SANNA SILTANEN

Contribution of the *ARLTS1* Gene to Prostate Cancer Susceptibility

ACADEMIC DISSERTATION
To be presented, with the permission of
the Board of the Institute of Biomedical Technology of the University of Tampere,
for public discussion in the Auditorium of Finn-Medi 5,
Biokatu 12, Tampere, on September 13th, 2013, at 12 o'clock.

UNIVERSITY OF TAMPERE

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, Institute of Biomedical Technology
BioMediTech
Fimlab Laboratories
Tampere Graduate Program in Biomedicine and Biotechnology (TGPBB)
Finland

Supervised by
Professor Johanna Schleutker
University of Tampere
Finland
Docent Tiina Wahlfors
University of Tampere
Finland

Reviewed by
Docent Auli Karhu
University of Helsinki
Finland
Docent Antti Rannikko
University of Helsinki
Finland

Cover design by
Mikko Reinikka

Acta Universitatis Tamperensis 1845
ISBN 978-951-44-9193-1 (print)
ISSN-L 1455-1616
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1325
ISBN 978-951-44-9194-8 (pdf)
ISSN 1456-954X
http://tampub.uta.fi

Suomen Yliopistopaino Oy – Juvenes Print
Tampere 2013

CONTENTS

LIST OF ORIGINAL COMMUNICATIONS

This thesis is based on the following communications, referenced in the text by their Roman numerals (I-III):

I **Siltanen S.**, Syrjäkoski K., Fagerholm R., Ikonen T., Lipman P., Mallott J., Holli K., Tammela T.L., Järvinen H.J., Mecklin J.P., Aittomäki K., Blomqvist C., Bailey-Wilson J.E., Nevanlinna H., Aaltonen L.A., Schleutker J., Vahteristo P. 2008, "*ARLTS1* germline variants and the risk for breast, prostate and colorectal cancer", *Eur J Hum Genet*. vol. 16, no. 8, pp. 983-91.

II **Siltanen S.***, Wahlfors T.*, Schindler M., Saramäki O.R., Mpindi J.P., Latonen L., Vessella R.L., Tammela T.L., Kallioniemi O., Visakorpi T., Schleutker J. 2011, "Contribution of the *ARLTS1* Cys148Arg (T442C) variant with prostate cancer risk and *ARLTS1* function in prostate cancer cells", *PLoS ONE*, vol. 6, no. 10, pp. e26595. *Authors contributed equally to this work.

III **Siltanen S.**, Fischer D., Rantapero T., Laitinen V., Mpindi J.P., Kallioniemi O., Wahlfors T., Schleutker J. 2013, "*ARLTS1* and prostate cancer risk – analysis of expression and regulation", *PLoS ONE*, vol. 8, no. 8, pp. e72040.

ABBREVIATIONS

AI	Allelic imbalance
ARF	ADP-ribosylation factor
ARF6	*ADP-ribosylation factor 6*
Arg	Arginine
ARL	ARF-like
ARL11	*ADP-ribosylation factor-like 11*
ARL14	*ADP-ribosylation factor-like 14*
ARL4C	*ADP-ribosylation factor-like 4C*
ARL4D	*ADP-ribosylation factor-like 4D*
ARLTS1	*ADP-ribosylation factor-like tumour suppressor 1*
ASAP	Acinar cell proliferation
BCC	Basal cell carcinoma of the skin
BCL2	*B-cell CLL/lymphoma 2*
BPH	Benign prostatic hyperplasia
BRCA	Breast cancer
BRCA1	*Breast cancer, early onset*
BRCA2	*Breast cancer 2, early onset*
BMI	Body mass index
C13ORF15	*Chromosome 13 open reading frame 15*
CCND1	*Cyclin D1*
CFA	Cancer familial aggregation
CGH	Comparative genomic hybridisation
CHC1.L	*Chromosome condensation 1-like*
CI	Confidence interval
CLL	Chronic lymphocytic leukaemia
c-MYC	*Proto-oncogene c-Myc*
CRC	Colorectal cancer
Cys	Cysteine
DLEU2	*Deleted in lymphocytic leukaemia 2*
DLEU7	*Deleted in lymphocytic leukaemia 7*
DNA	Deoxyribonucleic acid
ELAC2	*ElaC homolog 2*
ENCODE	The Encyclopedia of DNA Elements
EPHB2	*EPH receptor B2*
eQTL	Expression quantitative trait loci
ERBB2	*V-erb-b2 avian erythroblastic leukaemia viral oncogene homolog 2*
ERG	*V-ets erythroblastosis virus E26 oncogene homolog (avian)*
FFPE	Formalin-fixed paraffin embedded
FOXO1A	*Forkhead box protein O1A*
GAP	GTPase-activating protein
GDP	Guanosine diphosphate
GEF	Guanine nucleotide exchange factor
GTP	Guanosine-5'-triphosphate
GWAS	Genome-wide association study

HOXB13	*Homeobox B13*
HPC1	*Hereditary prostate cancer 1*
HPC3	*Prostate cancer, hereditary, 3*
HPCX	*Hereditary prostate cancer, X-linked*
ICAM	Intercellular adhesion molecule
IL-8	*Interleukin 8 gene*
IL-10	*Interleukin 10 gene*
KCNRG	*Potassium channel regulator*
KLF5	*Kruppel-like factor 5*
KPNA3	*Karyopherin alpha 3, importin alpha 4*
LCL	Lymphoblastoid cell line
LOD	Logarithm of the odds
LOH	Loss of heterozygosity
MIC-1	*Macrophage inhibitory cytokine-1*
miRNA	MicroRNA
MIR-15A	*MicroRNA gene 15a*
MIR16-1	*MicroRNA gene 16-1*
MLNR	*Motilin receptor*
mRNA	Messenger RNA
MSR1	*Macrophage scavenger receptor* 1
NGS	Next-generation sequencing
NKX3.1	*NK3 homeobox 1*
OR	Odd's ratio
PCAP	*Predisposing for prostate cancer*
PHF11	*PHD finger protein 11*
PIA	Proliferative inflammatory atrophy
PIN	Prostate intraepithelial neoplasia
Pro	Proline
PSA	Prostate specific antigen
PTEN	*Phosphate and tensin homolog*
QTL	Quantitative trait loci
RB1	*Retinoblastoma 1*
RCBTB1	*Regulator of chromosome condensation [RCC1] and BTB [POZ] domain containing protein 1*
RNA	Ribonucleic acid
RNASEL	*Ribonuclease L*
SAR	Secretion-associated and Ras-related
Ser	Serine
SETDB2	*SET domain, bifurcated 2*
SIAH3	*Seven in absentia homolog 3*
SNP	Single nucleotide polymorphism
SPRYD7	*SPRY domain containing 7*
ST13	*Suppression of tumorigenity 13 protein*
Stop	Stop codon
TMPRSS2	*Transmembrane protease, serine 2*
Trp	Tryptophan
TSG	Tumour suppressor gene
TURP	Transurethral resections of prostate
Val	Valine
VEGF	*Vascular endothelial growth factor*

WES	Whole exome sequencing
WGS	Whole genome sequencing
WNT3A	*Wingless-type MMTV integration site family, member 3A*

YHTEENVETO

Eturauhassyöpä on miesten yleisin syöpä aiheuttaen teollisuusmaissa toiseksi eniten miesten syöpäkuolemia. Eturauhassyöpä on merkittävä kansanterveydellinen haaste ja sen ilmaantuvuus eli insidenssi on ollut jatkuvassa kasvussa 2000-luvun loppupuolelle asti. Vuonna 2010 maassamme diagnosoitiin 4697 uutta eturauhassyöpätapausta ja ikävakioitu insidenssi on nykyään 89.4 tapausta 100 000 henkeä kohti (http://www.cancer.fi/syoparekisteri/). Epidemiologiset tutkimukset ovat selvästi osoittaneet, että suurin yksittäinen eturauhassyövän riskitekijä on taudin esiintyminen lähisukulaisissa. Perinnöllisten tekijöiden lisäksi eturauhassyöpäriskiä lisääviin tekijöihin luetaan korkea ikä ja etninen alkuperä. Uusien tapausten voimakkaaseen kasvuun johtavia tekijöitä ovat PSA-seulonnan lisääntyminen ja väestön ikääntyminen, mutta taustalla vaikuttaa osin myös vielä tuntemattomia tekijöitä, joita ei ole intensiivisestä tutkimuksesta huolimatta pystytty täysin identifioimaan. Suurin kliininen haaste on taudin käyttäytymisen ennustaminen eli prognostiikka. Eturauhassyövän geneettisen taustan tutkimus on tuottanut monta syöpäriskiin assosioituvaa riskialttiusgeeniä. Usein löydetyt geenimuutokset edustavat kuitenkin niin sanottuja matalan penetranssin muutoksia, jotka selittävät vain osan perinnöllisestä syöpäriskistä. Yhtenä selityksenä pidetään yksittäisten muutosten yhteisvaikutusta, joka on osoittautunut haastavaksi tutkimuskohteeksi.

Eturauhassyöpään kytkeytyvän kromosomialueen 13q14 on myös todettu deletoituneen sekä perinnöllisessä että selekoimattomassa eturauhassyövässä ja solukuoleman säätelyssä toimiva *ARLTS1* (*ADP-ribosylation factor like tumour suppressor 1*, /*ARL11*, *ADP-ribosylation factor-like 11*) on samalla alueella sijaitseva syövänestogeeni, jonka on osoitettu olevan osallisena monien syöpien syntyyn. Tämän väitöskirjatutkimuksen tavoitteena oli selvittää *ARLTS1*-geenin muutoksia ja molekulaarisia mekanismeja eturauhassyövän kehittymisen ja syöpäalttiuden taustalla, etenkin eturauhassyöpäperheiden potilailla.

