Validation of susceptibility locus 2q37.3 mutations in prostate cancer

Master's thesis
Elisa Pellinen
University of Tampere
Institute of Biomedical Technology (IBT)
April 2012

ACKNOWLEDGEMENTS

This thesis was carried out in the Genetic Predisposition to Cancer research group of Professor Johanna Schleutker in the Institute of Biosciences and Medical Technology, University of Tampere. I would like to thank her for the opportunity to carry out this work in her group. I would also like to thank PhD. Tiina Wahlfors and MSc. Virpi Laitinen for guidance and advice in all aspects of this thesis. For Riina Liikala, thanks for all the practical advice in laboratory work. A special thanks to the following present members of the Genetic Predisposition to Cancer research group: Tiina Wahlfors, Virpi Laitinen, Riina Liikala, Riikka Nurminen, Sanna Siltanen and Kirsi Kuusisto, for welcoming me to the group and being great friends during the time I spent working on my thesis.

I would also like to thank my family and friends for support in my study years and my fellow students for all the great events and making my time in Tampere University Biotechnology program memorable. The greatest thanks go to my soon-to-be husband Marko Vuorinen for putting up with all my stress-induced tantrums and other side-effects of being a university student for all these years. ☺

Tampere, March 2012

Elisa Pellinen

PRO GRADU -TUTKIELMA

Paikka:	TAMPEREEN YLIOPISTO Biolääketieteellisen teknologian yksikkö (IBT)
Tekijä:	PELLINEN, ELISA MIRJAMI
Otsikko:	Eturauhassyövälle altistavan lokuksen 2q37.3 mutaatioiden validointi
Sivumäärä:	68
Ohjaajat:	FT Tiina Wahlfors, FM Virpi Laitinen, Professori Johanna Schleutker
Tarkastajat:	FT Tiina Wahlfors, Professori Markku Kulomaa
Päiväys:	Huhtikuu 2012

Tiivistelmä

Tutkimuksen tausta ja tavoitteet: Eturauhassyöpä on miesten yleisin syöpä Suomessa ja muissa teollistuneissa maissa. Taudin syntyyn johtavista tekijöistä tiedetään edelleen melko vähän. Perimä on kuitenkin yksi tunnetuimmista eturauhassyövälle altistavista riskitekijöistä. Eturauhassyövän taudinkulku on vaikeasti ennalta määritettävissä: osa tapauksista pysyy oireettomina kymmeniä vuosia, kun taas osassa tapauksia tauti etenee nopeasti etäpesäkkeitä muodostavaksi. Jotta diagnosointia ja taudinkulun ennustettavuutta voitaisiin parantaa, tarvitaan parempia ennustustyökaluja. Geneettiset markkerit, kuten yhden emäksen polymorfiat, ovat tällä hetkellä lupaavimpia kandidaatteja. Tämän tutkimuksen tavoitteena oli tutkia, ovatko tietyt kromosomista 2q37.3 löydetyt mutaatiot assosioituneet eturauhassyöpään joko yleisellä tasolla tai voidaanko ne liittää tietynlaiseen seerumin PSA-pitoisuuteen. Tutkimuksessa arvioitiin myös uutta teknologiaa hyödyntävää KASPar genotyyppausteknologiaa mahdollisena kustannustehokkaana korvaajana laajasti käytetylle TaqMan -genotyyppausmenetelmälle.
Tutkimusmenetelmät: Tutkimuksessa käytettiin perhetaustaisilta ja satunnaisilta eturauhassyöpäpotilailta Pirkanmaan alueelta kerättyjä DNA-näytteitä. Kontrolleina oli Suomen punaisen ristin terveiltä verenluovuttajilta kerättyjä DNA-näytteitä. Näytteet genotyypattiin kaikkien valittujen mutaatioiden suhteen yhdellä neljästä genotyyppausmenetelmästä: TaqMan-genotyyppaus, KASPar-genotyyppays, high-resolution melt -analyysi (HRM) tai suora sekvensointi Sangerin menetelmällä. Osa mutaatioista genotyypattiin sekä TaqMan- että KASPar-menetelmällä, jotta menetelmiä voitiin vertailla. Genotyyppaustulosten perusteella laskettiin kunkin mutaation yhteys eturauhassyöpäriskiin.
Tutkimustulokset: Yksi kiinnostuksen kohteena olleista mutaatioista, *IQCA1* 237247036insA, assosioituu sekä perinnöllisen että satunnaisen eturauhassyövän kanssa. Muutoksen havaittiin olevan yleisempi syöpätapauksissa kuin kontrolleissa, joten kyseisellä mutaatiolla on mahdollisesti suojaava vaikutus eturauhassyövältä. Toinen mutaatio, *SCLY-UBE2F* 239002480delTTG, assosioitui perinnöllisen eturauhassyövän kanssa. Kyseisen mutaation kohdalla genotyyppauksessa havaittiin kuitenkin ongelmia, joten tämän muutoksen kohdalla tulos on vielä varmistettava. KASPar-genotyyppausteknologia havaittiin yhtä vaivattomaksi ja tehokkaaksi kuin TaqMan-teknologia.
Johtopäätökset: Tässä työssä saadut tulokset osoittavat, että *IQCA1* 237247036insA assosioituu perinnöllisen ja satunnaisesti syntyneen eturauhassyövän kanssa, ja on mahdollisesti taudilta suojaava tekijä. *SCLY-UBE2F* 239002480delTTG havaittiin assosioituvan perinnöllisen eturauhassyövän kanssa, tosin kyseinen tulos vaatii vielä varmistelua. Altistavien mutaatioiden löytyminen kromosomaalisesta lokuksesta 2q37 tukee tämän alueen kiinnostavuutta eturauhassyövälle altistavana. Tämän tutkimuksen seurauksena Johanna Schleutkerin johtama perinnöllisen syöpäalttiuden tutkimusryhmä omaksui KASPar-genotyyppausmenetelmän yhdeksi vaintoehdoksi tutkimuksessaan.

MASTER'S THESIS

Place:	UNIVERSITY OF TAMPERE Institute of Biomedical Technology (IBT)
Author:	PELLINEN, ELISA MIRJAMI
Title:	Validation of susceptibility locus 2q37.3 mutations in prostate cancer
Pages:	68
Supervisors:	Tiina Wahlfors, PhD; Virpi Laitinen, MSc; Professor Johanna Schleutker
Reviewers:	Tiina Wahlfors, PhD; Professor Markku Kulomaa
Date:	April 2012

Abstract

Background and aims: Prostate cancer is the most commonly diagnosed cancer in men in Finland and other industrialized countries. The causes of prostate cancer are still relatively poorly known, but family history is one of the most established risk factors. The progression and clinical picture of prostate cancer is hard to predict as some prostate cancer cases remain indolent for decades while others progress and metastasize rapidly. To improve the diagnostics and allow more accurate prognosis, better predictive tools are needed to recognize the rapidly progressing cases from the indolent ones. Genetic markers such as single nucleotide polymorphisms are the most promising candidates. The aim of this study was to see if certain chromosome 2q37.3 mutations are associated with prostate cancer in general or with a certain clinical characteristic of the disease, in this case serum PSA value. Secondary aim of this study was to evaluate the novel KASPar genotyping technology as a possible replacement for TaqMan genotyping.
Methods: Analyzed samples were DNA samples from both familial and sporadic prostate cancer cases. As controls, samples from healthy blood donors from the Finnish Red Cross were used. Each sample was genotyped for each mutation using one of four genotyping methods: TaqMan genotyping, KASPar genotyping, high-resolution melt analysis (HRM) or direct sequencing by Sanger method. Some of the mutations were genotyped with both TaqMan and KASPar assays to compare the two technologies. Based on the genotyping data, association of each mutation with prostate cancer was calculated.
Results: One of the studied mutations, *IQCA1* 237247036insA, was found to be associated with sporadic and familial prostate cancer. The mutation was more common in controls than in cases, which suggests it has some kind of protective effect from prostate cancer. Another mutation of interest, *SCLY-UBE2F* 239002480delTTG, was associated with familial prostate cancer. A problem in genotyping was observed, however, so the association of this mutation with prostate cancer will be further confirmed. KASPar genotyping technology was found to be as effective and reliable as TaqMan system.
Conclusions: This study suggests that *IQCA1* 237247036insA is associated with familial and sporadic prostate cancer and is possibly a protective factor. *SCLY-UBE2F* 239002480delTTG was found to be associated with familial prostate cancer, although a problem in genotyping was observed and further confirmation of the result is in order. The existence of predisposing variants in chromosomal locus 2q37.3 confirms that this region is interesting in regard to prostate cancer susceptibility. As a result of this study, the Genetic Predisposition to Cancer research group of Johanna Schleutker adopted the novel KASPar genotyping technology as one accepted tool for genotyping.

Contents

Abbreviations

AAA+	ATPases associated with a wide variety of cellular Activities
ADT	Androgen deprivation therapy
BPH	Benign prostatic hyperplasia
CGRP	Calcitonin-gene-related peptide
CRL	Cullin-RING ligase
CRLR	Calcitonin-receptor-like receptor
CRPC	Castration-resistant prostate cancer
CSN	COP9 signalosome
CZ	Central zone of prostate
EBRT	External beam radiotherapy
GWAS	Genome-wide association study
PIN	Prostatic intraepithelial neoplasia
PrCa	Prostate cancer
TZ	Transition zone of prostate
HLOD	Heterogeneity logarithm of odds
HPC	Hereditary prostate cancer
HRM	High-resolution melting
KASP	Kbioscience's Allele Specific PCR
PSA	Prostate specific antigen
PZ	Peripheral zone of prostate
RP	Radical prostatectomy
TNF	Tumor necrosis factor
TNFR	Tumor necrosis factor receptor
TRAF	Tumor necrosis factor receptor associated factor

1 Introduction

Prostate cancer is the most frequently diagnosed cancer among men in Western countries and the second leading cause of cancer-related deaths (Finnish Cancer Registry, 2012). The incidence of prostate cancer has increased rapidly through the 1990s, partly due to the widespread use of serum prostate-specific antigen (PSA) as tumor marker in the screening for prostate cancer. Prostate cancer is a disease with many different clinical courses: some cases remain dormant for decades without any effect on the life quality of the patient. However, a significant subset of prostate cancer cases progress rapidly to a metastasizing form, leading to death in a few years.

The precise etiology of prostate cancer is still quite poorly understood. The most established risk factor is age: the incidence of prostate cancer rises exponentially with age (Hsing and Chokkalingam, 2006) and up to 85 % of diagnosed cases are diagnosed in people over the age of 65 (Grönberg, 2003). Ethnicity is another factor that clearly increases risk of developing prostate cancer: the highest incidence of PrCa is found among African-American people living in the USA (Hsing *et al.*, 2000) and the lowest incidence is in Asia (Grönberg, 2003). The strongest factor after age and ethnicity is positive family history. A man with an affected first-degree relative (a father of a brother) has 2- to 3-fold risk of PrCa (Stanford and Ostrander, 2001). Twin studies have suggested a strong heritable component for prostate cancer (Ahlbom *et al.*, 1997; Lichtenstein *et al.*, 2000).

After the establishment of family history as a significant risk factor, a series of linkage studies was performed in an attempt to reveal rare, highly penetrant, dominant alleles that had previously been recognized in other cancer types such as cancers of the breast and colon. Several loci, including 1p36 (*CAPB*), 1q23–25 (*HPC1*) and 8p22–23 showed linkage to prostate cancer (Schaid, 2004). However, the results have been somewhat irreproducible. Some promising candidate genes were analyzed in the identified loci; however, no clear definition of the involvement of mutations in these genes in hereditary prostate cancer has emerged.

When linkage studies did not yield promising susceptibility genes, another view was taken: perhaps genetic susceptibility to prostate cancer is conferred by multiple recessive, interacting loci. Genome-wide association analyses were performed to identify these. GWASs were successful in recognizing multiple SNPs and loci that are associated with PrCa (Easton and Eeles, 2008). These studies implicate that a large number of common gene variants is involved in conferring prostate cancer risk (Eeles *et al.*, 2008) and the effect of multiple mutations in one person must be taken into account.

Prostate cancer represents a substantial burden to the health care system, especially as the number of elderly people is expected to increase. As the clinical outcome of a newly diagnosed prostate cancer cannot be precisely predicted and an accurate prognosis is hard to give, better prediction tools for the diagnosis and progress of the disease should be found to reduce overdiagnosis and overtreatment. The aim of recent research has been to identify some of the common gene variants that could be used to predict either a person's risk of developing prostate cancer or the clinical course of the disease. The ultimate goal is to develop a gene test panel for prostate cancer susceptibility.

The purpose of this study was to evaluate if any of the studied mutations found in the deep sequencing of chromosome 2q37.3 linkage region outcome found by Cropp *et al.* (2011) is associated with prostate cancer. Also, the association with a certain clinical picture of the disease, in this case serum PSA value was studied.

2 Review of the literature

2.1. The prostate

The prostate is the largest gland in the male reproductive system. Its purpose is to produce and store seminal fluid that provides nutrients for spermatozoa. Weighing around 20 grams, the gland is about the size of a walnut. The prostate is located just below the bladder, surrounding the urethra as it exits the bladder (Figure 1). The human prostate has two kinds of structures: fibromuscular and glandular (Timms, 2008). The anterior and ventral parts are fibromuscular and rich in collagen while the posterior part is mostly glandular and surrounds the ejaculatory ducts as they enter the urethra (Timms, 2008). The normal human prostate has a high degree of cellular organization and is composed of three distinct cell populations: basal layer populated by prostate epithelial stem cells and transiently proliferating/amplifying cells, secretory luminal layer composed of differentiated prostatic epithelial cells, and neuroendocrine cells that are scattered throughout the prostatic gland (Schalken, 2003). The prostate can be divided into three zones: the central zone (CZ), transition zone (TZ), and peripheral zone (PZ) (Lee *et al.*, 2011). These zones differ in their susceptibility to disorders: TZ is the predominant initiation area of benign prostatic hyperplasia (BPH) while most prostate cancer cases arise in the PZ (Lee *et al.*, 2011).

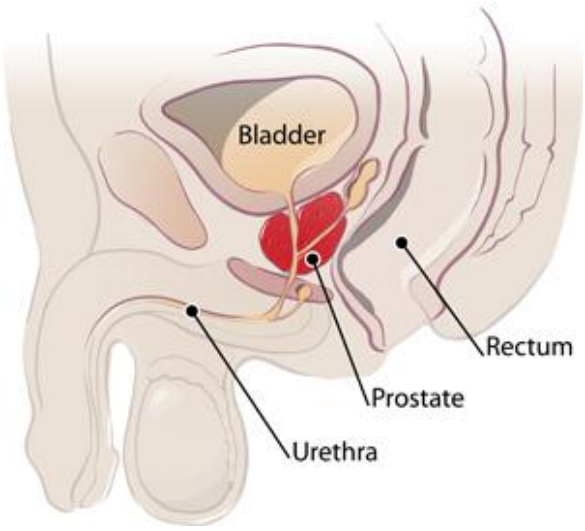

Figure 1. The prostate (http://www.miamiurologyconsultants.com/services-procedures/robotic-prostatectomy-surgery-prostate-cancer-miami-aventura-fl.php 5.1.2011)

2.1.1. Disorders of prostate

Prostatitis. Prostatitis means the inflammation of the prostate. Prostatitis is very common: according to studies up to 16 % of American men suffer symptoms of prostatitis at some point of their life (Collins *et al.*, 2002). There are four distinct classifications of the disease: acute bacterial, chronic bacterial, chronic prostatitis/chronic pelvic pain syndrome, and asymptomatic (Sharp *et al.*, 2010). Acute bacterial prostatitis and chronic bacterial prostatitis are caused by a bacterium, most commonly *E.coli*, but other Gram-negative bacteria as well as enterococcus have been documented as infectious agents (Krieger, 1999). Acute bacterial prostatitis is presented as acute symptoms of urinary tract infection while chronic bacterial prostatitis causes recurring urinary tract infections that alternate with asymptomatic periods (Krieger, 1999). Chronic prostatitis/chronic pelvic pain syndrome describes conditions where urological pain complaints are a primary component but where certain exclusion criteria (e.g. presence of active urethritis, urogenital cancer or urinary tract disease) have been applied (Krieger, 1999). Chronic prostatitis/chronic pelvic pain syndrome can be subdivided to inflammatory and non-inflammatory forms which are distinguished from each other by the presence or absence of leukocytes is secretions, respectively (Sharp *et al.*, 2010). The etiology of this syndrome is largely unknown (Nickel, 2011). Asymptomatic prostatitis is characterized by prostate infection in absence of genitourinary tract symptoms (Krieger, 1999). Asymptomatic prostatitis is usually diagnosed during urologic evaluation for other reasons (Sharp *et al.*, 2010). The infectious forms of the disease can be treated with antimicrobial agents such as antibiotics, but the treatment of chronic pelvic pain syndrome is not as straightforward as inflammation is not always present (Nickel, 2011).

Benign prostatic hyperplasia. Benign prostatic hyperplasia (BPH) is the most common disorder of the prostate. It is initiated around the age of 30 and progressively increases with age, affecting over 90 % of men in their 80s (Paolone, 2010). BPH is a proliferative abnormality that is initiated in the TZ that normally makes up about 5 % of the glandular prostate (Lee and Peehl, 2004). During BPH, the volume of the prostate expands via budding and branching of epithelial glandular tissue and to a lesser degree by glandular unit enlargement or proliferation of prostatic stromal elements (i.e. smooth muscle and fibroblasts) (Lee and Peehl, 2004). The enlargement of the prostate can cause partial obstruction of the urethra (Nurmi, 2011), which leads to urinary symptoms such as the need to urinate often and nocturia (Tammela, 2006). At worst, BPH can lead to total obstruction of the urethra and inhibit urination altogether (Tammela, 2006). BPH can usually be treated with medication by e.g. α1-adrenergic receptor-blocking agents, but in

some cases more invasive solutions such as transurethral microwave therapy or transurethral resection of the prostate are needed to alleviate the symptoms and improve the patient's quality of life (Paolone, 2010).

Prostate cancer. Prostate cancer is the direst of prostate disorders. It is a malignancy that arises from the glandular tissue of peripheral zone of the prostate (Schulz *et al.*, 2003) and is thus defined as adenocarcinoma (epithelial cancer that originates in glandular tissue). The earliest precursors of prostate cancer are prostatic intraepithelial neoplasia (PIN) that are characterized by thickening of the epithelial layer and loss of distinction between basal and secretory luminal layers (Schulz *et al.*, 2003). To transform into PrCa, a PIN must lose its basal layer (Lee *et al.*, 2011).

As a disease, PrCa is a heterogeneous one with many clinical courses. In most cases, PrCa is present as an asymptomatic, latent entity that does not result in clinical symptoms during the lifetime of a patient (Schulz *et al.*, 2003). However, approximately 30 % of PrCa cases become invasive, spreading beyond the prostatic tissue capsule and metastasize, usually to local lymph nodes and distal organs, typically lung, liver and bone (Schulz *et al.*, 2003), leading to death within years of the diagnosis. The clinical course of the disease cannot be predicted at the time of diagnosis by current methods.

