigene interactions. This should include a formal statistical evaluation of genotypes. Larger sample sizes may therefore be required to evaluate the potential interaction effects.

Developing a polygenetic model for CaP by incorporating multiple loci from individual genes could maximise the ability to identify individuals who are at high risk of developing carcinoma of the prostate. Little is known about molecular markers that may help early diagnosis of this carcinoma and the identification through genetic variation, of population subgroups that are at elevated risk of developing malignant prostate tumours could be of importance.

Alternative mutation detection technique for SNP identification

Denaturing high-performance liquid chromatography (DHPLC) is an efficient method for detection of mutations involving a single or few number of nucleotides, and it has been successfully used for mutation detection in disease-related genes (Mogensen *et al*, 2003). DHPLC analysis is cost-effective and allows a rapid throughput of large numbers of samples in minimal time.

Therefore it should be considered when possible large-scale population screening is likely in the event of a variant posing a high-risk.

Haplotype analysis and functional studies

Unfortunately, due to the limited sample size and lack of material from families, haplotype analysis could not be performed to "track" the transmission of variant alleles. Functional studies on the variants identified in this thesis could also not be performed by collaborators (The Norris Comprehensive Cancer Centre, University of Southern California), due to the restricted sample size. However, this study is being performed on a larger cohort by Dr Vanessa Hayes (Project Leader: Genitourinary Cancers, Garvan Institute of Medical Research, Australia), for publication purposes.

Animal models

Transgenic mouse models are proving to be invaluable in the effort to understand the molecular basis of metastatic CaP. In general, transgenic models feature higher penetrance, faster progression, and greater overall reproducibility of metastatic properties compared to spontaneous or induced models. The autochthonous transgenic adenocarcinoma of the mouse prostate (TRAMP) model, which spontaneously develops metastatic CaP, is one such model that mimics progressive forms of human disease. A study performed by Gupta and colleagues (2001), employing male TRAMP mice, showed that oral infusion of a polyphenolic fraction isolated from green tea (GTP) at a human achievable dose

(equivalent to 6 cups of green tea per day), significantly inhibited CaP development and increased survival in these mice. In order to elucidate the molecular mechanisms of CaP metastasis, it will be necessary to compare gene and protein expression patterns and biochemical analyses of clinical metastatic disease with data obtained from current models.

A consideration in the design of new models would be the ability to track the progression and spread of metastases to distant sites in live animals using new reporter genes and improved imaging technology. Assaying classic reporter genes such as galactosidase and alkaline phosphatase commonly requires the sacrifice of the animal. However, because metastatic spread occurs at variable time points and targets a variety of secondary sites, new live imaging technologies will be advantageous in eliminating some of the guesswork in animal sacrifice and tissue collection, hereby allowing investigators to use each animal to its full potential (Winter *et al*, 2003).

Genomic approaches to the development of CaP diagnostics

Combining proteomic technology with artificial intelligence based bioinformatics may be a powerful tool in the detection and diagnosis of CaP. A technique still under evaluation, based on the analysis of protein patterns, will be useful in the future for deciding whether men with marginally elevated PSA levels should undergo biopsy (http://www.cancer.gov, 2004). This diagnostic test relies on computer software that detects key patterns of small proteins in the blood. Mass

spectroscopy is used to sort proteins and other molecules based on their weight and electrical charge. Artificial intelligence is then used to train a computer to identify patterns of proteins that differed between patients with CaP and those in which a biopsy had found no evidence of disease. This new approach will prove useful in detecting and diagnosing many other cancers and diseases in the future.

Using bioinformatic and genomic approaches, several candidate genes have been identified that are differentially expressed in CaP (Ali S *et al*, 2003). Candidate genes were chosen that showed a strong specificity for expression in prostate tissue and also a significant upregulation in cancerous tissue compared with normal tissue. Expression analysis of these candidate genes further determined the tissue-specific expression and whether higher levels of expression are indicative of CaP. Expression analysis data using RNA from different tissues and disease states helped rank these candidates, the majority of which have been cloned and expressed in bacteria. These candidates, either individually or grouped in panels have the potential to become new CaP biomarkers for early detection, differential diagnosis, disease monitoring and surveillance.

REFERENCES

- ABRAMS ES, MURDAUGH SE and LERMAN LS: Comprehensive detection of single base changes in human genomic DNA using denaturing gradient gel electrophoresis and a GC clamp. Genomics 1990; 7: 463 - 475.

- ADLER A, DANIELSEN M and ROBINS D: Androgen-specific gene activation via a consensus glucocorticoid response element is determined by interaction with nonreceptor factors. Proc Natl Acad Sci 1992; 89: 11660 - 11663.

- ADLERCREUTZ H, HARKONEN M, KUOPPASALMI K, NAVERI H, HUHTANIEMI I and TIKKANEN H: Effect of training on plasma anabolic and catabolic steroid hormones and their response during physical exercise. Int J Sports Med 1986; 7(suppl): 27 - 28.

- ALERS JC, ROCHAT J. KRIJTENBURG PJ, HOP WC, KRANSE R, ROSENBERG C, TANKE HJ, SCHRÖDER FH and VAN DEKKEN H: Identification of genetic markers for prostatic cancer progression. Lab Invest 2000; 80: 931 - 942.

- ALEXANDER NJ, FREE MJ, PAULSEN CA, BUSHBOM R and FULGHAM DL: A comparison of blood chemistry, reproductive hormones, and the development of antisperm antibodies after vasectomy in men. J Androl 1980; 1: 40 - 50.

- ALI S, SUN Y, DRUMRIGHT C, CARRETERO L, TRAN C, NGUYEN A, MEDYNSKI D, LIU Y, HOANG VV, LIANG B, RECIPON H, CAFFERKEY R and MACINA R: Genomic approaches to the development of prostate cancer diagnostics.http://www.nature.com/cgitaf/DynaPage..urnal/v26/n4s/full/ng0401supp_38b.html 2003; pg 1.

- ALITALO K and SCHWAB M: Oncogene amplification in tumor cells. Adv Cancer Res 1986; 47: 235 - 281.

- AMATO AA, PRIOR TW, BAROHN RJ, SNYDER P, PAPP A and MENDELL JR: Kennedy's disease: a clinicopathologic correlation with mutations in the androgen receptor gene. Neurology 1993; 43: 791 - 794.

- ANDERSSON SO, BARON J, WOLK A, LINDGREN C, BERGSTROM R and ADAMI HO: Early life risk factors for prostate cancer: a population-based case-control study in Sweden. Cancer Epidemiol Biomarkers Prev 1995; 4: 187-192.

- ANWAR K, NAKAKUKI K, SHIRAISHI T, NAIKI H, YATANI R and INUZUKI: Presence of ras oncogene mutations and human papillomavirus DNA in human prostate carcinomas. Cancer Res 1992; 52: 5991 - 5996.

- ARBIZU T, SANTAMARIA J, GOMEZ JM, QUILEZ A and SERRA JP: A family with adult spinal and bulbar muscular atrophy, X-linked inheritance and associated testicular failure. J Neurol Sci 1983; 59: 371 - 382.

- AZZOUZI AR, COCHAND-PRIOLLET B, MANGIN P, FOURNIER G, BERTHON P, LATIL A and CUSSENOT O: Impact of constitutional genetic variation in androgen/oestrogen-regulating genes on age-related change in human prostate. Eur J Endocrinol 2002; 147(4): 479-484.

- BALL SE, SCATINA J, KAO J, FERRON GM, FRUNCILLO R, MAYER P, WEINRYB I, GUIDA M, HOPKINS PJ, WARNER N and HALL J: Population distribution and effects on drug metabolism of a genetic variant in the 5' promoter region of CYP3A4. Clin Pharmacol Ther 1999; 66: 288 - 294.

- BARETTINO D, VIVANCO RUIZ MM and STUNNENBERG HG: Characterization of the ligand-dependent transactivation domain of thyroid hormone receptor. EMBO J 1994; 13: 3039 - 3049.

- BATCH JA, WILLIAMS DM, DAVIES HR, BROWN BD, EVANS BAJ, HUGHES IA and PATTERSON MN: Androgen receptor gene mutations identified by SSCP in

fourteen subjects with androgen insensitivity syndrome. Hum Mol Genet 1992; 1: 497 - 503.

- BAXTER J and TOMKINS G: Specific cytoplasmic glucocorticoid hormone receptors in hepatoma tissue culture cells . Proc Natl Acad Sci USA 1971; 68: 932 - 937.

- BEATO M: Gene regulation of steroid hormones. Cell 1989; 56: 335 - 344.

- BEATO M, CHAVEZ S and TRUSS M: Transcriptional regulation by steroid hormones. Steroids 1996; 61: 240 - 251.

- BELSHAM DD, PEREIRA F, GREENBERG CR, LIAO S and WROGEMAN K: Leu-676-Pro mutation of the androgen receptor causes complete androgen insensitivity syndrome in a large Hutterite kindred. Hum Mutat 1995; 5: 28 - 33.

- BENTLEY JM and TILLEY WD: Androgen receptors in prostate cancer. J Endocrinol 1996; 151: 1 - 11.

- BERG JW: Can nutrition explain the pattern of international epidemiology of hormone-dependant cancers. Cancer Res 1975; 35: 3345 - 3350.

- BERG JW: DNA-binding specificity of steroid receptors. Cell 1989; 57: 1065 - 1068.

- BERKOVITZ GD, BROWN TR and MIGEON CJ: Androgen receptors. Clinics in Endocrinology and Metabolism 1983; 12: 155 - 173.

- BERNSTEIN L and ROSS RK: Cancer in Los Angeles County. A portrait of incidence and mortality, 1972-1987. Kennith Norris Jr Comprehensive Cancer Center, University of Southern California, Los Angeles 1991.

- BERREVOETS CA, DOESBURG P, STEKTEE K, TRAPMAN J and BRINKMANN AO: Functional interactions of the AF-2 activation domain core region of the human androgen receptor with the amino-terminal domain and with the transcriptional coactivator TIF(-transcriptional intermediary factor 2). Mol Endocrinol 1998; 12: 1172-1183.

- BERRY SJ, COFFEY DS, WALSH PC and EWING LL: The development of human benign prostatic hyperplasia with age. Journal of Urology 1984; 132: 474 - 479.

- BOEHMER ALM, BRINKMANN AO, NIERMEIJER MF, BAKKER L, HALLEY DJJ and DROP SLS: Germ-line and somatic mosaicism in the androgen insensitivity syndrome: Implications for genetic counseling. Am J Hum Genet 1997; 60: 1003 - 1006.

- BOOKSTEIN R, MACGROGAN D, HILSENBECK SG, SHARKEY F and ALLRED DC: p53 is mutated in a subset of advanced-stage prostate cancers. Cancer Res 1993; 53: 3369 - 3373.

- BORING CC, SQUIRES TS and HEALTH CW Jr. Cancer statistics for African-Americans. CA Cancer J Clin 1992; 42: 7 - 17.

- BOSLAND MC: The etiopathogenesis of prostate cancer with special reference to environmental factors. Adv Cancer Res 1988; 50: 1 - 106.

- BOSLAND MC: The male reproductive system. In : Waalkes MP, Ward JM, editors. Carcinogenicity. New York (NY): Raven Press 1994; 339 - 402.

- BOSLAND MC, OAKLEY-GIRVAN I and WHITTEMORE AS: Dietary fat, calories, and prostate cancer risk. J Natl Cancer Inst 1999; 91: 489 - 491.

- BOURGUET W, RUFF M, CHAMBON P, GRONEMEYER H and MORAS D: Crystal structure of the ligand binding domain of the human nuclear receptor RXR-alpha. Nature 1995; 375: 377 - 382.

- BRATT O, BORG A, KRISTOFFERSSON U, LUNDGREN R, ZHANG Q-X and OLSSON H: CAG repeat length in the androgen receptor is related to age at diagnosis of prostate cancer and response to endocrine therapy, but not to prostate cancer risk. Br J Cancer 1999; 81: 672 - 676.

- BRAWLEY OW, KNOPF K and MERRILL R: The epidemiology of prostate cancer part I: descriptive epidemiology. Semin Urol Oncol 1998a; 16: 187 - 192.

- BRAWLEY OW, KNOPF K and THOMPSON I: The epidemiology of prostate cancer part II: the risk factors. Semin Urol Oncol 1998b; 16: 193 - 201.

- BRAWN PN, JOHNSON EH, KUHL DH, RIGGS MW, SPEIGHTS VO, JOHNSON CF, PANDYA PP, LIND ML and BELL NF: Stage at presentation and survival of white and black patients with prostate carcinoma. Cancer 1993; 71: 2569 - 2573.

- BRINKMANN AO and TRAPMAN J: Androgen receptor mutants that affect normal growth and development. Cancer Surv 1992; 14: 95 - 111.

- BRINKMANN AO, JENSTER G, KUIPER GG, RIS C, VAN LAAR JH, VAN DER KORPUT JA, DEGENHART HJ, TRIFIRO MA, PINSKY L and ROMALO G: The human androgen receptor: structure/function relationship in normal and pathological situations. J Steroid Biochem Mol Biol 1992; 41: 361 - 368.

- BROWN CJ, GOSS SJ, LUBAHN DB, JOSEPH DR, WILSON EM, FRENCH FS and WILLARD HF: Androgen receptor locus on the human X chromosome: regional localization to Xq11-12 and description of a DNA polymorphism. Am J Hum Genet 1989; 44: 264 - 269.

- BROOKES AJ: The essence of SNPs: Gene 1999; 234: 177 - 186.

- BRUNING PF and BONFRER JMG: Free fatty acid concentrations correlated with the available fraction of estradiol in human plasma. Cancer Res 1989; 46: 2606 - 2609.

- BRZOZOWSKI AM, PIKE ACW, DAUTER Z, HUBBARD RE, BONN T, ENGSTROM O, OHMAN L, GREEN GL, GUSTAFSSON JA and CARLQUIST M: Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 1997; 389: 753 - 758.

- BUBENDORF L, KONONEN J, KOIVISTO P, SCHRAMI P, MOCH H, GASSER T, WILLI N, MIHATSCH M, SAUTER G and KALLIONEMI OP: Survey of gene amplification during prostate cancer progression by high-throughput flourescence in situ hybridization on tissue microarrays. Cancer Res 1999; 59: 803 - 806.

- BUCHANAN G, IRVINE RA, COETZEE GA and TILLEY WD: Contribution of the androgen receptor to prostate cancer predisposition and progression. Cancer Metastasis Rev 2001; 20(3-4): 207 -223.

- BURDETTE WJ and GEHAN EA: Planning and analysis of clinical studies (Thomas CC ed): Springfield, Illinois, 1970: p9.

- BURGLIN TR: The yeast regulatory gene PHO2 encodes a homeo box. Cell 1988; 53: 339 - 340.

- CAIRNS P, POLASCIK TJ, EBY Y, TOKINO K, CALIFANO J, MERLO A, MAO L, HERATH J, JENKINS R, WESTRA W, RUTTER JL, BUCKLER A, GABRIELSON E, TOCKMAN M, CHO KR, HEDRICK L, BOVA GS, ISSACS W, SCHWAB D and SIDRANSKY D: Frequency of homozygous deletion at p16/CDKN2 in primary human tumors. Nat Genet 1995; 11: 210 - 212.

- CAIRNS P, OKAMI K, HALACHMI S, HALACHMI N, ESTELLER M, HERMAN JG, JEN J, ISAACS WB, BOVA GS and SIDRANSKY D : Frequent inactivation of PTEN/MMAC1 in primary prostate cancer. Cancer Res 1997; 57: 4997 - 5000.

- CANOVATCHEL WJ and IMPERATO-MCGINLEY J: Benign prostatic hyperplasia: pathogenesis and the role of medical management. Geriatr Nephrol Urol 1992; 2: 1 - 23.