ARLTS1-geenin koodaavan alueen riskivarianttien seulonta toteutettiin vuodesta 1995 alkaen kerätystä laajasta potilasaineistosta suorasekvensointimenetelmällä ja

genotyyppaamalla. Analyyseissä oli mukana yhteensä 2060 eturauhassyöpäpotilaan DNA-näytettä, 375 hyperplasiatapausta ja 760 SPR:n anonyymiä verenluovuttajakontrollia. Tutkimuksissa oli mukana myös rintasyöpänäytteitä (n=1242), kolorektaalisyöpänäytteitä (n=241), tuumorikudosnäytteitä (n=53), ksenografteja (n=19) ja viisi eturauhassyöpäsolulinjaa. Eturauhassyöpäriskiin merkittävästi vaikuttavien geenimuutosten löytyessä niiden vaikutukset *ARLTS1*:n ilmentymiseen karakterisoitiin käyttäen Q-RT-PCR-menetelmää.

Tämän väitöskirjatutkimuksen tulosten mukaan *ARLTS1* variantit G194T (Gly65Val) ja T442C (Cys148Arg) lisäävät tilastollisesti merkitsevästi riskiä sairastua eturauhassyöpään. Lisäksi havaitsimme, että T442C variantin homotsygoottimuutos CC assosioituu taudin aggressiiviseen muotoon, ja muutos oli yleisin tai jopa ainoa *ARLTS1* muutos tutkituissa näytteissä. Lisäksi havaitsimme *ARLTS1* lähetti-RNA:n ali-ilmentymisen kudosnäytteissä, mikä tukee hypoteesia *ARLTS1*:n toiminnasta syövänestogeeninä. *ARLTS1*:n ali-ilmentyminen jo ituradan tasolla havaittiin tutkimalla eturauhassyöpäperhenäytteiden kokoverestä eristettyä RNA:ta. Jatkotutkimuksissa havaittiin lähialueen geenimuutosten vaikuttavan ilmentymiseen. Bioinformatiivisten analyysien mukaan *ARLTS1*:n ilmentyminen on vahvasti yhteydessä immuunijärjestelmän prosesseihin. Koska kroonisen tulehduksen on esitetty olevan osallisena eturauhassyövän synnyssä, jatkotyönä seulottiin tulehdussignaalireiteillä vaikuttavien geenien SNP-muutoksia laajassa aineistossa, yhdistettynä potilasnäytteiden *ARLTS1*-genotyyppiin. Tarkoituksena oli karakterisoida geeni-geeni -interaktioita ja niiden korrelaatioita geenien ilmentymistasoon, mutta assosiaatioita ei kuitenkaan löytynyt.

Eturauhassyöpä on myös kansantaloudellisesti merkittävä tauti, joten sen taustalla vaikuttavien geenien identifiointi sekä täsmentäisi taudin diagnosointia että helpottaisi ennaltaehkäisyä ja ennusteen laatimista ja mahdollisesti myös hoitoa. Tämä väitöskirja tukee aiempia havaintoja eturauhassyövän perinnöllisyydestä ja tarjoaa osaltaan lisätietoa ituradan ja solutason tapahtumista eturauhassyövän taustalla.

ABSTRACT

Prostate cancer is the most common malignancy among men in industrialised countries, including Finland. It is a heterogeneous disease with definitive risk factors that include age, ethnic origin and family history. In Finland, the incidence of prostate cancer is 89.4/100,000 and in 2010, 4,697 new prostate cancer cases were diagnosed (http://www.cancer.fi/syoparekisteri/en/). Despite extensive research over the last decade, the aetiological risk factors and genes underlying the genetic susceptibility to prostate cancer remain largely unclear, which has hampered effective prevention and the development of better treatments.

Mutations to known high-penetrance cancer predisposition genes can only explain a small fraction of cancer cases, and therefore, a polygenic model has been proposed in which several low-penetrance alleles may have a multiplicative and/or modifying effect that accounts for a substantial proportion of the familial aggregation of cancer. One such candidate gene is *ARLTS1* (*ADP-ribosylation factor like tumour suppressor 1*, also known as *ARL11, ADP-ribosylation factor-like 11*), a tumour suppressor gene that is known to function in many human cancers and to play a role in apoptotic signalling. Interestingly, the chromosomal area 13q14, where *ARLTS1* resides, was identified in a multi-centre genome-wide linkage search for new prostate cancer susceptibility genes in families with at least five affected members. In addition, the 13q14 locus is among the most frequently deleted chromosomal regions in prostate cancer and hereditary prostate cancer, suggesting that *ARLTS1* could be a target for both germline and somatic mutations.

The aim of this thesis study was to provide important new knowledge on the genetics and function of the putative tumour suppressor gene, *ARLTS1*, in prostate cancer, and to determine the molecular mechanisms involved in prostate cancer development and susceptibility, especially in patients with a family history of prostate cancer. The outcome of this study can be used to establish a basis for novel strategies for the diagnostics, prevention, prognosis and screening of prostate cancer, including possible novel therapeutic strategies.

We performed direct sequencing and genotyping analyses on both germline DNA from prostate cancer patients and on clinical tumours, xenograft samples and prostate cancer cell lines. Additionally, colorectal and breast cancer samples were analysed. The whole coding region of *ARLTS1* was analysed in Finnish familial cancer patients who had at least one affected 1st or 2nd degree relative. The results of this thesis suggest that *ARLTS1* polymorphisms, G194T (Gly65Val) and T442C (Cys148Arg), increase the risk of prostate cancer. Further studies of the T442C variant revealed a statistically significant association with an increased risk for homozygous CC carriers both in unselected and familial prostate cancer cases. T442C CC variant carriers had an elevated risk of developing more aggressive disease. Additionally, T442C was the only variant observed with a greater frequency among malignant tissue samples, prostate cancer cell lines and xenografts, supporting its role in prostate carcinogenesis. Interestingly, the T442C variant was the only *ARLTS1* variant found in any of the prostate cancer cell lines tested.

Q-RT-PCR revealed that *ARLTS1* is widely down-regulated in prostate cancer cell lines, xenografts and tumours. Most of the down-regulated samples also had the CC genotype for the T442C variant. The CC risk genotype was also associated with decreased *ARLTS1* expression in immunologically relevant lymphoblastoid cell lines. In addition, the *ARLTS1* co-expression signature confirmed the low expression of *ARLTS1* in prostate cancer, and *ARLTS1* expression was strongly associated with immune system processes. RNA expression analysis of whole-blood derived samples from 84 familial prostate cancer cases and 15 controls showed a significant decrease in *ARLTS1* expression levels in prostate cancer cases, indicating an effect of germline alteration on *ARLTS1* expression levels. The regulation of *ARLTS1* expression was analysed with the expression quantitative trait loci (eQTL) method, and altogether, we identified 14 significant cis-eQTLs located in regulatory regions of the 13q14 locus that affected *ARLTS1* expression levels.

INTRODUCTION

Cancer is a group of diseases characterised by uncontrolled growth and the accumulation of abnormal cells in a tissue. In the worst case, the uncontrolled growth may result in patient death. Cancer is caused by both external factors (tobacco, infectious organisms, chemicals and radiation) and internal factors (inherited mutations, hormones, immune conditions and mutations that occur from metabolism), presenting as a polygenic disorder. These causal factors may act together or in sequence to initiate or promote the development of cancer.

According to Hanahan and Weinberg (2000), cancer can be characterised by six hallmarks, including sustaining proliferative signalling, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis and activating invasion and metastasis. Two emerging hallmarks, reprogramming of energy metabolism and evading immune destruction, have been added to the list (Hanahan, Weinberg 2011). Beyond these properties of cancerous cells, genomic instability stands alone, creating genetic diversity and inflammation.

Prostate cancer is the most frequently diagnosed malignancy in men, and its incidence has been increasing over the last decades in many parts of the world, including Finland (Finnish Cancer Registry). The discovery and screening of prostate specific antigen, PSA, has contributed to the earlier diagnosis and treatment of prostate cancer. Although many screen-detected prostate cancers may be indolent, prostate cancer remains a major contributor to mortality in men.

Definite risk factors for prostate cancer include age, ethnicity and family history (Crawford 2003). As shown in several familial and twin studies, there is an inherited component that predisposes an individual to this disease. Indeed, up to 42% of the risk for prostate cancer is due to hereditary components (Lichtenstein *et al*. 2000). Linkage studies performed on high-risk families have provided genomic regions linked to prostate cancer, such as the *hereditary prostate cancer 1* (*HPC1*) locus at chromosome 1q25, the *elaC homolog 2* (*ELAC2*) gene at chromosome 17p and the *macrophage scavenger receptor 1* (*MSR1*) at 8p22 (Carpten *et al*. 2002, Tavtigian *et al*. 2001, Xu *et al*. 2002). Other susceptibility loci identified by linkage analyses

include 1p36, 1q42.2-43, 8p21-23, 8q24, 16q23, 17p11, 19q13, 20q13 and Xq27-28. Genome-wide association studies (GWAS) have identified candidate loci that are common in the population, resulting in over 70 susceptibility loci (Goh *et al*. 2012, Eeles *et al*. 2013). To date, the GWAS studies performed reported variants with moderate effects on prostate cancer risk, and the susceptibility loci revealed by linkage studies are not consistent between studies. Taken together, no clear, high risk locus or gene for prostate cancer has been identified, and few loci propose clear biological significance or function. Taken together, these results suggest the multifactorial and polygenetic nature of the disease.

Prostate cancer exhibits a tremendous challenge to public health systems throughout the world. However, the explicit genetic risk factors remain undefined, and the disease is starting to be considered a combination of different prostate cancer subtypes. The search for susceptibility genes has revealed only low penetrant alleles, most likely due to the complex structure and behaviour of this disease. Thus, there is an indication and need to identify new candidate genes and therapeutic strategies.

REVIEW OF THE LITERATURE

1. Prostate cancer

Prostate cancer is a major health burden all over the world. It is the most common cancer in males and the second leading cause of cancer related deaths. The most recognised risk factors for prostate cancer include age, race and family history. The overall risk for prostate cancer is one in six (American Cancer Society 2012).