2.2. Prostate cancer

2.2.1. Biomarkers for prostate cancer

Prostate-specific antigen (PSA) is a serine protease that is regulated by androgens (Balk *et al.*, 2003). It is secreted by the epithelial cells of the PZ and its major function is to liquefy semen in the seminal coagulum, to allow the free motility of sperm in the ejaculate (Tosoian and Loeb, 2010). Prostate cells secrete PSA in all conditions, whether they are normal, benign or malignant (Duffy, 2011). Serum level of PSA is the most widely used tumor marker worldwide in the detection of prostate cancer, as the increase in serum PSA correlates with the risk of PrCa, risk of high-grade disease as well as tumor stage (Parekh *et al.*, 2007), especially with higher PSA concentrations. The advantage of PSA testing is that it is more efficient in detecting PrCa in its early stage, as compared to the former screening method, digital rectal exam (Hoffman, 2011).

Questions about the validity of PSA as PrCa biomarker have risen lately. PSA might not be an ideal marker for prostate cancer screening because it is *prostate specific*, not *prostate cancer specific*. Many other factors, including other disorders of the prostate such as BPH and

inflammation (prostatitis), or even physical exercise such as bicycle riding, might lead to temporarily elevated PSA values (Barry, 2001; Duffy, 2011; Oesterling, 1991). Second, a substantial overlap exists between the serum levels of malignancy and non-malignancy. A cut-off level of 4 µg/l has traditionally been used, but there are healthy individuals with higher PSA level and high-grade PrCa is not uncommon in men with PSA value under 4 µg/l (Duffy, 2011) Screening for PSA might lead to overdiagnosis and overtreatment of indolent tumors that possibly would not cause any harm to these men in their lifetime (Duffy, 2011). Overdiagnosis and overtreatment can in turn lead to severe consequences varying from wide economical questions to unnecessary side-effects to the patient, such as impotence and incontinence (Duffy, 2011).

Due to the limitations of PSA in PrCa detection, novel and more specific biomarkers are under intensive investigation. Some are direct derivatives of PSA testing, such as the percentage of free serum PSA versus protein-bound form, PSA kinetics or PSA density (the ratio of PSA concentration to prostate volume) (Tosoian and Loeb, 2010). Other biomarkers that have been evaluated in PrCa detection are proteins that are overexpressed during the disease, such as alpha-methylacyl coenzyme A racemase (AMACR) (Tosoian and Loeb, 2010; Parekh *et al*., 2007). Non-coding transcripts that are overexpressed in PrCa, like prostate-cancer antigen 3 (PCA3), have also shown promise (Tosoian and Loeb, 2010; Parekh *et al*., 2007).

The interest in genome-wide association studies in PrCa has brought forward perhaps the most promising biomarkers: single nucleotide polymorphisms (SNPs). Many SNPs that show association with PrCa susceptibility have been identified (Tosoian and Loeb, 2010), and the human genome probably harbors many more. With the emergence of PrCa predisposing SNPs, the possibility that multiple SNPs might interact and together affect the risk of developing the disease must be taken into account. SNPs as PrCa biomarkers broaden the horizon of PrCa detection as they bring forward the possibility of genetic testing. The future perspectives iclude a test panel that could be used to identify PrCa susceptibility SNPs in a person's genome and perhaps to predict the clinical course of the disease.

2.2.2. Incidence and mortality

Prostate cancer is the most common cancer in the USA, with an estimated 241 740 new cases in 2012 that account for 28 % of all the new cancer cases diagnosed in men (American Cancer Society, 2012). In Finland in 2010, 4 712 new cases were diagnosed, accounting for 31 % of all

male cancers (Finnish Cancer Registry, 2012). The incidence of PrCa in Finland has been increasing annually since the 1960s, with a very rapid increase in the 1990s (Finnish Cancer Registry, 2012) (Figure 2). This radical increase in incidence is probably due to better health care and the establishment of frequent PSA testing (Kvale *et al.*, 2007) and means that PrCa cases are diagnosed more efficiently and asymptomatic cases that would not be diagnosed without PSA testing can be detected, rather than that the amount of PrCa cases has actually exploded. However, all of the change cannot be explained by PSA testing, so some unknown factors exist.

Prostate cancer is the second leading cause of cancer-related deaths in men in the USA and Finland. An estimated 28 170 prostate cancer deaths will occur in the USA in 2012 (American Cancer Society, 2012). In Finland, 847 deaths occurred due to PrCa in 2010 (Finnish Cancer Registry, 2012). Even though the incidence has increased substantially, the mortality of PrCa in Finland has remained quite steady, with a small incline detectable after 2002 (Figure 2). The slight decline in mortality can be explained by PSA testing as a larger proportion of cancers are detected at an earlier phase and thus can be treated, resulting in lower mortality rates relative to incidence. However, the incline in mortality is not quite as steep as one might have hoped after the introduction of PSA testing, and raises questions about the necessity of testing asymptomatic men.

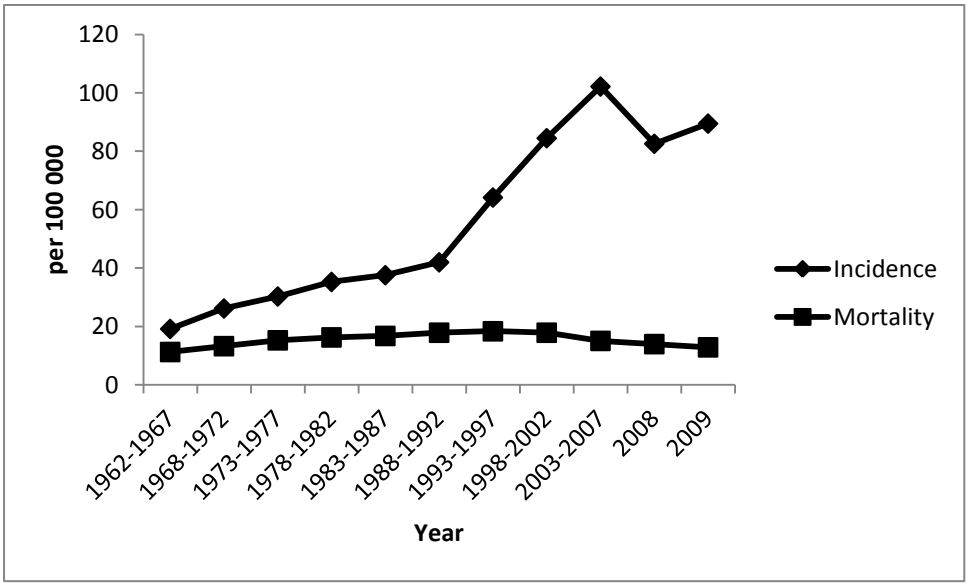

Figure 2. Incidence and mortality of prostate cancer in Finland 1962–2009. Data modified from Finnish Cancer Registry (2012).

2.2.3. Risk factors

Age and ethnicity. The risk of developing PrCa increases with age, and age is the most established risk factor of PrCa. Very few people are diagnosed with prostate cancer under the age of 50 (Hsing and Chokkalingam, 2006) and up to 85 % of diagnosed cases are in people over the age of 65 (Grönberg, 2003). The incidence of prostate cancer rises exponentially with age (Hsing and Chokkalingam, 2006) and autopsy studies suggest that most men over 85 have histological evidence of PrCa (Sakr *et al.*, 1993). Indeed, it is a common belief that every man would eventually develop prostate cancer given the chance to live long enough - most of them die of other causes, without ever having any symptoms.

Ethnic origin seems to be another risk factor, as prostate cancer incidence varies among countries and between ethnic groups. The highest incidence of PrCa is among African-American people living in the USA (Hsing *et al.*, 2000) and the lowest incidence can be found in Asia (Grönberg, 2003). Some of the differences in PrCa incidence between different groups of people may be accounted for by bias resulting from differences in PrCa screening, access to health care and cancer registration between countries. However, studies have shown that PrCa is very common in Africa (Grönberg, 2003), so it seems that African people are in a higher risk whether they are living in Africa or in America. Also, migration studies have shown that when Japanese people migrate from low-risk Japan to high-risk USA, the incidence of prostate cancer in these people increases, but the increased rate of incidence is still lower than in white people or African-Americans (Grönberg, 2003). These factors seem to point to real differences in PrCa susceptibility in different races and not just differences in other predisposing factors such as environment. However, the incidence of PrCa in low-risk countries, such as those in Asia, has been rising (Hsing *et al.*, 2000). Furthermore, the highest incidences seem to concentrate in countries with a western lifestyle. This might indicate that western culture is one predisposing factor, and the westernization of Asian countries is one cause of increase in incidence in previously low-risk countries (Hsing *et al.*, 2000). Interestingly, the amount of latent prostate cancer cases is quite the same in low-risk and high-risk populations (Sakr *et al.*, 1993), and the difference in incidence ratios is due to the fact that latent PrCa cases become clinically manifest more frequently in high-risk countries, supporting the view that environmental factors have something to do with PrCa incidence.

Family history. Family history is one of the most well-documented risk factors of prostate cancer. Most PrCa cases are sporadic, but there is an important subset of about 10 % of patients who have a positive family history of PrCa (Whittemore *et al.*, 1995). A male having a

first-degree relative (a father or a brother) with prostate cancer history has 2- to 3-fold risk of PrCa (Stanford and Ostrander, 2001), and an affected brother confers a larger risk than an affected father (Grönberg, 2003). Furthermore, the risk of developing PrCa increases as the number of affected individuals in the family increases (Grönberg, 2003), and risk is also increased if the age of diagnosis in a relative is lower (early-onset PrCa) (Stanford and Ostrander, 2001). A study of Nordic twins by Liechtenstein *et al*. (2000) showed that even 42 % of PrCa risk can be explained by heritable factors. This is the highest proportion documented for a common malignancy (Mattila, 2009). Other twin studies have also indicated that prostate cancer is more concordant in monozygotic twins than in dizygotic twins (Ahlbom *et al.*, 1997), supporting the view that prostate cancer is indeed to some extent a heritable disease. Of course, families often share environmental and or dietary risk factors which could explain some of the clustering of PrCa cases; however, these factors are not enough to explain all of the cases.

Other. Many studies have taken effort to reveal other factors contributing to prostate cancer risk. These factors include hormones, diet and nutrition, inflammation, environmental factors and viruses.

Hormones, especially androgens, are important factors in the development, maintenance and function of prostatic tissue (Hsing and Chokkalingam, 2006) . These hormones mediate their effect by binding to nuclear receptors (androgen receptors) that in turn bind to DNA sequences called hormone response elements. Nuclear receptors repress or stimulate transcription by contacting the basal transcription machinery or via recruitment of corepressors and coactivators. Prostate cancer is absent in castrated men (Hsing and Chokkalingam, 2006), and androgen deprivation therapy by surgical or medical castration is a well-known treatment of advanced prostate cancer (Tammela, 2004). These facts seem to hint that androgens have some role in the epidemiology of PrCa. However, the results from studies evaluating the effect of androgens on PrCa have been inconclusive (Eaton *et al.*, 1999; Hsing, 2001). The role of estrogen in PrCa development has been proposed and especially estrogen to androgen ratio has been hypothesized to contribute to PrCa risk, as the incidence of PrCa increases exponentially in older men in whom the ratio increases due to a decline in testicular function and increase in aromatization of adrenal androgens (Griffiths, 2000). The clinical evidence of estrogen/androgen ratio has also remained inconclusive (Eaton *et al.*, 1999).

A variety of dietary factors has been suggested in the development of PrCa. Especially Western lifestyle and the high intake of dietary fats have been linked to higher PrCa risk (Hsing and Chokkalingam, 2006). High intake of dietary calcium has been suggested as risk factor (Grönberg, 2003), whereas high consumption of lycopene in tomatoes has been indicated as a

possible protective agent (Ilic, 2011). Lycopene is an antioxidant present in many red fruits and vegetables and is perhaps the most effective quencher of reactive oxygen species (Levy *et al.*, 1995). Other dietary factors that have been studied include selenium, zinc and vitamin E, of which selenium and vitamin E seem to have protective effects whereas excess of zinc might have positive association (Grönberg, 2003; Hsing and Chokkalingam, 2006). The effects of single dietary factors remain somewhat controversial, but it has been proposed that a healthy lifestyle and diet drawn up according to the proposed guidelines could decrease the incidence of PrCa (Divisi *et al.*, 2006).

Chronic prostatitis as a potential PrCa risk factor has recently been given a lot of attention. The association between chronic inflammations and cancer has been long established (Coussens and Werb, 2002) and some estimations declare that 20 % of all human cancers could result from inflammations (De Marzo *et al.*, 2007). Chronic inflammation may lead to tumorigenesis via DNA damage by radical oxygen and nitrogen species, reduced DNA repair and enhanced cell proliferation and angiogenesis (Coussens and Werb, 2002). Evidence from epidemiological, histopathological and molecular pathological studies seems to confirm that chronic inflammation might also be an important PrCa risk factor (De Marzo *et al.*, 2007).

2.2.4. Treatment strategies

Prostate cancer is a heterogeneous disease with many clinical manifestations and the treatment strategies for PrCa should be chosen according to the cancer characteristics. Localized PrCa that is still inside the prostatic capsule can be treated with different methods than cancer that has spread outside the capsule and possibly metastasized. Also, a difference should be made between low- and high-risk localized PrCa.

Active surveillance. Due to widespread PSA screening, the incidence of prostate cancer has increased. PSA testing has also led to the detection of early-stage, dormant tumors that would not become clinically significant and thus would not be detected in the person's lifetime. As a result, there is a substantial population of men that would ultimately die of other causes without ever knowing they had PrCa, and can thus be defined as overdiagnosed (Cooperberg *et al.*, 2011). The overdiagnosis rate associated with PrCa screening has been estimated to be 23–42 % in the USA (Draisma *et al.*, 2009). In contemporary medicine, a diagnosis tends to lead to treatment; as a result, overdiagnosis leads to overtreatment. This is also the case with PrCa: most

men are treated aggressively even though they would be good candidates for active surveillance (Zerbib *et al.*, 2008).

Active surveillance as a treatment strategy takes advantage of the usually slow progression of low-risk prostate tumors. Low-risk PrCa is defined as having low volume, stage and grade, and the lead time from diagnosis to disease progression is usually very long for these tumors (Cooperberg *et al.,* 2011). Men under active surveillance are observed carefully with serial PSA assessments, repeated biopsies and other tests that are intended to detect any early signs of disease progression (Cooperberg *et al.*, 2011). If the disease shows signs of progression, aggressive treatment can be initiated. Active surveillance minimizes the overtreatment of indolent tumors (Zerbib *et al.,* 2008) and avoids the often radical side-effects of aggressive therapy. Even if aggressive treatment is required at a later stage, the patient has been provided with more healthy years but still can be treated within the window of opportunity to cure (Cooperberg *et al.,* 2011). However, active surveillance does have its own shortcomings. Identification of risk progression in men undergoing active surveillance can be challenging (Cooperberg *et al.*, 2011), frequent testing and biopsies might be stressful and, the anxiety of living with untreated cancer can be overwhelming (Zerbib *et al.*, 2008).

Aggressive treatment. Removal of the prostate gland or radical prostatectomy (RP) is one of the most employed strategies to treat localized prostate cancer that has a high risk of progressing, and also those low-risk tumors that cannot be managed with active surveillance due to the patient's wishes (Picard *et al.*, 2009). The main goal of RP is to remove all cancerous tissue that is still confined inside the prostatic capsule (Zerbib *et al.*, 2008). RP provides good long-tem cancer control and accurate prediction of prognosis via analysis of pathologic cancer features (Zerbib *et al.,* 2008). Substantial proportion of patients with localized disease can be cured by surgery (Picard *et al.,* 2009). RP is also associated with reduction in PrCa specific deaths when compared with active surveillance (Wilt *et al.*, 2008). However, RP is a major surgical procedure and associated with complications. Typical complications are urinary incontinence and erectile dysfunction that are usually temporary but sometimes permanent, in which case they affect the patient's quality of life (Zerbib *et al.,* 2008). In the study by Wilt *et al.* (2008), 35 % of patients who had undergone RP experienced urinary leakage daily whereas only 7 % of men under active surveillance had the same problem. In the same study, 58 % of RP patients had no erections at all while the corresponding number among active surveillance group was 32 % (Wilt *et al.*, 2008). Due to the side-effects, RP should be conserved to treating high-risk PrCa cases, with low-risk cases managed with less invasive techniques

Radiation therapy is another widely used method in treating clinically localized prostate cancer (Picard *et al.*, 2009). The most commonly used types of radiation therapy in PrCa management are external beam radiotherapy (EBRT) and brachytherapy. In EBRT, an external beam of radiation is directed to the cancerous tissue, whereas brachytherapy is referred to as internal radiation therapy In brachytherapy, a radiation source is placed inside or next to the area requiring treatment. Radiation has been used as a monotherapy, but studies support its use in combination with another therapy form (Picard *et al.*, 2009). The adverse effects of radiation therapy include acute radiation toxicity and temporary symptoms of the urinary tract (Zerbib *et al.*, 2008).

Hormonal therapy, usually androgen deprivation therapy (ADT) with either medical or surgical castration is the initial treatment for patients with metastatic PrCa, and also for clinically localized PrCa with an aggressive prognosis (Picard *et al.*, 2009). Also, anti-androgen therapy with androgen agonists is used (Picard *et al.*, 2009). However, ADT is not recommended to men with low-risk localized PrCa who could benefit from surgery or radiotherapy (Zerbib *et al.*, 2008). The purpose of ADT is to lower serum testosterone levels and hence limit the growth of the tumor. More recently, another form of ADT has been proposed: intermittent ADT, in which therapy is interrupted when serum PSA level drop and resumed again when PSA level increases (Niraula and Tannock, 2011). Intermittent ADT has been shown to be as effective as continuous ADT, with lesser side-effects (Niraula and Tannock, 2011). Usually the initial effect of ADT is positive, with improvement of symptoms and remission of the disease (Ong and Winquist, 2011). The effective period of ADT is quite short, though. After the initial period of 12 to 18 months on average the cells become resistant and disease progression occurs (Lassi, 2010), resulting in castration-resistant prostate cancer (CRPC). The development of CRPC is a result of high androgen receptor expression in PrCa cells (Lassi, 2010). CRPC is a step towards a worse prognosis, as the median survival time is one to two years (Lassi, 2010).

Chemotherapy is usually the last line of treatment for prostate cancer, and is not usually used before the development of castration-resistant stage of the disease. Mitoxantrone was the first chemotherapeutic agent to be approved in the palliative treatment of CRPC (Niraula and Tannock, 2011). Docetaxel was introduced later, providing palliative care with survival benefits (Ong and Winquist, 2011). In some patients the disease progresses despite docetaxel treatment, thus representing an unmet medical need that many ongoing research projects are trying to address with new chemotherapeutic agents (Bracarda *et al.*, 2011; Niraula and Tannock, 2011; Ong and Winquist, 2011; Picard *et al.*, 2009).

The most recent progress in prostate cancer therapy is immunotherapy. Therapeutic vaccines have been shown to be effective treatment modalities against castration-resistant prostate cancer in recent clinical trials (Bilusic *et al.*, 2011; Gulley and Drake, 2011). These vaccines teach the immune system to recognize and attack certain cancer-associated proteins (Bilusic *et al.*, 2011). Therapeutic vaccines have been shown to improve survival and clinical outcomes among patients with metastatic PrCa, producing minimal side effects (Gulley and Drake, 2011), and one therapeutic vaccine has been FDA-approved for the treatment of PrCa (Bilusic *et al.*, 2011; Gulley and Drake, 2011). Ongoing trials are currently estimating the effectiveness of immunotherapy in combination with other, more conservative treatment strategies (Bilusic *et al.*, 2011; Gulley and Drake, 2011).