- CAREY AH, WATERWORTH D, PATEL K, WHITE D, LITTLE J, NOVELLI P, FRANK S and WILLIAMSON R: Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet 1994; 3: 1873 - 1876.

- CARGILL M, ALTSHULER D, IRELAND J, SKLAR P, ARDLIE K, PATIL N, SHAW N, LANE CR, LIM EP, KALYANARAMAN N, NEMESH J, ZIAGURA L, FRIEDLAND L, ROLFE A, WARRINGTON J, LIPSHUTZ R, DALEY GQ and LANDER ES: Characterisation of single-nucleotide polymorphisms in coding regions of human genes. Nat Genet 1999; 22: 231 - 238.

- CARSON-JURICA MA, SCHRADER WT and O'MALLEY BW: Steroid receptor family: structure and functions. Endocr Rev 1990; 11: 201 - 220.

- CARTER BS, EPSTEIN JI, ISAACS WB: RAS gene mutations in human prostate cancer. Cancer Res 1990; 50: 6830 - 6832.

- CARTER BS, BEATY TH, STEINBERG GD, CHILDS B and WALSH PC: Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 1992; 89: 3367 - 3371.

- CATALONA WJ, SMITH DS, RATLIFF TL, DODDS KM, COPLEN DE, YUAN JJJ, PETROS JA and ANDRIOLE GL: Measurement of prostate-specific antigen in serum as a screening test for prostate cancer. N Engl J Med 1991; 324: 1156 - 1161.

- CATALONA WJ, HUDSON MA, SCARDINO PT, RICHIE JP, AHMANN FR, FLANIGAN RC, DEKERNION JB, RATLIFF TL, KAVOUSSI LR, DALKIN BL, WATERS WB, MACFARLANE MT and SOUTHWICK PC: Selection of optimal

prostate specific antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves. J Urol 1994; 152: 2037 - 2042.

- CAVAILLES V, DAUVOIS S, L'HORSET F, LOPEZ G, HOARE S, KUSHNER PJ and PARKER MG: Nuclear factor RIP140 modulates transcriptional activation by the estrogen receptor. EMBO J 1995; 14: 3741 - 3751.

- CHAKRAVARTI A: It's raining SNPs, hallelujah? Nat Genet 1998; 19: 216 - 217.

- CHAMBERLAIN NL, DRIVER ED and MIESFELD RL: The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids 1994; 22: 3181 - 3186.

- CHER ML, MACGROGAN D, BOOKSTEIN R, BROWN JA, JENKINS RB and JENSEN RH: Comparative genomic hybridization, allelic imbalance, and flourescence in situ hybridization on chromosome 8 in prostate cancer. Genes Chromosomes Cancer 1994; 11: 153 - 162.

- CHER ML, BOVA GS, MOORE DH, SMALL EJ, CARROLL PR, PIN SS, EPSTEIN JI, ISAACS WB and JENSEN RH: Genetic alterations in untreated metastases and androgen independent prostate cancer detected by comparative genomic hybridization and allelotyping. Cancer Res 1996; 56: 3091 - 3092.

- CHER ML, LEWIS PE, BANERJEE M, HURLEY PM, SAKR W, GRIGNON DJ and POWELL IJ: A similar pattern of chromosomal alterations in prostate cancers from African-Americans and Caucasian Americans. Clin Cancer Res 1998; 4: 1273 - 1278.

- CHI SG, DE VERE WHITE RW, MEYERS FJ, SIDERS DB, LEE F and GUMERLOCK PH: P53 in prostate cancer: frequent expressed transition mutations. J Natl Cancer Inst 1994; 86: 926 - 933.

- CHOONG CS and WILSON EM: Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. J Mol Endocrinol 1998; 21: 235 - 257.

- CHOONG CS, KEMPPAINEN JA and WILSON EM: Evolution of the primate androgen receptor: a structural basis for disease. J Mol Evol 1998; 47: 334 -342.

- CHUNG B, PICADO-LEONARD J, HANIU M, BIENKKOWSKI M, HALL PF, SHIVELY JE and MILLER WL: Cytochrome P450c17 (steroid 17α-hydroxylase/17,20-lyase): cloning of human adrenal and testes cDNAs indicates the same gene is expressed in both tissues. Proc Natl Acad Sci USA 1987; 84:407 - 411.

- CLARK LC, DALKIN B, KRONGRAD A, COMBS G, TURNBULL BW, SLATE EH, WITHERINGTON R, HERLONG JH, JANOSKO E, CARPENTER D, BOROSSO C, FALKS and ROUNDER J: Decreased incidence of prostate cancer with selenium supplementation: results of a double-blind cancer prevention trial. Br J Urol 1998; 81: 730 - 734.

- COETZEE GA and ROSS RK: Prostate cancer and the androgen receptor. J Natl Cancer Inst 1994; 86: 872 - 873.

- COFFEY DS: Physiological control of prostatic growth, in Prostate Cancer, An Overview. UICC workshop on prostatic cancer 1978. Technical Report Series, Geneva, International Union Against Cancer 1979; 48: 4 - 23.

- COFFEY DS: Androgen action and the sex accessory tissues; in Knobil E, Neill JJ and Ewing LL (eds): The Physiology of Reproduction, New York, Raven Press, 1988: 1081 – 1118.

- COFFEY DS: The molecular biology of the prostate; in Lepor H and Lawson RK (eds): Prostate Disease. Philadelphia, Saunders, 1993: 28 - 56.

- COLLINS A, LONJOU C and MORTON NE: Genetic epidemiology of single-nucleotide polymorphisms. Proc Natl Acad Sci USA 1999; 96: 15173 - 15177.

- COLLINS FS, GUYER MS and CHARKRAVARTI A: Variations on a theme: cataloging human DNA sequence variation. Science 1997; 278: 1580 - 1581.

- COLLINS FS, BROOKS LD and CHAKRAVARTI A: A DNA polymorphism discovery resource for research on human genetic variation. Genome Res 1998; 8: 1229 - 1231.

- COLOMBEL M, SYMMANS F, GIL S, OTOOLE KM, CHOPIN D, BENSON M, OLSSON CA, KORSMEYER S and BUTTYAN R: Detection of the apoptosis-suppressing oncoprotein bcl-2 in hormone-refractory human prostate cancers. Am J Pathol 1993; 143: 390 – 400.

- COOK PJ, DOLL R and FELLINGHAM SA: A mathematical model for the age distribution of cancer in man. Int J Cancer 1969; 4: 93 - 112.

- COOKE P, YOUNG P and CUNHA G: Androgen receptor expression in developing male reproductive organs. Endocrinology 1991; 128: 2867 - 2873.

- CORREA-CERRO, WOHR G, HAUSSLER J, BERTHON P, DRELON E, MAHGIN P, FOURNIER G, CUSSENOT O, KRAUS P, JUST W, PAISS T, CANTU JM and VOGEL W: (CAG)nCAA, and GGN repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. Eur J Hum Genet 1999; 7: 357 - 362.

- COTE RJ, SHI Y, GROSHEN S, FENG AC, CORDON-CARDO C, SKINNER D and LIESKOVOSKY G: Association of p27Kip1 levels with recurrence and survival in patients with stage C prostate carcinoma. J Natl Cancer Inst 1998; 90: 916 - 920.

- COX DR and HINKLEY DV: Theoretical statistics (Chapman and Hall eds). London, 1974: p66.

- CRAWFORD ED, EISENBERGER M, MCLEOD D, SPAULDING J, BENSON R, DORR F, BLUMENSTEIN BA, DAVIS MA and GOODMAN PJ: A controlled trial of leuprolide with and without flutamide in prostatic carcinoma. N Engl J Med 1989; 321: 419 - 424.

- CRAWFORD ED, DE-ANTONI EP, ETZIONI R, SCHAEFER VC, OLSON RM and ROSS CA: Serum-prostate specific antigen and digital rectal examination for early detection of prostate cancer in a national community based program. The Prostate Cancer Education Coucil. Urology 1996; 47: 863 - 869.

- CUDE KJ, DIXON SC, GOU Y, LISELLA J and FIGG WD: The androgen receptor: genetic considerations in the development and treatment of prostate cancer. J Mol Med 1999; 77: 419 - 426.

- CULIG Z, HOBISCH A, CRONAUER MV, CATO ACB, HITTMAIR A, RADMAYR C, EBERLE J, BARTSCH G and KLOCKER H: Mutant androgen receptor detected in an advanced-stage prostatic carcinoma is activated by adrenal androgens and progesterone. Mol Endocrinol 1993; 7: 1541 - 1550.

- CULIG Z, HOBISCH A, HITTMAIR A: Expression, structure and function of androgen receptor in advanced prostatic carcinoma. Prostate 1998; 35: 63 -70.

- CUTLER GB, GLENN M, BUSH M, HODGEN GD, GRAHAM CE and LORIAUX DL: Adrenarche: a survey of rodents, domestic animals and primates. Endocrinology 1978; 103: 2112 - 2118.

- CUTLER SJ and YOUNG JL. Third National Cancer Survey: incidence data. NCI Monogr 1975; 41: 106 - 111.

- DANIELIAN PS, WHITE R, LEES JA and PARKER MG: Identification of a conserved region required for hormone dependent transcriptional activation by steroid hormone receptors. EMBO J 1992; 11: 1025 - 1033.

- DAVID VA and MENOTTI-RAYMOND M: Molecular genetic analysis of populations (Hoelzel AR ed): Chapter 10, Automated DNA detection with fluorescence-based technologies. Oxford University Press, New York, 1998: p337- 348.

- DENMARK-WAHNEFRIED W, LESKO SM, CONAWAY MR, ROBERTSON CN, CLARK RV , LOBAUGH B, MATHIAS BJ, STRIGO TS and PAULSON DF: Serum androgens: associations with prostate cancer risk and hair patterning. J Androl 1997; 18: 495 - 500.

- DERISI J, PENLAND L, BROWN PO, BITTNER ML, MELTZER PS, RAY M, CHEN Y, SU YA and TRENT JM: Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996; 14: 457 - 460.

- DEVGAN SA, HENDERSON BE, YU MC, SHI CY, PIKE MC, ROSS RK and REICHARDT JK: Genetic variation of 3β-hydroxysteroid dehydrogenase type II in three racial/ethnic groups: implications for prostate cancer risk. Prostate 1997; 33: 9 - 12.

- DINJENS WN, VAN DER WEIDEN MM, SCHRÖDER FH, BOSMAN FT and TRAPMAN J: Frequency and characterization of p53 mutations in primary and metastatic human prostate cancer. Int J Cancer 1994; 56: 630 - 633.

- DOESBURG P, KUIL CW, BERREVOETS CA, STEKETEE K, FABER PW, MULDER E, BRINKMANN AO and TRAPMAN J: Functional in vivo interaction between the amino-terminal transactivation domain and the ligand binding domain of the androgen receptor. Biochemistry 1997; 36: 1052 - 1054.

- DONG JT, LAMB PW, RINKER-SCHAEFFER CW, VUKANOVIC J, ISAACS JT and BARRETT JC: KAI-1, A metastasis suppressor gene for prostate cancer on human chromosome 11p11.2. Science 1995; 268: 884 - 886.

- DONG JT, SIPE TW, HYYTINEN ER, LI CL, HEISE C, MCCLINTOCK DE, GRANT CD, CHUNG LW and FRIERSON HF JR: PTEN/MMAC1 is infrequently mutated in pT2 and pT3 carcinomas of the prostate. Oncogene 1998; 17: 1979 - 1982.

- EASTON DF, SCHAID DJ, WHITTEMORE AS, ISAACS WJ: Where are the prostate cancer genes? – A summary of eight genome wide searches. Prostate 2003; 57(4): 261 - 269.

- EDITORIAL: Freely associating. Nat Genet 1999; 22: 1 - 2.

- EDWARDS A, HAMMOND HA, JIN L, CASKEY CT and CHAKRABORTY R: Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 1992; 12: 241 - 253.

- EDWARDS SM, BADZIOCH MD, MINTER R, HAMOUDI R, COLLINS N, ARDERN-JONES A, DOWE A, OSBORNE S, KELLY J, SHEARER R, EASTON DF, SAUNDERS GF, DEARNALEY DP and EELES RA: Androgen receptor polymorphisms: association with prostate cancer risk, relapse and overall survival. Int J Cancer 1999; 84: 458 - 465.

- EELES RA, DUROCHER F, EDWARDS S, TEARE D, BADZIOCH M, HAMOUDI R, GILL S, BIGGS P, DEARNALEY D, ANDERN-JONES A, DOWE A, SHEARER R, MCLENNAN DL, NORMAN RL, GHADIRIAN P, APRIKIAN A, FORD D, AMOS C, KING TM, THE CANCER RESEARCH CAMPAIGN/BRITISH PROSTATE GROUP UK FAMILIAL PROSTATE CANCER STUDY COLLABORATORS, LABRIE F, SIMARD J, NAROD SA, EASTON D and FOULKES WD: Linkage analysis of chromosome 1q markers in 136 prostate cancer families. Am J Hum Genet 1997; 62: 653 - 658.

- ELO JP, KVIST L, LEINONEN K, ISOMAA V, HENTTU P, LUKKARINEN O and VIKHO P: Mutated human androgen receptor gene detected in a prostatic cancer patient is also activated by estradiol. J Clin Endocrinol Metab 1995; 80: 3494 - 3500.

- EMAHAZION T, FEUK L, JOBS M, SAWYER SL, FREDMAN D, ST CLAIR D, PRINCE JA and BROOKES AJ: SNP association studies in Alzheimer's disease highlight problems for complex disease analysis. Trends in Genetics 2001; 17(7): 407 - 413.

- EVANS BA, YUN ZX, CLOSE JA, TREGEAR GW, KITAMURA N, NAKANISHI S, CALLEN DF, BAKER E, HYLAND VJ, SUTHERLAND GR and RICHARDS RI: Structure and chromosomal localization of the renal kallikrein gene. Biochemistry 1988; 27(9): 3124 - 3129.

- EVANS BA, HARPER ME, DANIELLS CE,WATTS CE, MATENHELIA S, GREEN J and GRIFFITHS K : Low incidence of androgen receptor gene mutations in human prostatic tumors using single strand conformation polymorphism analysis. Prostate 1996; 28: 162 - 171.

- FABER PW, KUIPER GGJM, VAN ROOIJ HCJ, VAN DER KORPUT JAGM, BRINKMANN AO and TRAPMAN J: The N-terminal domain of the human androgen receptor is encoded by one, large exon. Molec Cell Endocr 1989; 61: 257 - 262.

- FABER PW, VAN ROOIJ HCJ, VAN DER KORPUT HAGM, BAARENDS WM, BRINKMANN AO, GROOTEGOED JA and TRAPMAN J: Characterization of the human androgen receptor transcription unit. J Biol Chem 1991; 266: 10743 - 10749.

- FABER PW, VAN ROOIJ HCJ, SCHIPPER HJ, BRINKMANN AO and TRAPMAN J: Two different, overlapping pathways of transcription initiation are active on the

TATA-less human androgen receptor promoter: the role of Sp1. J Biol Chem 1993; 268: 9296 - 9301.

- FAN YS, SASI R, LEE C, WINTER JS, WATERMAN MR and LIN CC: Localization of the human CYP17 gene (cytochrome P450 (17 alpha)) to 10q24.3 by fluorescence *in situ* hybridization and simultaneous chromosome banding. Genomics 1992; 14: 1110 - 1111.

- FANG S, ANDERSON KM and LIAO S: Receptor proteins for androgens: on the role of specific proteins in selective retention of 17-beta-hydroxy-5-alpha-androstan-3-one by rat ventral prostate in vivo and in vitro. J Biol Chem 1969; 244: 6584 - 6595.