1.1 Clinical features

The human prostate is a lobular organ consisting of three types of epithelial cells: basal, luminal and neuroendocrine cells (Shen, Abate-Shen 2010). Prostate cancer can arise from both basal and luminal cells and the basal cells may also operate as the cellular origin for metastatic prostate cancer (Xin 2012). Gene expression profiles from human prostate cancer specimens have been generated (Taylor *et al*. 2010, Glinsky *et al*. 2004), but distinct tumour subtypes have not been classified. The same prostate encompasses independent tumour loci of independent clonal origin (Kobayashi *et al*. 2008). Thus, prostate cancer is a very heterogeneous disease at both the individual and molecular levels.

Localised prostate cancer usually does not cause any symptoms, whereas benign prostatic hyperplasia (BPH), or the increased size of the prostatic gland due to tumour growth, can obstruct the urethra or bladder, causing more painful or increased frequency of urination. The diagnosis of prostate cancer is based on testing for prostate specific antigen (PSA) and/or a digital rectal examination. PSA is a kallikrein-related serine protease that is released into the blood, and it is increased after the disruption of the normal prostate architecture (Lilja, Ulmert & Vickers 2008). Elevated PSA levels are usually followed by a needle biopsy, which is used for histological grading of the disease by the Gleason score 1-5 (most to least

differentiated), which is determined by the combination score of the sum of the two most prevalent tissue patterns (Epstein 2010). Primary tumours are defined as being organ-confined to fully invasive (T1-4), with or without lymph node involvement (N0 or 1) and the presence of metastases (M0 and 1a-c) (Ohori, Wheeler & Scardino 1994). Recent studies suggest that PSA screening should be reduced as it has led to increased diagnosis and the overtreatment of clinically indolent tumours that could have been managed with active surveillance. Thus, the major clinical challenge is to distinguish indolent prostate cancer from aggressive tumours.

To better detect cancer patients at risk and target their treatment, several commercial gene tests have been developed and are in use. More than a dozen companies offer tests with advanced techniques to measure multiple genes or molecular markers. The genetic tests that are currently available or coming into the market are summarised in Table 1.

Nevertheless, these tests have not been adequately validated, and their sales have been low because the prices are quite high. Even after careful testing, with a possible outcome of very minor cancer risk, many men do not want to live with cancer and request a prostatectomy. Thus, better and more accurate molecular markers and genetic variants are needed to overcome the clinical challenges.

Table 1. Commercial prostate cancer gene tests

	Test	Company	Available	Sample	Measures
Screening	Prostate Health Index	Beckman Coulter	2013	Blood	Variants of PSA
Reduce the number of unnecessary	4Kscore	Opko Health	2013	Blood	Variants of PSA and a related compound
(negative) biopsies spurred	Prostarix	Metabolon	2013	Urine	Sarcosine and other chemicals
by elevated PSA levels					associated with cancer
After a positive biopsy	Prolaris	Myriad Genetics	2012	Biopsy	Gene activity pattern (46 genes)
Distinguish aggressive cancers	Oncotype DX test	Genomic Health	2013	Biopsy	Gene activity pattern
that need treatment from	ProstaVysion	Bostwick Laboratories	2011	Biopsy	*ERG* fusion/translocation and the loss of *PTEN*
non-aggressive ones that do not	Metamark Biopsy Test	Metamark Genetics	2013	Biopsy	Marker proteins
After a negative biopsy	Progensa PCA3	Gen-Probe (Hologic)	2012	Urine	Expression of *PCA3* gene
If PSA levels stay high	Prostate Core Mitomic Test	Mitomics	2011	Biopsy	Deletions of mitochondrial DNA
despite a negative biopsy,	ConfirmMDx	MDxHealth	2012	Biopsy	Methylation of *GSTP1*, *APC* and *RASSF1*
is a repeat biopsy needed?					
After surgery	Decipher	GenomeDx Biosciences	2013	Tissue	Gene activity pattern
Is additional treatment necessary if pathology shows	Nadia ProsVue	Iris International	2012	Blood	Increase over time of minuscule PSA levels
risk after surgery?					

adapted from www.pcf.org

Treatment options for prostate cancer include radical prostatectomy (57% of patients under 65 years of age), radiation therapy (25%) and active surveillance (18%) (Siegel *et al*. 2012), although there is wide variance in the treatment options provided by different health care districts and health care systems. Active surveillance is more commonly used in patients diagnosed at an older age. For advanced stage prostate cancer, androgen deprivation therapy, chemotherapy, bone-directed therapy, radiation therapy or a combination of these treatments is used. Most prostate cancers are discovered in the local or regional stage, and the 5-year survival rate reaches almost 100%, with at 15-year survival rate of over 90% (Siegel *et al*. 2012, Siegel, Naishadham & Jemal 2012). For distant stage prostate cancer, the survival is only 29% within all races. According to an estimate by Siegel R *et al*. (2012), prostate cancer will remain the most common cancer among men in 2022, and the percentage of cancer survivors will increase from 43% to 45%.

1.2 Epidemiology

The worldwide morbidity of prostate cancer is high. Prostate cancer is one of the major causes of cancer-related mortality among men. It is estimated that in the United States, a total of 238,590 new prostate cancer cases will be diagnosed in 2013 and that the disease will account for 29,720 deaths (American Cancer Society 2013). Most men diagnosed with prostate cancer die from other symptoms and diseases than from prostate cancer itself. In Finland, the age-adjusted incidence of prostate cancer was 89.1/100 000 in 2010; 4,697 new cases were diagnosed, and 845 men died for prostate cancer (Finnish Cancer Registry). The incidence rates in the United States and the number of new cases in Finland are displayed in Figure 1.

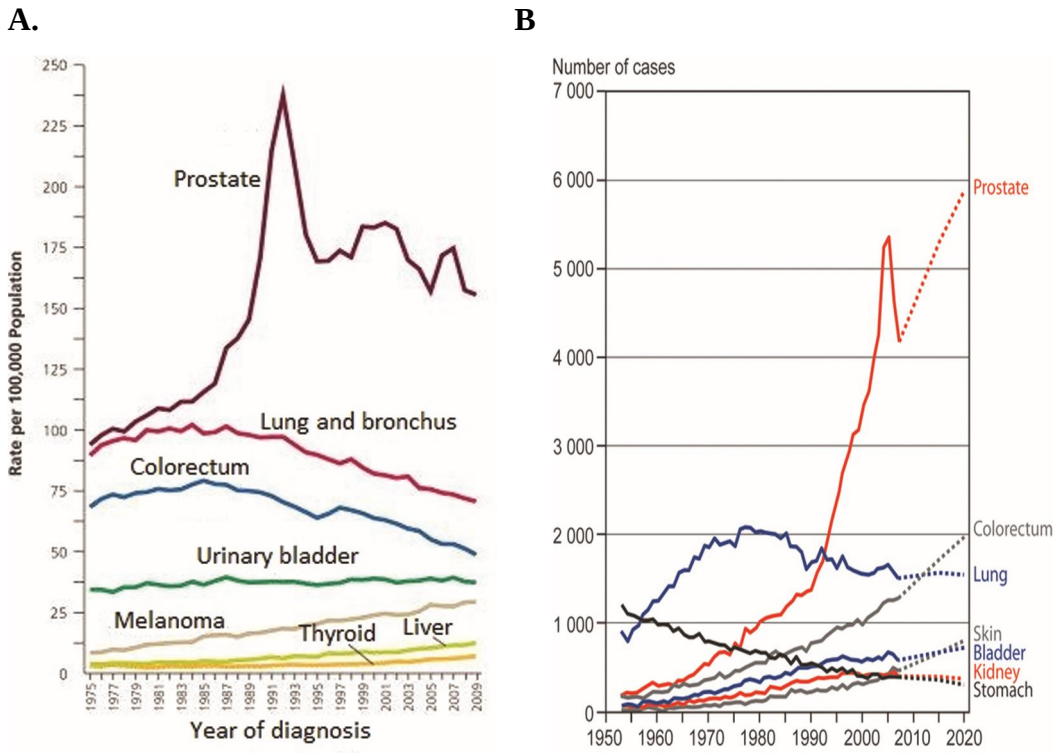


Figure 1. A; Trends in incidence rates for selected cancers in the United States (adapted from Siegel, Naishadham & Jemal 2012) and **B**; the number of new cases of common cancer types in Finland (Finnish Cancer Registry).

1.3 Risk factors

Age. The median age at prostate cancer diagnosis is 67 years (Howlader *et al*. 2012). Prostate cancer frequency increases rapidly with aging, and the disease is rarely detected in men younger than 40 years of age. Early onset prostate cancer is diagnosed at ≤55 years. In Finland, from 2007-2011, there was an average of only six new prostate cancer cases within the 40-44 year age group, and the highest annual number of new prostate cancer cases was 819 among men aged 70-74 years (Finnish Cancer Registry).

Ethnicity. Prostate cancer is diagnosed more often in Europeans and Americans than in Asian men, and African American men have the highest incidence (Siegel,

Naishadham & Jemal 2012). Incidence rates varied 24-fold, with the highest incidence in developed countries and the lowest in Asia and Africa (Figure 2). This variation is in part due to different diagnostic procedures, PSA screening and environmental factors, including genetic factors. In addition to the variation in incidence in diverse ethnic groups, there is a difference in mortality rates (Center *et al*. 2012). For example, African American men die from prostate cancer more often.