2.2.5. The economic burden of prostate cancer

Being the most prevalent malignancy in men in developed countries, prostate cancer is a major health concern. In the USA, the average life-time cost of prostate cancer was 34 000 dollars (appr. 26 000 Euros) per patient in 2008 (Stokes *et al.*, 2011) and total prevalence-based treatment cost was estimated at 9,8 billion dollars in 2006 (Roehrborn and Black, 2011). For European countries, no studies estimating the total life-time cost of prostate cancer could be found, but the cost of diagnosis, treatment and five-year follow-up period was 10 000 Euros per patient, ranging from 8 100 Euros in the UK to 12 800 Euros in Germany (Roehrborn and Black, 2011). The costs of PrCa can be divided to three phases: the initial phase (12 months after diagnosis), the continuing care phase (between initial phase and end of life) and the end of life phase (last 12 months of life) (Skolarus *et al.*, 2010). The cost of each phase per patient depends on the stage at which cancer is diagnosed, survival time and treatment strategy (Roehrborn and Black, 2011). As the number of men over 65 in the world is expected to increase from 400 million in 2000 to 1,5 billion individuals by 2050 (Lunenfeld, 2002) and considering the rise in PrCa incidence, slow decline in mortality and the fact that more and more men are diagnosed at earlier stage due to PSA screening and thus survive longer, the economic burden of PrCa towards the health care system is going to multiply in the near future. This is one reason why better tools to distinguish aggressive PrCa from indolent PrCa are required. Most PrCa cases are slow-progressing and do not become clinically significant during the patient's lifetime and could be managed with watchful waiting. There is, however, a subset of PrCa cases that are aggressive, proceed rapidly and metastasize. These are the cases that should be caught.

2.3. Genetic susceptibility to prostate cancer

Over decades, genetic epidemiological evidence has accumulated in favor of a strong hereditary component of prostate cancer. The familial aggregation of prostate cancer was recognized as early as the 1950s (Morganti *et al.*, 1956), and since then the view that PrCa is a heritable disease has been confirmed by many studies. Familial PrCa is defined as a disease where there are two first-degree relatives (father, brother or son) and at least two second degree relatives (e.g. grandfather, nephew, uncle) diagnosed with PrCa (Stanford and Ostrander, 2001). This definition accounts for 10–20 % of all PrCa cases in the general population (Stanford and Ostrander, 2001). A more specifically defined subset of familial prostate cancers is termed hereditary prostate cancer (HPC). HPC is characterized by at least one of the next three criteria: 1) three or more first-degree relatives with PrCa, 2) three successive generations with prostate cancer, either through paternal or maternal lineage; or 3) two first-degree relatives with prostate cancer diagnosed at a relatively young age (e.g., <55 years) (Carter *et al.*, 1993). HPC is estimated to account for 5–10 % of PrCa in the general population (Stanford and Ostrander, 2001). However, a larger proportion or PrCa cases in younger men can probably be explained by hereditary form of prostate cancer (Stanford and Ostrander, 2001). As no definite mutations causing HPC have been identified, the definition of this disease is based solely on pedigree analysis (Pakkanen, 2010), and distinguishing sporadic PrCa from familial form is not possible. Furthermore, it is impossible to predict whether an inherited cancer is indolent or fast-growing, aggressive form (Aly *et al.*, 2011). For these reasons, among others, there is a great need for prediction tools that can be used to predict an overall susceptibility to PrCa and a predisposition to aggressive form of the disease. Genetic variants serve as promising tools for this purpose.

2.3.1. Knudson's two-hit hypothesis of hereditary cancer

The identity and function of each cell is dependent on its genetic material, DNA. Changes in a cell's DNA may influence the function of genes that guide these basic aspects. As a result, a situation may arise where a certain change in a cell's genome provides the cell a survival advantage over its neighboring cells (Stratton *et al.*, 2009). This selection advantage leads to abnormalities e.g. in cell division and ultimately to cancer. Usually mutations leading to cancer are somatic and are passed on only to the cancerous cell's progeny but not to the individual's offspring (Stratton *et al.*, 2009). There are two classes of genes that are frequently mutated in

cancers: cancer-promoting oncogenes are usually activated and tumor suppressor genes are inactivated (Mattila, 2009).

Forty years ago, Knudson studied retinoblastoma, a rare form of cancer that develops in the retina. He created a theory that tumor formation requires loss-of-function mutations in certain regulatory genes that control cell cycle, apoptosis and proliferation (tumor suppressor genes) (Knudson, 1971). His "two-hit" hypothesis suggests that two "hits" or mutation events are required to inactivate both alleles of a tumor suppressor gene (Figure 2). The first hit can be sporadic, resulting from random mutagenesis, or it can be inherited in which case it is present in all cells of the body. The second hit is usually somatic, and when the second copy of the tumor suppressor gene is inactivated, allowing the cell to undergo transformation to malignant tumor cell. People that have inherited the first hit only need to accumulate one more and hence have a higher risk of developing cancer in their lifetime than people who initially have two healthy alleles of the tumor suppressor gene.

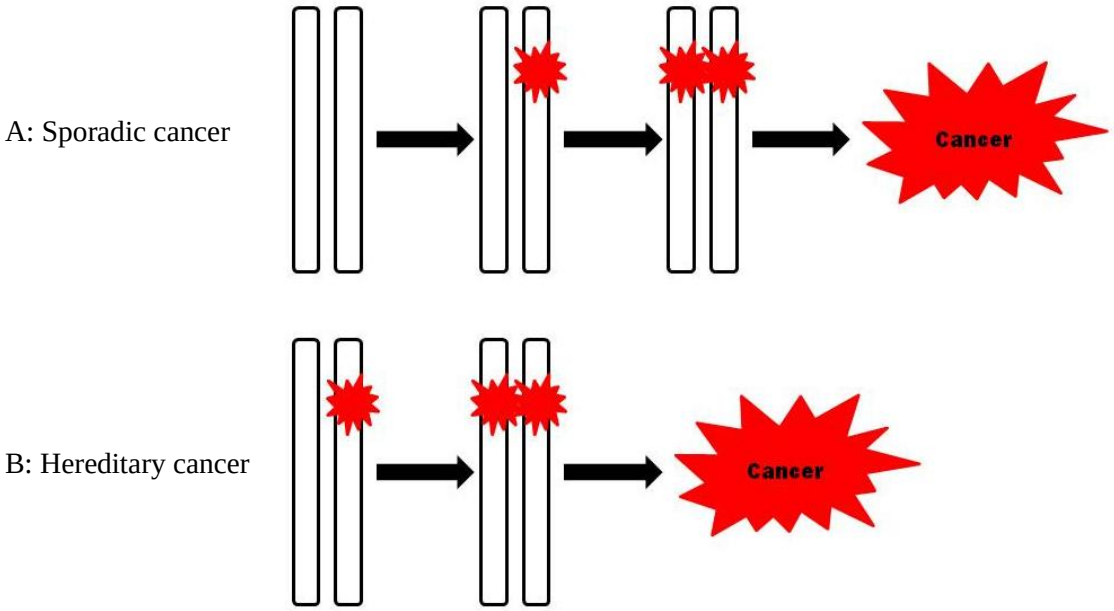

Figure 2. Knudson's two-hit hypothesis for cancer formation. A: In the sporadic form of the cancer, the individual initially has two healthy alleles of a tumor suppressor gene that both have to be mutated (red star) in the same somatic cell for cancer to develop. B: In hereditary form, the individual has inherited one mutated allele of the tumor suppressor gene. The mutation is present in all their cells. Thus, only one somatic mutation is required for tumorigenesis and individuals carrying one mutation in their germline have a higher risk of developing cancer in their lifetime. (Knudson, 1971)

2.3.2. Linkage studies

Early segregation analyses have suggested a Mendelian inheritance pattern for hereditary prostate cancer predisposition. This means that a subset of PrCa cases could be caused by mutations in a few major genes. In a segregation analysis performed by Carter and coworkers (1992) the results suggested that a rare, autosomal dominant and highly penetrant allele could explain hereditary predisposition to prostate cancer. Other studies have confirmed this mode of inheritance as a possibility (Schaid *et al.*, 1998). In these studies, the model for inheritance fitted best when applied to early-onset prostate cancer cases. However, other modes of inheritance have been indicated in other segregation analyses: autosomal recessive (Pakkanen *et al.*, 2007) and X-linked (Monroe *et al.*, 1995).

The unraveling of inheritance modes by segregation analyses has led to large efforts to carry out linkage studies on families with multiple PrCa cases to find the chromosomal regions behind the inheritance. These studies screen the genome using genetic markers to identify genetic loci that cosegregate with the disease (Pomerantz, 2010), ultimately aiming to map the disease causing loci on chromosomes to identify the genetic basis of the disease. Genes that are close to each other on chromosomes tend to segregate together during meiotic recombination, enabling the recognition of cancer gene locations by their close proximity to a specific marker gene (Altshuler *et al.*, 2008). Linkage studies have been useful in identifying rare and highly penetrant susceptibility loci for many common cancers, such as the cancers of lung and colon. However, even though over a dozen PrCa linkage analyses have been performed, the results have been somewhat inconclusive. Several predisposition loci like 1p36 (*CAPB*), 1q23–25 (*HPC1*) and 8p22–23 have been reported but these findings could not be confirmed in other studies (Schaid, 2004). Furthermore, the results of different linkage scans have been shown to vary depending on the population in which the linkage scan was performed. For example, the susceptibility loci 1q25 (*HPC1*) recognized in other populations were found to have no significant effect in on prostate cancer predisposition in Finnish population (Rökman *et al.*, 2002). On the other hand, the Finnish population seems to have additional susceptibility loci, including Xq27 – 28 (*HPCX*) (Xu *et al.*, 1998) and 3p25–26 (Schleutker *et al.*, 2003).

Overall, the confirmation of the susceptibility loci in different study populations has been inconsistent and despite extensive effort, no major predisposing prostate cancer genes have been found thus far by further analysis of linkage study results (Schaid, 2004). This is most probably due to the large possibility that hereditary prostate cancer risk is rather mediated by many common, low-penetrance mutations and their cumulative effects (common disease-common

variants –model) than a few rare, high-penetrance mutations (common disease-rare variants –model). The common disease-common variants –model is supported further by the fact that PrCa is a very heterogeneous disease with multiple clinical manifestations. Furthermore, the inconsistency in linkage scan results might be due to the fact that they have been carried out in different populations that might have totally different predisposing variants. Despite the shortcomings of linkage studies, they are an important tool in identifying new predisposing loci.

Linkage studies are not dead yet. They are an important tool in recognizing novel susceptibility loci that can be further analyzed. A recent genome-wide linkage scan of 69 Finnish HPC families by Cropp *et al*. (2011) confirmed a previously found linkage to 17q21–22 and revealed a novel, strong linkage signal at 2q37.3 (Figure 3).

2.3.3. Association studies

As the linkage studies have indicated that any major genes are unlikely to explain genetic prostate cancer predisposition, genome-wide association studies (GWAS) have emerged as an another tool to identify genes that might confer susceptibility to PrCa. Association studies aim to identify genes that have more common, but weaker and low-penetrance risk alleles (Pakkanen 2010). Association studies are usually performed with case-control setting (McCarthy *et al.*, 2008). For prostate cancer, the study populations are traditionally men with or without PrCa. Usually the PrCa cases used in GWASs are sporadic with no regard to family history (Ishak and Giri, 2011). GWASs scan for general associations between a genetic variant (typically a SNP) and a phenotypic trait (e.g. prostate cancer) without prior knowledge of function or position (Easton and Eeles, 2008). The frequency of a polymorphic allele in cases vs. controls can rapidly provide information about the disease risk of a certain gene. GWASs have been used to successfully recognize multiple SNPs and loci that are associated with PrCa (Easton and Eeles, 2008). These studies implicate that a large number of gene variants is involved in conferring prostate cancer risk (Eeles *et al.*, 2008) –a result that is consistent with the findings of linkage studies. As these variants are low-risk alleles but their frequencies are quite high, the cumulative effect of many variants in one person must be taken into account. Other studies have tried to reveal associations between different SNPs and aggressive or dormant form of PrCa (Aly *et al.*, 2011). Generally, the results have not yielded encouraging results.

In a very recent GWAS study by Schumacher *et al*. (2011) the 2q37.3 susceptibility region found in a previous linkage scan by Cropp *et al*. (2011) was confirmed (Schumacher *et*

al., 2011). They found a SNP in the 2q37.3 area that was very strongly associated with PrCa (rs2292884; $p= 4.3 \times 10^{-8}$) (Schumacher *et al.*, 2011). This seems to point to the fact that the chromosomal area in 2q37.3 does indeed contain something interesting in regard to PrCa since the area has emerged in both linkage studies and GWASs.

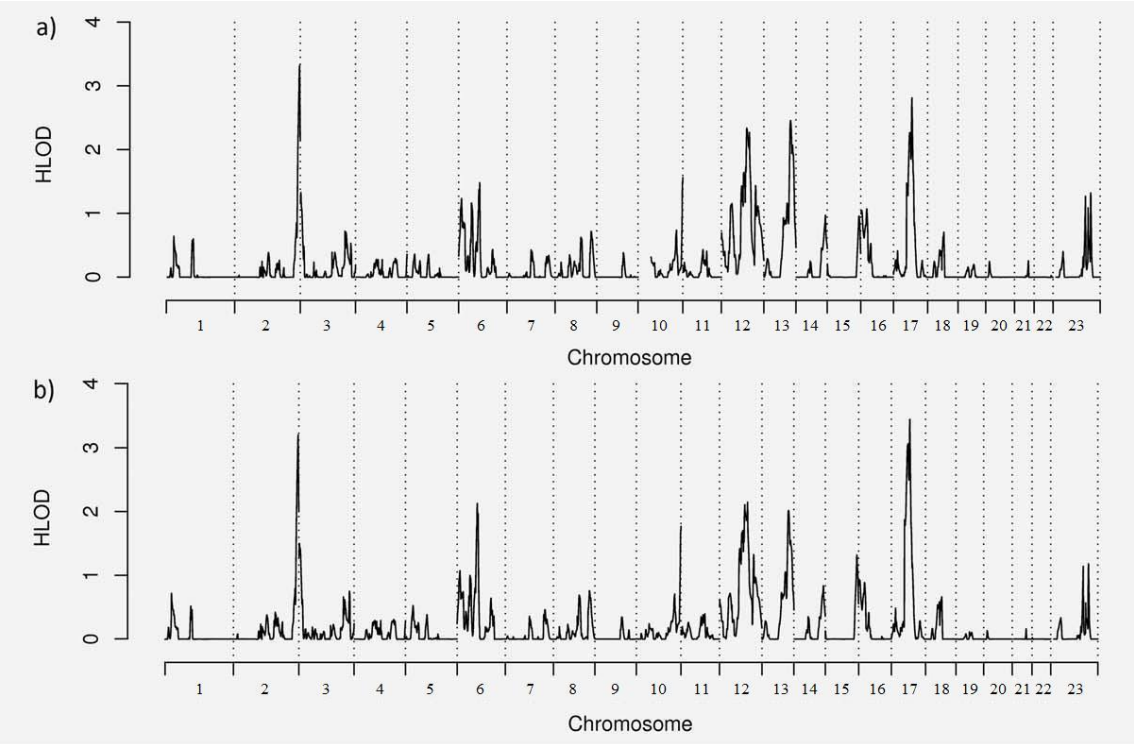

Figure 3. Whole genome graphical images for linkage results from a recent linkage study with 69 Finnish prostate cancer families. Heterogeneity logarithm of odds (HLOD) linkage results are shown for (a) a high penetrance dominant affected-only model and (b) a reduced penetrance dominant affected-only model using GENEHUNTER-PLUS. (Cropp *et al.* 2011)

2.4. Theory of the methods

2.4.1. High-resolution melting

High-resolution melting (HRM) is a detection method for DNA sequence variants such as mutations, single-nucleotide polymorphisms and epigenetic differences in double-stranded DNA, first described almost a decade ago (Wittwer *et al.*, 2003). It is based on the fact that a double-stranded DNA molecule has a unique melting temperature that is dependent on its base composition. Two dsDNA molecules that differ in their base composition at even one base have different melting temperatures. In HRM, the differential melting profile of DNA molecules is

detected, enabling the recognition of DNA molecules with sequence variations.HRM analysis is initiated with real-time PCR reaction in which the DNA sequence of interest is amplified. A fluorescent, DNA intercalating dye is included in this PCR reaction in amounts great enough to reach saturation. These dyes fluoresce at high intensity when they are not bound to DNA and at a low intensity when they are bound. After PCR amplification, the melting analysis begins: the temperature of the PCR mixture is increased slowly, typically in 0,2°C increments from around 50°C to 95°C, where all of the dsDNA molecules have melted. As the dsDNA duplexes are melted to ssDNA, the fluorescent dye is released and the increase in fluorescence during each temperature increment step is detected. This change in fluorescence can be transformed to melting curves. For DNA molecules with different base composition and hence different melting temperature, the curves are different (Figure 4).

HRM has been described in the literature as simple, easy to use, flexible, nondestructive, sensitive and specific (Vossen *et al.*, 2009). Being more cost-effective than direct sequencing by Sanger method, the golden standard of sequence variant detection, HRM has become the genotyping method of choice for many laboratories (Vossen *et al.*, 2009).

2.4.2. TaqMan® genotyping

TaqMan®-genotyping technology by Applied Biosystems employs the 5′→3′ exonuclease activity of the thermostable DNA polymerase of *Thermus aquaticus*, a bacterium living in hot springs, in a PCR product detection system. An allele-specific oligonucleotide probe labeled at 5′ end with a fluorophore and with a quencher at 3′ end is designed to anneal with a sequence of interest and introduced to a PCR reaction with unlabeled forward and reverse primers (Holland *et al.*, 1991). A different fluorophore is used for each allele. The annealing of the probe with template strand creates a suitable substrate for Taq polymerase exonuclease activity as it extends a DNA strand (Holland *et al.*, 1991). The probe is degraded, releasing the fluorophore from the quencher and thus producing a fluorescent signal that can be detected with a detection device post-PCR. If an individual is homozygote wild-type, only the fluorophore attached to the wild-type allele-specific probe will produce a signal; if the individual has a homozygous mutation of interest, only the mutation-specific probe will fluoresce. Heterozygous individuals will produce both signals. TaqMan® genotyping is a single-step procedure, easy to use, effective, and Applied Biosystems has a wide variety of ready-to-order assays for many markers. This is why this genotyping method is used widely around the world.