- FAWELL SE, LEES JA, WHITE R and PARKER MG: Characterization and colocalization of steroid binding and dimerization activities in the mouse estrogen receptor. Cell 1990; 60: 953 - 962.

- FEBBO PG, KANTOFF PW, PLATZ EA, CASEY D, BATTER S, GIOVANUCCI E, HENNEKENS CS and STAMPFER JM: The V89L polymorphism in the 5α-reductase type 2 gene and risk of prostate cancer. Cancer Res 1999; 59: 5878 - 5881.

- FEIGELSON HS, COETZEE GA, KOLONEL L, ROSS RK and HENDERSON BE: A polymorphism in the *CYP17* gene increases the risk of breast cancer. Cancer Res 1997; 57: 1063 - 1065.

- FEILOTTER HE, NAGAI MA, BOAG AH, ENG C and MULLIGAN LM: Analysis of PTEN and the 10q23 region in primary prostate carcinomas. Oncogene 1998; 16: 1743- 1748.

- FENG W, RIBEIRO RCJ, WAGNER RL, NGUYEN H, APRILETTI JW, FLETTERICK RJ, BAXTER JD, KUSHNER, PJ and WEST BL: Hormone-

dependent coactivator binding to a hydrophobic cleft on nuclear receptors. Science 1998; 280: 1747 - 1749.

- FENTON MA, SHUSTER TD, FERTIG AM, TAPLIN ME, KOLVENBERG G, BUBLEY GJ and BALK SP: Functional characterization of mutant androgen receptors from androgen-independent prostate cancer. Clin Cancer Res 1997; 3: 1383 - 1388.

- FINCHAM SM, HILL GB, HANSON J and WIJAYASINGHE C. Epidemiology of prostate cancer: a case-control study. Prostate 1990; 17: 189 - 206.

- FISCHER SG and LERMAN LS: DNA fragments differing by single base-pair substitutions are seperated in denaturing gradient gels: correspondence with melting theory. Proc Natl Acad Sci USA 1983; 80: 1579 - 1583.

- FISHER RA: Statistical methods for research workers (Hafner, ed): New York, 1958: p80.

- FREEDMAN LP: Anatomy of the steroid receptor zinc finger region. Endocrine Reviews 1992; 13: 129 - 145.

- FRONSDAL K, ENGEDAL N, SLAGSVOLD T and SAATCIOGLU F: CREB binding protein is a coactivator for the androgen receptor and mediates cross-talk with AP-1. J Biol Chem 1994; 273: 31 853 - 31 859.

- GADDIPATI JP, MCLEOD DG, HEIDENBERG HB, SESTERHENN IA, FINGER MJ, MOUL JW and SRIVASTAVA S: Frequent detection of codon 877 mutation in the androgen receptor gene in advanced prostate cancers. Cancer Research 1994; 54: 2861-2864.

- GANN PH, HENNEKENS CH, MA J, LONGCOPE C and STAMPFER MJ: Prospective study of sex hormone levels and risk of prostate cancer. J Natl Cancer Inst 1996; 88: 1118 - 1126.

- GANN PH, MA J, GIOVANNUCCI E, WILLET W, SACKS F, HENNEKENS CH and STAMPFER MJ: Lower prostate cancer risk in men with elevated plasma lycopene levels: results of a prospective analysis. Cancer Res 1999; 59: 1225 - 1230.

- GAO T, MARCELLI M and MCPHAUL MJ: Transcriptional activation and transient expression of the human androgen rceptor. J Steroid Biochem Mol Biol 1996; 59: 9 -20.

- GELLER DH, AUCHUS RJ, MENDONCA BB and MILLER WL: The genetic and functional basis of isolated 17,20-lyase deficiency. Nature Genetics 1997; 17: 201 - 205.

- GIOVANNUCCI E, TOSTESON TD, SPEIZER FE, VESSEY MP, COLDITZ GA: A long-term study of mortality in men who have undergone vasectomy. N Engl J Med 1992; 326: 1392 - 1398.

- GIOVANNUCCI E, TOSTESON TD, SPEIGER FE, ASCHERIO A, VESSEY MP and COLDITZ GA. A retrospective cohort study of vasectomy and prostate cancer in US men. JAMA 1993; 269: 878 - 882.

- GIOVANNUCCI E, ASCHERIO A, RIMM EB, STAMPFER MJ, COLDITZ GA and WILLET WC: Intake of carotenoids and retinol in relation to risk of prostate cancer. J Natl Cancer Inst 1995; 87: 1767 - 1776.

- GIOVANNUCCI E, STAMPFER MJ, KRITHIVAS K, BROWN M, DAHL D, BRUFSKY A, TALCOTT J, HENNEKENS CH and KANTOFF PW: The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 1997; 94: 3320 - 3323.

- GIOVANNUCCI E, LEITZMANN M, SPEIGELMAN D, RIMM EB, COLDITZ GA, CAMPFER MJ, STAMPFER MJ and WILLETT WL: A prospective study of

physical activity and prostate cancer in male health professionals. Cancer Res 1998; 58: 5117 - 5122.

- GIOVANNUCCI E, PLATZ EA, STAMPFER MJ, CHAN A, KRITHIVAS K, KAWACHI I, WILLETT WC and KANTOFF PW: The CAG repeat within the androgen receptor gene and benign prostatic hyperplasia. Urology 1999; 53: 121 - 125.

- GIBBS M, STANFORD JL, MCINDOE RA, JARVIK GP, KOLB S, GOODE EL, CHAKRABARTI L, SCHUSTER EF, BUCKLEY VA, MILLER EL, BRANDZEL S, LI S, HOOD L and OSTRANDER EA: Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 1999; 64: 776 - 787.

- GOTTLIEB B, LEHVASLAIHO H, BEITEL LK, LUMBROSO R, PINSKY L, and TRIFIRO M: The androgen receptor gene mutations database. Nucleic Acids Res 1998; 26: 234 - 238.

- GOTTLIEB B, BEITEL LK, LUMBROSO R, PINSKY L and TRIFIRO M: Update of the androgen receptor gene mutations database. Human Mutation 1999; 14: 103 - 114.

- GOVERNMENT COMMUNICATION AND INFORMATION SYSTEM: South African Yearbook. Universal Printers, Durban, 2001; pp. 1 – 41.

- GRAY IC, CAMPBELL DA and SPURR NK: Single nucleotide polymorphisms as tools in human genetics. Human Molecular Genetics 2000; 9(16): 2403 - 2408.

- GREEN S, KUMAR V, THEULAZ I, WAHLI W and CHAMBON P: The N-terminal DNA-binding "zinc finger" of the oestrogen and glucocorticoid receptors determines target gene specificity. EMBO J 1988; 7: 3037 - 3044.

- GRINO PB, GRIFFIN JE and WILSON JD: Testosterone at high concentrations interacts with the human androgen receptor similarly to dihydrotestosterone. Endocrinology 1990; 126: 1165 - 1172.

- GRONBERG H, XU J, SMITH JR, CARPTEN JD, ISAACS SD, FREIJE D, BOVA GS, DANBER J-E, BERGH A. WALSH PC, COLLINS FS, TRENT JM, MEYERS DA and ISAACS WB: Early age at diagnosis in families providing evidence of linkage to the hereditary prostate cancer locus (HPC1) on chromosome 1. Cancer Res 1997; 57: 4707 - 4709.

- GRONEMEYER H and LAUDET V: Transcription factors 3: nuclear receptors. Protein Profile 1995; 11: 1173 - 1308.

- GROVER P and ODELL W: Correlation of in vivo and in vitro activities of some naturally occurring androgens using a radioreceptor assay for 5α-dihydrotestosterone with rat prostate cytosol receptor protein. J Steroid Biochem 1975; 6: 1373 - 1379.

- GRÜTTER G and VAN ZYL DJ: The story of South Africa. Human and Rousseau (Pty) Ltd, (Cape Town, Pretoria, Johannesburg) 1981; pp. 12 – 62.

- GSUR A, BERNHOFER G, HINTEREGGER S, HAIDINGER G, SCHATZL G, MADERSBACHER S, MARBERGER M, VUTUC C and MICKSCHE M: A polymorphism in the *CYP17* gene is associated with prostate cancer risk. Int J Cancer 2000; 87: 434 - 437.

- GULDBERG P and GUTTLER F: A simple method for identification of point mutations using denaturing gradient gel electrophoresis. Nucleic Acids Res 1993; 21: 2261 - 2262.

- GUMERLOCK PH, POONAMALLEE UR, MEYERS FJ and DE VERE WHITE RW: Activated ras alleles in human carcinoma of the prostate are rare. Cancer Res 1991; 51: 1632 - 1637.

- GUPTA S, HASTAK K, AHMAD N, LEWIN JS and MUKHTAR H: Inhibition of prostate carcinogenesis in TRAMP mice by oral infusion of green tea polyphenols. Proc Natl Acad Sci USA 2001; 98(18) : 10350-10355.

- HABUCHI T, LIQING Z, SUZUKI T, SASAKI R, TSUCHIYA N, TACHIKI H, SHIMODA N, SATOH S, SATO K, KAKEHI Y, KAMOTO T, OGAWA O and KATO T: Increased risk of prostate cancer and benign prostatic hyperplasia associated with a *CYP17* gene polymorphism with a gene dosage effect. Cancer Res 2000; 60: 5710 - 5713.

- HAKIMI JM, SCHOENBERG MP, RONDINELLI RH, PIANTADOSI S and BARRACK ER: Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. Clin Cancer Res 1997; 3: 1599 - 1608.

- HALUSHKA MK, FAN JB, BENTLEY K, HSIE L and SHEN N: Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis. Nat Genet 1999; 22: 239 - 247.

- HAMALAINEN E, ADLERCREUTZ H, PUSKA P and PIETINEN P: Diet and serum hormones in healthy men. J Steroid Biochem 1984; 20: 459 - 464.

- HANLEY NA, RAINEY WE, WILSON DI, BALL SG and PARKER KL: Expression profiles of SF-1, DAX1, and CYP17 in the human fetal adrenal gland: potential interactions in gene regulation. Mol Endocrinol 2001; 15: 57 - 68.

- HARD T, KELLENBACH E, BOELENS R, MALER BA, DAHLMAN K, FREEDMAN LP, CARLSTEDT-DUKE J, YAMAMOTO KR, GUSTAFSSON J-A and KAPTEIN R: Solution structure of the glucocorticoid receptor DNA-binding domain. Science 1990; 249: 157 - 160.

- HARDY DO, SCHER HI, BOGENREIDER T, SABBATINI P, ZHANG Z-F, NANUS DM and CATTERALL JF: Androgen receptor CAG repeat lengths in prostate

cancer: correlation with age of onset. J Clin Endocrinol Metab 1997; 81: 4400 - 4405.

- HARRIS SE, RONG Z, HARRIS MA and LUBAHN DD: Androgen receptor in human prostate carcinoma LNCaP/ADEP cells contains a mutation which alters the specificity of the steroid-dependent transcriptional activation region. Endocrinology 1990; 126(Suppl): 93.

- HAYES RB, DE JONG FH, RAATGEVER J, BOGDANOVICZ J, SCHRÖDER FH, VAN DER MAAS P, OISHI K and YOSHIDA O: Physical characteristics and factors related to sexual development and behaviour and the risk for prostate cancer. Eur J Cancer Prev 1992; 1: 239 - 245.

- HAYES RB, POTTERN LM, GREENBERG R, SCHOENBERG J, SWANSON GM, LIFF J, SCHWARTZ AG, BROWN LM and HOOVER RN: Vasectomy and prostate cancer in US blacks and whites. Am J Epidemiol 1993; 137: 263 - 269.

- HAYES VM, WU Y, OSINGA J, MULDER IM, VAN DER VLIES P, ELFFERICH P, BUYS CHCM and HOFSTRA RMW: Improvements in gel composition and electrophoretic conditions for broad-range mutation analysis by denaturing gradient gel electrophoresis. Nucleic Acids Res 1999; 27: e29.

- HAYNES SR, REBBERT ML , MOZER BA, FORQUIGNON F and DAWID IB: Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a Drosophila cDNA homologous to the rat helix destabilizing protein. Proc Natl Acad Sci USA 1987; 84: 1819 - 1923.

- HEINONEN OP, ALBANES D, VIRTAMO J, TAYLOR PR, HUTTUNEN JK, HARTMAN AM, HAAPAKOSKI J, MALILA N, RAUTALAHTI M, RIPATTI S, MAENPAA H, TEERENHOVI L, KOSS L, VIROLAINEN M and EDWARDS BK: Prostate cancer and supplementation with alpha-tocopherol and beta-carotene: incidence and mortality in a controlled trial. J Natl Cancer Inst 1998; 90: 440 - 446.

- HENDERSON BE, BERNSTEIN L, ROSS RK, DE PUE RH and JUDD HL: The early in-utero estrogen and testosterone environment of blacks and whites: potential effects on male offspring. Br J Cancer 1988a; 57: 216 - 218.

- HENDERSON BE, ROSS RK and BERNSTEIN L: Estrogens as a cause of human cancer: the Richard and Hinda Rosenthal Foundation Award Lecture. Cancer Res 1988b; 48: 246 - 253.

- HEYNS CF, NAUDE AM, VISSER AJ, MARAIS DC, STOPFORTH HB, NYARKO JK and STELLMACHER GA: Early diagnosis of prostate cancer in the Western Cape. S Afr Med J 2001; 91: 679 - 684.

- HILL P, WYNDER EL, GARBACZEWSKI L, GARNES H and WALKER AR: Diet and urinary steroids in black and white North American and black South African men. Cancer Res 1979; 39: 5101 - 5105.

- HILL P, WYNDER EL, GARNES H and WALKER AR: Environmental factors, hormone status, and prostate cancer. Prev Med 1980; 9: 657 - 666.

- HONDA GD, BERNSTEIN L, ROSS RK, GREENLAND S, GERKINS V and HENDERSON BE: Vasectomy, cigarette smoking, and age at first sexual intercourse as risk factors for prostate cancer in middle-aged men. Br J Cancer 1988; 57: 326 - 331.

- HONG H, KOHLI K, TRIVEDI A, JOHNSON DL and STALLCUP MR: GRIP1, a novel mouse protein that serves as a transcriptional coactivator in yeast for the hormone binding domains of steroid receptors. Proc Natl Acad Sci USA 1996; 93: 4948 - 4952.

- HOUSLEY PR, SANCHEZ ER, DANIELSEN M, RINGOLD GM and PRATT WB: Evidence that the conserved region in the steroid binding domain of the glucocorticoid receptor is required for both optimal binding of hsp90 and

protection from proteolytic cleavage. A two site model for hsp90 binding to the steroid-binding domain. J Biol Chem 1990; 265: 12778 - 12781.

- HOVIG E, SMITH-SORENSEN B, GERHARD MC, RYBERG D, LOTHE R and BORRESEN AL: No alterations in exon 21 of the RB1 gene in sarcomas and carcinomas of the breast, colon and lung. Genes Chromosomes Cancer 1992; 5:97 - 103.

- HOWARDS SS and PETERSON HB: Vasectomy and prostate cancer. Chance, bias, or a causal relationship? JAMA 1993; 269: 913 - 914.

- HTTP://WWW.CANCER.GOV/NEWSCENTER/PROSTATE/PROTEOMICS: 2004; pg 1-2.

- HTTP://WWW.CANSA.ORG.ZA/PREVENTION/CANCER_OF_THE_PROSTATE: 2000; pg 1.

- HTTP://WWW.FOTW.CA/FLAGS/ZA: 2000; pg 1-2.

- HTTP://WWW.PHOENIX5.ORG/GRADING.HTML: 2000; pg 1 - 3.

- HUGGINS C and HODGES CV: Studies on prostatic cancer: The effects of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. Cancer Res 1941; 1: 293 - 297.