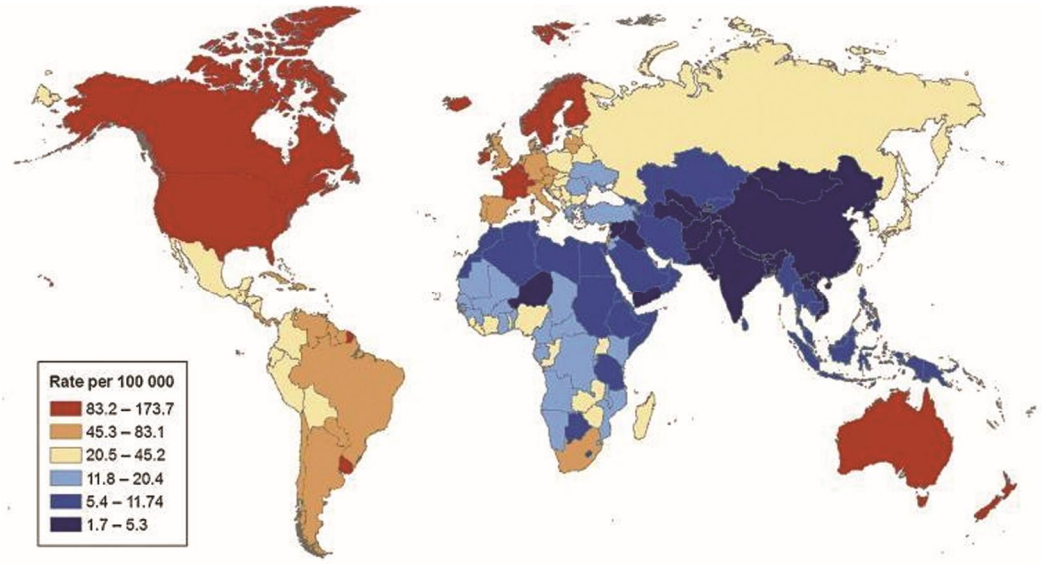


Figure 2. International variation in age-standardised prostate cancer incidence rates (Center *et al.* 2012).

Family history. Approximately 10-15% of men with prostate cancer have at least one affected relative (Nakagawa *et al.* 2012). Family history has been identified as a major risk factor for prostate cancer in many studies, indicating that inherited factors play an important role in the disease carcinogenesis. There is a great amount of evidence supporting the existence of prostate cancer susceptibility genes (Schaid 2004, Zeegers, Jellema & Ostrer 2003). Inherited genetic risk factors, or susceptibility genes, account for approximately 5% to 10% of prostate cancer cases. According to a highly cited twin study by Lichtenstein *et al.* (2000), the contribution of hereditary factors in prostate cancer was 42% (95% CI 29-50). Furthermore, having a brother or father with prostate cancer increases the risk of developing the disease, and the risk is greater when a brother is affected (Zeegers, Jellema & Ostrer

2003). Prostate cancer risk increases even more if the affected brother was diagnosed before 55 years of age or when both brother and father were affected (Hemminki, Czene 2002). A significant genetic component correlated with prostate cancer associated genetic variants was also confirmed in early onset (before age 55) prostate cancer (Lange *et al.* 2012). Familial aggregation may also be the result of increased screening in prostate cancer families (Bratt *et al.* 2010).

Segregation analysis performed in prostate cancer families suggested the presence of a highly penetrant gene with an autosomal dominant mode of transmission (Carter *et al.* 1992, Gronberg *et al.* 1997, Schaid 1998). Additionally, a recessive model has been proposed (Pakkanen *et al.* 2007).

Environmental factors. The other widely studied, but not fully confirmed, risk factors are diet and obesity. There is some evidence linking red and processed meat and dairy product consumption to prostate cancer development, but according to a meta-analysis, no clear positive association was found between red/processed meat intake and prostate cancer (Alexander *et al.* 2010). The consumption of tomato-based products has been shown to have a protective influence on prostate cancer. This effect is likely due to lycopene, a carotenoid present in tomatoes that has strong anti-oxidative properties (Giovannucci 2005, Wei, Giovannucci 2012). Other studied nutritional factors that may influence prostate cancer include micronutrients and vitamins (carotenoids, retinoids, vitamins C, D and E), fruit and vegetable intake, minerals (calcium and selenium), and phytoestrogens (isoflavonoids, flavonoids and lignans) (Schultz, Meier & Schmid 2011). Vitamin D has been shown to promote the differentiation of prostate cancer cells and to protect against tumorigenesis (Hollis *et al.* 2012). However, according to a review by JR Marshall (2012), all of the reported non-experimental epidemiologic studies linking dietary components with prostate cancer risk and prevention are inconsistent. In regard to alcohol consumption, mild daily consumption does not affect prostate cancer risk, whereas heavy consumption of seven or more drinks per day may be associated with increased prostate cancer risk (Rizos *et al.* 2010). Body mass index (BMI) and obesity are correlated with increased prostate cancer metastasis and mortality (Hsing, Sakoda & Chua 2007, Gong *et al.* 2007). Additionally, cholesterol levels are associated with prostate cancer risk, and it was reported that statins have a

protective effect on prostate cancer (Platz, Clinton & Giovannucci 2008, Jacobs *et al.* 2007, Murtola *et al.* 2007, Platz *et al.* 2006).

Asian populations have the lowest incidence rates of prostate cancer. Lifestyle changes, such as a shift towards westernised diet, may be one reason for the rapid increase in prostate cancer incidence in Japan (Matsuda, Saika 2009). When the strongest risk factors, age, ethnicity and family history, are not modifiable, lifestyle risk factors could be used to guide prostate cancer prevention.

Inflammation. Chronic inflammation has repeatedly been proposed as a risk factor for prostate cancer (Sfanos, De Marzo 2012, De Marzo *et al.* 2007). Several genes acting in innate immunity and inflammatory pathways are associated with prostate cancer, supporting the role of inflammation in the aetiology of prostate cancer (Zheng *et al.* 2006, Kwon *et al.* 2011). These inflammatory genes include the ones identified via linkage analysis (*ribonuclease L*, *RNASEL*, and *macrophage scavenger receptor 1*, *MSR1*), and multiple genes identified in case-control studies, such as cytokines (e.g., *macrophage inhibitory cytokine-1*, *MIC-1*), interleukins (*IL-8, IL-10*), the intercellular adhesion molecule (ICAM), *vascular endothelial growth factor* (*VEGF*) and the Toll-like receptor gene cluster (Sun *et al.* 2007).

Increased prostate epithelial proliferation, referred to as proliferative inflammatory atrophy (PIA), has been proposed to precede high-grade prostatic intraepithelial neoplasia (PIN) and, eventually, prostate adenocarcinoma (De Marzo *et al.* 1999, De Marzo *et al.* 2003). In addition, the prevalence of asymptomatic prostatic inflammation (i.e., histological prostatitis) is high. Potential causes of prostate inflammation include infectious agents (sexually transmitted organisms, non-sexually transmitted bacteria and viruses), urine reflux inducing chemicals and physical trauma and dietary factors (Sfanos, De Marzo 2012, De Marzo *et al.* 2007). Inflammation is a very complex phenomenon of biochemical processes, including increased DNA damage, oxidative stress, proliferation, angiogenesis, invasion and metastasis. It has been proposed that gene-gene interaction studies may be able to find the true causal variants and verify the connection between inflammation and prostate cancer (Sun *et al.* 2007, Kundu, Surh 2012).

2. Prostate cancer genetics

2.1 Overview of cancer genetics

Cancer is a disease of abnormal gene functions that is caused by the accumulation of genetic changes. The genetic information within cancerous cells, DNA, provides a tool to identify the factors affecting the initiation and progression of cancer. During the last decades, advances in biomedical and genetic research have furthered our understanding of the human genome structure and the genetic code. Since the end of the nineteenth century, our knowledge of genomics has risen remarkably, as Mendel discovered the laws of genetics, followed by the description of the structure of DNA by Watson and Crick in 1953 and the emergence of recombinant DNA-technology and molecular genetics in the 1970s. The Human Genome Project (HGP) started in 1990, and the first phase finished in the year 2001, unmasking the genetic code and offering science a scale of new possibilities to uncover the pathogenesis beyond common diseases, such as cancer.

Usually, the accumulation of genetic changes in tissue happens over decades. In adult human tumours, approximately 33 to 66 genes are somatically mutated, the average in prostate cancer being 41 (Vogelstein *et al.* 2013). Some mutations, termed driver mutations, provide a growth advantage for tumour cells. Other mutations that do not have any direct or indirect effect on cell growth are called passenger mutations, and they are mainly located in non-coding regions of the genome (Vogelstein *et al.* 2013).

For common human disease traits, over 1,200 associated genetic loci have been found, but for cancer, the majority of the heritable traits remain unknown (Lander 2011). Likewise, this "phantom/missing heritability" may be due to genetic interactions and greater biological complexity among the pathways underlying cancer (Zuk *et al.* 2012).

Tumour suppressor genes (TSGs). Germline mutations are inherited, whereas somatic mutations are acquired throughout the lifespan. In tumour biology, mutations are commonly classified as tumour suppressor genes or oncogenes.

Mutations leading to a loss of function usually occur in tumour suppressor genes, which prevent the cell from undergoing cancerous changes. According to the Knudson's two-hit theory, two mutations are needed to inactivate a tumour suppressor gene, either two acquired mutations in the case of sporadic cancer or one germline and one somatic mutation in a hereditary form of cancer (Knudson 1971). Knudson's investigation of retinoblastoma patients led to the identification of the first tumour suppressor gene, retinoblastoma 1 (*RB1*), located at chromosomal region 13q14. Another widely studied example of a tumour suppressor gene is *TP53*, which encodes for the p53 protein, which is crucial for controlling the cell cycle (Matlashewski *et al.* 1984). Tumour suppressor genes studied in prostate cancer include *BRCA1* at 13q14, *KLF5* at 13q21, *PTEN* at 10q23 and *NKX3.1* at 8p21, among which, *PTEN* is the most frequently mutated tumour suppressor gene in prostate cancer, and *KLF5* shows the most frequent hemizygous deletions and loss of expression (Dong 2001). Tumour suppressor genes can also be referred to as "gatekeeper" genes.

Oncogenes. Many human genes regulating cell growth and differentiation normally function as proto-oncogenes. After a single gain-of-function mutation, a proto-oncogene turns into a cancer contributing oncogene that causes uncontrolled cell division (Todd, Wong 1999). Increased protein expression can also be a sign of the activation of a proto-oncogene into an oncogene. Additionally, proto-oncogenes can be activated by chromosomal translocation and gain or amplification of the chromosomal locus. Oncogenes found to act in prostate cancer include *c-MYC* at 8q24, which plays a crucial role in the regulation of cellular proliferation, differentiation and apoptosis, *ERBB2/HER2* at 17q12 and the apoptotic inhibitor, *BCL2* at 18q21 (Isaacs, Kainu 2001).