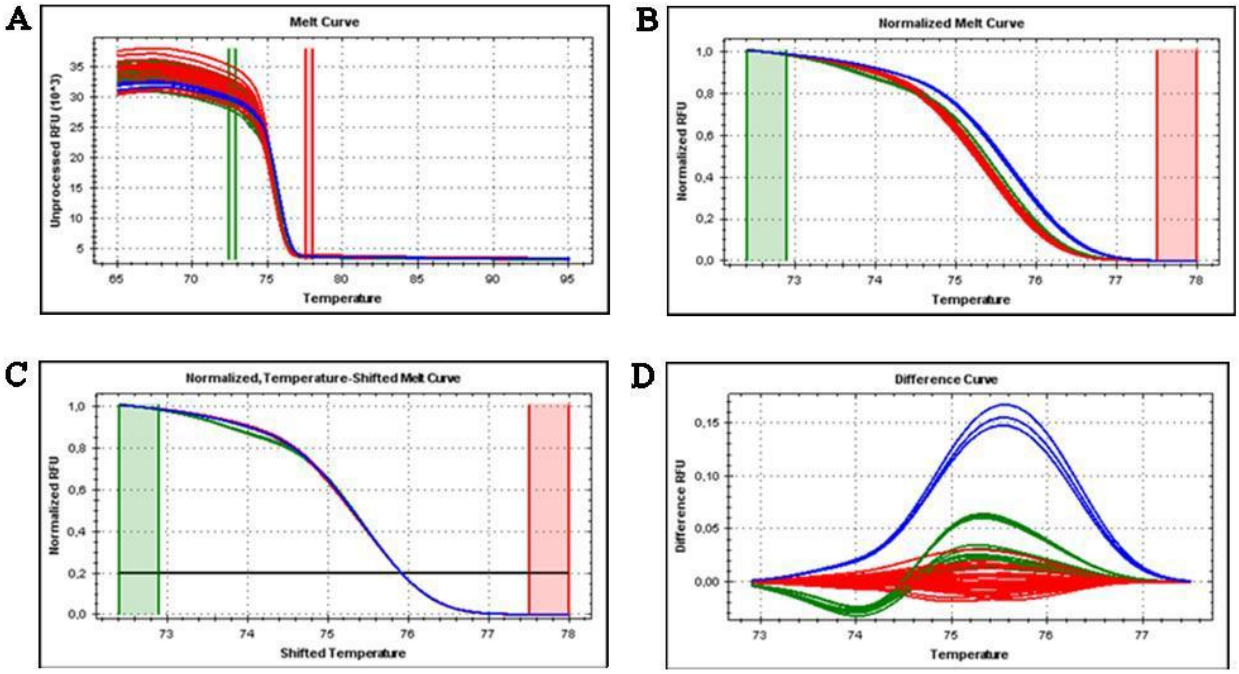

Figure 4. Various visualization modes available through Bio-Rad Precision Melt Analysis™ Software, the HRM analysis software used in this study. Three different genotype clusters are marked in red (wild-type), blue (rare homozygote) and green (heterozygote). The change is a single-nucleotide polymorphism (C>T).
A: Original melting curve acquired through gradually increasing the temperature of PCR product and monitoring the fluorescence; B: Normalized melting curve compensates for small differences in the amount of PCR product that might arise due e.g. to different amounts of template; C: Normalized, temperature-shifted melting curve allows easy visualization of heterozygotes; D: Difference curve accentuates differences between the three genotypes so that they are all easily distinguishable.

2.4.3. KASP genotyping

KBioscience has recently developed a competitive allele specific PCR system for SNP genotyping, called Kbioscience's Allele Specific PCR (KASP, or KASPar assays). These assays use an innovative fluorescent genotyping system: the assay consists of two allele-specific unlabeled primers, one for each genotype, with a tail sequence that does not bind to the template, a common reverse primer and a fluorescent FRET (Fluorescent Resonance Energy Transfer) cassette that is initially bound to a quencher (www.kbioscience.co.uk). During PCR reaction, double-stranded DNA molecules with a tail sequence from forward primer are eventually produced. The FRET cassettes bind to the tail sequences and the fluorophore is released from the quencher, producing a signal that can again be detected.

According to the manufacturer, KASPar assays are simple, reproducible and furthermore are more cost-effective and have higher SNP-to-assay conversion rate than TaqMan® technology (www.kbioscience.co.uk). The plate reading can be performed with same system as TaqMan PCR plates, so no new devices are needed for those who already use TaqMan

technology. All this makes the KASPar technology attractive option for TaqMan® genotyping for laboratories that use SNP genotyping on a regular basis. However, being a new method, this system needs to be validated so that a researcher can be sure they get as good results as with a more costly method. Furthermore, the practical aspects such as the ease of ordering and the simplicity of use protocol of KASPar assays have to be considered.

3 Aims of the research

The aims of this research were to

1. Validate the association of the most promising variants found by deep sequencing of the PrCa linked region in chromosome 2 locus q37.3 with prostate cancer.
2. Define the association with clinical features of PrCa, in this case serum PSA value.
3. Compare Applied Biosystems' TaqMan® SNP genotyping assays and KBioscience's KASP genotyping assays to validate KASP assays as possible replacement for TaqMan genotyping in our laboratory.

4 Materials and methods

4.1. Mutations chosen for validation

Ten of the mutations found in chromosome 2 linkage peak deep sequencing were chosen for this study. The chosen mutations were all possibly pathogenic according to predictive programs and were located preferably in exonic regions of the gene, although a few intronic mutations were included as well. The mutations chosen for this study, their genomic locations, locations in the gene and genotyping methods are presented in Table 1.

Table 1. Names of the genes where the mutations are found, genomic locations, mutation types, SNP ID:s, locations within the gene and genotyping methods of the mutations chosen for this study.

Gene	Genomic location and type of the mutation	SNP ID (if known)	Location in the gene	Genotyping method
AGAP1	1733insCAGG	rs142341634	CDS	HRM
ASB1	239344663T>A	rs11904390	Intron	KASPar
COPS8	149C>T	-	5'UTR	KASPar
IQCA1	237247036insA	rs111440161	Intron	HRM
MYEOV2	241075809C>T	rs13406410	5'UTR	TaqMan
MYEOV2	241075991A>C	rs13411615	5'UTR	KASPar
RAB17	258T>G	rs78523256	5'UTR	Sequencing
RAMP1	238767662C>G	rs3754699	5'UTR	TaqMan & KASPar
SCLY-UBE2F	239002480delTTG	rs72316729	Intron	HRM
TRAF3IP1	536G>A	rs61742338	CDS	TaqMan & KASPar

4.2. Study subjects

All mutations were validated in a sample set including 1 505 unselected PrCa cases and 190 familial PrCa cases, except *RAB17* 258T>G (rs78523256), which was validated using 653 unselected and 190 familial PrCa cases. The unselected cases were collected from patients living in Pirkanmaa Hospital District and diagnosed in the Department of Urology at Tampere University Hospital from 1996 to 2010, resulting in a population-based collection of unselected PrCa cases. Of the familial PrCa samples, 150 samples were collected from the youngest affected patient from Finnish PrCa families with three or more affected members and 40 samples from the youngest affected patient from Finnish PrCa families with two affected members. An informed consent has been collected from each of the patients. As control group, 162 samples

from male blood donors at the Finnish Red Cross were used. All control individuals were healthy at the time of blood donation.

4.3. DNA extraction

All samples were obtained as EDTA blood tubes. Genomic DNA was extracted from peripheral blood lymphocytes using a commercial extraction kit (Puregene, Gentra Systems, Inc., Minneapolis, MN, USA & Wizard®, Promega Corporation, Madison, WI, USA) according to manufacturer's instructions (. DNA yields were quantified using a ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

4.4. High-resolution melt analysis

HRM as a genotyping method was first described by Wittwer *et al*. (2003). HRM was used as genotyping method for *AGAP1* 1733insCAGG, *IQCA1* 237247036insA (rs111440161) and *SCLY-UBE2F* 239002480delTTG (rs72316729). The primers used in HRM analysis were designed with Beacon Designer 7 software (PREMIER Biosoft, Palo Alto, CA, USA) in SYBR® Green Design mode. As a template, reference sequences for each of the genes from UCSC Genome Browser (http://genome.ucsc.edu/) were used. During primer design, homology and primer-dimer forming regions identified by the software were avoided as well as possible to ensure specificity. The amplicon lengths were set to 150 - 250 bp. The primers were obtained from Sigma Aldrich Finland Oy (Helsinki, Finland) as 100 µM solution in H_2O and diluted with water to use concentration of 10 µM. The sequences of primers, their amplicon lengths and annealing temperatures used in HRM are presented in Table 2.

Table 2. Sequences, annealing temperatures and amplicon lengths of primers used in HRM analysis and direct sequencing.

Gene	Forward (5'>3')	Reverse (5'>3')	T_a (°C)	Amplicon length (bp)
AGAP1[a]	CCATAACCCTCATCAGATAAAG	TAACTCCCACCGCTAATTC	59.6	242
IQCA1[a]	AAATTGTGTTTGTCAGAGCTTTGC	TTTCAGGAGTTTAGGTTCATTCTG	58.4	149
IQCA1[b]	AAATTGTGTTTGTCAGAGCTTTGC	TTTCAGGAGTTTAGGTTCATTCTG	63.0	222
RAB17[b]	TTCTGGGCTGTTGGTTTTTC	AGGTCAGGCCTCTTCTCCTC	60.0	300
SCLY-UBE2F[a]	CAGGAGAATGAGCAGTTTC	TCATAAGCCTCGCAGTTC	57.5	146
SCLY-UBE2F[b]	TTTGTCCGTCTGCCTCTC	GCTCACTTCCAGCCTCTC	60.0	375

[a]Primers for HRM, [b]Primers for sequencing

Primer annealing conditions were optimized by performing an initial temperature gradient (52–62°C) PCR run with Bio-Rad CFX384 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). The optimal temperature was selected based on C(t) values (the number of cycles before the amplification curve reaches a set threshold) and final amount of product (final fluorescence). All amplicons were tested for primed-dimer and unspecific product formation by performing a PCR run in the optimal primer annealing temperatures with a rising template concentration (0,1-1-10-25-50 ng) and analyzing the melting profiles. Additionally, an agarose gel electrophoresis was performed to ensure no unspecific product was formed.

Bio-Rad CFX384 Real-Time PCR Detection System was used to perform real-time PCR and a following high-resolution melting analysis. RT-PCR reactions were run in 10 µl reactions containing 10-25 ng template, 500 nM forward and reverse primers, and 1x SsoFast™ EvaGreen® Supermix (Bio-Rad Laboratories). All reactions were run in duplicates. The amounts of reagents for 1x reaction mix are presented in Table 3. The cycling conditions were: initial denaturation at 98°C for 3 minutes, followed by 30 cycles of denaturation at 98°C for 3 sec and annealing and extension at optimal annealing temperature for each of the primer pairs for 30 seconds. Fluorescence was measured after each cycle to monitor the accumulation of products. After cycling, final denaturation and dissociation was performed at 98°C for 1 minute followed by rapid cooling at 50°C for 1 minute. After this, melt curves were obtained by rising the temperature in 0,2°C increments lasting 10 seconds. Melt curves were analyzed with Bio-Rad Precision Melt Analysis™ (Bio-Rad Laboratories) software and the genotypes of the clusters and aberrant samples were confirmed with direct sequencing of representative samples.

Table 3. 1x HRM reaction mix

Reagent	1x (µl)
SsoFast™ EvaGreen® Supermix (Bio-Rad Laboratories)	4
10 µM forward primer	0,50
10 µM reverse primer	0,50
sterile H_2O	5
Total reaction volume	10

4.5. Direct sequencing

Direct sequencing by Sanger method (Sanger *et al*. 1977) was used to confirm the genotypes of different clusters of HRM analysis, genotyping of samples that were not amplified properly

during HRM and hence could not be genotyped by HRM, and as genotyping method for *RAB17* 258T>G (rs78523256).

Sequencing primers were designed using free Primer3 program (http://frodo.wi.mit.edu/) and the reference sequence from UCSC Genome Browser (http://genome.ucsc.edu/). The primer annealing temperatures were optimized in an initial temperature gradient run (55–62°C) and a following agarose gel electrophoresis run. Primer sequences, their amplicon lengths and annealing temperatures in used in sequencing are presented in Table 2.

For sequencing, DNA was first amplified by PCR using 25 µl reactions that contained 1x HotStart PCR buffer (Fermentas, Glen Burnie, MD, USA), 1.5 mM $MgCl_2$, 200 nM dNTPs, 600 nM forward and reverse primers, 0.75 U Maxima HotStart Taq Polymerase (Fermentas) and 50-100 ng template DNA. For *AGAP1* 1733-1734insCAGG, HRM PCR products were used as a direct template for sequencing. The amounts of reagents for 1x reaction mix are presented in Table 4. The cycling conditions were: initial denaturation at 95°C for 10 minutes, 34 cycles of denaturation at 95°C for 30 sec, annealing at optimal annealing temperature for each of the primer pairs for 30 seconds and extension at 72°C for 45 seconds, followed by final extension at 72°C for 5 min. Agarose gel electrophoresis was used to confirm that only right-sized product was amplified.

PCR products were purified from excess primers and nucleotides by ExoSAP purification: 10 µl of ExoSAP mix containing 0.025 µl of Exonuclease 1 (New England Biolabs, Ipswich, MA, USA), 0.25 µl of rAPid Alkaline Phosphatase (Roche Diagnostics GmbH, Mannheim, Germany) and 9.725 µl of sterile H_2O was added to each PCR reaction, incubated at 37°C for 30 minutes and at 95°C for 5 minutes.

Purified products were sequenced using the BigDye Terminator Cycle Sequencing Reaction Kit v3.1 (Applied Biosystems, Foster City, CA, USA). A sequencing PCR was performed to amplified DNA in 10 µl reaction volumes. One reaction contained 3 µl of purified PCR product, 3 µl of 2.5x sequencing buffer, 160 nM primer, 1 µl of BigDye™ terminator reaction mix (Applied Biosystems) and 1µl of sterile H_2O. The cycling conditions of sequencing-PCR were: 26 cycles of denaturation at 96°C for 30 sec, annealing at 50°C for 20 sec and elongation at 60°C for 4 min. After cycling, the reaction mix was initially cooled at 11°C for 1 min and then to 4°C.

The sequencing PCR reactions were purified from excess materials using ethanol precipitation: 26 µl of 1:25 mixture of 3M sodium acetate and absolute ethanol was added to each sequencing reaction and incubated for 10 minutes in RT and subsequently centrifuged in 2 000 x g for 30 minutes. Supernatant was removed and DNA precipitate was washed with 100 µl

of 70 % ethanol and 2 000 x g for 10 minutes. After washing, the samples were allowed to dry and 12 µl of HiDi™ formamide (Applied Biosystems) was added. Sequencing was performed with ABI Prism™ 3130XL sequencer (Applied Biosystems).

Table 4. 1x PCR reaction mix

Reagent	**1x (µl)**
10 x HotStart PCR Buffer (Fermentas)	2,5
25 mM $MgCl_2$ (Fermentas)	1,5
10 mM dNTP mix (Fermentas)	0,5
10 µM forward primer	1,5
10 µM reverse primer	1,5
Maxima HotStart Taq DNA polymerase	0,15
Template DNA	50 ng
sterile H_2O	to 25 µl
Total reaction volume	25 µl

4.6. TaqMan® genotyping

TaqMan genotyping technology was first developed by Holland *et al*. (1991). Mutations RAMP1 238767662C>G (rs3754699), TRAF3IP1 536G>A (rs61742338) and MYEOV2 241075809G>T (rs13406410) were genotyped using available TaqMan® SNP Genotyping Assays (Applied Biosystems, Foster City, CA, USA). PCR reactions were run in 5 µl reactions containing 10-25 ng DNA, 1x TaqMan Universal Master Mix and 1x assay mix. Four microliters of 1x assay mix-1x TaqMan Universal Master mix –mixture was pipetted to 384-well plate by electronic pipettor and 1 µl of sample DNA was subsequently added using TECAN Freedom EVO® 75 pipetting robot (Tecan Group Ltd., Männerdorf, Swizerland). The cycling conditions were the following: initial denaturation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 92°C for 15 seconds and annealing and extension at 60°C for 1 minute. For rs13406410, a PCR program with 45 denaturation and annealing cycles was used. Post-PCR reading of the plates was performed with ABI Prism® 7900HT sequence detection system (Life Technologies Corporation, Carlsbad, CA, USA) according to manufacturer's instructions.

4.7. KASP genotyping

KASP genotyping is a novel genotyping technology and a protocol created by the manufacturer was used without any deviations (KBioscience). Mutations RAMP1 238767662C>G (rs3754699), TRAF3IP1 536G>A (rs61742338), ASB1 239344663T>A (rs11904390), MYEOV2 241075991A>C (rs13411615) and *COPS8* 149C>T were genotyped using KASP-on-Demand genotyping assays (KBioscience, Hoddesdon, Hertfordshire, England). PCR was performed in 5 µl reactions containing 10-25 ng DNA, 1x KASP reaction mix and 1x assay mix. Four microliters of 1x KASP reaction mix-1x assay mix –mixture was pipetted to a 384-well plate by electronic pipettor and 1 µl of sample DNA was subsequently added using TECAN Freedom EVO® 75 pipetting robot (Tecan Group Ltd., Männerdorf, Swizerland). The cycling conditions were following: initial denaturation at 94°C for 15 minutes, 10 cycles of denaturation at 94°C for 20 seconds and annealing and extension at 61-55°C for 1 minute, dropping -0,6°C/cycle. Final elongation was performed with 26 cycles of denaturation at 94°C for 10 seconds and annealing and extension at 55°C for 1 minute. Post-PCR plate reading was performed with ABI Prism® 7900HT sequence detection system (Life Technologies Corporation, Carlsbad, CA, USA) according to manufacturer's instructions.

4.8. Data analysis

HRM melt curves were analyzed using Bio-Rad Precision Melt Analysis™ software (Bio-Rad Laboratories, Hercules, CA, USA). Sequencing data was analyzed with Sequencher 4.10.1 software (Gene Codes Corporation, Ann Arbor, MI, USA). KASPar and TaqMan data obtained from plate reading was analyzed with SDS v2.2. software (Life Technologies Corporation, Carlsbad, CA, USA). Association analysis and basic case-control association analysis of the genotypes were performed using PLINK tool for genome-wide association analysis (Purcell *et al.*, 2007). The analyses were performed for all samples versus controls, for sporadic cases versus controls and for familial samples versus controls. For mutations that reached significant values in association analysis and case-control analysis, a PLINK quantitave trait analysis was performed to see if these mutations could be associated with clinical characteristics of prostate cancer (PSA value). The association was further analyzed with GraphPad Prism 5.02 software (GraphPad Software Inc. La Jolla, CA, USA) using Kruskall-Wallis and Mann-Whitney nonparametric tests. Also, Hardy-Weinberg equilibrium calculation was performed with PLINK tool to see if the genotype frequencies were in Hardy-Weinberg equilibrium (HWE) in respect to

the studied loci, for deviations from HWE could indicate errors in genotyping as well as association with PrCa.

5 Results

All of the ten studied mutations were found to be real mutations and not some artifact of deep sequencing. Mutations were present in sporadic cases as well as familial cases, and both heterozygotes and homozygotes were found for most of the mutations (Table 5). Most of the mutations were quite common in the Finnish population: most heterozygotes had percentages around 30–50. *COPS8* 149C>T and *TRAF3IP1* 536G>A were the rarest of the studied mutations with heterozygote frequency around 3 % and wild-type frequency of about 97 %. The two mutations of *MYEOV2* were the most common with wild-type allele frequencies of only 25 %.

For *SCLY-UBE2F* 239002480delTTG and *TRAF3IP1* 536G>A, no rare homozygotes were found. For *SCLY-UBE2F* 239002480delTTG, this result is some kind of false negative: the mutation was genotyped with HRM, and the analysis suggested that only two genotypes were present: heterozygotes and wild-types. However, confirmation sequencing of some samples from both clusters revealed that also rare homozygotes were present but they were clustered together with the wild-type samples. Thus, the rare homozygotes exist, but they could not be distinguished from the wild-type samples for some reason and thus were seen as wild-types due to analysis method.