- ICHIKAWA T, ICHIKAWA Y, DONG J, HAWKINS AL, GRIFFIN CA, ISAACS WB, OSHIMURA M, BARRETT JC and ISAACS JT: Localization of metastasis suppressor gene(s) for prostatic cancer to the short arm of human chromosome 11. Cancer Res 1992; 52: 3486 - 3490.

- IGARASHI S, TANNO Y, ONODERA O, YAMAZAKI M, SATO S, ISHIKAWA A, MIYATANI N, NAGASHIMA M, ISHIKAWA Y, SAHASHI K, IBI T, MIYATAKE T and TSUJI S: Strong correlation between the number of CAG repeats in androgen

receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. Neurology 1992; 42: 2300 - 2302.

- IKONEN T, PALVIMO JJ, KALLIO PJ: Stimulation of androgen-regulated transactivaton by modulators of protein phosphorylation. Endocrinology 1994; 135: 1359 - 1366.

- INGLES SA, ROSS RK, YU MC, IRVINE RA, LA PERA G, HAILE RW, COETZEE GA: Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. J Natl Cancer Inst 1997; 89: 166 - 170.

- IRVINE RA, YU MC, ROSS RK and COETZEE GA: The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with CaP. Cancer Res 1995; 55: 1937 - 1940.

- ISAACS WB, BOVA GS, MORTON RA, BUSSEMAKERS MJG, BROOKS JD and EWING CM: Molecular biology of prostate cancer. Semin Oncol 1994; 21: 514 - 521.

- ITTMANN MM and WIECZOREK R: Alterations of the retinoblastoma gene in clinically localized, stage B prostate adenocarcinomas. Hum Pathol 1996; 27: 28 - 34.

- JAMES AJ, MARCELLI M, LAMB DJ and WIEGEL NL: Functional analysis of androgen receptor mutants identified in human prostate cancer. In: Program of the 79th Endocrine Society Meeting, Abstr 1997; 2 - 484.

- JARRAD DF, BOVA GS, EWING CM, PIN SS, NGUYEN SH, BAYLIN SB, CAIRNS P, SIDRANSKY D, HERMAN JG and ISAACS WB : Deletional, mutational, and methylation analyses of CDKN2 (p16/MTS1) in primary and metastatic prostate cancer. Genes Chromosomes Cancer 1997; 19: 90 - 96.

- JENSEN E, SUJUKI T, KAWASHIMA T, STUMPF W, JUNGBLUT P and DE SOMBRE E: A two-step mechanism for the interaction of estradiol with rat uterus. Proc Natl Acad Sci USA 1968; 59: 632 - 638.

- JENSTER G, VAN DER KORPUT HAGM, VAN VROONHOVEN C, VAN DER KWAST TH, TRAPMAN J and BRINKMANN AO: Domains of the human androgen receptor involved in steroid binding, transcriptional activition and nuclear binding. Mol Endocrinol 1991; 5: 1396 - 1404.

- JENSTER G, VAN DER KORPUT HA, TRAPMAN J and BRINKMANN AO: Functional domains of the human androgen receptor. J Steroid Mol Biol 1992; 41: 671-675.

- JENSTER G, DE RUITER PE, VAN DER KORPUT HAGM, KUIPER GGJM, TRAPMAN J and BRINKMANN AO: Changes in abundance of androgen receptor isotypes: Effects of ligand treatment, glutamine-stretch variation and mutation of putative phosphorylation sites. Biochemistry 1994; 33: 14064 - 14072.

- JENSTER G, VAN DER KORPUT JAGM, TRAPMAN J and BRINKMANN AO: Identification of two transcription units in the N-terminal domain of the human androgen receptor. J Biol Chem 1995; 270: 7341 - 7346.

- JOHN EM, WHITTEMORE AS, WU AH, KOLONEL LN, HISLOP TG, HOWE GR, WEST DW, HANKIN J, DREON DM, TEH CZ, BURCH JD and PAFFENBERGER RS JR: Vasectomy and prostate cancer: results from a multiethnic case-control study. J Natl Cancer Inst 1995; 87: 662 - 669.

- JOOS S, BERGERHEIM US, PAN Y, MATSUYUMA H, BENTZ M, DU MANOIR S and LICHTER P: Mapping of chromosomal gains and losses in prostate cancer by comparative genomic hybridization . Genes Chromosomes Cancer 1995; 14: 267 - 276.

- KAGIMOTO M, WINTER JSD, KAGIMOTO K, SIMPSON ER and WATERMAN MR: Structural characterization of normal and mutant human steroid 17α-hydroxylase genes: molecular basis of one example of combined 17α-hydroxylase/17,20 lyase deficiency. Mol Endocrinol 1988; 2: 564 - 570.

- KALLIO PJ, PALVIMO JJ, MEHTO M and JANNE OA: Analysis of androgen receptor-DNA interactions with receptor proteins produced in insect cells. J Biol Chem 1994; 269: 11514 - 11522.

- KALLIONEMI OP and VISAKORPI T: Genetic basis and clonal evolution of human prostate cancer. Cancer Res 1996; 68: 225 - 255.

- KAUL L, HESHMAT MY, KOVI J, JACKSON MA, JACKSON AG, JONES GW, EDSON M, ENTERLINE JP, WORRELL RG and PERRY SL: The role of diet in prostate cancer. Nutr Cancer 1987; 9: 123 - 128.

- KAWASAKI T, USHIYAMA T, UEYAMA H, INOUE K, MORI K, OHKUBO I and HUKUDA S: Polymorphic CAG repeats of the androgen receptor gene and rheumatoid arthritis. Ann Rheum Dis 1999; 58: 500 - 502.

- KAZEMI-ESFARJANI P, TRIFIRO MA and PINSKY L: Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: Possible pathogenic relevance for the (CAG)n-expanded neuropathies. Hum Mol Genet 1995; 4: 523 - 527.

- KEAVENEY M, KLUG J, DAWSON MT, NESTER PV, NEILAN JG, FORDE RC and GANNON F: Evidence for a previously unidentified upstream exon in the human oestrogen receptor gene. J Mol Endocrinol 1991; 6: 111 - 115.

- KEHINDE EO: The geography of prostate cancer and its treatment in Africa. Cancer Surv 1995; 23: 281 - 286.

- KEIZER H, JANSSEN GME, MENHEERE P and KRANENBURG G: Changes in basal plasma testosterone, cortisol, and dehydroepiandrosterone sulfate in previously untrained males and females preparing for a marathon. Int J Sports Med 1989; 10(suppl 3): S139 - S145.

- KELSEY J and BERNSTEIN L: Epidemiology and prevention of breast cancer. Annu Rev Public Health 1996; 17: 47 - 67.

- KENNEDY WR, ALTER M and SUNG JH: Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. Neurology 1968; 18: 671 - 680.

- KEOHAVONG P, ZHU D, MELACRINOS AC, DEMICHELE MAA, WEYANT RJ, LUKETICH JD, TESTA JR, FEDDER M and SIEGFRIED JM: Detection of low-fraction K-ras mutations in primary lung tumours using a sensitive method. Int J Cancer Res 1997; 74: 162 - 170.

- KIARIS H, ELIOPOULOS AG, SIVRIDIS E, ERGAZAKI M and SPANDIDOS DA: Activaiting mutations at ras family genes in prostate cancer. Oncol Rep 1995; 2: 427 - 430.

- KITTLES RA, PANGULURI RK, CHEN W, MASSAC A, AHAGHOTU C, JACKSON A, UKOLI F, ADAMS-CAMPBELL L, ISAACS W and DUNSTON GM: CYP17 promoter variant associated with prostate cancer aggressiveness in African Americans. Cancer Epidemiol Biomarkers Prev 2001a; 10: 943 - 947.

- KITTLES RA, YOUNG D, WEINRICH S, HUDSON J, ARGYROPOULOS G, UKOLI F, ADAMS-CAMPBELL L and DUNSTON GM: Extent of linkage disequilibrium between the androgen receptor gene CAG and GGC repeats in human populations: implications for prostate cancer risk. Hum Genet 2001b; 109: 253 - 261.

- KLEE GG, GOODMANSON MK, JACOBSON SJ, YOUNG CY, FINLAY JA, RITTENHOUSE HG, WOLRERT RL and TINDALL DJ: Highly sensitive automated chemiluminometric asay for measuring free human glandular kallekrein-2. Clin Chem 1999; 45: 800 - 806.

- KLUG WS and CUMMINGS MR: Concepts of genetics (Snavely SL ed): Chapter 24, Population genetics. 1997a: p659 - 663.

- KLUG WS and CUMMINGS MR: Concepts of genetics (Snavely SL ed): Chapter 15, Recombinant DNA technology. 1997b: 450 - 451.

- KOBRINSKY NL, WINTER JS, REYES FI and FAIMAN C: Endocrine effects of vasectomy in man. Fertil Steril 1976; 27: 152 - 156.

- KOIVISTO P, KONONEN J, PALMBERG C, TAMMELA T, HYYTINEN E, ISOLA J, TRAPMAN J, CLEUTJENS K, NOORDZIJ A, VISAKORPI T and KALLIONEMI OP: Androgen receptor gene amlification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. Cancer Res 1997; 57: 314 - 319.

- KOLONEL LN: Nutrition and prostate cancer. Cancer Causes Control 1996; 7: 83 - 94.

- KOLONEL LN, NOMURA AMY, HINDS MW, HIROHATA T, HANKIN JH and LEE J: Role of diet in cancer incidence in Hawaii. Cancer Res 1983 (suppl); 43: 2397 - 2402.

- KONISHI N, ENOMOTO T, BUZARD G, OHSHIMA M, WARD JM and RICE JM: K-ras activation and ras p21 expression in latent prostatic carcinomas in Japanese men. Cancer 1992; 69: 2293 - 2299.

- KONISHI N, HIASA Y, TSUZUKI T, TAO M, ENOMOTO T and MILLER GJ: Comparison of ras activation in prostate carcinoma in Japanese and American men. Prostate 1997; 30: 53 - 57.

- KRAIN LS: Some epidemiologic variables in prostate carcinoma in California. Prev Med 1974; 3: 154 - 159.

- KRUGLYAK L: Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet 1999; 22: 139 - 144.

- KUIL CW and MULDER E: Mechanism of antiandrogen action: Conformational changes of the receptor. Mol Cell Endocrinol 1994; 102: R1 - R5.

- KUIL CW, BERREVOETS CA and MULDER E: Ligand-induced conformational alterations of the androgen receptor analyzed by limited trypsinization — Studies on the mechanism of antiandrogen action. J Biol Chem 1995; 270: 27569 - 27576.

- KUIPER GGJM, FABER PW, VAN ROOIJ HCJ, VAN DER KORPUT JAGM, RIS-STALPERS C, KLAASEN P, TRAPMAN J and BRINKMANN AO: Structural organization of the human androgen receptor gene. J Mol Endocrinol 1989; 2: R1 - R4.

- KUUKASJARVI T, KONONEN J, HELIN H, HOLLI K and ISOLA J: Loss of estrogen receptor in recurrent breast cancer is associated with poor response to endocrine therapy. J Clin Oncol 1996; 14: 2584 - 2589.

- KWOK PY, DENG Q, ZAKERI H, TAYLOR SL and NICKERSON DA: Increasing the information content of STS-based genome maps: identifying polymorphisms in mapped STSs. Genomics 1996; 31: 123 - 126.

- LABRIE F, DUPONT A, SUBURU R, CUSAN L, TREMBLAY M, GOMEZ JL and EDMOND J: Serum prostate specific antigen as pre-screening test for prostate cancer. J Urol 1992a; 147 (3 Pt 2): 846 - 851.

- LABRIE F, SUGIMOTO Y, LUU-THE V, SIMARD J, LACHANCE Y, BACHVAROV D, LEBLANC G, DUROCHER F and PAQUET N: Structure of the human type II 5α-reductase gene. Endocrinol 1992b; 131: 1571 - 1573.

- LANDER ES and SCHORK NJ: Genetic dissection of complex traits. Science 1994; 265: 2037 - 2047.

- LANDER ES, LINTON LM, BIRREN B, NUSBAUM C and ZODY MC: Initial sequencing and analysis of the human genome. Nature 2001; 409: 860 -921.

- LANDRY GL and PRIMOS WA JR: Anabolic steroid abuse. Adv Pediatr 1990; 37: 185 - 205.

- LANGLEY E, ZHOU ZX and WILSON EM: Evidence for an anti-parallel orientation of the ligand-activated human androgen receptor dimer. J Biol Chem 1995; 270: 29983 - 29990.

- LANGLEY E, KEMPPAINEN JA and WILSON EM: Intermolecular NH_2-/carboxy-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. J Biol Chem 1997; 273: 92 - 101.

- LANZ RB and RUSCONI S: A conserved carboxy-terminal subdomain is important for ligand interpretation and transactivation by nuclear receptors. Endocrinology 1994; 135: 2183 - 2195.

- LA SPADA AR, WILSON EM, LUBAHN DB, HARDING AE and FISCHBECK KH: Androgen receptor gene mutations in X-linked spinl and bulbar muscular atrophy. Nature 1991; 352: 77 - 79.

- LA SPADA AR, ROLING DB, HARDING AE, WARNER CL, SPIEGEL R, HAUSMANOWA PETRUSEWICZ I, YEE WC and FISCHBECK K: Meiotic instability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. Nature Genet 1992; 2: 301 - 304.

- LE DOUARIN B, ZECHEL C, GARNIER J-M, LUTZ Y, TORA L, PIERRAT P, HEERY D, GRONEMEYER H, CHAMBON P and LOSSON R: The N-terminal part of TIF 1, a putative mediator of ligand-dependent activation function (AF-2) of nuclear receptors, is fused to b-raf in the oncogene protein T18. EMBO J 1995; 14: 2020 - 2033.

- LEE C, KOZLOWSKI J and GRAYHACK J: Intrinsic and extrinsic factors controlling benign prostatic growth. Prostate 1997a; 31: 2131 - 2138.

- LEE WH, MORTON RA, EPSTEIN JI, BROOKS JD, CAMPBELL PA, BOVA GS, HSIEH WS, ISAACS WS and NELSON WG: Cytidine methylation of regulatory sequences near the π-class glutathione S-transferase gene accompanies human prostatic carcinogenesis. Proc Natl Acad Sci USA 1994; 91: 11733 - 11737.

- LEE WH, ISAACS WB, BOVA GS and NELSON WG: CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. Cancer Epidemiol Biomarkers Prev 1997b; 6: 443 - 450.

- LEW EA and GARFINKEL L: Variations in mortality by weight among 750,000 men and women. J Chron Dis 1979; 32: 563 - 576.

- LERMAN LS and SILVERSTEIN K: Computational simulation of DNA melting and its application to denaturing gel electrophoresis. Meth Enzymol 1987; 155: 482 - 501.

- LERNMARK A and OTT J: Sometimes it's hot, sometimes it's not. Nat Genet 1998; 19: 213 - 214.

- LIEBERMAN B, BONA BJ, EDWARDS DP and NORDEEN SK: The constitution of a progesterone response element. Mol Endrocrinol 1993; 7: 515 - 527.

- LI SL, TING SS, LINDEMAN R, FRENCH R and ZIEGLER JB: Carrier identification in X-linked immunodeficiency diseases. J Paediatr Child Health 1998; 34: 273 - 279.

- LIN CJ, MARTENS JWM and MILLER WL: NF-1C, Sp1, and Sp3 are essential for transcription of the human gene for P450c17 (steroid 17-alpha-hydroxylase) in human adrenal NC1-H295A cells. Mol Endocrinol 2001; 15: 1277 - 1293.