Genomic alterations. Chromosomal aberrations are widely detected in tumour cells. Tumour cells may also harbour homozygous or heterozygous deletions in tumour suppressor genes, amplification of oncogenic regions, inversions, translocations and other genetic abnormalities (Frohling, Dohner 2008). Using comparative genomic hybridisation (CGH), several alterations have been identified in prostate cancer, including gains at 8q and losses at 3p, 8p, 10q, 13q and 17p (Taylor *et al.* 2010, Dong 2001, Lapointe *et al.* 2007). The most common genomic rearrangement in

prostate cancer is located at 21q.22.3, where all coding exons of the *ERG* are fused to androgen-regulated sequences of the promoter of the prostate-specific *TMPRSS2* gene (Tomlins *et al.* 2005), which contributes to various *TMPRSS2-ERG* fusion transcripts. Although this fusion is frequently detected in prostate tumours, the functional significance of this rearrangement is not fully understood.

Epigenetics. In addition to mutations, human tumours contain many epigenetic changes, including histone and chromatin modifications (Bird 2007). The DNA methylation status changes between different cell types, development stages and patient age (Pelizzola, Ecker 2011), so these unstable events may also represent reasons for the missing heritability in many cancers and may provide new, useful biomarkers for disease progression.

Genetic heterogeneity. Genetic heterogeneity is an important component in cancer progression. It is observed at three levels: the genome level, the gene level and the epigenetic level, with interactions between the different levels (Stevens *et al.* 2001). Vogelstein *et al.* (2013) extended the description of heterogeneity to include intratumoral, intermetastatic, intrametastatic and interpatient levels. Together with epistatic interactions, the investigation of common human traits, such as cancer, has become far more complicated. Novel, more robust methods are needed to explain the heritable factors behind patient risk and to understand the pathways and genetic interactions underlying the disease to develop strategies for personalised diagnosis, therapy and prevention.

2.2 Linkage studies

To uncover the landscapes of genetic variation and inheritance models of quantitative traits, several methodologies have been exploited. In human medical genetics, linkage analysis is used to map the genetic loci of disease susceptibility genes. Evidence for the linkage of a particular chromosomal region is reported as a logarithm of the odds (LOD) (Dawn Teare, Barrett 2005) by statistically comparing the genotypes of affected and unaffected family members. The mutations identified in linkage studies are usually highly penetrant within families and rare in the general population.

Families that are at high-risk for prostate cancer, those with several cases of prostate cancer, have been exploited in linkage studies. The first candidate genes identified by linkage analysis were *ribonuclease L* (*RNASEL*) – *hereditary prostate cancer 1* (*HPC1*) at chromosome 1q25, the *elaC homolog 2* (*ELAC2*) gene at chromosome 17p and the *macrophage scavenger receptor 1* (*MSR1*) at 8p22 (Carpten *et al.* 2002, Tavtigian *et al.* 2001, Xu *et al.* 2002). Since these discoveries, vast linkage analyses have been performed, reporting several susceptibility loci, including 1p36 (*CABP/EPHB2*), 1q42.2-43 (*PCAP*), 8p21-23, 8q24, 16q23, 17p11, 19q13, 20q13 (*HPC20/HPC3*) and Xq27-28 (*HPCX*). However, many of the linkage study results remain controversial, with conflicting evidence due to differential procedures, disease criterions, data limitations and the heterogeneity of prostate cancer. Thus, genetic alterations contributing to hereditary prostate cancer have been extremely hard to identify.

One promising new prostate cancer susceptibility gene, *HOXB13* (*homeobox B13*) lies at chromosomal region 17q21-22, and it was identified by a fine-mapping linkage analysis (Lange *et al.* 2003, Lange *et al.* 2007). By further scanning the genes in this region, Ewing *et al.* (2012) identified a rare but recurrent mutation (G84E) that is important for prostate cancer development. In Finland, an increased risk for prostate cancer was also associated with G84E mutation carriers (Laitinen *et al.* 2013). Shang Z *et al.* (2013) completed a meta-analysis of G84E mutations in 24,213 prostate cancer cases and 73,631 control men, revealing a significant association between the mutation and prostate cancer, and as reported previously,

the mutation frequency was even higher among patients with a younger age at onset. However, the clinical utility of *HOXB13* mutations has not yet been defined.

2.3 Genome-wide association studies

Genome-wide association studies (GWAS) are based on a case-control approach, in where millions of markers are genotyped at a single time (Chung *et al.* 2010, Varghese, Easton 2010). GWAS have been used to investigate common, low-penetrant polymorphisms associated with the risk of complex diseases or other QTL. GWAS assume that genetic predisposition is governed by one allele conferring modest risk. In the case of several alleles, the alleles are usually located near each other in the same chromosome, showing strong correlation due to linkage disequilibrium. This phenomenon allows researchers to study the whole genome without testing all of the known variations.

For prostate cancer, over 77 associating variants or loci have been discovered so far (Amundadottir *et al.* 2006, Gudmundsson *et al.* 2007, Yeager *et al.* 2007, Haiman *et al.* 2007, Gudmundsson *et al.* 2008, Thomas *et al.* 2008, Eeles *et al.* 2008, Yeager *et al.* 2009, Al Olama *et al.* 2009, Gudmundsson *et al.* 2009, Eeles *et al.* 2009, Takata *et al.* 2010, Kote-Jarai *et al.* 2011, Haiman *et al.* 2011, Schumacher *et al.* 2011, Xu *et al.* 2012, Akamatsu *et al.* 2012, Gudmundsson *et al.* 2012, Eeles *et al.* 2013). The most consistent association has been reported in chromosomal region 8q24. However, the biological significance of this approximately 1 Mb region remains unclear (Nakagawa *et al.* 2012).

GWAS reported that variants that are common in nature do not contribute substantially to cancer risk (the odds ratio usually being less than 1.5), and the genetic polymorphisms that have been identified exhibit limited use in predicting risk, response to treatment or response to toxicity. To develop a screening tool to predict prostate cancer risk, Zheng *et al.* (2008) designed a study using 16 SNPs from five previously reported chromosomal regions, in addition to family history, and they determined cumulative associations with prostate cancer. Additionally, a study with nine risk-variants (Helfand *et al.* 2010) and three variants in the Chinese population confirmed the cumulative risk in a dose-dependent manner (Wang *et al.* 2012).

2.4 Next-generation sequencing

Despite the fact that many publications have reported genetic risk factors for prostate cancer; a large proportion of the genetic risk remains unexplained. With the evolution of next-generation, high-throughput technologies, additional highly penetrant genetic mutations will likely be identified for subsets of hereditary prostate cancer families. Next-generation sequencing (NGS) technologies processed by massive parallel sequencing offer a dramatic increase in the speed and volume of genetic data as these sequencing technologies become cheaper and more accessible (Dong, Wang 2012). NGS technologies performed by whole genome (WGS) or whole exome sequencing (WES) are providing novel insights into cancer genomes. The concept of personalised medicine has gained attention throughout the oncology field, aiming for individualised cancer diagnosis and therapy. Technological advances have enabled large international studies, such as the 1,000 genomes project. The aim of this project is to sequence the whole genome of approximately 2,000 individuals from different populations worldwide, including over 100 Finnish individuals, identifying a majority of the genetic variation occurring at a population frequency >1%. The results of this study will help to more precisely identify and study the variation amongst disease progression (Via, Gignoux & Burchard 2010, 1000 Genomes Project Consortium *et al.* 2012).

For prostate cancer, RNA-sequencing has been used to identify novel, recurrent gene fusions, and these genomic rearrangements could serve as personalised biomarkers used for monitoring disease progression (Barbieri, Demichelis & Rubin 2012). Using the WGS data of 32.5 million variants studied from 1,795 Icelanders, a new low-frequency variant (rs188140481) at 8q24 was found that is associated with prostate cancer in European populations (Gudmundsson *et al.* 2012). Targeted NGS has been used to detect genomic alterations and to determine potential therapeutic implications within archival formalin-fixed paraffin embedded (FFPE) tissue biopsies from advanced prostate cancer patients (Beltran *et al.* 2013). Likewise, the study by Ewing *et al.* (2012) that reported an association between the *HOXB13* gene variant and hereditary prostate cancer was originally performed via targeted re-sequencing of the 17q21-22 regions. The same region at 17q, in addition to 2q, has also been re-sequenced within Finnish familial prostate cancer samples (Laitinen V, personal communication).

Efforts to improve early diagnoses and to develop novel therapies have been confounded by significant patient heterogeneity, and there is a need for more accurate NGS methods focusing on a personalised point of view. Recent studies have defined novel molecular subtypes and cancer promoting variants and have characterised extensive genomic aberration underlying disease initiation and progression (Wyatt *et al.* 2013). It is now clear that the heterogeneity observed in the clinic is supported by a complex molecular landscape.

2.5 Integrative analysis

Complex diseases, such as cancer, are caused by a combination of multiple genetic interactions and environmental factors, which are hard to detect and are linked to each other. For prostate cancer, no clear, reproducible variant or locus has been reported, and only a few variants or genes propose clear biological significance or function.

One of the major challenges in the "post-GWAS" era is to elucidate the functional consequences of the variants and loci associated with the disease, drawing attention to detailed mapping technologies, epigenetic landscapes and proper cell and tissue models (Freedman *et al.* 2011). Importantly, most of the reported SNPs found by GWAS reside in intergenic regions, intronic regions, regulatory regions and splice sites (Freedman *et al.* 2011). In particular, the GWAS SNPs located at intronic or intergenic regions show only association and no pathogenicity. This is also the case for prostate cancer (Catalog of Published Genome-Wide Association Studies); thus, it is important to determine the functional consequences of these associated loci. Today, more emphasis is placed on possible regulatory elements and transcriptional activation sites within the genome. A list of all of the functional elements in the human genome was recently published by an international collaboration study, The Encyclopedia of DNA Elements (ENCODE). The results are freely available for download and analysis at http://genome.ucsc.edu/ENCODE/, and hopefully, they will benefit prostate cancer research. The results of the ENCODE Consortium have encountered criticism as scientists have delineated several logical and methodological misconducts involved in determining the functionality of the human genome (Graur *et al.* 2013).