Table 5. Genotype frequencies of studied mutations in all genotyped samples (sporadic cases, familial cases, controls)

Gene variant	**Genotype frequencies in all genotyped samples (homozg/heterozg/wild-type)**	**Percentage of genotypes (%)**		
		Homozg	**Heterozg**	**WT**
AGAP1 1733insCAGG	100/597/1135	5,5	32,6	62,0
ASB1 239344663T>A	57/529/1209	3,2	29,5	67,4
COPS8 149C>T	2/56/1737	0,1	3,1	96,8
IQCA1 237247036insA	195/880/756	10,6	48,1	41,3
MYEOV2 241075809C>T	354/711/360	24,8	49,9	25,3
MYEOV2 241075991A>C	442/891/444	24,9	50,1	25,0
RAB17 258T>G	10/133/859	1,00	13,3	85,7
RAMP1 238767662C>G	78/581/1147	4,3	32,2	63,5
SCLY-UBE2F 239002480delTTG	0/983/849	0	53,8	46,4
TRAF3IP1 536G>A	0/38/1428	0	2,6	97,4

5.1. Hardy-Weinberg equilibrium

A majority of the studied mutations were found to be in Hardy-Weinberg equilibrium (HWE) in the study population. Only *AGAP1* 1733insCAGG, *IQCA1* 237247036insA and *SCLY-UBE2F* 239002480delTTG had deviations from the equilibrium. In *AGAP1* 1733insCAGG and *IQCA1* 237247036insA the disequilibrium was confined to the familial cases (p=0,04 and 2.06 x 10^{-13}, respectively) and not the sporadic cases (p=0,28 and 0,38, respectively) or the controls (p=0,81 and 1, respectively). Because the disequilibrium was found only in the familial cases and not the controls, a mistake in genotyping is unlikely.

For *SCLY-UBE2F* 239002480delTTG the HW disequilibrium was found to be strong in both familial and sporadic cases (p= 8.22 x 10^{-20} and 1.54 x 10^{-59}, respectively) and also in controls (p= 1.32 x 10^{-05}). Since the disequilibrium was found in controls as well, it seems that for this mutation there has been an error in genotyping. This is likely also in light of the fact that as mentioned earlier, HRM analysis could not distinguish the rare homozygotes from wild-types. This caused the rare homozygotes to be seen as wild types and their frequency to be zero even though, in reality, it is not. This is obviously a mistake in genotyping and causes problems with HWE calculations. The results of Hardy-Weinberg equilibrium calculations are presented in Table 6.

Table 6. Results of PLINK Hardy-Weinberg equilibrium calculations. Statistically significant disequilibrium p values are in bold.

Gene variant	**p for Hardy-Weinberg equilibrium**			
	All samples	**Sporadic samples**	**Familial samples**	**Controls**
AGAP1 1733insCAGG	0.07	0,28	**0,04**	0,81
ASB1 239344663T>A	1	0,30	1	1
COPS8 149C>T	0.09	0,20	1	1
IQCA1 237247036insA	0.01	0,38	**2,06 x 10^{-13}**	1
MYEOV2 241075809C>T	0.92	0,81	0,24	0,87
MYEOV2 241075991A>C	0.96	0,70	0,52	1
RAB17 258T>G	0.07	0,43	0,10	0,22
RAMP1 238767662C>G	0.66	0,39	0,37	0,61
SCLY-UBE2F 239002480delTTG	**0.00**	**1,54 x 10^{-59}**	**8,22 x 10^{-20}**	**1,32 x 10^{-5}**
TRAF3IP1 536G>A	1	1	1	1

5.2. Allelic association analysis

An allelic association analysis was performed for all samples versus controls, sporadic samples versus controls and familial samples versus controls to see if any of the mutations were associated with prostate cancer. For most of the mutations, no association was found. *IQCA1* 237247036insA reached significant association in all three analysis modes ($p<0,05$). This indicates that this mutation has a causative effect in PrCa. Also, *SCLY-UBE2F* 239002480delTTG was significantly associated with familial prostate cancer ($p=1,50 \times 10^{-4}$). The results are presented in Table 7.

Table 7. Allelic association of studied gene variants with prostate cancer. Mutations with statistically significant association are in bold. OR=odds ratio (ratio of the odds of an event occurring in one group to the odds of it occurring in another group), CI=confidence interval (measure of reliability of the estimate).

	Gene variant	**OR**	**95 % CI**	**p-value**
All samples	*AGAP1* 1733insCAGG	1,11	0,84–1,48	0,45
	ASB1 239344663T>A	1,12	0,82–1,52	0,47
	COPS8 149C>T	1,41	0,51–3,90	0,51
	***IQCA1* 237247036insA**	**2,1**	**1,67–2.64**	**0**
	MYEOV2 241075809C>T	1,03	0,82–1,31	0,78
	MYEOV2 241075991A>C	1,07	0,85–1,35	0,54
	RAB17 258T>G	1,06	0,67–1,66	0,81
	RAMP1 238767662C>G	1,13	0,84–1,52	0,41
	SCLY-UBE2F 239002480delTTG	1,19	0,92–1,56	0,19
	TRAF3IP1 536G>A	0,67	0,28–1,61	0,36
Sporadic cases	*AGAP1* 1733insCAGG	1,10	0,83–1,46	0,51
	ASB1 239344663T>A	1,10	0,80–1,48	0,58
	COPS8 149C>T	1,47	0,53–4,10	0,46
	***IQCA1* 237247036insA**	**2,12**	**1,68–2,67**	**0**
	MYEOV2 241075809C>T	1,06	0,84–1,34	0,61
	MYEOV2 241075991A>C	1,10	0,87–1,38	0,44
	RAB17 258T>G	1,11	0,70–1,77	0,65
	RAMP1 238767662C>G	1,12	0,84–1,51	0,44
	SCLY-UBE2F 239002480delTTG	1,13	0,86–1,47	0,38
	TRAF3IP1 536G>A	0,73	0,28–1,92	0,52
Familial cases	*AGAP1* 1733insCAGG	1,16	0,81–1,67	0,41
	ASB1 239344663T>A	1,30	0,87–1,92	0,21
	COPS8 149C>T	0,75	0,17–3,37	0,70
	***IQCA1* 237247036insA**	**1,94**	**1,43–2,62**	**$1,60 \times 10^{-5}$**
	MYEOV2 241075809C>T	0,94	0,70–1,27	0,70
	MYEOV2 241075991A>C	1,00	0,73–1,36	1
	RAB17 258T>G	0,84	0,47–1,52	0,57
	RAMP1 238767662C>G	1,12	0,77–1,63	0,55
	***SCLY-UBE2F* 239002480delTTG**	**1,88**	**1,35–2,61**	**$1,50 \times 10^{-4}$**
	TRAF3IP1 536G>A	1,00	0,33–3,01	1,00

5.3. Case-control analysis

A case-control association analysis was performed to further study the association of the chosen mutations with PrCa. For most of the mutations, the frequencies in cases did not differ from the frequencies in controls in either sporadic cases or familial cases and hence these mutations are probably not associated with PrCa ($p>0,05$). *IQCA1* 237247036insA was the only mutation that reached significant association with both sporadic and familial PrCa ($p<0,05$). For some unknown reason, no case-control analysis could be performed to *SCLY-UBE2F* 239002480delTTG. This is probably due to some problems with the genotype frequencies. The results are presented in Table 8.

For IQCA1, case-control analysis shows that the frequency of the mutation of interest is higher in controls than in cases. This seems to suggest that *IQCA1* 237247036insA has some kind of protective effect from prostate cancer and individuals having the insertion are at a lower risk of developing the disease.

5.4. Association with serum PSA values

For the two mutations that reached significant association with prostate cancer in previous analyses, *IQCA1* 237247036insA and *SCLY-UBE2F* 239002480delTTG, a further analysis was performed to see if they could be associated with serum PSA levels and hence the aggressiveness of PrCa. For *IQCA1* 237247036insA, a non-parametric Kruskal-Wallis test was performed with GraphPad Prism 5.02 software (GraphPad Software Inc. La Jolla, CA, USA). The results showed that persons having homozygous mutation had significantly lower PSA values than persons having heterozygous mutation or those who were wild-type in respect of *IQCA1* 237247036insA ($p=0,04$) (Figure 5A). For *SCLY-UBE2F* 239002480delTTG a Mann-Whitney test was performed with the same software to evaluate the association with PSA value. An inverse trend could be detected: wild-type patients had lower PSA values than those heterozygous in respect to the mutation (Figure 5B). As no rare homozygotes could be recognized for this mutation with the genotyping method used, the mean PSA value of homozygous individuals remains unknown. The differences were not statistically significant with the sample size used, though ($p=0,25$).

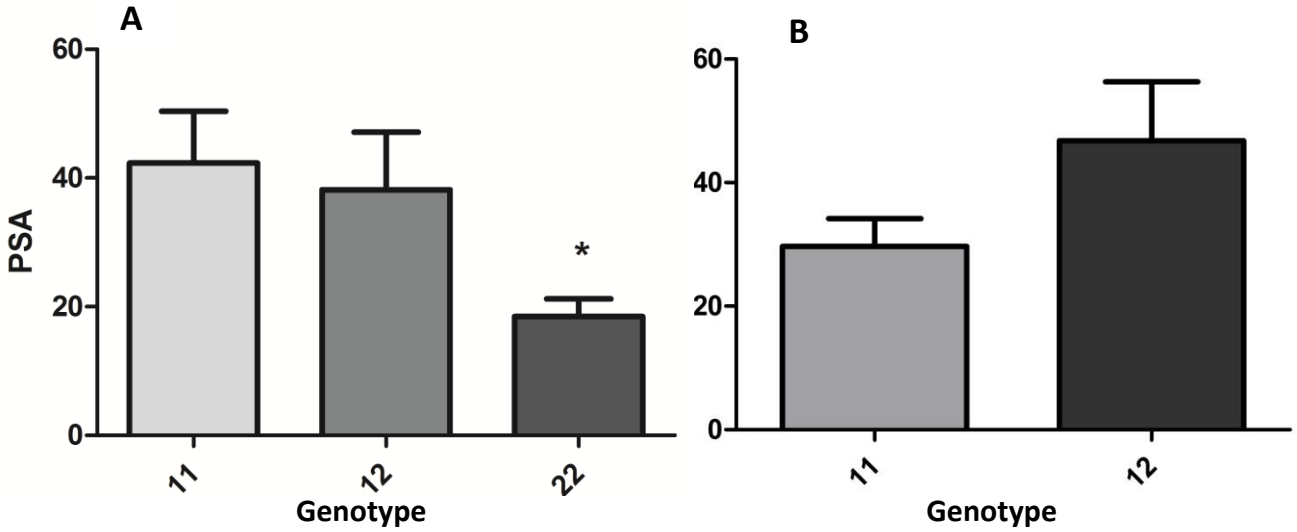

Figure 5. Effect of *IQCA1* 237247036insA and *SCLY-UBE2F* 239002480delTTG (A and B, respectively) mutations on serum PSA values. Genotypes: 11=wild-type, 12=heterozygote, 22=homozygote). Persons carrying a homozygous *IQCA1* 237247036insA has significantly lower serum PSA values that wild-types or heterozygotes (p=0,04). The corresponding bar is marked with an asterisk. For *SCLY-UBE2F* 239002480delTTG, an inverse trend was detected. The differences in PSA values between the genotypes are not statistically significant, however.

5.5. Comparison of TaqMan® and KASPar assays

TaqMan® genotyping is a widely used genotyping method. As KASP genotyping system is more cost-effective, it is an attractive option to reduce the costs of genotyping. To validate KASPar assays for future use, genotyping of two of the mutations, *RAMP1* 238767662C>G (rs3754699) and *TRAF3IP1* 536G>A (rs61742338), was performed with both TaqMan ready-to-order assays and KASP-on-Demand assays.

The results of the two different assays are well in concordance (Table 9). The differences in allele calls result from differences in the callrates of the assays. In other words, the amount of samples that has an undetermined genotype, i.e. somehow failed in PCR due to pipetting errors or some other factor, varies between the assays. The genotype calls, when made, were always the same between the two assays.

The practical aspects of KASPar assays, when compared with TaqMan assays, were twofold: on the other hand, the performing of the assays was just as easy and simple. The conditions for KASP-on-Demand assays were optimized by manufacturer before delivery, so there was no need for time-consuming optimization. The assays functioned well in each case. However, the ordering protocol of KASPar assays was not as easy as for TaqMan assays and their delivery was slower.

Table 8. Case-control analysis of the association of studied gene variants with prostate cancer. Mutations with statistically significant association are in bold. f=frequency

Samples	Gene variant	f in cases	f in controls	OR	95 % CI	p-value
All	*AGAP1* 1733insCAGG	0,22	0,20	1,11	0,84–1,47	0,46
	ASB1 239344663T>A	0,19	0,16	1,19	0,86–1,65	0,29
	COPS8 149C>T	0,02	0,01	1,36	0,47–3,91	0,57
	***IQCA1* 237247036insA**	**0,33**	**0,51**	**0,48**	**0,38–0,60**	**$9{,}54 \times 10^{-11}$**
	MYEOV2 241075809C>T	0,49	0,48	1,01	0,80–1,29	0,92
	MYEOV2 241075991A>C	0,50	0,49	1,03	0,81–1,31	0,80
	RAB17 258T>G	0,08	0,08	1,04	0,66–1,64	0,87
	RAMP1 238767662C>G	0,20	0,19	1,05	0,78–1,42	0,73
	SCLY-UBE2F 239002480delTTG					
	TRAF3IP1 536G>A	0,01	0,02	0,65	0,27–1,58	0,34
Sporadic cases	*AGAP1* 1733insCAGG	0,22	0,20	1,10	0,82–1,46	0,53
	ASB1 239344663T>A	0,19	0,17	1,16	0,83–1,61	0,39
	COPS8 149C>T	0,02	0,01	1,49	0,51–4,33	0,46
	***IQCA1* 237247036insA**	**0,33**	**0,51**	**0,47**	**0,37–0,59**	**$7{,}64 \times 10^{-11}$**
	MYEOV2 241075809C>T	0,50	0,51	0,94	0,74–1,20	0,63
	MYEOV2 241075991A>C	0,49	0,48	1,02	0,80–1,31	0,86
	RAB17 258T>G	0,08	0,08	1,09	0,69–1,73	0,72
	RAMP1 238767662C>G	0,20	0,19	1,04	0,77–1,41	0,80
	SCLY-UBE2F 239002480delTTG					
	TRAF3IP1 536G>A	0,01	0,02	0,72	0,27–1,89	0,50
Familial cases	*AGAP1* 1733insCAGG	0,23	0,20	1,16	0,81–1,67	0,41
	ASB1 239344663T>A	0,20	0,16	1,31	0,87–1,96	0,20
	COPS8 149C>T	0,01	0,01	0,77	0,17–3,46	0,73
	***IQCA1* 237247036insA**	**0,35**	**0,51**	**0,52**	**0,38–0,70**	**$1{,}61 \times 10^{-5}$**
	MYEOV2 241075809C>T	0,48	0,49	0,94	0,70–1,28	0,71
	MYEOV2 241075991A>C	0,49	0,48	1,03	0,75–1,40	0,88
	RAB17 258T>G	0,06	0,08	0,80	0,43–1,49	0,48
	RAMP1 238767662C>G	0,21	0,19	1,09	0,75–1,59	0,66
	SCLY-UBE2F 239002480delTTG					
	TRAF3IP1 536G>A	0,02	0,02	0,97	0,32–2,92	0,96

Table 9. Results of TaqMan and KASPar assay comparison.

Mutation	Call	TaqMan (n)	KASPar (n)	Concordance
RAMP1 238767662C>G (rs3754699)	C	935	945	98 %
	G	61	61	
	Both	486	489	
	Undetermined	23	10	
	Total	1505	1505	
	Callrate	98 %	99 %	
TRAF3IP1 536G>A (rs61742338)	G	1822	1785	97 %
	A	0	0	
	Both	55	49	
	Undetermined	22	65	
	Total	1899	1899	
	Callrate	99 %	97 %	

6 Discussion

The aims of this study were to evaluate the association of the chosen markers with prostate cancer, define the association of these mutations with serum PSA values of prostate cancer patients and to compare TaqMan and KASPar genotyping technologies to validate KASPar assays as possible replacement of TaqMan genotyping in future use. The experiments showed that *IQCA1* 237247036insA was associated with both sporadic and familial forms of PrCa. The mutation was found to have a protective effect as individuals carrying the mutation were at lower risk of developing the disease. *IQCA1* 237247036insA associated also serum PSA values in a manner that homozygous individuals had lower PSA values. Another mutation, *SCLY-UBE2F* 239002480delTTG associated with familial form of the disease but an error in genotyping was detected, so the results require further confirmation. TaqMan and KASPar assays were found to yield results that were well in concordance.

6.1. Evaluation of the used methods

HRM. HRM as mutation screening and genotyping method has been praised as a cost-effective method to reduce the need for direct sequencing, which is still the golden standard of genotyping. With HRM, SNPs as well as indels can in theory be detected. Indeed, it might be an effective means to screen for new mutations as the sample amounts that can be handled with minimal effort are much greater than in sequencing: in a simple one-step PCR run, 384 samples can be handled at once. When compared with sequencing which requires the amplification of target DNA by PCR, purification of the PCR products, sequencing PCR and precipitation of the sequenced products, the hands-on time and effort are indeed minimal. HRM is also an effective genotyping method for some mutations that have been previously characterized. However, during this study, several problems concerning HRM were encountered.

To start from the beginning, the designing of HRM primers may be hard, or even impossible in some cases. The method is so sensitive that the primers must be excellent to function. One mutation that was planned to be included in this study had to be excluded because the area around the mutation was so homologous with other parts of the genome that no primers could be designed – the best two pairs were ordered but neither worked. HRM is also very sensitive to the amount of template and the amount of reaction mix in each reaction. Very slight pipetting errors can affect the results. In this study, the reactions were always run in duplicates to ensure better specificity, but this did not always work either. Very often aberrant melt curves were obtained, and these had to be sequenced to determine the genotype with certainty. Also, the type of a mutation cannot be recognized by the sole

inspection of the HRM curves, so confirmation sequencing is required to identify the variation unambiguously. As advantage, HRM products can usually be purified and used as template in sequencing.

The greatest problem of HRM was encountered was the capability of the method to distinguish between different genotypes. The system functions precisely for some mutations, i.e. *AGAP1* 1733insCAGG could be genotyped with certainty by HRM, and the different genotypes stood out clearly. For the two other mutations genotyped with HRM, the situation was not as clear. *IQCA1* 237247036insA is an insertion of one A after a repetition of nine other A:s, and HRM has problems distinguishing these kinds of variations where the number of long run of same bases is altered by one. The differences between the three genotype clusters were sometimes minimal, and lots of confirmation sequencing had to be performed.