- LIN D, HARIKRISHNA JA, MOORE CCD, JONES KL and MILLER WL: Missense mutation $Ser^{106} \rightarrow Pro$ causes 17α-hydroxylase deficiency. J Biol Chem 1991; 266: 15992 - 15998.

- LIN D, BLACK SM, NAGAHAMA Y and MILLER WL: Steroid 17α-hydroxylase and 17,20-lyase activities of P450c17: contributions of $serine^{106}$ and P450 reductase. Endocrinology 1993; 132: 2498 - 2506.

- LINDZEY J, KUMAR MV, GROSSMAN M, YOUNG C and TINDALL DJ: Molecular mechanisms of androgen action. Vitam Horm 1994; 49: 383 - 432.

- LOOKINGBILL DP, DEMERS LM, WANG C, LEUNG A, RITTMASTER RS and SANTEN RJ: Clinical and biochemical parameters of androgen action in normal and healthy Caucasian versus Chinese subjects. J Clin Endocrinol Metab 1991; 72: 1242 - 1248.

- LUBAHN DB, JOSEPH DR, SAR M, TAN J-A, HIGGS HN, LARSON RE, FRENCH FS and WILSON EM: The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. Mol Endocrinol 1988b; 2: 1265 - 1275.

- LUBAHN DB, JOSEPH DR, SULLIVAN PM, WILLARD HF, FRENCH FS and WILSON EM: Cloning of human androgen receptor complementary DNA and localization to the X chromosome. Science 1988a; 240: 327 - 330.

- LUBAHN DB, BROWN TR, SIMENTAL JA, HIGGS HN, MIGEON CJ, WILSON EM and FRENCH TS: 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 USA 1989; 86: 9534 - 9538.

- LUISI BF, XU WX, OTWINOWSKI Z, FREEDMAN LP, YAMAMOTO KR and SIGLER PB: Crystallographic analysis of the interaction of the glucocortioid receptor with DNA. Nature 1991; 352: 497 - 505.

- LUNDWALL: Characterization of the gene for prostate-specific antigen, a human glandular kallikrein. Biochem Biophys Res Commun 1989; 161: 1151 - 1159.

- LUNN RM, BELL DA, MOHLER JL and TAYLOR JA: Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). Carcinogenesis (Lond) 1999; 20: 1727 - 1731.

- LUU-THE V, SUGIMOTO Y, PUY L, LABRIE Y, LOPEZ I, SINGH M and LABRIE F: Charaterization expression and immunohistochemical localization of 5α-reductase in human skin. J Invest Dermatol 1994; 102: 221 – 226.

- MACOSKA JA, MICALE MA, SAKR WA, BENSON PD and WOLMAN SR: Extensive genetic alterations in prostate cancer revealed by dual PCR and FISH analysis. Genes Chrom Cancer 1993; 8: 88 - 97.

- MAKRIDAKIS N, ROSS RK, PIKE MC, CHANG L, STANCZYK FZ, KOLONEL LN, SHI CY, YU MC, HENDERSON BE and REICHARDT JKV: A prevalent missense substitution that modulates activity of prostatic steroid 5α-reductase. Cancer Res 1997; 57: 1020 -1022.

- MAKRIDAKIS NM, ROSS RK, PIKE MC, CROCITTO LE, KOLONEL LN, PEARCE CL, HENDERSON BE and REICHARDT JK: Association of mis-sense substitution in *SRD5A2* gene with prostate cancer in African-American and Hispanic men in Los Angeles, USA. Lancet 1999; 354: 975 - 978.

- MANDEL JS and SCHUMAN LM: Sexual factors and prostatic cancer: results from a case-control study. J Gerontol 1987; 42: 259 - 264.

- MARCELLI M, ITTMANN M, MARIANI S, SUTHERLAND R, NIGAM R, MURTHY L, ZHAO Y, DICONCINI E, PUXEDDU E, ESEN A, EASTHAM J, WEIGEL NL and LAMB DJ: Androgen receptor mutations in prostate cancer. Cancer Res 2000; 60: 944 - 949.

- MARTH G, YEH R, MINTON M, DONALDSON R, LI Q, DUAN S, DAVENPORT R, MILLER RD and KWOK PY: Single-nucleotide polymorphisms in the public domain: how useful are they? Nat Genet 2001; 27: 371 - 372.

- MATTESON KJ, PICADO-LEONARD J, CHUNG BC, MOHANDAS TK and MILLER WL: Assignment of the gene for adrenal P450C17 (steroid 17α-hydroxylase/17,20 lyase) to human chromosome10. J Clin Endocrinol Metab 1986; 63: 789 - 791.

- MCDONNELL TJ, TRONCOSO P, BRISBAY SM, LOGOTHETIS C, CHUNG LW, HSIEH JT, TU SM and CAMPBELL ML: Expression of the protooncogene Bcl-2 in the prostate and its association with emergence of androgen-independent prostate cancer. Cancer Res 1992; 52: 6940 - 6944.

- MCNEMAR Q: Psychological statistics (Wiley J ed): New York, 1969.

- MCPHAUL MJ, MARCELLI M, TILLEY WD, GRIFFIN JE and WILSON JD: Androgen resistance caused by mutations in the androgen receptor gene. FASEB J 1991; 5: 2910 - 2915.

- MEAD G: Odyssey illustrated guide to South Africa. The Guidebook company Ltd, Hong Kong, First edition 1997; pp. 65 – 73.

- MEBANE C, GIBBS T and HORM J: Current status of prostate cancer in North American black males. J Natl Med Assoc 1990; 82: 782 - 788.

- MEIKLE AW, SMITH J and WEST D: Familial factors affecting prostate cancer risk and plasma sex-steroid levels. Prostate 1985; 6 : 121 - 128.

- METTLIN C, NATARAJAN N and HUBEN R: Vasectomy and prostate cancer risk. Am J Epidemiol 1990; 132: 1056 - 1061.

- METTLIN C: Recent developments in the epidemiology of prostate cancer. Eur J Cancer 1997; 33: 340 - 347.

- MILLER RD and KWOK P: The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. Human Molecular Genetics 2001; 10: 2195 - 2198.

- MILLER RD, TAILLON-MILLER P and KWOK PY: Regions of low single-nucleotide polymorphism incidence in human and orangutang Xq: deserts and recent coalescences. Genomics 2001; 71: 78 - 88.

- MILLER WL: Congenital lipoid adrenal hyperplasia: the human gene knockout of the steroidogenic acute regulatory protein. Journal of Molecular Endocrinology 1997; 17: 227 - 240.

- MILLER WL, AUCHUS RJ and GELLER DH: The regulation of 17,20 lyase activity. Steroids 1997; 62: 133 - 142.

- MIZOKAMI A, YEH SY and CHANG C: Identification of 3', 5'-cyclic adenosine monophosphate response element and other cis-acting elements in the human androgen receptor gene promoter. Mol Endorinol 1994; 8: 77 - 88.

- MOGENSEN J, BAHL A, KUBO T, ELANKO N, TAYLOR R and McKENNA WJ: Comparison of fluorescent SSCP and denaturing HPLC analysis with direct sequencing for mutation screening in hypertrophic cardiomyopathy. J Med Genet 2003; 40:e59 (http://www.jmedgenet.com/cgi/content/full/40/5/e59).

- MOILANEN AM, POUKKA H, KARVONEN U, HAKLI M, JANNE OA and PALVILMO JJ: Identification of a novel RING finger protein as a coregulator in steroid receptor-mediated gene transcription. Mol Cell Biol 1998; 18: 5128 - 5139.

- MOILANEN AM, POUKKA H, KARVONEN U, HAKLI M, JANNE OA and PALVILMO JJ: Identification of a novel RING finger protein as a coregulator in steroid receptor-mediated gene transcription. Mol Cell Biol 1998; 18: 5128 - 5139.

- MOILANEN AM, KARVONEN U, POUKKA H, YAN W, TOPPARI J, JANNE OA and PALVILMO JJ: A testis-specific androgen receptor coregulator that belongs to a novel family of nuclear proteins. J Biol Chem 1999; 274: 3700 - 3704.

- MO ZN, HUANG X, ZHANG SC and YANG JR: Early and late long-term effects of vasectomy on serum testosterone,dihydrotestosterone, luteinizing hormone and follicle-stimulating hormone levels. J Urol 1995; 154 : 2065 - 2069.

- MONROE KR, YU MC, KOLONEL LN, COETZEE GA, WILKENS LR, ROSS RK and HENDERSON BE: Evidence of an X-linked or recessive genetic component to prostate cancer risk. Nat Med 1995; 1: 827 - 829.

- MONTANO MM, EKENA K, KRUEGER KD, KELLER AL and KATZENELLENBOGEN BS: Human estrogen receptor ligand activity inversion mutants: receptors that interpret antiestrogens as estrogens and discriminate among different antiestrogens. Mol Endocrinol 1996; 10: 230 - 242.

- MONTIE JE and PIENTA KJ: A unifying model to explain the increased incidence and higher mortality of prostate cancer in black men. Urology 1999; 53: 1073 - 1076.

- MORTON NE, ZHANG W, TAILLON-MILLER P, ENNIS S, KWOK PY and COLLINS A: The optimal measure of allelic association. Proc Natl Acad Sci USA 2001; 98: 5217 - 5221.

- MORTON RA Jr: Racial differences in adenocarcinoma of the prostate in North American men. Urology 1994; 44: 637 - 645.

- MYERS RM, FISCHER SG, LERMAN LS and MANIATIS T: Nearly all single base substitutions in DNA fragments joined to a GC-clamp can be detected by denaturing gradient gel electrophoresis. Nucleic Acids Res 1985; 13: 3131 - 3145.

- NACHMAN MW: Single nucleotide polymorphisms and recombination rate in humans. Trends Genet 2001; 17: 481 - 485.

- NANCE MA: Clinical aspects of CAG repeat diseases. Brain Pathol 1997; 7: 881 - 900.

- NAROD SA, DUPONT A, CUSAN L, DIAMOND P, GOMEZ JL, SUBURU R and LABRIE F: The impact of family history on early detection of prostate cancer. Nat Med 1995; 1: 99 - 101.

- NATIONAL CANCER REGISTRY OF SOUTH AFRICA, 1992.

- NAVONE NM, TRONCOSO P, PISTERS LL, GOODROW TL, PALMER JL, NICHOLS WW, ESCHENBACH AC and CONTI CJ: p53 protein accumulation and gene mutation in the progression of human prostate carcinoma. J Natl Cancer Inst 1993; 85: 1657 - 1669.

- NELSON DR, KAMATAKI T, WAXMAN DJ, GUENGERICH FP, ESTABROOK RW, FEYEREISEN R, GONZALEZ FJ, COON MJ, GUNSALUS IC, GOTOH O, OKUDA K and NEBERT W: The P450 superfamily: update on new sequences,

gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol 1993; 12: 1 - 51.

- NEWMARK JR, HARDY DO, TONB DC, CARTER BS, EPSTEIN JI, ISAACS WB, BROWN TR and BARRACK ER: Androgen receptor gene mutations in human prostate cancer. Proc Natl Acad Sci 1992; 89: 6319 - 6323.

- NOBEL RL: The development of prostatic adenocarcinoma in Nb rats following prolonged sex hormone administration. Cancer Res 1959; 19: 1125 - 1139.

- NOBEL RL: The development of prostatic adenocarcinoma in Nb rats following prolonged sex hormone administration. Cancer Res 1977; 37: 1929 - 1933.

- NOMURA AM and KOLONEL LN: Prostate cancer: a current perspective. Am J Epidemiol 1991; 13: 200 - 227.

- NORDEEN SK, SUH BJ, KUHNEL B and HUTCHISON CA: Structural determinants of a glucocorticoid response element. Mol Endocrinol 1990; 4: 1866 - 1873.

- NUPPONEN NN, KAKKOLA L, KOIVISTO P and VISAKORPI T: Genetic alterations in hormone-refractory recurrent prostate carcinomas. Am J Pathol 1998; 153: 141 - 148.

- O'MALLEY B: The steroid receptor superfamily: more excitement predicted for the future. Mol Endocrinol 1990; 4: 363 - 369.

- O'MALLEY BW and TSAI MJ: Molecular pathways of steroid receptor action. Biol Reprod 1992; 46: 163 - 167.

- O'MALLEY BW, TSAI SU, BAGCHI M, WERGEL NL, SCHRADER WT and TSAI MJ: Molecular mechanism of action of a steroid hormone receptor. Recent Progress in Hormone Research 1991; 47: 1 - 26.

- ONATE SA, TSAI SY, TSAI M-J and O'MALLEY BW: Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. Science 1995; 270: 1354 - 1357.

- OPTENBERG SA, THOMPSON IM, FRIEDRICHS P, WOJCIK B, STEIN CR and KRAMER B: Race, treatment, and long-term survival from prostate cancer in an equal-access medical care delivery system. JAMA 1995; 274: 1599 - 1605.

- PALMBERG C, KOIVISTO P, HYYTINEN E, ISOLA J, VISAKORPI T, KALLIONEMI OP and TAMMELA T: Androgen receptor gene amplification in a recurrent prostate cancer after monotherapy with nonsteroidal potent antiandrogen "Casodex" (bicalutamide) with a subsequent favorable response to maximal androgen blockade. Eur Urol 1997; 31: 216 - 219.

- PALMBERG C, KOIVISTO P, KAKKOLA L, TAMMELA TLJ, KALLIONEMI OP and VISAKORPI T: Androgen receptor gene amplification at the time of primary progression predicts response to combined androgen blockade as a second-line therapy in advanced prostate cancer. J Urol 2000; 164: 1992 - 1995.

- PAN Y, MATSUYAMA N, WANG N, YOSHIHIRO S, HAGGARTH L, LI C, TRIBUKAIT B, EKMAN P and BERGERHEIM US: Chromosome 16q24 deletion and decreased E-cadherin expression: possible association with metastatic potential in prostate cancer. Prostate 1998; 36: 31 - 38.

- PARIS F, WEINBAUER GF, BLUM V and NIESCHLAG F: The effect of androgens and antiandrogens on the immunohistochemical localization of the androgen receptor in accessory reproductive organs of male rats. J Steroid Biochem Mol Biol 1994; 48: 129-137.

- PARIS PL, KUPELIAN PA, HALL JM, WILLIAMS TL, LEVIN H, KLEIN EA, CASEY G and WITTE JS: Association between a *CYP3A4* genetic variant and clinical presentation in African-American prostate cancer patients. Cancer Epidemiol Biomark Prev 1999; 8: 901 - 905.

- PARKER SL, TONG T, BOLDEN S and WINGO PA: Cancer Statistics. CA Cancer J Clin 1997; 47: 5 - 27.

- PARKER M, WEBB P, MILLS J, et al: Analysis of regulatory sequences in androgen-responsive genes. J Steroid Biochem 1988; 30: 47 - 51.

- PAZ A, LINDNER A, ZISMAN A, et al: A genetic sequence change in the 3'-noncoding region of the androgen receptor gene in prostate carcinoma. Eur Urol 1997; 31: 209 - 215.

- PEELING WB: Phase III studies to compare goserelin (zoladex) with orchiectomy and with diethylstilbestrol in treatment of prostatic carcinoma. Urology 1989; 13(suppl 5): 45 - 52.

- PESCHE S, LATIL A, MUZEAU F, CUSSENOT O, FOURNIER G, LONGY M, ENG L and LIDERLAU R: PTEN/MMAC1/TEP1 involvement in primary prostate cancers. Oncogene 1998; 16: 2879 - 2883.

- PETERSON DE, REMINGTON PL and ANDERSON HA: Re: "Vasectomy and the risk of prostate cancer". Am J Epidemiol 1992; 135: 324 - 325.

- PERLMAN JA, SPIRTAS R and KELAGHAN J: Vasectomy and the risk of prostate cancer. Am J Epidemiol 1991; 134: 107 - 108.