To date, GWAS results explain only a portion of genetic risk, proposing a single-gene model. One emerging factor contributing to the "missing heritability" is the joint effect and interaction among genetic variants of multiple genes or loci, a phenomenon called epistasis (Manolio *et al.* 2009), and gene-gene interactions are thought to assist in the pathogenesis of many complex diseases (Cordell 2009). To detect genetic interactions in complex diseases, it is crucial to include hundreds of thousands of cases and controls to gain enough statistical power (Zuk *et al.* 2012). These gene-gene interaction studies have become more frequent in recent years.

Using a multifactor dimensionality reduction (MDR) analysis program, Xu *et al.* (2005b) identified a four-SNP interaction between Toll-like receptor signalling pathway genes and several cytokines that significantly predicted prostate cancer risk. In another high-dimensional effort, a two-stage, genome-wide gene-gene interaction scan was performed with a novel statistical method, resulting in 16 significant SNP pairs in a population of 1,964 prostate cancer cases and 3,172 controls subjects; however, none of the pairs remained significant after further testing, suggesting the need for repetitive studies (Tao *et al.* 2012). An interaction between apoptotic SNPs and aggressive prostate cancer was studied by different modes of the MDR program, discovering a positive association that was not significant after multiple comparisons (Lavender *et al.* 2012). Moreover, a large-scale interaction study by Ciampa *et al.* (2011) suggested that follow-up studies on the two-gene interaction finding at the 8q24 region are warranted. Furthermore, a recent study revealed that five SNP-SNP interactions within genes that regulate angiogenesis are associated with aggressive prostate cancer (Lin *et al.* 2013). Many methodologies have been developed to diminish the computational burden of detecting gene-gene interactions at the genome-wide scale (Li, Lou & Lu 2012).

To date, the findings of these studies account for only a fraction of the heritable prostate cancer risk, and additional attention has turned to the idea of distal gene regulation. Genome-wide expression quantitative trait loci (eQTL) analysis is a method for studying gene regulation in complex traits, and it is able to detect associations between genotypic and expression data (Cookson *et al.* 2009, Michaelson, Loguercio & Beyer 2009). eQTL analysis is widely used to gain insight into the role of how SNPs affect transcript levels. Gene-SNP associations in *cis* are depicted as a 1 Mb window upstream or downstream of the gene of interest, as most of the *cis*-eQTLs are located within or close to the gene of interest (Veyrieras *et al.*

2008). For prostate cancer, no published data using eQTL analysis have been reported, but there are several ongoing projects (e.g., Laitinen V, personal communication).

2.6 Chromosomal region 13q14 and prostate cancer

One of the most reported sites of the loss of heterozygosity (LOH) in prostate cancer is chromosomal region 13q14 (Dong, Boyd & Frierson 2001, Hyytinen *et al.* 1999, Yin *et al.* 1999, Chen *et al.* 2001). The association of this region with prostate cancer was also reported in a genome-wide association study (Xu *et al.* 2005a). This region also harbors deletions in Finnish sporadic and familial prostate cancer patients (Visakorpi *et al.* 1995, Rokman *et al.* 2001). The LOH of the 13q14-31 region was also reported in 44% of multifocal prostate cancer foci of Japanese patients by comparative genomic hybridisation (CGH) (Kobayashi *et al.* 2008). Using samples from intermediate-risk prostate cancer patients, Ishkanian *et al.* (2009) also identified a deletion in the 13q14.2 region using high-resolution array-CGH. Allelic imbalance (AI) was also detected at the 13q14.2-q14.3 region in patients harbouring localised prostate cancer, and this phenomenon was reported to be associated with biochemical relapse (Brookman-Amissah *et al.* 2007). Allelic imbalance/LOH at 13q14 was found to be different between Japanese and Caucasian men, and the frequency of AI was significantly higher in high Gleason score tumours, suggesting a different biological behaviour of prostate cancer in different populations (Misumi *et al.* 2010). Additionally, it was shown that loss of heterozygosity increases with prostate cancer progression (Lu *et al.* 2006). A region close to 13q14, chromosomal locus 13q13.3, was also found to have a strong linkage to prostate cancer in the Finnish population (Cropp *et al.* 2011).

The identification of deleterious region(s) in the band q14 of chromosome 13 suggests the presence of one or more tumour suppressor genes. The potential candidates involved in prostate cancer that have been reported so far include *CHCL-1* (*chromosomal condensation 1-like*) (Latil *et al.* 2002), which is located at 13q14.2, and a pseudogene, *FAM10A4* (Sossey-Alaoui *et al.* 2002), at 13q14.3 that is homologous to the *ST13* tumour suppressor gene. A study by Dong *et al.* (2006) identified another candidate at 13q14.11, a member of the forkhead transcription factor family, *FOXO1A* (*forkhead box protein O1A*), which was frequently deleted and down-regulated in tumorigenic samples and prostate cancer cell lines. The same area, 13q14.11, as well as region 13q14.12, was focally, homozygously deleted in metastatic prostate tumours, harbouring key genes, including *SIAH3* (*seven in*

absentia homolog 3) and *C13ORF15* (*chromosome 13 open reading frame 15*) (Robbins *et al.* 2011).

MicroRNAs at 13q14. MicroRNAs (miRNAs) are small, typically 20-23 nucleotide RNA molecules, that function in the post-transcriptional regulation of their target gene expression (Bartel 2004). MicroRNAs are encoded by nuclear DNA and are transcribed via precursors into mature molecules. Down-regulation of the target gene is mediated by complementary base pairing with the messenger RNAs of the target genes. The Sanger Institute maintains a database of miRNAs, which now comprises 21,264 different miRNAs (www.mirbase.org, August 2012). A majority of miRNA genes have been mapped to regions of putative cancer susceptibility and at fragile sites, supporting the role of miRNAs in carcinogenesis (Garzon, Calin & Croce 2009, Sevignani *et al.* 2006, Calin *et al.* 2004).

Two microRNAs, miR-15a and miR-16-1, were identified on chromosomal region 13q14 via a deletion in the region 13q14 in chronic lymphocytic leukaemia (CLL) patients (Calin *et al.* 2002). These miRNAs were also frequently deleted or down-regulated in CLL samples. The target gene of these miRNAs is the oncogene, *BCL2* (Cimmino *et al.* 2005). In prostate cancer, *CCND1* (*encoding cyclin D1*) and *WNT3A* (*wingless-type MMTV integration site family, member 3A*), which promote tumour survival, proliferation and invasion, are reported to be target genes for the miR-15a-miR-16-1 cluster (Bonci *et al.* 2008). This miRNA cluster also exhibited a tumour suppressor function, its expression levels were significantly decreased, and it inversely regulates the expression of target genes that are expressed in advanced prostate tumours. Further confirmation of the role of the miR-15a-miR-16-1 locus as a putative tumour suppressor was performed in Finland by detecting homozygous deletions in a subset of prostate cancers (Porkka *et al.* 2011). This cluster resides in the proximity of the *retinoblastoma 1* (*RB1*) gene, but there has been no correlation between the loss of the *RB1* and deregulation of miR-15a and miR-16-1, indicating the independent function of the cluster (Aqeilan, Calin & Croce 2010).

3. The *ARLTS1* gene

A promising candidate gene for prostate cancer is *ARLTS1* (*ADP-ribosylation factor-like tumour suppressor gene 1*), also known as *ARL11* (*ADP-ribosylation factor-like 11*), which is located at locus 13q14. *ARLTS1* is composed of two exons, and the open reading frame of the second exon codes for a 196 amino acid protein, with a molecular mass of 21 kDa. The *ARLTS1* gene was originally identified by Calin GA *et al.* (2005) while investigating at chromosomal region linked to chronic lymphocytic leukaemia (Bullrich *et al.* 2001). The *ARLTS1* gene harbours genetic polymorphisms, including the nonsense mutation, G446A (Trp149Stop), which is strongly associated with the propagation of general familial cancers (Calin *et al.* 2005). This substitution to a stop codon prematurely terminates translation, resulting in a truncated protein of 148 amino acids. The first paper on this gene was a comprehensive investigation of *ARLTS1* variants and its association with many cancers in different populations. Since then, the contribution of *ARLTS1* variants to various cancers in different cohorts has been investigated. The results regarding the most promising risk predictor variants G446A and T442C (Cys148Arg), are summarised in Table 2 and Table 3, respectively.

The G446A variant is significantly associated with breast cancer (Akisik, Yazici & Dalay 2011). The *ARLTS1* variant, T442C, is significantly associated with high-risk familial breast cancer (Frank *et al.* 2006b), both sporadic and familial colorectal cancer (Castellvi-Bel *et al.* 2007) and familial ovarian cancer (Yang *et al.* 2009). Moreover, in the ovarian cancer study, the investigators found higher rates of advanced stage tumours in *ARLTS1* T442C carriers. The *ARLTS1* T442C variant is also associated with an increased risk of early-onset ovarian cancer (Yang, Yu & Peng 2009). In colorectal cancer, chronic lymphocytic leukaemia or basal cell carcinoma of the skin, no association was found with any of the *ARLTS1* variants (Frank *et al.* 2006a, Sellick, Catovsky & Houlston 2006b, Ng *et al.* 2007, Li *et al.* 2007, Sellick, Catovsky & Houlston 2006a). For melanoma, G446A did not affect risk, but the heterozygous T442C form was associated with a statistically significant risk of cancer (Frank *et al.* 2006). A study by Masojc *et al.* (2006) reported no differences in the frequency of the G446A variant in breast, prostate, laryngeal,

thyroid papillary cancer or melanoma within cases and controls in Poland, and the G446A polymorphism was not significantly more frequent in CFA (cancer familial aggregation) cases, except for families in which the proband had melanoma. The G446A mutation was not found to be a low-risk ovarian cancer susceptibility allele, but *in silico ARF6* (*ADP-ribosylation factor 6*) homology protein modelling with the Trp149Stop resulted in a dysfunctional, truncated protein product that lacked the fragment responsible for GTP binding (Masojc *et al.* 2007). The other *ARLTS1* polymorphisms found in the studies were C65T (Ser22Leu), G297A (Ser99Ser), G396C (Lys132Lys), G490A (Glu164Lys), A580C (Lys195Gln) and C392T (Pro131Leu), but none of them were associated with familial cancer in any study. *ARLTS1* haplotypes have not been established either in CLL (Sellick, Catovsky & Houlston 2006b), colorectal cancer (Castellvi-Bel *et al.* 2007) or basal cell carcinoma of the skin (BCC) (Li *et al.* 2007). According to *in silico* predictions, the *ARLTS1* variants, especially C392T and T442C, have putative phenotypic effects (Castellvi-Bel *et al.* 2007).