For *SCLY-UBE2F* 239002480delTTG, the situation was even more challenging. Heterozygous samples stood out clearly as one cluster, but wild-types and rare homozygotes were clustered together without any difference in melting curve shapes or heights. Confirmation sequencing of some samples from both clusters confirmed that the rare homozygotes do exist, but they could not be distinguished from the wild-types. As a result, there exists a clear genotyping error for *SCLY -UBE2F* 239002480delTTG, since the two homozygote genotypes are all seen as wild-type. This is probably partly the reason why this mutation reached significant associations in all analysis modes and the result is in all probability a false positive. The seeming lack of rare homozygotes is likely also the reason for the Hardy-Weinberg disequilibrium detected for this mutation. The reason why HRM could not detect the mutation remains unclear. If HRM is capable of detecting single-base substitutions, it should be able to see a three-base deletion as well. As a three-base deletion the mutation is such a large change that it should affect the melting temperature of the DNA and hence be detected by HRM. However, several Tm calculators were tested to evaluate the melting temperature of the amplicons with and without the mutation. All calculators gave results that suggested that the amplicon has the same melting temperature whether it has the mutation or not. This would explain why the heterozygotes did not stand out from the wild-types in HRM, but the next question is, how can a 200 bp DNA strand have the same melting temperature as a 197 DNA strand? The only explanation that was presented was that an insertion occurs somewhere in the amplicon and somehow negates the effect the TTG deletion has on the melting temperature. This could not be confirmed, however. Furthermore, this does not explain the fact that the Tm calculators could not see a difference between the two amplicons. The difficulties of HRM in detecting insertions and deletions has also been described in literature (Wittwer, 2009) and one possible solution would be to mix the samples to be genotyped with equal amount of

known wild-type DNA so that during cycling, the homozygous variants are converted to heterozygous and their melting curves could be detected more easily (Wittwer, 2009).

Overall, HRM as a method might be effective for genotyping of some mutations, and it probably has earned its good reputation in the literature. The system is clearly vulnerable, though, as it cannot detect all kinds of mutations. When one is considering using HRM as genotyping method, an initial test run and following confirmation sequencing might be in order.

Direct sequencing. Direct sequencing was for a long the only method to genotype DNA samples and really detect what variations are present. Although it is the golden standard of genotyping, the method is not infallible. The course of sequencing requires many phases that are done manually, increasing the risk of errors and contaminations. Even though most of the work phases are done in a laminar to reduce the risk of DNA contamination, sometimes contamination events cannot be avoided. One major source of contamination is mixing of the samples that are on the same 96-well plate, resulting in sequence from negative control sample. Sequence obtained from a negative control can be an indicator of more extensive mixing of the samples on the original sample plates and always requires a more careful evaluation. Sometimes the quality of template DNA affects the reaction and the resulting sequence is not of good quality, in which case the making of genotype calls might be difficult. As a method, direct sequencing is time-consuming and expensive, so it is no wonder that extensive research has been concentrating on finding possible replacements to reduce the need for sequencing. In this study, direct sequencing was used mainly in parallel with HRM to confirm the genotypes of different clusters and samples with aberrant melting curves.

TaqMan® and KASPar. TaqMan- and KASPar assays were perhaps the most effortless and reliable methods used in this study. The preparing of the PCR reactions was easy as the DNA samples could be pipetted by automated robot, and 384 samples could be genotyped in about two hours per each available thermal cycler. The problems encountered with these methods were mostly due to the pipetting robot. The Freedom EVO® 75 pipetting robot (Tecan Group Ltd., Männerdorf, Swizerland) is sensitive to the liquid level present in the source container, in this case the 96-well plate containing the DNA samples to be pipetted to 384-well plate containing the genotyping master mix. If the amount of liquid in a well is lower than 30 µl, the robot might not be able to draw a sufficient amount of template DNA and the subsequent PCR reaction may fail. This can be seen during plate reading as samples that cluster together with negative controls and thus have an undetermined genotype. Some problems not related to liquid level in source containers were also encountered with the pipetting robot. In some cases, a tip in certain position had poor attachment, resulting again in insufficient amount of template DNA is PCR reaction mix and undetermined genotype. These cases could be usually recognized by inspection of the plate reading file. If the undetermined samples were located on

the plate in a pattern related to the pipetting pattern, any other mistake in genotyping could be excluded. The pipetting robot was used despite its occasional problems, for in the longer run it is still more accurate than pipetting the samples manually, not to mention the time and effort saved. The undetermined genotypes might also result from the poor quality of template DNA or differing amounts of PCR master mix in wells. After all, the PCR master mix was pipetted to the plates by a manual pipettor, which leaves opportunities for human error.

6.2. Comparison of TaqMan® and KASPar assays

The results obtained using KASPar assays were as good as with TaqMan assays and the two different genotyping technologies were well in concordance, i.e. the genotype calls for each sample were the same with both systems. A recent study by Pruvost *et al*. (2012) compared pyrosequencing and KASPar assays in SNP genotyping and reached similar results, concluding that the one-step protocol of KASPar genotyping and its cost-effectiveness is advantageous in large-scale genotyping. It seems that the novel technology of KASPar assays functions and is reliable in the genotyping of different SNPs. However, one limitation of KASP genotyping is that it cannot be used to detect indels. We encountered this limitation when designing this study, and Pruvost and coworkers reported the same problem. This is a big limitation as small indels make up a large part of mutations that must by genotyped e.g. in fine-mapping of chromosomes.

The use protocol of KASPar assays was as easy as the protocol of TaqMan system: a master mix of reaction mix and assay mix was prepared and pipette to a 384-well plate, sample DNA was added and then thermal cycling and subsequent plate reading could be performed. The genotyping by these methods is easy and fast, enabling the genotyping of 384 samples in about two hours per each available thermal cycler. Thus, the everyday lab work is as fluid with KASPar assays as with TaqMan assays.

The ordering protocol of KASPar assays is, however, not as easy as the ordering of TaqMan assays. KBioscience does not have a “shopping cart” kind of ordering system in their website that would require the customer to fill out all important details of their order, and all orders have to be made via e-mail. The result often some essential piece of information is forgotten and multiple e-mails have to be exchanged before an order is successfully placed. They do have an ordering form that can be downloaded from their website, but even though the form is used, an exchange of e-mails was required. To make ordering of the products easier and hence improve the simplicity of using their products, KBioscience should improve the ordering system. Furthermore, the delivery times are

slightly longer for KASP assays. The standard turnaround delivery time is four to six weeks. A faster turnaround time (delivery in two to three weeks) can be chosen but the faster delivery time increases the cost of an assay by 50 %. However, the assays are even up to 36 times cheaper than TaqMan assays, so even with faster delivery time, the costs probably remain smaller. Despite the minor inconveniences of a more complicated ordering protocol and slightly longer delivery times, any laboratory performing large-scale genotyping with TaqMan assays might want to overlook these factors as the savings made when using KASPar genotyping can be substantial.

Another con for the KASPar assays was that a couple of times the manufacturer forgot to send a customer conditions form along with the assays. The customer conditions form is essential for the performance of the assays, as it gives the ready-optimized cycling conditions for the assay in question, plus tells you that which allele is labeled with which fluorophore, so that the genotypes can be assigned correctly during plate reading. The form was obtained via e-mail after an enquiry to KBioscience, so no harm was done. However, this was another exchange of e-mails that could have been easily avoided and that delayed the performing of the assays. In conclusion, the KBioscience's KASPar assays function well, are cost-effective and easy to perform but the practical aspects of communicating with KBioscience, such as order placement, should be improved so that the convenience of using these assays would be more convenient.

6.3. Relevance of the studied genes and their variants

Bioinformatic tools are convenient in the initial analysis of promising mutants. The mutations chosen for this study were analyzed with predictive programs. All of them were found to be possibly pathogenic according to these predictions. Bioinformatic tools can get as so far, but at some point one must stop staring at the outcomes of such analyses and turn to the biological relevance of the genes in which the mutations are found.

AGAP1. The gene product of *AGAP1*, AGAP1 protein is the first identified member of ASAP family ADP-ribosylation factor GTPase-activating proteins (Arf GAPs). The Arfs are a subfamily of GTP binding proteins within the Ras superfamily and in normal cell physiology they regulate membrane traffic and actin cytoskeleton dynamics (Nie *et al.*, 2002). Arfs require the controlled binding and hydrolysis of GTP for efficient function. AGAP1 functions in the endosomal-lysosomal system, more specifically in the Golgi apparatus, regulating Arf1 as GTPase-activating protein (Nie *et al.*, 2002). AGAP1 induces and localizes to Rab4/AP1 containing structures presumed to be endosomes and specifically alters stress fibers (Nie *et al.*, 2002). Members of the Ras family have been

shown to be frequently mutated in human cancers and are hence also called Ras oncogenes (Bos, 1989). Thus, it is not impossible that AGAP1, as a Ras family member, could be associated with cancer in general. *AGAP1* has previously been studied as an autism susceptibility gene (Wassink *et al.*, 2005) and it has also been shown to be expressed in human small cell lung cancer cells (Cuttitta *et al.*, 1988). Furthermore, AGAP2, a close relative of AGAP1, has been linked to promotion of prostate cancer progression (Cai *et al.*, 2009). This supports the view that also AGAP1 could be associated with PrCa. However, in this study, no association was found with either sporadic PrCa or familial form of the disease; the frequency of *AGAP1* 1733insCAGG was the same in cases versus controls. The mutation is located in coding sequence in one transcript of the gene and the insertion is large (four bases), causing truncation of the gene product due to premature stop codon. It seems likely that the mutation should have some pathologic effect, at least in homozygous individuals. The pathological effect might be manifested as something else than prostate cancer.

ASB1. Ankyrin repeat and SOCS box protein 1 (ASB1) is a member of the ankyrin repeat and suppressor of cytokine signaling (SOCS) box family of proteins. These proteins belong to the suppressor of cytokine signaling protein superfamily (Kohroki *et al.*, 2005). The ASB family proteins have been shown to bind accessory proteins and together this complex functions as an E3 type ubiquitin-ligase (Kohroki *et al.*, 2005). The ubiquitin-proteasome pathway is one of the most important protein degradation pathways of cellular proteins involved in cell-cycle progression and signal transduction pathways, and proteins destined for this degradation pathway are targeted to it by ubiquitinylation. The targeted degradation of proteins controls many cellular processes and alterations in the ubiquitin pathway are associated with human diseases, including cancer (Guardavaccaro and Pagano, 2004). The association of ASB proteins with PrCa has not been previously investigated, at least not to a larger extent. The results of this study suggest that *ASB1* 239344663T>A is not associated with PrCa. One explanation for this might be that the other members of ASB family replace the function of a dysfunctional or inadequate ASB1 in cells. The ASB family is very conserved and the proteins have overlapping functions, thus enabling the replacement of a missing member with another. Even though *ASB1* 239344663T>A is not associated with PrCa, other mutations in the gene or in other members of the family might emerge. ASB6, ASB8 and ASB9 have been shown to be related with oral carcinoma, lung carcinoma and colorectal cancer, respectively (Da Silva, 2011), so it might be likely that associations with other cancer types should be found. In addition, the ASB1 mutation included in this study is a single base substitution in an intronic region of the gene. Even though it is a transversion from pyrimidine to purine, the effects probably remain small in the intronic region and it is not perhaps unexpected that it does not have any significant effect.

COPS8. COPS8 is the subunit number 8 of the COP9 signalosome (CSN), a highly conserved protein complex that functions as an important regulator of many signaling pathways. CSN is essential for the function of a subtype of E3 ubiquitin ligases, the Cullin RING ligases (CRL) (Choo *et al.*, 2011). The activity of CRLs is controlled by the ligation and removal of an ubiquitin-like protein Nedd8 (neddylation and de-neddylation, respectively) (Choo *et al.*, 2011). CSN functions in the de-neddylation of CRLs, thus regulating the CRL activity and hence the E3 ligase-dependent protein degradation pathway (Choo *et al.*, 2011). The COP9 signalosome has been suggested to play a role in cancer, as the products of many tumor suppressor and oncogenes are regulated by ubiquitination- and proteasome-mediated protein degradation (Lee *et al.*, 2011). Especially subunit 5 might regulate oncogenic and tumor suppressive functions independently of, or coordinately with, the CSN holoenzyme (Richardson and Zundel, 2005). No association between *COPS8* 149C>T and PrCa was found in this study. Perhaps the subunit 8 is not most essential in the function of the holoenzyme and other subunits confer the pathogenic properties. The mutation in *COPS8* is a C to T transition, which is a quite small change as the only thing that distinguishes these two bases is the amino group of cytosine in the 4-carbon position compared with the keto group of thymidine. The mutation was very rare with heterozygote frequency of 3 % and homozygote only 0,1 %. This suggests that the area where the mutation is located is under effective repair systems and most mutations are corrected. Areas like this are usually very important to the function of the protein and hence conserved. Being located in the 5' UTR, this area might represent an important regulatory element.

IQCA1. The product of *IQCA1* is a protein IQ and AAA domain-containing protein 1. As the name suggests, it contains a calmodulin-binding IQ domain and an AAA+ (ATPases Associated with a wide variety of cellular Activities) domain. Calmodulin is a major calcium censor in cells, regulating many cellular proteins via direct interaction with them (Rhoads and Friedberg, 1997). IQ domains are evolutionarily conserved, calcium-independent calmodulin binding motifs that are generally found in myosins, but have been recognized in a wide variety of other proteins that have not been demonstrated to bind calmodulin; these proteins include e.g. calcium-independent subfamily of protein kinase C, sodium channel proteins, and multidrug resistance protein (Rhoads and Friedberg, 1997). The AAA+ domain containing proteins form a large and functionally diverse superfamily of NTPases (Snider *et al.*, 2008). The AAA+ proteins are involved in variety of cellular activities in which the energy of ATP hydrolysis is used to molecular remodeling events, such as protein unfolding and degradation, peroxisome biogenesis, bacteriochlorophyll biosynthesis and DNA recombination, replication and repair (Snider *et al.*, 2008). One AAA+ domain containing protein, ATPase family AAA domain containing 3A (ATAD3A), has been suggested an anti-apoptotic factor and a PSA secretion regulator in PrCa (Huang *et al.*, 2011). Considering this and the important functions of

the AAA+ domain proteins in processes like DNA repair, it seems likely that mutations in these proteins are associated with pathological effects. Indeed, in this study, *IQCA1* 237247036insA was found to be significantly associated with PrCa in both allelic association analysis and case-control analysis (Tables 7 and 8, respectively).

IQCA1 237247036insA was found to be in Hardy-Weinberg equilibrium in sporadic cancer cases and controls, but not in familial cases. The equilibrium in controls suggests that the association with PrCa is real and not a false positive resulting from a mistake in genotyping. The disequilibrium observed in familial cases could result from the relatively small sample size (n=188). However, the disequilibrium in familial cases might also suggest that one allele has some kind of selective advantage over the others in familial prostate cancer.

The case-control analysis suggests that the mutation has a protective effect from prostate cancer, as the mutation is more common in controls than in cases. Hence, the individuals having the mutation are at lower risk of developing PrCa. The analysis of the mutation in respect to serum PSA values seems to support this view: persons having *IQCA1* 237247036insA as homozygous mutation have lower serum PSA values than wild-type individuals or heterozygotes. In light of this result, it seems that like ATAD3A, IQCA1 protein could also function as PSA secretion regulator. The role of *IQCA1* in PrCa has not been previously studied, but the very significant p values of the gene's association with PrCa support further investigations.

MYEOV2. MYEOV2 is a gene encoding a protein myeloma overexpressed 2 protein (NCBI Gene/MYEOV2). The protein is a helicase/primase complex (NCBI Gene/MYEOV2)). Helicases are a group of enzymes that participate in processes such as DNA replication and repair by separating the two strands of DNA from each other by utilizing the energy of ATP hydrolysis. Primases on the other hand are enzymes that catalyze the synthesis of a RNA primer so that DNA replication has a template to proceed on. The protein has two isoforms resulting from alternative splicing (NCBI Gene/MYEOV2). The function of this gene has not been investigated as no articles about the function or other aspects, such as possible pathogenesis, could be found in PubMed or other databases. However, considering that the protein is a helicase/primase complex, it supposedly has a great role in the maintenance of genomic integrity, as DNA repair and replication are essential for the survival of cells. It is impossible to say how important this one helicase/primase is. However, two mutations found in *MYEOV2* and located some 200 bp apart in the 5' UTR of the gene, *MYEOV2* 241075809C>T and *MYEOV2* 241075991A>C were validated in this study. Both were quite common in Finnish population: only about 25 % of wild-types were present for both mutations. Neither mutation was found to be associated with PrCa. *MYEOV2* 241075809C>T is again a C to T transversion, a quite small change as discussed earlier. *MYEOV2* 241075991A>C on the other hand is a tranversion where a

purine is converted to a pyrimidine. This is a bigger change in base composition, so it could affect the function of the gene. However, the effect, if present, is probably something other than effect on PrCa susceptibility.

RAB17. The product of the *RAB17* gene is RAB17, member of the Ras oncogene family protein. The protein is a member of the rab subfamily of small GTPases localized in epithelial cells (Lutcke *et al.*, 1993). It has an important role in the regulation of membrane trafficking in eukaryotic cells (Lutcke *et al.*, 1993), most probably in transcytosis (Hunziker and Peters, 1998). As a Ras oncogene family member, RAB17 might also have some indications for cancer pathogenesis. The studied mutation, *RAB17* 258T>G is a pyrimidine-to-purine transversion located in the 5'UTR of the gene. RAB17 has not been previously indicated to have some association with PrCa. In a study by Schumacher *et al* (2011) a mutation in *MLPH* located very near *RAB17* was found to be associated with PrCa, and *RAB17* was found to be in linkage disequilibrium with the *MLPH* mutation. This might suggest that also *RAB17* could have some influence to PrCa predisposition, maybe via interaction with *MLPH*. However, in this study, the *RAB17* 258T>G mutation was not associated with PrCa. The mutation was validated in a smaller sporadic case sample set than the other mutations as the genotyping method was direct sequencing which is time-consuming and expensive. This might affect the power of the statistical analysis. Also, the sequenced ~400 bp area around *RAB17* 258T>G was found to harbor four other mutations in addition to the mutation of interest, all single-base substitutions. This indicates that the gene is for some reason a hotspot for mutations. The other mutations in this area remain to be analyzed as they are not included in this study; future analysis of these mutations might still reveal *RAB17* as a PrCa susceptibility gene.

RAMP1. Receptor (G-protein coupled) activity modifying protein 1 (RAMP1) is a member of the RAMP family of receptor (calcitonin) activity modifying proteins. RAMPs are type I transmembrane proteins with one transmembrane domain, an extracellular N-terminus and a cytoplasmic C-terminus (NCBI Gene). RAMPs regulate the transport and ligand specificity of calcitonin-receptor-like receptor (CRLR), which can function as receptor for calcitonin-gene-related peptide (CGRP) or adrenomedullin, depending on the RAMP proteins associated with it (McLatchie *et al.*, 1998). CGRP and adrenomedullin are related peptides that have distinct pharmacological profiles: CGRP functions in the central nervous system, regulating neuromuscular junctions, antigen presentation within the immune system, vascular tone and sensory neurotransmission; adrenomedullin an effective vasodilatory agent (McLatchie *et al.*, 1998), but it has also been shown to promote cell growth and survival and angiogenesis (Hay *et al.*, 2010). When associated with RAMP1, CRLR is presented to the cell surface as a CGRP receptor (McLatchie *et al.*, 1998).