- PHILLIPS SM, BARTON CM, LEE SJ, MORTON DG, WALLACE DM, LEMOINE NR and NEOPTOLEMOS: Loss of the retinoblastoma susceptibility gene (Rb1) is a frequent and early event in prostatic tumorigenesis. Br J Cancer 1994; 70: 1252 - 1257.

- PICADO-LEONARD J and MILLER WL: Cloning and sequence of the human gene for P450c17 (steroid 17α-hydroxylase/17,20 lyase): similarity with the gene for P450C21. DNA 1987; 6: 439 - 448.

- PIENTA KJ, DEMERS R, HOFF M, KAU TY, MONTIE JE and SEVERSON RK: Effect of age and race on the survival of men with prostate cancer in the metropolitan Detroit Tri-County area, 1973 to 1987. Urology 1995; 45: 93 - 102.

- PIKE MC, KRAILO MD, HENDERSON BE, CASAGRANDE JT and HOEL DG: Hormonal risk factors, breast tissue age and the incidence of breast cancer. Nature (Lond) 1983; 303: 767 - 770.

- PLATZ EA, GIOVANNUCCI E, DAHL DM, KRITHIVAS K, HENNEKENS CH, BROWN M, STAMPFER MJ, and KANTOFF PW: The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol Biomark Prev 1998; 7: 379 - 384.

- POWELL IJ: Prostate cancer in the African-American: is this a different disease? Semin Urol Oncol 1998; 16: 221 - 226.

- POWELL IJ, SCHWARTZ K and HUSSAIN M: Removal of the financial burden of health care: Does it impact on prostate cancer at presentation and survival? A comparative study between black and white men in a Veterans Affairs system. Urology 1995; 46: 825 - 830.

- PRESTON-MARTIN S, PIKE MC, ROSS RK and HENDERSON BE: Epidemiologic evidence for the increased cell proliferation model of carcinogenesis. Environ Health Perspect 1993; 101 (Suppl5): 137 - 138.

- PRINS GS: Molecular biology of the androgen receptor. Supplement to the Mayo Clin Proc 2000; 75: S32 -S35.

- PRITCHARD JK: Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001; 69: 124 - 137.

- PTASHNE M: How eukaryotic transcriptional activators work. Nature, London 1988; 335: 683 - 689.

- PURVIS K, SAKSENA SK, CEKAN Z, DIEZFALUSY CK and GINGER J: Endocrine effects of vasectomy. Clin Endocrinol (Oxf) 1976; 5: 263 - 272.

- QUIGLEY CA, DE BELLIS A, MARSCHKE KB, AL-AWADY MK, WILSON EM and FRENCH FS: Androgen receptor defects: Historical, clinical and molecular perspectives. Endocrine Rev 1995; 16: 271 - 321.

- RAIVIO T, SANTTI H, SCHATZL G, GSUR A, HAIDINGER G, PALVILMO JJ, JANN OA, MADERSBACHER S: Reduced circulating androgen bioactivity in patients with CaP. Prostate 2003; 55(3): 194-198.

- RANDALL VA: Role of 5α-reductase in health and disease. Bailliere's Clinical Endocrinology and Metabolism 1994; 8, No 2: 405 - 431.

- REBBECK TR: More about: modification of clinical presentation of prostate tumours by a novel genetic variant in CYP3A4. J Natl Cancer Inst 2000; 92: 76.

- REBBECK TR, JAFFE JM, WALKER AH, WEIN AJ and MALKOWICZ SB: Modification of clinical presentation of prostate tumors by a novel genetic variant in CYP3A4. J Natl Cancer Inst 1998; 90: 1225 - 1229.

- REBBECK TR, KANTOFF PW, KRITHIVAS K, NEUHAUSEN S, BLACKWOOD MA, GODWIN AK, DALY MB, NAROD SA, GARBER JE, LYNCH HT, WEBER BL and BROWN M: Modification of BRCA1-associated breast cancer risk by the polymorphic androgen receptor CAG repeat. Am J Hum Genet 1999; 64: 1371 - 1377.

- REICHARDT JKV, MAKRIDAKIS N, HENDERSON BE, YU MC, PIKE MC and ROSS RK: Genetic variability of the human SRD5A2 gene: implications for prostate cancer risk. Cancer Res 1995; 55: 3973 - 3975.

- REICHARDT J, MAKRIDAKIS N, PIKE M, CROCITTO L, KOLONEL L, ROSS R and HENDERSON B: The A49T mutation in the *SRD5A2* gene increases risk for

aggressive prostate cancer and prostatic steroid 5-α reductase activity. Am J Hum Genet 1997; 61: A209

- RENAUD JP, ROCHEL N, RUFF M, VIVAT V, CHAMBON P, GRONEMEYER H and MORAS D: Crystal structure of the RAR-γ ligand-binding domain bound to all-trans retinoic acid. Nature 1995; 378: 681 - 689.

- RENNIE P, BRUCHOVSKY N, LECO K, SHEPPARD PC, McQUEEN SA, CHENG H, SNOEK R, HAMEL A, BOCK ME, MACDONALD BS, NICKEL BE, LIAO CCS, CATTINI PA and MATUSIK RJ: Characterization of two cis-acting DNA elements involved in the androgen regulation of the probasin gene. Mol Endocrinol 1993; 7: 23 - 36.

- RIETBERGEN JBW, KRANSE R, KIRKELS WJ, DE KONING HJ and SCHRÖDER FH: Evaluation of PSA, DRE and TRUS in population based screening for prostate cancer: suggestions to improve efficacy of early detection. Br J Urol 1997; 79 (Suppl 2): 57 -63.

- RISCH N and MERIKANGAS K: The future of genetic studies of complex human diseases. Science 1996; 273: 1516 - 1517.

- RINKER-SCHAEFFER CW, HAWKINS AL, RU N, DONG J, STOICA G, GRIFFIN CA, ICHIKAWA T, BARRETT JC and ISAACS JT: Differential suppression of mammary and prostate cancer metastasis by human chromosomes 17 and 11. Cancer Res 1994; 54: 6249 - 6256.

- ROCHE PJ, HOARE SA and PARKER MG: A consensus DNA-binding site for the androgen receptor. Mol Endocrinol 1992; 6: 2229 - 2235.

- ROSENBERG L, PALMER JR, ZAUBER AG, WARSHAUER ME, STOLLEY PD, SHAPIRO S: Vasectomy and the risk of prostate cancer. Am J Epidemiol 1990; 132: 1051 - 1055.

- ROSS RK, BERNSTEIN L, JUDD H, HANISCH R, PIKE M and HENDERSON B: Serum testosterone levels in healthy young black and white men. J Natl Cancer Inst 1986; 76: 45 - 48.

- ROSS RK, BERNSTEIN L, LOBO RA, SHIMIZU H, STANCZYK FZ, PIKE MC and HENDERSON BE: 5-α Reductase activity and risk of prostate cancer among Japanese and US white and black males. Lancet 1992; 339: 887 - 889.

- ROSS RK, SHIMIZU H, PAGANINI-HILL A, HONDA G and HENDERSON BE. Case-control studies of prostate cancer in blacks and whites in southern California. J Natl Cancer Inst 1987; 78: 869 - 874.

- ROSS RK and COETZEE GA: The epidemiology and etiology of prostate cancer; in Petrovich Z, Baert L, Brady LW (eds) Diagnostic Imaging and Radiation Oncology. Berlin, Springer, 1996: 1 – 11.

- ROSS RK and SCHOTTENFELD D: Prostate Cancer. In Schottenfeld D and Fraumeni JF (eds). Cancer Epidemiology and Prevention 1996: 1180 - 1206, New York University Press.

- ROSS RK, PIKE MC, COETZEE GA, REICHARDT JKV, YU MC, FEIGELSON H, STANCZYK FZ, KOLONEL LN and HENDERSON BE: Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res 1998; 58: 4497 - 4504.

- ROSS RK, COETZEE GA, PEARCE CL, REICHARDT JKV, BRETSKY P, KOLONEL LN, HENDERSON BE, LANDER E, ALTSHULER D and DALEY G: Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Eur Urol 1999; 35: 355 – 361.

- RUSSELL DW, BERMAN DM, BRYANT JT, CALA KM, DAVIS DL, LANDRUM CP, PRIHODA JS, SILVER RI, THIGPEN AE and WIGLEY WC: The molecular genetics of steroid 5α-reductases. Recent Prog Horm Res 1994; 49: 275 - 284.

- RUIZEVELD DE WINTER JA, JANSEN PJ, SLEDDENS HMEB, VERLEUN-MOOIJMAN MCT, TRAPMAN J, BRINKMANN AO, SANTERSE AB, SCHRÖDER FH and VAN DER KWAST TH: Androgen receptor status in localised and locally progresive hormone refractory human prostate cancer. Am J Pathol 1994; 144: 735 - 746.

- SANDBERG AA: Chromosomal abnormalities and related events in prostate cancer. Hum Pathol 1992; 23: 368 - 380.

- SATA F, SAPONE A, ELIZONDO G, STOCKER P, MILLER VP, ZHENG W, RAUNIO H, CRESPI CL and GONZALEZ FJ: CYP3A4 allelic variants with amino acid substitutions in exons 7 and 12: evidence for an allelic variant with altered catalytic activity. Clin Pharmacol Ther 2000; 67: 48 - 56.

- SATTLER HP, ROHDE V, BONKHOFF H, ZWERGEL T and WULLICH B: Comparative genomic hybridization reveals DNA copy number gains to frequently occur in human prostate cancer. Prostate 1999; 39: 79 - 86.

- SAWAYA ME and SHALITA AR: Androgen receptor polymorphisms (CAG repeat lengths) in androgenic alopecia, hirsuitism, and acne. J Cutan Med Surg 1998; 3: 9 - 15.

- SCHAID DJ, MCDONELL S, BLUTE M and THIBODEAU S: Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 1998; 62 : 1425 - 1438.

- SCHEDLICH L, BENNETTS BH and MORRIS BJ: DNA 1987; 6: 429 - 437.

- SCHLECHTE H, LENK SV, LONING T, SCHNORR D, RUDOLPH BD, DITSCHERLEIN G and LOENING SA: P53 tumor suppressor gene mutations in benign prostatic hyperplasia and prostate cancer. Eur Urol 1998; 34: 433 - 440.

- SCHOENBERG MP, HAKIMI JM, WANG S, BOVA GS, EPSTEIN JI, FISCHBECK KH, ISAACS WB, WALSH PC and BARRACK ER: Microsatellite mutation (CAG_{24-18}) in the androgen receptor gene in human prostate cancer. Biochem Biophys Res Commun 1994; 198: 74 - 80.

- SCHWABE JWR, CHAPMAN L, FINCH JT and RHODES D: The crystal structure of the estrogen receptor DNA-binding domain bound to DNA: how receptors discriminate between their response elements. Cell 1993; 75: 567 - 578.

- SEVERSON RK, GROVE JS, NOMURA AM, and STEMMERMANN GN: Body mass and prostatic cancer: A prospective study. BMJ 1988; 297 : 713 - 715.

- SEVERSON RK, NOMURA AMY, GROVE JS and STEMMERMANN GN: A prospective study of demographics, diet, and prostate cancer among men of Japanese ancestry in Hawaii. Cancer Res 1989; 49: 1857 - 1860.

- SHEFFIELD VC, COX DR, LERMAN LS and MYERS RM: Attachment of a 40 base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single base changes. Proc Natl Acad Sci USA 1989; 86: 232 - 236.

- SHILLINGTON K: History of Southern Africa. Longman Group, UK Limited, 1987; pp. 7 – 31.

- SHIMADA N, SOBUE G, DOYU M, YAMAMOTO K, YASUDA T, MUKAI E, KACHI T and MITSUMA T: X-linked recessive bulbospinal neuronopathy: clinical phenotypes and CAG repeat size in androgen receptor gene. Muscle Nerve 1995; 18: 1378 - 1384.

- SHIMIZU H, ROSS RK and BERNSTEIN L: Underestimation of the incidence rate of prostate cancer in Japan. Jpn J Cancer Res 1991; 82: 483 - 485.

and Pathophysiologic Effects in Animals and Man. New York (NY): Academic Press 1979: 183 - 197.

- SMITH DF and TOFT DO: Steroid receptors and their associated proteins. Mol Endocrinol 1993; 7: 4 - 11.

- SPITZ MR, CURRIER R, FUEGER J, BABAIAN R and NEWELL G: Familial patterns of prostate cancer: a case-control analysis. J Urol 1991; 146: 1305 - 1307.

- SPITZ MR, FUEGER JJ, BABAIAN RJ and NEWELL GR: Vasectomy and the risk of prostate cancer. Am J Epidemiol 1991; 134 : 108 - 109.

- STANFORD JL, JUST JJ, GIBBS M, WICKLUND KG, NEAL CL, BLUMENSTEIN BA and OSTRANDER EA: Polymorphic repeats in the androgen receptor gene: Molecular markers of prostate cancer risk. Caner Res 1997; 57: 1194 - 1198.

- STEINBERG GD, CARTER B, BEATY T, CHILDS B and WALSH P: Family history and the risk of prostate cancer. Prostate 1990; 17: 337 - 347.

- STOCCO DM and CLARK BJ: Regulation of the acute production of steroids in steroidogenic cells. Endocrine Review 1996; 17: 221 - 244.

- STRAUB B, MULLER M, KRAUSE H, SCHRADER M and MILLER K: Real-time quantitative reverse transcriptase-polymerase chain reaction for luteinizing hormone-releasing hormone receptor gene mRNA expression in human prostate cancer. Urology 2003; 62(1): 172-176.

- SUPAKAR PC, SONG CS, JUNG MH, SLOMCZYNSKA MA, KINI JM, VERLANOWETH RL, CHATTERJEE B and ROY AK: A novel regulatory element asociated with age-dependent expression of the rat androgen receptor gene. J Biol Chem 1993; 268: 26400 - 26408.

- SUZUKI H, SATO N, WATABE Y, MASAI M, SEINO S and SHIMAZAKI J: Androgen receptor gene mutations in human prostate cancer. J Steroid Biochem Mol Biol 1993; 46: 759 - 765.

- SUZUKI H, AKAKURA K, KOMIYA A, AIDA S, AKIMOTO S and SHIMAZAKI J: Codon 877 in the androgen receptor gene in advanced prostate cancer: relation to antiandrogen withdrawal syndrome. Prostate 1996; 29: 153 - 158.

- SUZUKI H, FREIJE D, NUSSKERN DR, OKAMI K, CAIRNS P, SIDRANSKY D, ISAACS WB and BOVA GS: Interfocal heterogeneity of PTEN/MMAC1 gene alterations in multiple metastatic prostate cancer tissues. Cancer Res 1998; 58: 204 - 209.

- SYVANEN AC, LANDEGREN U, ISAKSSON A, GYLLENSTEN U and BROOKES AJ: Enthusiasm mixed with skepticism about single-nucleotide polymorphism markers for dissecting complex disorders. Eur J Hum Genet 1999; 7: 98 - 101.

- TAILLON-MILLER P, GU ZJ, LI Q, HILLIER L and KWOK PY: Overlapping genomic sequences – a treasure trove of single-nucleotide polymorphisms. Genome Res 1998; 8: 748 - 754.

- TAILLON-MILLER P, PIERNOT EE and KWOK PY: Efficient approach to unique single-nucleotide polymorphism discovery. Genome Res 1999; 9: 499 - 505.

- TAKAHASHI H, FURUSATO M, ALLSBROOK WC JR, NISHII H, WAKUI S, BARRET JC and BOYD J: Prevalence of androgen receptor gene mutations in latent prostatic carcinomas from Japanese men. Cancer Res 1995; 55: 1621 - 1624.