ARLTS1 belongs to the family of ARF/ARL (ADP-ribosylation factor/ARF-like) guanine-nucleotide-binding proteins from the Ras superfamily, and it acts in multiple regulatory pathways (Wennerberg, Rossman & Der 2005). The *ARLTS1* sequence harbours significant homology with other ARF and ARL members, with the closest homology to *ARL14*, *ARL4D* and *ARL4C* (Yendamuri *et al.* 2007). The ARF family contains at least 30 members and three classes of proteins: the ARFs, the ARLs and SARs (Secretion-associated and Ras-related) (Yendamuri, Trapasso & Calin 2008). In mammalian cells, six ARFs have been identified. These small, approximately 20 kDA GTPases function on membrane surfaces as molecular switches between two forms; the active guanosine-5'-triphosphate (GTP) and the inactive guanosine diphosphate (GDP), which are regulated by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) (Bernards 2003). ARF-related GTPases regulate membrane and vesicular trafficking (Chavrier, Goud 1999). The inactive form of ARF is soluble and associates weakly with membranes as the active ARF binds tightly to the membrane (Donaldson, Jackson 2000). ARF proteins are highly conserved and have been found on all eukaryotic organisms, suggesting an important cellular function. All ARFs harbour a conserved G box GTP/GDP binding element. These proteins apply a two residue shift to reach an active conformation, which is the most important functional feature of the family

(Kahn *et al.* 2006). *ARLTS1* is expressed most abundantly in lymphatic tissues, such as lymph nodes, spleen and bone marrow.

ARLTS1 acts as a tumour suppressor gene. It functions in apoptotic signalling, increasing apoptosis, inducing growth arrest and inhibiting proliferation upon the re-expression of a fully functional protein product in cancer cells (Castellvi-Bel *et al.* 2007, Yendamuri *et al.* 2007, Petrocca *et al.* 2006, Yang *et al.* 2009). The tumour suppressor function of *ARLTS1* was demonstrated by Calin GA *et al.* (2005), who found that the transduction of full-length *ARLTS1* into A549 cells in Nu/Nu mice decreased tumour growth compared to the empty vector. The ability of *ARLTS1* to suppress tumour formation in preclinical models was also observed with ovarian (Petrocca *et al.* 2006) and lung cancer cells (Yendamuri *et al.* 2007). *ARLTS1* is widely down-regulated in chronic lymphocytic leukaemia (Calin *et al.* 2005), ovarian cancer (Petrocca *et al.* 2006) and lung cancer samples (Yendamuri *et al.* 2007). In a study of ovarian carcinomas and ovarian and breast cancer cell lines, the majority of *ARLTS1* down-regulation was due to DNA methylation of its promoter region (Petrocca *et al.* 2006). Additionally, the same phenomenon was found in lung carcinomas and lung cancer cell lines (Yendamuri *et al.* 2007). The 5' promoter region CpG islands are differentially methylated in tumour and normal tissues, and methylation correlated with *ARLTS1* expression levels. When the tumour cell lines were treated with a demethylating agent (decitabine), *ARLTS1* expression increased (Yendamuri, Trapasso & Calin 2008). The apoptotic function of *ARLTS1* and its ability to regulate apoptosis-related proteins was confirmed by the fact that it synergises with chemotherapeutic agents in the ovarian cancer cell line SKOV3 (Yang, Yu & Peng 2011).

Inherited susceptibility constitutes most of the familial aggregation of cancer, but highly penetrant mutations in known genes do not exclusively account for the risk. Low-penetrant alleles, such as those observed in the *ARLTS1* gene, may cause some genetic susceptibility, and any gene with relevant tumour promoting functions is a strong candidate as a low-penetrance susceptibility allele (Houlston, Peto 2004). Thus, it is important to choose a candidate gene based on the current knowledge of its role in cancer biology. However, there might also be unidentified pathways that influence tumour development.

Table 2. Summary of the *ARLTS1* variant G446A association studies

Cancer type	Country; ethnicity	Size*	OR (95% CI)	p-value	Assoc.	Reference
Familial cancers	France, USA; Caucasian	109; 475	5.7 (1.3 - 24.8)	0.02 [sporadic] vs. [familial]	Yes	Calin GA et al., NEJM 2005
breast, prostate+melanoma, CLL, pancreatic/melanoma						
Sporadic tumors	Italy, USA, Denmark, Romania; Caucasian	219; 475	-	-	No	Calin GA et al., NEJM 2005
thyroid, colorectal, BRCA, CLL, lung, idiopathic pancytopenia						
Familial BRCA	German; Caucasian	482; 530	1.58 (0.75 - 3.34)	0.23 [GA + AA] vs. [GG]	No	Frank B et al., Int J Cancer 2006
High-risk familial BRCA	German; Caucasian	305; 530	2.20 (1.02 - 4.74)	0.04	Yes	Frank B et al., NEJM 2005
Bilateral BRCA	German; Caucasian	46; 530	4.11 (1.27 - 13.31)	0.011 [GA] vs. [GG]	Yes	Frank B et al., Int J Cancer 2006
Breast	Poland; Caucasian	9; 8	1.26 (0.5 - 3.3)	0.81	No	Masoj´c B et al., Breast Cancer Res 2006
Prostate		5; 8	1.0 (0. 3 - 3.1)	1.0		
Melanoma		6; 8	1.2 (0.4 - 3.5)	0.78		
Thyroid		3; 8	1.1 (0.3 - 4.3)	0.74		
Laryngeal		8; 8	1.9 (0.7 - 5.0)	0.29		
Breast cancer	Turkey; Caucasian	147; 120	-	<0.00001 [GA] vs. [GG]	Yes	Akisik E et al., Mol Biol Rep 2011
CLL	British; Caucasian	413; 471	2.61 (0.87 - 7.90)	- [GA+AA] vs. [GG]	No	Sellick GA et al., Leuk Res 2006
CLL	USA; Caucasian	31; 100	6.83 (0.6 - 78)	0.14 [GA+AA] vs. [GG]	No	Ng et al., Br J Haematol. 2007
BCC	Hungary, Romania, Slovakia; Caucasian	528; 533	1.5 (0.6 - 4.0)	0.41	No	Li X et al., Hered Cancer Clin Pract. 2007
Colorectal cancer	German; Caucasian	611; 538	0.97 (0.46 - 2.04)	0.94 [GA+AA] vs. [GG]	No	Frank B et al., Cancer Letters 2006
Colorectal cancer	Spain; Caucasian	515; 515	1.20 (0.52 - 2.81)	0.66 [GA+AA] vs. [GG]	No	Castellvi-Bel S et al., Carcinogenesis 2007
Melanoma	German; Caucasian	351; 804	0.83 (0.37 - 1.88)	0.65 [GA+AA] vs. [GG]	No	Frank B et al., Int J Cancer 2006
Ovarian cancer	Poland; Caucasian	250; 552	0.83 (0.2 - 3.1)	1.0 [GA+AA] vs. [GG]	No	Masoj´c B et al., Cancer Res 2007
**Cases; controls*						

Table 3. Summary of the *ARLTS1* variant T442C association studies

Cancer type	Country; ethnicity	Size*	OR (95% CI)	p-value	Assoc.	Reference
Familial cancers	France, USA; Caucasian	109, 272	-	-	No	Calin GA et al, NEJM 2005
breast, prostate+melanoma, CLL, pancreatic/melanoma						
Sporadic tumors	Italy, USA, Denmark, Romania; Caucasian	216; 272	-	-	No	Calin GA et al, NEJM 2005
thyroid, colorectal, BRCA, CLL, lung, idiopathic pancytopenia						
Familial BRCA	German; Caucasian	482; 530	1.48 (1.10–1.99)	0.009 [CC + TC] vs. [TT]	Yes	Frank B et al., Int J Cancer 2006
			1.72 (1.20–2.45)	0.003 [CC] vs. [TT]	Yes	
High-risk familial BRCA	German; Caucasian	305; 530	1.47 (1.04–2.06)	0.03 [CC + TC] vs. [TT]	Yes	Frank B et al., Int J Cancer 2006
			1.75 (1.16–2.63)	0.07 [CC] vs. [TT]	No	
Bilateral BRCA	German; Caucasian	46; 530	1.29 (0.63–2.67), 0.49	0.49 [CC + TC] vs. [TT]	No	Frank B et al., Int J Cancer 2006
CLL	German; Caucasian	416; 471	0.96 (0.70 - 1.32)	- [CT] + [TT] vs. [CC]	No	Sellick GA et al. Leuk Res 2006
CLL	USA; Caucasian	31; 100	0.68 (0.3 - 1.56)	0.39 [CC+TC] vs. [TT]	No	Ng et al., Br J Haematol. 2007
BCC	Hungary, Romania, Slovakia; Caucasian	528; 533	0.9 (0.6 - 1.2)	0.64	No	Li X et al., Hered Cancer Clin Pract. 2007
Colorectal cancer	German; Caucasian	611; 538	1.17 (0.89 - 1.53)	0.27 [CC + TC] vs. [TT]	No	Frank B et al. Cancer Letters 2006
Colorectal cancer	Spain	515; 515	1.45 (1.13 - 1.86)	0.003 [TT + CC] vs. [TC]	Yes	Castellvi-Bel S et al., Carcinogenesis 2007
			1.47 (1.10 - 1.98)	0.009 [CC+ TC] vs. [TT]	Yes	
Melanoma	German	351; 804	1.43 (1.95 - 1.95)	0.02 [TC] vs. [TT]	Yes	Frank B et al., Int J Cancer 2006
			1.32 (0.99 - 1.77)	0.06 [CC + TC] vs. [TT]	No	
Ovarian cancer	Chinese	165; 120	-	0.0031 [familial] vs. [sporadic]	Yes	Yang XY et al., Int J Gynecol Cancer 2009
			-	0.012 [familial] vs. [controls]	Yes	Yang XY et al., 2009
Ovarian cancer, early onset	Chinese				Yes	(Sichuan Da Xue Xue Bao Yi Xue Ban.)