CGRP and adrenomedullin can act as growth or survival factors for a number of tumors via e.g. stimulation of angiogenesis in endocrine-related cancers (Hay *et al.*, 2010). Furthermore, adrenomedullin receptors have roles in tumor growth and progression (Hay *et al.*, 2010). RAMP1 could have a role in cancer development via its regulation of the receptors of CGRP and adrenomedullin. In a study by Romanuik *et al*. (2009) *RAMP1* expression was found to be significantly decreased in metastatic castration-recurrent PrCa compared with androgen-dependent primary prostate cancer (Romanuik *et al.*, 2009). However, in a study conducted in mice by Toda *et al*. (2008), the inhibition of CGRP function with gene knock-out or antagonist infusion slowed Lewis lung carcinoma tumor growth (Toda *et al.*, 2008). This seems quite inconsistent with the findings of Romanuik *et al*.: if RAMP1 expression is decreased in prostate cancer, shouldn't the amount of CGRP be also decreased in PrCa cells, leading to the inhibition of tumor growth? The study by Toda *et al*. was performed with another cancer type, so perhaps the results are not directly comparable. Another possibility is that the decrease in RAMP1 levels in CRPC causes the CGRP receptor to associate with greater probability with other RAMP proteins and be presented to cell surface more often as an adrenomedullin receptor, hence stimulating the advancement of the disease to the castration-resistant mode. *RAMP1* 238767662C>G was not found to be associated with PrCa in this study.

***SCLY/SCLY-UBE2F*.** The *SCLY-UBE2F* gene represents a natural readthrough transcription event, where the neighboring selenocysteine lyase (SCLY) and ubiquitin-conjugating enzyme E2F (UBE2F) are transcribed into one mRNA. The mutation of interest actually lies in the area of *SCLY* gene, but since the *UBE2F* gene is located upstream of the *SCLY* gene, the mutation is also present in the readthrough transcript. UBE2F is again an ubiquitin ligase that conjugates Nedd8 to target proteins (Huang *et al.*, 2009). SCLY on the other hand catalyzes the conversion of the amino acid L-selenocysteine to L-alanine and elemental selenium (Mihara *et al.*, 2000). Selenium has been suggested to protect from PrCa. The role of SCLY in PrCa could hence be mediated through its participation in selenium metabolism. UBE2F, being an ubiquitin ligase, would probably have similar functions as ASB1, functioning in the proteasome-degradation pathway. The mutation found in *SCLY-UBE2F* is a three-base deletion, so the mutation is quite large.

In this study, *SCLY-UBE2F* 239002480delTTG was found to be associated with familial PrCa ($p=1{,}50 \times 10^{-4}$, Table 7). For some reason, the PLINK tool could not perform a case-control analysis of the obtained data for this mutation. This is probably due to some problem in the genotype frequencies as no rare homozygotes could be recognized with HRM even though confirmation sequencing showed that they do exist. The mutation in SCLY-UBE2F was found to be in Hardy-Weinberg disequilibrium in both cases and controls. This is probably also caused by the lack of rare homozygotes. As conclusion, a clear error exists in the genotyping of this mutation and the results should be regarded

with a reasonable doubt. To confirm the results of this study, the mutation should be genotyped with another method that reveals the rare homozygotes. This was impossible to do within the timeframe of this study. On the other hand, one could argue that if the heterozygotes alone confer significant association to PrCa, the association might be even more significant if the rare homozygotes could be distinguished. The effect might be also inverse, so only future studies will reveal if this mutation really is associated with PrCa.

TRAF3IP1. TNF receptor-associated factor 3 interacting protein 1 or TRAF3IP1 (also known as microtubule-interacting protein associated with TRAF3 or MIP-T3) is an evolutionarily conserved protein that interacts with actin filaments and microtubules (Guo *et al.*, 2010). These interactions suggest that the protein plays a role in the regulation of cytoskeleton dynamics (Guo *et al.*, 2010). Tumor necrosis factors (TNFs) are a group of cytokines that cause apoptosis in cells and mediate their effects to cells via TNF receptors (TNFRs) or death receptors. The signal transduction from TNFRs is mediated by the TNF receptor-associated factors (TRAFs). TRAF3IP1 is a protein that interacts with TRAF type 3. Considering that TRAF3IP1 interacts with TNF receptors that regulate apoptosis, it seems reasonable that mutations in the gene could impair the interaction and hence affect the function of TNFRs and this way somehow inhibit apoptosis and aid the progression of tumor growth. *TRAF3IP1* 536G>A was not found to be associated with PrCa in this study. The 536G>A mutation was very rare in the study population: no homozygotes were found and only 2,6 % of study subjects were heterozygous. Like the mutation in *COPS8,* also *TRAF3IP1* 536G>A seems to be under careful observation and repair. This is perhaps not surprising as the mutation is located in the coding sequence of the gene. Exonic regions are generally under more effective observation than intronic areas because mutations in them would probably have dire consequences. The mutation, if present, causes a missense mutation that changes arginine to glutamine. Arginine has a positively charged side-chain while glutamine is neutral. This change might have some effect on the protein structure: by removing a charged amino acid, the electrochemical properties of that area in the protein might change, resulting in changes in protein structure and function. For example, if the positive charge of arginine interacts with a negative change of another amino acid nearby, the disruption of this interaction might weaken the tertiary structure of the protein and hence alter the functional properties. If the change alters the protein function radically to non-functional direction, a negative selective pressure might be targeted towards the mutant genotype as the individuals with the mutation are at a disadvantage of survival due to a less functional protein. Both of the amino acids are polar, so the hydrophobicity profile of the protein is probably not changed. The effects of this mutation to the structure and function of the encoded protein were, however beyond the interest of this study.

6.4. Study population

The Finnish population has traditionally been considered a good study material for genetic studies, especially in cancer epidemiology. First of all, Finland has good registries, of population in general (Population Register Center) and of cancer patients (Finnish Cancer Registry). The registries makes it easy to obtain information about study subjects as the relations of family members and the characteristics of cancer cases can be confirmed from trustworthy sources. Second, we are considered a genetically homologous population due to a small founder population and long geographical isolation. However, evidence suggesting that Finns are not as genetically homologous as previously thought is emerging. According to the recent findings, Finns are divided into several subpopulations that differ genetically from each other quite a lot. The divide is especially clear in east-west axis of the country, with western Finns belonging to a different population cluster as the eastern Finns (Hannelius *et al.*, 2008; Jakkula *et al.*, 2008). The genetic homologousness of a population affects the results of e.g. linkage analyses and GWASs, both important tools for genetic epidemiology research. The results of these analysis modes are dependent on the population and analyses performed in different populations yield different results. This has been especially clear in linkage studies: the linkage peaks obtained in different studies have been observed in different chromosomes and the results have been poorly reproducible.

This study was performed using a samples from sporadic prostate cancer cases obtained from the Pirkanmaa area and familial samples from Finnish PrCa families all over the country. The Pirkanmaa area belongs to Western Finland and the people living in the area are a part of the western subpopulation cluster. As the sporadic samples are collected from the same area, they can be considered good sample material for association analysis. The amount of sporadic cases (n=1505) should also be adequate to detect any association of the studies SNPs with sporadic PrCa. Familial cohort is more challenging. These represent samples from the youngest affected member of Finnish prostate cancer families with three or more affected members. To obtain two full 96-well plates of familial samples for analysis, some samples from families with only two affected members were included. Still, the amount of familial samples remains quite low, as the total number of samples is 188. This is quite a small set, and it may not be adequate to detect statistically significant association, especially with mutations that are rare. In a small sample set, the statistical power might not be strong enough to detect association and the effect of chance remains large. However, the situation could not be improved as no more familial samples were available at the time of the study.

Another question arises from the adequacy and quality of used controls. The number of controls used in this study was 162, which is a quite low amount, especially when compared with the

amount of the sporadic cases. The small amount of controls leaves a lot space for the effect of chance when studying the number of SNPs of interest in these samples. Perhaps the results of this study could have been improved by using more controls. Another consideration is the area from which the control samples were collected. When performing association analyses, the control samples should be collected from the same region as cases, so that the sample individuals are genetically close to each other. Otherwise the controls could represent genetically another subpopulation. The control samples were obtained from Finnish Red Cross blood donors from the cities of Kuopio and Turku. Neither city is located in the Pirkanmaa area, from where the cases were. Especially Kuopio is located so far east from Pirkanmaa that the controls probably belong to another subpopulation cluster, affecting the results. Another aspect concerning the quality of the controls is that they were obtained from blood donors that were healthy at the time of donation. This does not rule out the possibility that these people will develop prostate cancer later on, especially when Finnish blood donors are 18–65 years old and PrCa is usually diagnosed over the age of 65. As long as no prediction tools for PrCa susceptibility exist, one cannot be entirely certain that the controls are truly negative controls without prostate cancer.

6.5. This study in respect to other studies

Although other parts of the chromosome 2 have been investigated in prostate cancer and some susceptibility variants have been recognized (Gudmundsson, 2008; Liu *et al.*, 2009), the 2q37.3 chromosomal area has not been extensively studied in this regard. The area was found to be linked with prostate cancer only recently (Cropp *et al.*, 2011), and hence not many studies have yet been published. Hence, this study is among the pioneering studies in this area. A study by Schumacher *et al*. (2011) is thus far only other study that has investigated this chromosomal locus. In the study in question, a SNP (rs2292884) located in the gene *MLPH* reached significant association with prostate cancer (Schumacher *et al.*, 2011). This gene is located very close to some of the genes that were of interest in this study (Figure 5).

Figure 5 also shows that in the study by Schumacher *et al*. some SNPs located in the genes of this study were genotyped, and for certain *RAB17* SNPs, mild linkage disequilibrium with rs2292884 mutation was established. This means that the *RAB17* SNPs in linkage disequilibrium with rs2292884 are inherited together with rs2292884 and could also be associated with PrCa. The confirmation of chromosome 2q37.3 as a PrCa susceptibility locus by Schumacher et al. reinforces the suspicions that this locus does indeed harbor something really interesting when it comes to prostate cancer. This locus

is currently under vigorous investigation and future years will show if any real candidate genes are found.

The *MLPH* mutation that rose to attention in the study by Schumacher *et al*. (2011) was not found in deep sequencing of the genomes of Finnish prostate cancer families that showed strong linkage to the 2q37 region (Rantapero, 2012). However, the same gene seems to harbor two other mutations in five Finnish PrCa families (Rantapero, 2012). Had the article by Schumacher et al. been published earlier, the mutations would have probably been included in this study. Now they remain in the future studies that concern the 2q37 area.

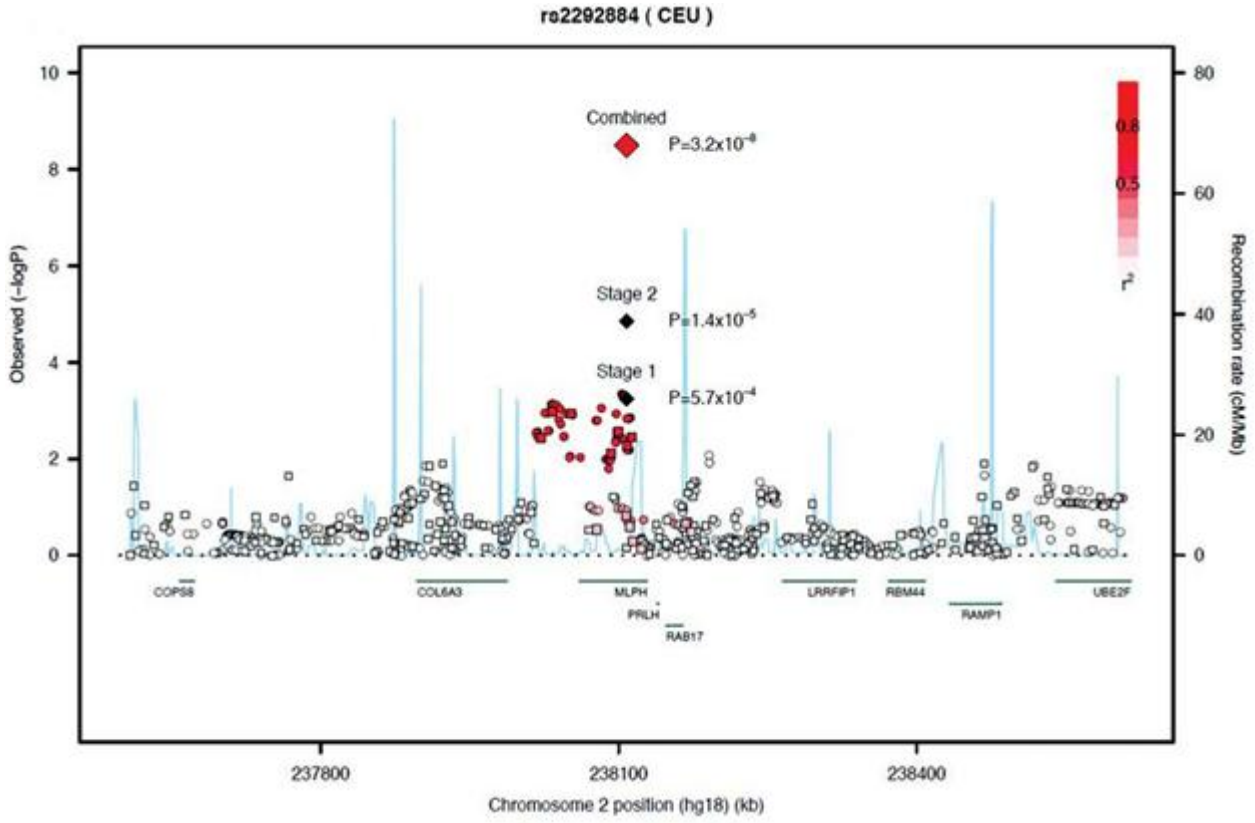

Figure 5. Regional association plots PrCa loci found by Schumacher et al. (2011). Plot shows the genomic regions of association with overall PrCa risk on chromosome 2q37.3. Shown are the –log10 association P-values of genotyped (square) and imputed (circle) SNPs in 2782 advanced PrCa cases and 4458 controls in Stage 1. The –log10 association P-values for the index SNP in Stage 1, Stage 2 and combined is shown (diamond). The intensity of red shading indicates the strength of linkage disequilibrium with the index SNP. Also shown are the SNP build 36 coordinates in kilobases (kb), recombination rates in centimorgans (cM) per megabase (Mb) (in blue) and genes in the region (in green). Adapted from Schumacher *et al*. 2011.

6.6. Future perspectives

The chromosomal area 2q37.3 has proven to be an interesting locus in respect to prostate cancer susceptibility. It has recently risen to attention in linkage studies and association analyses alike (Cropp *et al.*, 2011; Schumacher *et al.*, 2011), and further studies are ongoing (Laitinen *et al*., unpublished data). As the study by Schumacher *et al*. revealed a PrCa susceptibility SNP in 2q37.3 and this study unveiled two possible mutations more, the vigorous research is justified. The deep sequencing of 2q37.3 linkage region performed with Finninsh PrCa families was found to harbor many mutations (Laitinen *et al*., unpublished data), of which only a small fraction was chosen for this study. It is likely that some among the remaining mutations will prove to be mutations conferring susceptibility to PrCa as well. Some of the remaining mutations are planned to be sent to the Institute for Molecular Medicine Finland (FIMM) for genotyping to see if any association with PrCa can be found. The two mutations that reached significant association in this study are also included among the SNPs to be genotyped, so the results shall be confirmed.

It is interesting that both of the mutations that showed any association with PrCa in this study were located in intronic areas. In the past, the research has concentrated on the exonic regions. This is of course rational in the sense that there the exonic regions are significantly smaller that intronic regions. However, as this study points out, some of the important mutations might be found in intronic areas as well. Intronic regions might contain important regulatory elements that affect the expression of the gene via splicing cite changes or other methods. In the future, the mapping of intronic and regulatory areas should be increased if we really want to deduct the susceptibility loci for PrCa as well as other diseases.

This study concentrated only on the association of one mutation with prostate cancer. One interesting possible future direction for research would be to analyze whether some prostate cancer patients have more than one of the studied mutations in their genomes and to see if the presence of more than one mutation can be linked to prostate cancer susceptibility. According to present view, the interaction of many common, low-penetrant mutations might be an important factor in conferring PrCa predisposition.

7 Conclusions

The purpose of this study was to validate chromosome 2q37.3 mutations and see if they were associated with prostate cancer risk in Finnish population. The association with sporadic and familial prostate cancer was analyzed. Also, association of promising mutations with serum PSA level was evaluated. Secondary aim of this study was to validate novel competitive allele-specific PCR system KASPar as a possible replacement for widely used TaqMan genotyping technology.

All of the mutations chosen for this study were found to exist in the Finnish population and hence they were validated as real mutations and not some artifact of deep sequencing. Most of the mutations were quite common in the Finnish population. Of the studied mutations, two were associated with prostate cancer: *IQCA1* 237247036insA associated with both sporadic and familial PrCa while *SCLY-UBE2F* 239002480delTTG was associated with only familial form of the disease. *IQCA1* 237247036insA was more common in controls than in cases which suggests that individuals with the mutation have a lower risk of developing PrCa. There exists an uncertainty concerning the results: both mutations were genotyped using HRM, which proved to be less than optimal method for these two mutations. *IQCA1* 237247036insA is a one A insertion after a long run of adenines, and HRM documentetly has problems distinguishing between genotypes in these cases. In this case, some samples might have been interpreted wrong. For *SCLY-UBE2F* 239002480delTTG the situation was even worse as the rare homozygotes could not be distinguished from wild-types even though confirmation sequencing showed that they exist. To really ensure the associations, further studies are needed.

This study was pioneering considering that the 2q37.3 chromosomal area has not been mapped in regard to PrCa. The association of *IQCA1* 237247036insA with PrCa proved that the region indeed contains something interesting in regard to prostate cancer susceptibility. Further studies are ongoing to untangle the relationship of the many yet unstudied mutations found in 2q37. Perhaps the area becomes the first to yield some mutations that can be used to predict the clinical course of prostate cancer.

Validation of new mutations is always easy to start from the ones that cause functional changes in the encoded protein. This does not rule out the possibility that mutations that are not predicted to be pathogenic could have effects on PrCa pathogenesis. These changes might exert their effects by functioning as important regulators of other genes that possibly contain pathogenic mutations. Future expression quantitave trait loci (eQTL) analyses are a tool to study the changes in regulatory regions of genes.

As a result of this study, the KASPar genotyping assays were validated as a reliable method for genotyping of SNPs and the Genetic Predisposition to Prostate Cancer study group adopted the KASPar genotyping method as on option for genotyping.

8 References

Ahlbom A, Lichtenstein P, Malmström H, Feychting M, Hemminki K, Pedersen NL. (1997). Cancer in twins: Genetic and nongenetic familial risk factors. *J Natl Cancer Inst* **89:** 287-293.

Altshuler D, Daly MJ, Lander ES. (2008). Genetic mapping in human disease. *Science* **322:** 881-888.

Aly M, Wiklund F, Grönberg H. (2011). Early detection of prostate cancer with emphasis on genetic markers. *Acta Oncol* **50 Suppl 1:** 18-23.

American Cancer Society (2012). Cancer facts & figures 2012.

Balk SP, Ko YJ, Bubley GJ. (2003). Biology of prostate-specific antigen. *J Clin Oncol* **21:** 383-391.

Barry MJ. (2001). Prostate-Specific–Antigen testing for early diagnosis of prostate cancer. *N Engl J Med* **344:** 1373-1377.

Bilusic M, Heery C, Madan RA. (2011). Immunotherapy in prostate cancer: Emerging strategies against a formidable foe. *Vaccine* **29:** 6485-6497.

Bos JL. (1989). Ras oncogenes in human cancer: A review. *Cancer Res* **49:** 4682-4689.