- TAN J, SHARIEF Y, HAMIL KG, GREGORY CW, ZANG DY, SAR M, GUMERLOCK PH, DEVERE WHITE RW, PRETLOW TG, HARRIS SE, WILSON EM, MOHLER JL and FRENCH FS: Dehydroepiandrosterone activates mutant

androgen receptors expressed in the androgen-dependent human prostate cancer xenograft CWR22 and LNCaP cells. Mol Endocrinol 1997; 11: 450 - 459.

- TANENBAUM DM, WANG Y, WILLIAMS SP and SIGLER PB: Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. Proc Natl Acad Sci USA 1998; 95: 5998 - 6003.

- TAPLIN ME, BUBLEY GJ, SHUSTER TD, FRANTZ ME, SPOONER AE, OGATA GK, KEER HN and BALK SP: Mutation of the androgen-receptor gene in metastatic androgen-independent prostate cancer. N Engl J Med 1995; 332: 1393 - 1398.

- TAPLIN ME, BUBLEY GJ, KO YJ, SMALL EJ, UPTON M, RAJESHKUMAR B and BALK SP: Selection for androgen receptor mutations in prostate cancers treated with androgen antagonists. Cancer Res 1999; 59: 2511 - 2515.

- TAYLOR JG, CHOI E, FOSTER CB and CHANOCK SJ: Using genetic variation to study human disease. Trends in Mol Med 2001; 7(11): 507-512.

- TERWILLIGER JD and GORING HH: Gene mapping in the 20th and 21st centuries: statistical methods , data analysis, and experimental design. Hum Biol 2000; 72: 63 - 132.

- THE LEUPROLIDE STUDY GROUP: Leuprolide versus diethylstilbestrol for metastatic prostate cancer. N Engl J Med 1984; 311(20): 1281 - 1286.

- THIGPEN AE, SILVER RI, GUILEYARDO JM, CASEY ML, McCONNELL JD and RUSSELL DW: Tissue distribution and ontogeny of steroid 5α-reductase isozyme expression. J Clin Invest 1993; 92: 903 - 910.

- THOMPSON MW, MCINNES RR and WILLARD HF: Genetics in medicine: Chapter 7, Genes in populations. 1991: p143 - 150.

- THOMPSON TC and CHUNG LW: Extraction of nuclear androgen receptors by sodium molybdate from normal rat prostates and prostatic tumors. Cancer Res 1984; 44: 1019 - 1026.

- THOMPSON MW, McINNES RR and WILLARD HF: Patterns of single-gene inheritance. Genetics in Medicine 1991, 5th edition, WB Saunders Company; pp 53.

- TILLEY WD, MARCELLI M, WILSON JD and MCPHAUL: Characterization and expression of a xDNA encoding the human androgen receptor. Proc Natl Acad Sci USA 1989; 86: 327 – 331.

- TILLEY WD, MARCELLI M and MCPHAUL MJ: Expression of the human androgen receptor gene utilizes a common promoter in diverse human tissues and cell lines. J Biol Chem 1990; 265: 13776 - 13781.

- TILLEY WD, LIM-TIO SS, HORSFALL DJ, ASPINALL JO, MARSHALL VR and SKINNER JM: Detection of discrete androgen receptor epitopes in prostate cancer by immunostaining: measurement by color video image analysis. Cancer Res 1994; 54: 4096 - 4102.

- TILLEY WD, BUCHANAN G, HICKEY TE and BENTEL JM: Mutations in the androgen receptor gene are associated with progression of human prostate cancer to androgen independence. Clin Cancer Res 1996; 2: 277 - 285.

- TOFT D and GORSKY J: A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. Proc Natl Acad Sci USA 1966; 5: 1574 - 1581.

- TRAPMAN J and BRINKMANN AO: The androgen receptor in prostate cancer. Pathol Res Pract 1996; 192: 752 - 760.

- TRAPMAN J and CLEUTJENS KMJB: Androgen-regulated gene expression in prostate cancer. Cancer Biology 1997; 8: 29 – 36

- TRAPMAN J, KLAASSEN P, KUIPER GGJM, VAN DER KORPUT JAGM, FABER PW, VAN ROOIJ HCJ, GEURTS VAN KESSEL A, VOORHORST MM, MULDER E and BRINKMANN AO: Cloning, structure and expression of a cDNA encoding the human androgen receptor. Biochem Biophys Res Commun 1988; 153: 241 - 248.

- TRUSS M and BEATO M: Steroid hormone receptors: interaction with deoxyribonucleic acid and transcription factors. Endocrine Rev 1993; 14: 459 - 479.

- TSITOURAS PD, MARTIN CE and HARMAN SM: Relationship of serum testosterone to sexual activity in healthy elderly men. J Gerontol 1982; 37: 288 - 293.

- TUT TG, GHADESSY FJ, TRIFIRO MA, PINSKY L and YOUNG EL: Long polyglutamine tracts in the androgen receptor are associated with reduced trans-activation, impaired sperm production, and male infertility. J Clin Endocrinol Metab 1997; 82: 3777 - 3782.

- VELDSCHOLTE J, RIS-STALPERS C, KUIPER GGJM, JENSTER G, BERREVOETS C, CLAASSEN E, VAN ROOIJ HCJ, TRAPMAN J, BRINKMANN AO and MULDER E: A mutation in the ligand binding domain of the androgen receptor of human LNCaP cells affects steroid binding characteristics and response to anti-androgens. Biochem Biophys Res Commun 1990; 173: 534 - 540.

- VELDSCHOLTE J, BERREVOETS CA and ZEGERS ND: Hormone-induced dissociation of the androgen receptor-heat-shock protein complex: use of a new monoclonal antibody to distinguish transformed from nontransformed receptors. Biochemistry 1992; 31: 7422 - 7430.

- VISAKORPI T, HYYTINEN E, KOIVISTO P, TANNER M, KEINANEN R, PALMBERG C, PALOTIE A, TAMMELA T, ISOLA J KALLIONEMI OP: In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet 1995a; 9: 401 - 406.

- VISAKORPI T, KALLIONIEMI AH, SYVANEN AC, HYYTINEN ER, KARHU R, TAMMELA T, ISOLA JJ and KALLIONEMI OP: Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. Cancer Res 1995b; 55: 342 - 347.

- VLIESTRA RJ, VAN ALEWIJK DC, HERMANS KG, VAN STEENBRUGGE GJ, TRAPMAN J: Frequent inactivation of PTEN in prostate cancer cell lines and xenografts. Cancer Res 1998; 58: 2720 - 2723.

- VOEGEL JJ, HEINE MJS, ZECHEL C, CHAMBON P and GRONEMEYER H: TIF2, a 160kDa transcriptional mediator for the ligand-dependent activation function AF-2 of nuclear receptors. EMBO J 1996; 15: 3667 - 3675.

- VOELLER HJ, SUGARS LY, PRETLOW T and GELMANN EP: P53 oncogene mutations in human prostate cancer specimens. J Urol 1994; 151: 492 - 495.

- VOM BAUR E, ZECHEL C, HEERY D, HEINE MJ, GARNIER JM, VIVAT V, LE DOUARIN B, GRONEMEYER H, CHAMBON P and LOSSON R: Differential ligand-dependent interactions between the AF-2 activating domain of nuclear receptors and the putative transcriptional intermediary factors mSUG1 and TIF1. EMBO J 1996; 15: 110 - 124.

- WADELIUS M, ANDERSON S, JOHANSSON J, WADELIUS C and RANE A: Prostate cancer associated with *CYP17* genotype. Pharmacogenetics 1999; 9: 635 - 639.

- WAGNER RL, APRILETTI JW, MCGRATH ME, WEST BL, BAXTER JD and FLETTERICK RJ: A structural role for hormone in the thyroid hormone receptor. Nature 1995; 378: 690 - 697.

- WANG DG, FAN JB, SIAO CJ, BERNO A, YOUNG P, SAPOLSKY R, GHANDOUR G, PERKINS N, WINCHESTER E, SPENCER J, KRUGLYAK L, STEIN L, HSIE L, TOPALOGLOU T, HUBBELL E, ROBINSON E, MITTMAN M, MORRIS MS, SHEN N, KILBURN D, RIOUX J, NUSBAUM C, ROZEN S, HUDSON TJ, LANDER ES, LIPSHUTZ R and CHEE M: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998a; 280: 1077 - 1082.

- WANG MC, PAPSIDERO LD, KURIYAMA M, VALENZULA IA, MURPHY GP and CHU TM: Prostate antigen: a new potential marker for prostatic cancer. Prostate 1981; 2: 89 - 96.

- WANG Q, GHADESSY FJ and YONG EL: Analysis of the transactivation domain of the androgen receptor in patients with male infertility. Clin Genet 1998b; 54: 185 - 192.

- WANG SI, PARSONS R and ITTMAN M: Homozygous deletion of the PTEN tumor suppressor gene in a subset of prostate adenocarcinomas. Clin Cancer Res 1998c; 4: 811- 815.

- WATERMAN MR and KEENEY DS: Genes involved in androgen biosynthesis and the male phenotype. Horm Res 1992; 38: 217 - 221.

- WATT KWK, LEE PJ, TIMKULU T, CHAN WP and LOOR R: Proc Natl Acad Sci USA 1986; 83: 3166 - 3170.

- WEINBERGER C, HOLLENBERG S, ROSENFELD M and EVANS R: Domain structure of human glucocorticoid receptor and its relationship to the v-erbA oncogene product. Nature 1985; 318: 670 - 672.

- WERTZ IE, DEITCH AD, GUMERLOCK PH, GANDOUR-EDWARDS R, CHI SG and VERE WHITE RW: Correlation of genetic and immunodetection of TP53 mutations in malignant and benign prostate tissues. Hum Pathol 1996; 27(6): 573 - 580.

- WEST D, SLATTERY ML, ROBISON LM, FRENCH TK and MAHONEY AW: Adult dietary intake and prostate cancer in Utah: a case-control study with special emphasis on aggressive tumors. Cancer Causes Control 1991; 2: 85-94.

- WESTON A, PAN CF, BLEIWEISS IJ, KSIESKI HB, ROY N, MALONEY N and WOLF MS: *CYP17* genotype and breast cancer risk. Cancer Epidemiol Biomark Prev 1998; 7: 941 - 944.

- WHITBY RM, GORDON RD and BLAIR BR: The endocrine effects of vasectomy: a prospective five-year study. Fertil Steril 1979; 31: 518 - 520.

- WILLIAMS SP and SIGLER PB: Atomic structure of progesterone complexed with its receptor. Nature 1998; 393: 392 - 396.

- WILSON CM, GRIFFIN JE, WILSON JD, MARCELLI M, ZOPPI S and MCPHAUL MJ: Immunoreactive androgen receptor expression in subjects with androgen resistance. J Clin Endocr Metab 1992; 75: 1474 - 1478.

- WILSON EM and FRENCH FS: Binding properties of androgen receptors: Evidence for identical receptors in rat testis, epididymis and prostate. J Biol Chem 1976; 251: 5620 - 5629.

- WILSON JD: The pathogenesis of benign prostatic hyperplasia. Am J Med 1980; 68: 745 - 756.

- WILSON JD: Syndromes of androgen resistance. Biology of Reproduction 1992; 46: 168 - 173.

- WILSON JD: Role of dihydrotestosterone in androgen action. The Prostate Suppl 1996; 6: 88 - 92.

- WINTER SF, COOPER AB and GREENBERG NM. Models of metastatic prostate cancer: a transgenic perspective. Prostate Cancer Prostatic Dis 2003; 6(3): 204-211.

- WOLF DA, HERZINGER T, HERMEKING H, BLASCHKE D and HORZ W: Transcriptional and posttranscriptional regulation of human androgen receptor expression by androgen. Mol Endocrinol 1993; 7: 924 - 936.

- WOLF D, SCHULZ P and FITTLER F: Transcriptional regulation of prostate kallikrein-like genes by androgen. Mol Endocrinol 1992; 6: 753 - 762.

- WOOD RD, MITCHELL M, SGOUROS J, and LINDAHL T: Human DNA repair genes. Science 2001; 291: 1284 - 1289.

- WONG C-I, ZHOU Z-X, SAR M and WILSON EM: Steroid requirement for androgen receptor dimerization and DNA binding. Modulation by intramolecular interactions between the NH_2-terminal and steroid-binding domains. J Biol Chem 1993; 268: 19004 -19012.

- WU Y, HAYES VM, OSINGA J, MULDER IM, LOOMAN MWG, BUYS CHCM and HOFSTRA RMW: Improvement of fragment and primer selection for mutation detection by denaturing gradient gel electrophoresis. Nucleic Acids Res 1998; 26: 5432 -5440.

- WURTZ JM, BOURGUET W, RENAUD JP, VIVAT V, CHAMBON P and MORAS D: A canonical structure for the ligand-binding domain of nuclear receptors. Nature Struct 1996; 3: 87 - 94.

- XU J, MEYERS D, FREIJE D, ISAACS S, WILEY K, NUSSKERN D, EWING C, WILKENS E, BUJNOVSZKY P, BOVA GS, WALSH P, ISAACS W,

SCHLEUTKER J, MATIKAINEN M, TAMMELA T, VISAKORPI T, KALLIONIEMI OP, BERRY R, SCHAID D, FRENCH A, MCDONNELL S, SCHROEDER J, BLUTE M, THIBODEAU S and TRENT J: Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 1998; 20: 175 - 179.

- YEH S and CHANG C: Cloning and characterization of a specific coactivator, ARA70, for the androgen receptor in human prostate cells. Proc Natl Acad Sci USA 1996; 93: 5517 - 5521.

- YOSHIDA O, OISHI K, OHNO Y and SCHRÖDER FH: A comparative study on prostatic cancer in the Netherlands and Japan. In: Sasaki R, Aoki K, editors. Proceedings Monbusho International Symposium. Comparitive Study of Etiology & Prevention of Cancer. Nagoya (Japan): Nagoya Press 1990: 73 - 84.

- ZENG Z, ALLAN GF, THALLER C, COONEY AJ, TSAI SY, O'MALLEY BW and TSAI MJ: Detection of potential ligands for nuclear receptors in cellular extracts. Endocrinol 1994; 135: 248 - 252.

- ZHAO XY, MALLOY PJ, KRISHNAN AV, SWAMI S, NAVONE NM, PEEHL DM and FELDMAN D: Glucocorticoids can promote androgen-independent growth of prostate cancer cells through a mutated androgen receptor. Nat Med 2000; 6: 703 - 706.

- ZHOU ZX, SAR M, SIMENTAL JA, LANE MV and WILSON EM: A ligand-dependant bipartite nuclear targeting signal in the human androgen receptor. Requirement for the DNA-binding domain and modulation by NH_2-terminal and carboxy-terminal sequences. J Biol Chem 1994; 269: 13115 - 13123.

- ZHUANG YH, BLAUER M, PEKKI A and TUOHIMAA P: Subcellular location of androgen receptor in rat prostate, seminal vesicle and human osteosarcoma MG-63 cells. J Steroid Biochem Mol Biol 1992; 41: 693 - 696.

- ZONG J, ASHRAF J and THOMPSON EB: The promoter and first, untranslated exon of the human glucocorticoid receptor gene are GC-rich but lack consensus glucocorticoid receptor element sites. Mol Cell Biol 1990; 10: 5580 - 5585.

Notes to Appendix II:

Allele frequencies

Genetic comparisons of tumour and blood material of the CaP patients[36], and that obtained from the BPH (normal tissue), as well as blood samples from the general population[37] were performed, to determine the allele frequencies (by manual allele counting) of polymorphic sites. Control samples from the WPBTS consisted of males and females. The number of individuals referred to in allelic frequency tables are either homozygous (MM), heterozygous (MN) for a specific SNP, or without the SNP (NN).