Bracarda S, Logothetis C, Sternberg CN, Oudard S. (2011). Current and emerging treatment modalities for metastatic castration-resistant prostate cancer. *BJU Int* **107:** 13-20.

Cai Y, Wang J, Li R, Ayala G, Ittmann M, Liu M. (2009). GGAP2/PIKE-a directly activates both the akt and nuclear factor-kappaB pathways and promotes prostate cancer progression. *Cancer Res* **69:** 819-827.

Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC. (1992). Mendelian inheritance of familial prostate cancer. *Proc Natl Acad Sci U S A* **89:** 3367-3371.

Carter BS, Bova GS, Beaty TH, Steinberg GD, Childs B, Isaacs WB *et al*. (1993). Hereditary prostate cancer: Epidemiologic and clinical features. *J Urol* **150:** 797-802.

Choo YY, Boh BK, Lou JJ, Eng J, Leck YC, Anders B *et al*. (2011). Characterization of the role of COP9 signalosome in regulating cullin E3 ubiquitin ligase activity. *Mol Biol Cell* **22:** 4706-4715.

Collins MM, Meigs JB, Barry MJ, Walker Corkery E, Giovannucci E, Kawachi I. (2002). Prevalence and correlates of prostatitis in the health professionals follow-up study cohort. *J Urol* **167:** 1363-1366.

Cooperberg MR, Carroll PR, Klotz L. (2011). Active surveillance for prostate cancer: Progress and promise. *J Clin Oncol* **29:** 3669-3676.

Coussens LM, Werb Z. (2002). Inflammation and cancer. *Nature* **420:** 860-867.

Cropp CD, Simpson CL, Wahlfors T, Ha N, George A, Jones MS *et al*. (2011). Genome-wide linkage scan for prostate cancer susceptibility in finland: Evidence for a novel locus on 2q37.3 and confirmation of signal on 17q21-q22. *Int J Cancer* **129:** 2400-2407.

Cuttitta F, Fedorko J, Gu JA, Lebacq-Verheyden AM, Linnoila RI, Battey JF. (1988). Gastrin-releasing peptide gene-associated peptides are expressed in normal human fetal lung and small cell lung cancer: A novel peptide family found in man. *J Clin Endocrinol Metab* **67:** 576-583.

Da Silva MAS (2011). Action and function of ASB proteins in compartment size regulation. Academic dissertation. Erasmus University of Rotterdam, Wöhrmann Print Service.

De Marzo AM, Platz EA, Sutcliffe S, Xu J, Grönberg H, Drake CG *et al*. (2007). Inflammation in prostate carcinogenesis. *Nat Rev Cancer* **7:** 256-269.

Divisi D, Di Tommaso S, Salvemini S, Garramone M, Crisci R. (2006). Diet and cancer. *Acta Biomed* **77:** 118-123.

Draisma G, Etzioni R, Tsodikov A, Mariotto A, Wever E, Gulati R *et al*. (2009). Lead time and overdiagnosis in prostate-specific antigen screening: Importance of methods and context. *J Natl Cancer Inst* **101:** 374-383.

Duffy MJ. (2011). Prostate-specific antigen: Does the current evidence support its use in prostate cancer screening? *Ann Clin Biochem* **48:** 310-316.

Easton DF, Eeles RA. (2008). Genome-wide association studies in cancer. *Human Molecular Genetics* **17:** R109-R115.

Eaton NE, Reeves GK, Appleby PN, Key TJ. (1999). Endogenous sex hormones and prostate cancer: A quantitative review of prospective studies. *Br J Cancer* **80:** 930-934.

Eeles RA, Kote-Jarai Z, Giles GG, Olama AA, Guy M, Jugurnauth SK *et al*. (2008). Multiple newly identified loci associated with prostate cancer susceptibility. *Nat Genet* **40:** 316-321.

Finnish Cancer Registry (2011). Cancer Statistics at www.cancerregistry.fi, updated on 21.10.2011

Genome Browser: http://genome.ucsc.edu/, 2.6.2011

Griffiths K. (2000). Estrogens and prostatic disease. *Prostate* **45:** 87-100.

Grönberg H. (2003). Prostate cancer epidemiology. *Lancet* **361:** 859-864.

Guardavaccaro D, Pagano M. (2004). Oncogenic aberrations of cullin-dependent ubiquitin ligases. *Oncogene* **23:** 2037-2049.

Gudmundsson J. (2008). Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. *Nat Genet* **40:** 281-283.

Gulley JL, Drake CG. (2011). Immunotherapy for prostate cancer: Recent advances, lessons learned, and areas for further research. *Clin Cancer Res* **17:** 3884-3891.

Guo CW, Xiong S, Liu G, Wang YF, He QY, Zhang XE *et al*. (2010). Proteomic analysis reveals novel binding partners of MIP-T3 in human cells. *Proteomics* **10:** 2337-2347.

Hannelius U, Salmela E, Lappalainen T, Guillot G, Lindgren CM, von Dobeln U *et al*. (2008). Population substructure in finland and sweden revealed by the use of spatial coordinates and a small number of unlinked autosomal SNPs. *BMC Genet* **9:** 54.

Hay DL, Walker CS, Poyner DR. (2010). Adrenomedullin and calcitonin gene-related peptide receptors in endocrine-related cancers: Opportunities and challenges. *Endocr Relat Cancer* **18:** C1-14.

Hoffman RM. (2011). Screening for prostate cancer. *N Engl J Med* **365:** 2013-2019.

Holland PM, Abramson RD, Watson R, Gelfand DH. (1991). Detection of specific polymerase chain reaction product by utilizing the 5'----3' exonuclease activity of thermus aquaticus DNA polymerase. *Proc Natl Acad Sci U S A* **88:** 7276-7280.

Hsing AW, Chokkalingam AP. (2006). Prostate cancer epidemiology. *Frontiers in Bioscience* **11:** 1388-1413.

Hsing AW, Tsao L, Devesa SS. (2000). International trends and patterns of prostate cancer incidence and mortality. *Int J Cancer* **85:** 60-67.

Hsing AW. (2001). Hormones and prostate cancer: What's next? *Epidemiologic Reviews* **23:** 42-58.

Huang DT, Ayrault O, Hunt HW, Taherbhoy AM, Duda DM, Scott DC *et al*. (2009). E2-RING expansion of the NEDD8 cascade confers specificity to cullin modification. *Mol Cell* **33:** 483-495.

Huang KH, Chow KC, Chang HW, Lin TY, Lee MC. (2011). ATPase family AAA domain containing 3A is an anti-apoptotic factor and a secretion regulator of PSA in prostate cancer. *Int J Mol Med* **28:** 9-15.

Hunziker W, Peters PJ. (1998). Rab17 localizes to recycling endosomes and regulates receptor-mediated transcytosis in epithelial cells. *Journal of Biological Chemistry* **273:** 15734-15741.

Ilic D. (2011). Lycopene for the prevention of prostate cancer. *Cochrane Database of Systematic Reviews* .

Ishak MB, Giri VN. (2011). A systematic review of replication studies of prostate cancer susceptibility genetic variants in high-risk men originally identified from genome-wide association studies. *Cancer Epidemiol Biomarkers Prev* **20:** 1599-1610.

Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, Pedersen NL *et al*. (2008). The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet* **83:** 787-794.

KBioscience: http://www.kbioscience.co.uk/reagents/KASP/KASP.html, 22.1.2012

KBioscience, KASPar protocol: http://www.kbioscience.co.uk/download/KASP%20Manual%20V4%200.pdf 10.6.2011

Knudson AG,Jr. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68:** 820-823.

Kohroki J, Nishiyama T, Nakamura T, Masuho Y. (2005). ASB proteins interact with Cullin5 and Rbx2 to form E3 ubiquitin ligase complexes. *FEBS Lett* **579:** 6796-6802.

Krieger JN. (1999). NIH consensus definition and classification of prostatitis. *JAMA-JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION* **282:** 236-237.

Kvale R, Auvinen A, Adami HO, Klint A, Hernes E, Moller B *et al*. (2007). Interpreting trends in prostate cancer incidence and mortality in the five nordic countries. *J Natl Cancer Inst* **99:** 1881-1887.

Lassi K. (2010). Update on castrate-resistant prostate cancer: 2010. *Curr Opin Oncol* **22:** 263-7.

Lee CH, Akin-Olugbade O, Kirschenbaum A. (2011). Overview of prostate anatomy, histology, and pathology. *Endocrinol Metab Clin North Am* **40:** 565-75, viii-ix.

Lee KL, Peehl DM. (2004). Molecular and cellular pathogenesis of benign prostatic hyperplasia. *J Urol* **172:** 1784-1791.

Lee MH, Zhao R, Phan L, Yeung SC. (2011). Roles of COP9 signalosome in cancer. *Cell Cycle* **10:** 3057-3066.

Levy J, Bosin E, Feldman B, Giat Y, Miinster A, Danilenko M *et al*. (1995). Lycopene is a more potent inhibitor of human cancer cell proliferation than either alpha-carotene or beta-carotene. *Nutr Cancer* **24:** 257-266.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M *et al*. (2000). Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from sweden, denmark, and finland. *N Engl J Med* **343:** 78-85.

Liu W, Sun J, Li G, Zhu Y, Zhang S, Kim S- *et al*. (2009). Association of a germ-line copy number variation at 2p24.3 and risk for aggressive prostate cancer. *Cancer Res* **69:** 2176-2179.

Lunenfeld B. (2002). The ageing male: Demographics and challenges. *World J Urol* **20:** 11-16.

Lutcke A, Jansson S, Parton RG, Chavrier P, Valencia A, Huber LA *et al*. (1993). Rab17, a novel small GTPase, is specific for epithelial cells and is induced during cell polarization. *J Cell Biol* **121:** 553-564.

Mattila H. (2009). Hereditary prostate cancer in Finland – from genetic linkage to susceptibility genes. Academic dissertation. Tampere University Print Oy – Juvenes Print.

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP *et al*. (2008). Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet* **9:** 356-369.

McLatchie LM, Fraser NJ, Main MJ, Wise A, Brown J, Thompson N *et al*. (1998). RAMPs regulate the transport and ligand specificity of the calcitonin-receptor-like receptor. *Nature* **393:** 333-339.

Mihara H, Kurihara T, Watanabe T, Yoshimura T, Esaki N. (2000). cDNA cloning, purification, and characterization of mouse liver selenocysteine lyase. candidate for selenium delivery protein in selenoprotein synthesis. *J Biol Chem* **275:** 6195-6200.

Monroe KR, Yu MC, Kolonel LN, Coetzee GA, Wilkens LR, Ross RK *et al*. (1995). Evidence of an X-linked or recessive genetic component to prostate cancer risk. *Nat Med* **1:** 827-829.

Morganti G, Gianferrari L, Cresseri A, Arrigoni G, Lovati G. (1956). Clinico-statistical and genetic research on neoplasms of the prostate. *Acta Genet Stat Med* **6:** 304-305.

NCBI Gene/MYEOV2: http://www.ncbi.nlm.nih.gov/gene?term=MYEOV2, 16.2.2012

NCBI Gene/SCLY: http://www.ncbi.nlm.nih.gov/gene/51540, 18.2.2012

Nickel JC. (2011). Prostatitis. *Can Urol Assoc J* **5:** 306-315.

Nie Z, Stanley KT, Stauffer S, Jacques KM, Hirsch DS, Takei J *et al*. (2002). AGAP1, an endosome-associated, phosphoinositide-dependent ADP-ribosylation factor GTPase-activating protein that affects actin cytoskeleton. *J Biol Chem* **277:** 48965-48975.

Niraula S, Tannock IF. (2011). Broadening horizons in medical management of prostate cancer. *Acta Oncol* **50:** 141-147.

Nurmi M. (2011). Eturauhasen hyvänlaatuisen liikakasvun hoito. *Duodecim* **127,:** 601-606.

Oesterling JE. (1991). Prostate specific antigen: A critical assessment of the most useful tumor marker for adenocarcinoma of the prostate. *J Urol* **145:** 907-923.

Ong M, Winquist E. (2011). Recent advances in second-line treatment of castration-resistant prostate cancer. *Curr Opin Support Palliat Care* **5:** 199-205.

Pakkanen S. (2010). Profiling of high-risk prostate cancer families in Finland. Academic dissertation. Tampere University Print Oy – Juvenes Print.

Pakkanen S, Baffoe-Bonnie AB, Matikainen MP, Koivisto PA, Tammela TL, Deshmukh S *et al*. (2007). Segregation analysis of 1,546 prostate cancer families in finland shows recessive inheritance. *Hum Genet* **121:** 257-267.

Paolone DR. (2010). Benign prostatic hyperplasia. *Clin Geriatr Med* **26:** 223-239.

Parekh DJ, Ankerst DP, Troyer D, Srivastava S, Thompson IM. (2007). Biomarkers for prostate cancer detection. *J Urol* **178:** 2252-2259.

Picard JC, Golshayan A-, Marshall DT, Opfermann KJ, Keane TE. (2009). The multi-disciplinary management of high-risk prostate cancer. *Urologic Oncology: Seminars and Original Investigations* .

Pomerantz MM. (2010). Genetics of prostate cancer risk. *Mount Sinai Journal of Medicine* **77:** 643-54.

Primer3: http://frodo.wi.mit.edu/, 30.8.2011

Pruvost M, Reissmann M, Benecke N, Ludwig A. (2012). From genes to phenotypes - evaluation of two methods for the SNP analysis in archaeological remains: Pyrosequencing and competitive allele specific PCR (KASPar). *Ann Anat* **194:** 74-81.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81:** 559-575.

Rantapero T. (2012). Applying bioinformatics in the analysis of next-generation sequencing data. Master's thesis. Tampere University Print Oy - Juvenes Print.

Rhoads A, Friedberg F. (1997). Sequence motifs for calmodulin recognition. *The FASEB Journal* **11:** 331-340.

Richardson KS, Zundel W. (2005). The emerging role of the COP9 signalosome in cancer. *Molecular Cancer Research* **3:** 645-653.

Roehrborn CG, Black LK. (2011). The economic burden of prostate cancer. *BJU Int* **108:** 806-813.

Rokman A, Ikonen T, Seppala EH, Nupponen N, Autio V, Mononen N *et al*. (2002). Germline alterations of the RNASEL gene, a candidate HPC1 gene at 1q25, in patients and families with prostate cancer. *Am J Hum Genet* **70:** 1299-1304.

Romanuik TL, Ueda T, Le N, Haile S, Yong TM, Thomson T *et al*. (2009). Novel biomarkers for prostate cancer including noncoding transcripts. *Am J Pathol* **175:** 2264-2276.

Sakr WA, Haas GP, Cassin BF, Pontes JE, Crissman JD. (1993). The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *J Urol* **150:** 379-385.

Sanger F, Nicklen S, Coulson AR. (1977). DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A.* **74:** 5463-5467.

Schaid DJ. (2004). The complex genetic epidemiology of prostate cancer. *Hum Mol Genet* **13 Spec No 1:** R103-21.

Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN. (1998). Evidence for autosomal dominant inheritance of prostate cancer. *Am J Hum Genet* **62:** 1425-1438.

Schalken JA. (2003). Cellular and molecular biology of the prostate: Stem cell biology. *Urology* **62:** 11-20.

Schleutker J, Baffoe-Bonnie AB, Gillanders E, Kainu T, Jones MP, Freas-Lutz D *et al*. (2003). Genome-wide scan for linkage in finnish hereditary prostate cancer (HPC) families identifies novel susceptibility loci at 11q14 and 3p25-26. *Prostate* **57:** 280-289.

Schulz WA, Burchardt M, Cronauer MV. (2003). Molecular biology of prostate cancer. *Mol Hum Reprod* **9:** 437-448.

Schumacher FR, Berndt SI, Siddiq A, Jacobs KB, Wang Z, Lindstrom S *et al*. (2011). Genome-wide association study identifies new prostate cancer susceptibility loci. *Hum Mol Genet* **20:** 3867-3875.

Sharp VJ, Takacs EB, Powell CR. (2010). Prostatitis: Diagnosis and treatment. *Am Fam Physician* **82:** 397-406.

Skolarus TA, Zhang Y, Miller DC, Wei JT, Hollenbeck BK. (2010). The economic burden of prostate cancer survivorship care. *J Urol* **184:** 532-538.

Snider J, Thibault G, Houry WA. (2008). The AAA+ superfamily of functionally diverse proteins. *Genome Biol* **9:** 216.

Stanford JL, Ostrander EA. (2001). Familial prostate cancer. *Epidemiologic Reviews* **23:** 19-23.

Stokes ME, Ishak J, Proskorovsky I, Black LK, Huang Y. (2011). Lifetime economic burden of prostate cancer. *BMC Health Serv Res* **11:** 349.

Stratton MR, Campbell PJ, Futreal PA. (2009). The cancer genome. *Nature* **458:** 719-724.

Tammela T. (2004). Endocrine treatment of prostate cancer. *J Steroid Biochem Mol Biol* **92:** 287-295.

Tammela T. (2006). Eturauhasen hyvänlaatuinen liikakasvu : Käypä hoito -suosituksen tiivistelmä. *Duodecim* **122,:** 107-108.

Timms BG. (2008). Prostate development: A historical perspective. *Differentiation* **76:** 565-577.

Toda M, Suzuki T, Hosono K, Hayashi I, Hashiba S, Onuma Y *et al*. (2008). Neuronal system-dependent facilitation of tumor angiogenesis and tumor growth by calcitonin gene-related peptide. *Proceedings of the National Academy of Sciences* **105:** 13550-13555.

Tosoian J, Loeb S. (2010). PSA and beyond: The past, present, and future of investigative biomarkers for prostate cancer. *ScientificWorldJournal* **10:** 1919-1931.

Urological Consultants of Florida: http://www.miamiurologyconsultants.com/services-procedures/robotic-prostatectomy-surgery-prostate-cancer-miami-aventura-fl.php 5.1.2012

Vossen RH, Aten E, Roos A, den Dunnen JT. (2009). High-resolution melting analysis (HRMA): More than just sequence variant screening. *Hum Mutat* **30:** 860-866.

Wassink TH, Piven J, Vieland VJ, Jenkins L, Frantz R, Bartlett CW *et al*. (2005). Evaluation of the chromosome 2q37.3 gene CENTG2 as an autism susceptibility gene. *Am J Med Genet B Neuropsychiatr Genet* **136B:** 36-44.

Whittemore AS, Wu AH, Kolonel LN, John EM, Gallagher RP, Howe GR *et al*. (1995). Family history and prostate cancer risk in black, white, and asian men in the united states and canada. *Am J Epidemiol* **141:** 732-740.

Wilt TJ, MacDonald R, Rutks I, Shamliyan TA, Taylor BC, Kane RL. (2008). Systematic review: Comparative effectiveness and harms of treatments for clinically localized prostate cancer. *Ann Intern Med* **148:** 435-448.

Wittwer CT. (2009). High-resolution DNA melting analysis: Advancements and limitations. *Hum Mutat* **30:** 857-859.

Wittwer CT, Reed GH, Gundry CN, Vandersteen JG, Pryor RJ. (2003). High-resolution genotyping by amplicon melting analysis using LCGreen. *Clinical Chemistry* **49:** 853-860.

Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D *et al*. (1998). Evidence for a prostate cancer susceptibility locus on the X chromosome. *Nat Genet* **20:** 175-179.

Zerbib M, Zelefsky MJ, Higano CS, Carroll PR. (2008). Conventional treatments of localized prostate cancer. *Urology* **72:** S25-S35.