When calculating allelic frequencies it must be remembered that the *AR* gene is located on the X chromosome and males therefore only have a single copy of the gene, as opposed to females[38] who have two X-chromosomes and therefore two copies of the gene.

Testing for Hardy-Weinberg equilibrium

Allele frequencies determine genotype frequencies. p^2 is the measure of the frequency of MM homozygotes for a specific SNP in the ensuing generation. In a similar way, 2pq describes the frequency of the MN heterozygotes, and q^2 is a measure of the frequency of the NN " normals", i.e. individuals without SNPs. The distribution of homozygous and heterozygous genotypes in the next

[36] Including paraffin material for the *AR* E211 and *CYP17* IVS4+58G>C and IVS6-25C>A SNPs
[37] Obtained from the Western Province Blood Transfusion service (WPBTS)
[38] Controls from the WPBTS

generation can be expressed as p2+ 2pq+ q2= 1. The observed frequencies[39] varied significantly from the expected frequencies[40] and therefore Chi-square analyses were performed, to confirm that the population was in a state of nonequilibrium.

Chi-square test

Chi-square analyses were performed for the identified SNPs within the CaP and BPH samples, in all 3 population groups. *P* values of > 0.05 were obtained for all SNPs, giving no reason to reject the null hypothesis.

Fisher exact tests

Fisher exact tests were performed for each SNP to compare genotype distribution and allele frequencies among patient[41] and control groups[42], in order to determine any association with CaP risk. This test excluded any samples from the WPBTS and therefore included only samples from males. Since males have a single copy of the *AR* gene, total number of alleles for example, would be 47 if there were 47 individuals with CaP.

[39] From allelic frequency calculations for all SNPs
[40] From Hardy-Weinberg calculations for all SNPs

[41] Tumour and blood from grade 1, 2 or 3 prostate carcinoma, either with or without metastasis and either with or without hormone resistance. For the *AR* E211 and the *CYP17* IVS+58G>C and IVS6-25C>A SNPs, paraffin-embedded cancer material was also used.
[42] Either with or without prostatitis

APPENDIX I: SOLUTIONS

Solutions used for DGGE gels

20x TAE-buffer (10 litres) :

969 g Tris-HCL

544 g NaAc

74.5 g EDTA

Adjust the pH to 8.0 with 300-350 ml NaOH and fill up to 10 litres with distilled H_2O.

9% PAA (1 litre) :

25 ml 20x TAE

225 ml 40% PAA (acryl: bisacryl = 37.5: 1)

Fill up to 1 litre with distilled H_2O.

80% UF / 9% PAA (1 litre) :

25 ml 20x TAE

225 ml 40% PAA (acryl: bisacryl = 37.5: 1)

336 g urea

320 ml formamide (deionised)

Fill up to 1 litre with distilled H_2O.

OR

For a 90% UF / 9% PAA stock solution, 378 g of urea and 360 ml of deionised formamide can be used instead.

20% ammoniumpersulphate (APS) (1 litre) :

20 g APS filled up to 100 ml with distilled H_2O. Dispense into 1 ml aliquots and store at -20^{0}C.

10x Gel loading buffer (GLB) (100 ml) :

250 mg broomphenolblue

25 g ficoll

10 mM Na_2EDTA

Fill up to 100 ml with distilled H_2O.

Deionisation of formamide :

Mix 50 ml of formamide with 5 g of mixed-bed ion-exchange resin (Bio-Rad AG 501-X8, 20-50 mesh). Stir for 1 hr at room temperature. Filter once with Whatman No. 1 filter paper and store at -20^{0}C.

Solutions used for agarose gel electrophoresis

5 X TBE stock (1L) :

54,5 g Tris

27,8 g Boric acid

2,9 g EDTA

Make up to 1litre with dH_2O.

0,5 X TBE made from 5X stock solution (1L) :

0,5/5 X 1000ml ∴ → 100ml of stock made up to 1L

2% agarose gel :

2/100 X x/150 (gel tray holds maximum volume of 150ml)

use 3g of agarose (GIBCO BRL, Ultra Pure Agarose Electrophoresis Grade) + 150ml 0,5 X TBE and heat in microwave till dissolved, cool to room temperature, pour in gel tray and place gel combs in gel to form slots. Allow about 30 minutes to set.

Electrophoresis gel loading buffer :

22,7 ml formamide

0,168 g EDTA

0,0125 g xylene cyanol

0,0125 g bromophenol blue

Make up to 25ml with dH_2O.

APPENDIX II: Chi-square and Fisher exact tests comparing the genotype distribution and allele frequencies among patient and control groups for identified *AR* and *CYP17* SNPs

TABLE 8. Allele frequencies for the AR E211 polymorphism by allele counting

Genotype	MM	MN	NN	Total
Number of individuals	53	35	123	211
Number of M alleles	70	35	0	105
Number of N alleles	0	35	157	192
Total number of alleles	70	70	157	297

MM=homozygotes; MN=heterozygotes; NN=normal

Frequency of M in population: 105/297 = 0.35 = 35%

Frequency of N in population: 192/297 = 0.647 = 65%

Total genotypes for next generation: testing for Hardy-Weinberg equilibrium

Expected frequency of type M $= p^2$

$= (0.35)^2 = 0.1225$

$= 12.25\%$

Expected frequency of type MN = 2pq

$=2(0.35)(0.65) = 0.455$

$= 45.5\%$

Expected frequency of type N $= q^2$

$=(0.65)^2 = 0.4225$

$= 42.25\%$

TABLE 9. The comparison of two variables (CaP and BPH) for the *AR* E211 SNP in all races

	CaP	**BPH**	**Total**
♦Wild-type	35	19	54
	(41%)	(28%)	(69%)
#E211	17	7	24
	(24%)	(7%)	(31%)
***Total**	52	26	78
	(65%)	(35%)	(100%)

***P* = 0.1876**

*Total =number of alleles, for example; there are 50 individuals with CaP (Refer to Table 1: Materials and Methods; Section 2.1.2 where there are 30 parafifin-embedded CaP's from Blacks, 17 blood/tissue CaP's combined from Blacks and Coloureds and 3 CaP samples from Whites), therefore 50 alleles (males have a single copy of the gene), etc.
#E211 = number of mutant alleles
♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

TABLE 10. The Chi-square analysis of two variables (CaP and BPH) for the *AR* E211 SNP

Class	**Observed (O)**	**Expected (E)**
Blacks	21	9
Caucasians	11	2
Coloureds	26	13

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

= 1.53 with (3 - 1) = 2 *df*

***P* = 0.465**

TABLE 11. Allele frequencies for the *CYP17* H46 polymorphism by allele counting

Genotype	MM	MN	NN	Total
Number of individuals	20	41	36	97
Number of M alleles	40	41	0	81
Number of N alleles	0	41	72	113
Total number of alleles	40	82	72	194

MM=homozygotes; MN=heterozygotes; NN=normal

Frequency of M in population: 81/194 = 0.42 = 42%

Frequency of N in population: 113/194 = 0.58 = 58%

Total genotypes for next generation: testing for Hardy-Weinberg equilibrium

Expected frequency of type M $= p^2$

$= (0.42)^2 = 0.176$

$= 17.6\%$

Expected frequency of type MN = 2pq

$=2(0.42)(0.58) = 0.487$

$= 48.72\%$

Expected frequency of type N $= q^2$

$=(0.58)^2 = 0.336$

$= 33.64\%$

TABLE 12. The comparison of two variables (CaP and BPH) for the *CYP17* H46 SNP in all races

	CaP	**BPH**	**Total**
♦Wild-type	19	35	54
	(18%)	(34%)	(52%)
#H46	21	21	42
	(25%)	(23%)	(48%)
***Total**	40	56	96
	(43%)	(57%)	(100%)

***P* = 0.1408**

*Total =number of alleles, for example; there are 20 individuals with CaP, therefore 20X2 alleles, etc.

#H46 = number of mutant alleles

♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

TABLE 13. The Chi-square analysis of two variables (CaP and BPH) for the*CYP17* H46 SNP

Class	**Observed (O)**	**Expected (E)**
Blacks	2	3
Caucasians	5	8
Coloureds	13	21

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

= 0.0057 with (3 - 1) = 2 *df*

***P* = 0.997**

TABLE 14. Allele frequencies for the *CYP17* IVS4+58G>C polymorphism by allele counting

Genotype	MM	MN	NN	Total
Number of individuals	0	12	144	156
Number of M alleles	0	12	0	12
Number of N alleles	0	12	288	300
Total number of alleles	0	24	288	312

MM=homozygotes; MN=heterozygotes; NN=normal

Frequency of M in population: 12/312 = 0.038 = 4%

Frequency of N in population: 300/312 = 0.96= 96%

Total genotypes for next generation: testing for Hardy-Weinberg equilibrium

Expected frequency of type M $= p^2$

$= (0.04)^2 = 0.0016$

$= 0.16\%$

Expected frequency of type MN = 2pq

$=2(0.04)(0.96) = 0.768$

$= 7.68\%$

Expected frequency of type N $= q^2$

$=(0.96)^2 = 0.9216$

$= 92.16\%$

TABLE 15. The comparison of two variables (CaP and BPH) for the *CYP17* IVS4+58G>C SNP in all races

	CaP	**BPH**	**Total**
♦Wild-type	95	49	144
	(69%)	(26%)	(95%)
#IVS4+58G>C	3	3	6
	(3%)	(3%)	(5%)
***Total**	98	52	150
	(71%)	(29%)	(100%)

***P* = 0.3530**

*Total =number of alleles, for example; there are 42 individuals with CaP, therefore 42X2 alleles, etc.

#IVS4+58G>C = number of mutant alleles

♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

TABLE 16. The Chi-square analysis of two variables (CaP and BPH) in *CYP17* for the IVS4+58G>C SNP

Class	**Observed (O)**	**Expected (E)**
Blacks	32	2
Caucasians	11	0
Coloureds	25	4

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

= 2.45 with (3 - 1) = 2 *df*

***P* = 0.293**

TABLE 17. Allele frequencies for the *CYP17* IVS6-25C>A polymorphism by allele counting

Genotype	MM	MN	NN	Total
Number of individuals	0	17	105	122
Number of M alleles	0	17	0	17
Number of N alleles	0	17	210	227
Total number of alleles	0	34	210	244

MM=homozygotes; MN=heterozygotes; NN=normal

Frequency of M in population: 17/244 = 0.067 = 7%

Frequency of N in population: 227/244 = 0.937= 93%

Total genotypes for next generation: testing for Hardy-Weinberg equilibrium

Expected frequency of type M $= p^2$

$= (0.067)^2 = 0.004489$

$= 0.449\%$

Expected frequency of type MN = 2pq

$=2(0.067)(0.937) = 0.125558$

$= 12.556\%$

Expected frequency of type N $= q^2$

$=(0.937)^2 = 0.878$

$= 87.8\%$

TABLE 18. The comparison of two variables (CaP and BPH) for the *CYP17* IVS6-25C>A SNP in all races

	CaP	**BPH**	**Total**
♦Wild-type	93	48	141
	(62%)	(33%)	(95%)
#IVS6-25c/a	5	4	9
	(2%)	(3%)	(5%)
***Total**	98	52	150
	(64%)	(36%)	(100%)

***P* = 0.2545**

*Total =number of alleles, for example; there are 44 individuals with CaP, therefore 44X2 alleles, etc.

#IVS6-25c/a = number of mutant alleles

♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

TABLE 19. The Chi-square analysis of two variables (CaP and BPH) in *CYP17* for the IVS6-25C>A SNP

Class	**Observed (O)**	**Expected (E)**
Blacks	30	2
Caucasians	9	2
Coloureds	29	7

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

= 2.65 with (3 - 1) = 2 *df*

***P* = 0.265**

TABLE 20. Allele frequencies for the *CYP17* S65 polymorphism by allele counting

Genotype	MM	MN	NN	Total
Number of individuals	5	6	88	99
Number of M alleles	10	6	0	16
Number of N alleles	0	6	176	182
Total number of alleles	10	12	176	198

MM=homozygotes; MN=heterozygotes; NN=normal

Frequency of M in population: 16/198 = 0.071 = 8%

Frequency of N in population: 182/198 = 0.92= 92%

Total genotypes for next generation: testing for Hardy-Weinberg equilibrium

Expected frequency of type M $= p^2$

$= (0.071)^2 = 0.05$

$= 0.50\%$

Expected frequency of type MN = 2pq

$=2(0.071)(0.92) = 0.13064$

$= 13.064\%$

Expected frequency of type N $= q^2$

$=(0.92)^2 = 0.8464$

$= 84.64\%$

TABLE 21. The comparison of two variables (CaP and BPH) for the *CYP17* S65 SNP in all races

	CaP	BPH	Total
♦**Wild-type**	40	49	89
	(43%)	(49%)	(92%)
#S65	0	7	7
	(0%)	(8%)	(8%)
***Total**	40	56	96
	(43%)	(57%)	(100%)

***P* = 0.0174**

*Total =number of alleles, for example; there are 20 individuals with CaP, therefore 20X2 alleles, etc.

#S65 = number of mutant alleles

♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

TABLE 22. The Chi-square analysis of two variables (CaP and BPH) in *CYP17* for the S65 SNP

Class	Observed (O)	Expected (E)
Blacks	5	0
Caucasians	11	0
Coloureds	30	5

$$\chi^2 = \Sigma \frac{(O - E)^2}{E}$$

= 2.53 with (3 - 1) = 2 *df*

***P* = 0.282**

TABLE 23. The comparison of two variables (CaP and BPH) for the *CYP17* S65 SNP in Coloured patients

	CaP	**BPH**	**Total**
♦Wild-type	28	24	52
	(47%)	(40%)	(87%)
#S65	0	8	8
	(0%)	(13%)	(13%)
***Total**	28	32	60
	(47%)	(53%)	(100%)
		***P* = 0.0053**	

*Total =number of alleles, for example; there are 14 individuals with CaP, therefore 14X2 alleles, etc.

#S65 = number of mutant alleles

♦Wild-type =normal number of alleles (Total alleles - the mutant alleles)

ACKNOWLEDGEMENTS

Dr. VM Hayes (Dept of Urology), my promoter, for your guidance, good advice, encouragement and support throughout this study. My sincere gratitude for all your faith in me that enabled me to grow as a researcher.

Prof. CF Heyns (Head of Dept of Urology), my co-promoter, for allowing me to perform experimentation in your Department for the work presented in this dissertation. Thank you for your helpful manner in assisting with clinical information.

Dr. R Hillermann (Dept of Genetics), my co-promoter, thanks for your advice and interest in my work.

All the clinicians from the **Dept of Urology**, many thanks for providing tissue and blood samples for this study.

Mr M Petersen (Dept of Urology), for collection of clinical material.

Dr. W Bates (Dept of Anatomical Pathology), for the paraffin-embedded samples that were provided.

Dr AJ Visser (Dept of Urology), thank you for the valuable clinical information and the patient database.

Mr PM De Beer (Dept of Urology), for all your help and technical assistance.

The personnel of the Dept of Urology, for creating such a pleasant working environment and the interest that you showed in my project.

Mrs A Laten (Dept of Medical Virology), thank you so much for all the sequencing performed, sometimes at such short notice.

Prof. E Janse van Rensburg (Dept of Medical Virology), thank you for allowing me to take part in Research meetings and for the use of your facilities. Your sincere interest in my studies is appreciated.

The Hayes Foundation, for financial support.

The South African Medical Research Council and **The University of Stellenbosch** for the bursaries awarded for my MSc studies.

Fareed and Waagheda Hendricks, my wonderful parents, for their unconditional love, guidance, encouragement and inspiration that have helped shape my life. I am forever in your gratitude.

My brother, **Gibbaan Hendricks**, and to the rest of my family, although I cannot mention you all by name, my heartfelt gratitude for your interest in my studies and your amazing support.