**Cases; controls*

AIMS OF THE STUDY

The overall aim of this study was to provide new knowledge of prostate cancer genetics by studying a previously found cancer susceptibility gene and a prostate cancer locus.

The specific aims were as follows:

- To investigate the role of the potential tumour suppressor gene, *ARLTS1*, in inherited cancer susceptibility in Finland (I).

- To further validate the role of *ARLTS1* in prostate cancer susceptibility (II).

- To characterise the *ARLTS1* gene and study *ARLTS1* expression, regulation and interactions in prostate cancer (III).

MATERIALS AND METHODS

1. Human subjects (I-III)

All samples are of Finnish origin. The familial samples used had at least one affected first or second degree relative, and at least one patient per family was included in the genotyping or sequencing studies. The different samples used in Studies I-III are summarised in Table 4.

Table 4. Summary of cancer and control samples used in Studies I - III.

Sample group	**Study I**	**Study II**	**Study III**
Familial prostate cancer cases	164	212	102
° *Tumour tissue samples*	-	3	-
° *Lymphoblastoid cell lines*	-	24	-
Unaffected control subjects from prostate cancer families	-	-	33
Unselected prostate cancer patients	377	1848	-
° *Tumour tissue samples*	-	17	-
BPH patients	-	389	-
° *BPH tissue samples*	-	14	-
Familial breast cancer patients	598	-	-
Unselected breast cancer patients	644	-	-
Familial colorectal cancer patients	93	-	-
Unselected colorectal cancer cases	148	-	-
Male population controls	809	760	-
Prostate cancer cell lines	-	7	-
Normal prostate epithelial cell lines	-	2	-
Xenografts	-	19	-

1.1 Familial prostate cancer cases (I-III)

Prostate cancer family samples were collected at the University of Tampere and the Tampere University Hospital. Collection started in 1995, and the families were

identified by several methods, including a questionnaire, nation-wide registry based searches, referrals from urologists in Finland and advertisements in newspapers, television and radio (Schleutker *et al.* 2000). Diagnoses were confirmed by the Finnish Cancer Registry, regional hospitals and parish records. In Study II, additional 48 familial prostate cancer patient samples were added to the Study I analysis. In Study III, 24 prostate cancer cases were overlapping the Study II population.

1.2 Unselected prostate cancer cases (I, II)

Unselected, consecutive prostate cancer cases comprised the population-based collection of patients. These patients were diagnosed with prostate cancer between 1999 and 2005 in the Department of Urology at the Tampere University Hospital, which is a regional referral centre for all patients with prostate cancer. Clinical data were collected from hospital records.

1.3 Breast cancer cases (I)

Blood samples from newly diagnosed breast cancer cases were collected at the Department of Oncology at the Helsinki University Central Hospital (HUCH) (Finland) from 1985-1993 and 1997-1998, and samples were collected at the Tampere University Hospital (Finland) from 1997-1999 (Eerola *et al.* 2000, Syrjakoski *et al.* 2000). These cohorts encompassed 82% of all breast cancer patients who visited the aforementioned clinics during the study period. The genealogy of the families was confirmed through church parish registries and the Population Register Centre. The cancer diagnoses of the patients and all traced relatives were confirmed through hospital records from the Department of Oncology, HUCH and through the Finnish Cancer Registry. The patients were asked to fill out a questionnaire on breast, ovarian, and other cancers in their family. Only patients reporting breast and/or ovarian cancers in first- or second-degree relatives were considered to have a family history of the disease, and patients with a strong family history of breast or ovarian cancer were referred for genetic counselling and a possible mutation screening to exclude the *BRCA1* and *BRCA2* variants.

1.4 Colon cancer cases (I)

Familial and unselected DNAs from a colorectal adenocarcinoma patient set were collected between 1994 and 1998 at nine large regional hospitals in southeastern Finland (Aaltonen *et al.* 1998, Salovaara *et al.* 2000). The patients' family histories were gathered by identifying all first-degree relatives in the official population registries, and the cancer diagnoses in the relatives were verified through the Finnish Cancer Registry and the Finnish HNPCC registry.

1.5 Patients with benign prostatic hyperplasia (II)

The benign prostatic hyperplasia (BPH) samples were collected from the Pirkanmaa Hospital District. Diagnosis of the BPH cohort was based on lower-urinary tract symptoms, free uroflowmetry, and evidence of increased prostate size, as determined by palpation or transrectal ultrasound. If serum-based free prostate specific antigen (PSA) was elevated and if transrectal ultrasound or digital rectal examination showed any indications of cancer, the patients underwent biopsies. The biopsies were examined to exclude diagnoses of prostate cancer, high-grade prostate intraepithelial neoplasia (PIN), atypical small acinar cell proliferation (ASAP) or suspicion of malignancy. The cut-off for biopsy was total PSA of ≥4 ng/ml or total PSA of 3.0-3.9 mg/ml with the proportion of free PSA <6%.

1.6 Healthy control individuals (I-III)

Voluntary Finnish anonymous male blood donors served as healthy population controls in Studies I and II. Samples were obtained from the Blood Center of the Finnish Red Cross in Tampere and Helsinki. The EDTA blood samples used in Study II were from healthy anonymous donors. In Study III, the control individuals were healthy male family members belonging to the studied families (n=31).

1.7 Clinical tissue samples (II)

Clinical prostate specimens were obtained from the Tampere University Hospital (Tampere, Finland). Samples included freshly frozen prostate tumour specimens that

included benign prostate hyperplasia (BPH, n=14), androgen-dependent (n=14) and castration resistant (n=6) carcinomas. The BPH samples were obtained from prostatectomy cancer patients, and it was histologically verified that they did not contain any cancerous cells. Castration resistant carcinoma samples were obtained from transurethral resections of prostate (TURP) from patients experiencing urethral obstruction, regardless of on-going hormonal therapy. The time from the beginning of hormonal therapy to progression varied from 15 to 60 months. Histological examination and H&E staining were used to determine the presence of tumour cells. Only samples containing >60% cancerous or hyperplastic epithelial cells were selected for the analyses. For direct sequencing, 53 DNA samples from clinical prostate tumours were analysed.

1.8 Ethical considerations (I-III)

Patient information and samples were obtained from all of the participants who gave full written informed consent. Permission to collect the family data of prostate cancer families throughout Finland and to use Finnish Cancer Registry data was granted on 20.6.1995 by the Ministry of Social Affairs and Health (license number 59/08/95). Permission to collect and use prostate cancer patients' blood samples and clinical data in the Pirkanmaa Hospital District was granted by the Institutional Review Board of the Tampere University Hospital (license number 95062). The use of blood samples and clinical data from prostate cancer patients at the Hatanpää City Hospital was permitted by the Institutional Review Board of the City of Tampere (license number 8595/403/2005). Tumour sample collection and their use for medical research purposes was granted by the National Authority for Medicolegal Affairs (license number 5569/32/300/05). The breast and colon cancer studies in Helsinki were performed with appropriate research permissions from the Ethics Committees of the Departments of Obstetrics, Gynecology and Oncology (Helsinki University Central Hospital, Finland), as well as from the Ministry of Social Affairs and Health in Finland. Collaborative studies performed in the United States were granted by the Institutional Review Board of the National Human Genome Research Institute (National Institutes of Health, USA.)

2. Human cell lines and xenografts (II)

The human prostate cancer cell lines, LNCaP, DU145, PC-3, NCI-660 and 22Rv1, were obtained from the American Type Culture Collection (Manassas, VA, USA). The prostate cancer cell lines, LAPC-4 and VCaP, were kindly provided by Dr. Charles Sawyers (UCLA, Los Angeles, CA) and Dr. Jack Schalken (Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands), respectively. Primary prostate epithelial cells, PrECs, were obtained from Lonza (Walkersville, MD, USA). EP156T is an h-tert immortalised normal prostate epithelial cell line that was kindly provided by Dr. Olli Kallioniemi (Institute for Molecular Medicine Finland FIMM, Helsinki, Finland). All cell lines were maintained under their respective recommended cell culture conditions.

Lymphoblastoid cell lines were derived by Epstein-Barr virus transformation of peripheral mononuclear leucocytes from familial prostate cancer patients. Lymphoblastoid cell lines were grown in RPMI-1640 medium (Lonza, Walkersville, MD, USA) supplemented with 10% foetal bovine serum (Sigma-Aldrich, St. Louis, MO, USA) and antibiotics.

Human prostate cancer xenografts were kindly provided by Dr. Robert L. Vessella (Department of Urology, University of Washington/Puget Sound VA Medical System, Seattle, Washington, USA) (Laitinen *et al.* 2002, Saramaki *et al.* 2006b). Nineteen human prostate cancer xenografts of the LuCaP series were used for direct sequencing, and 16 of them were used in Q-RT-PCR experiments.

3. Mutation screening (I-III)

Mutation analyses of genomic DNA were performed by direct sequencing. The coding region of *ARLTS1* was sequenced in the genomic DNA from prostate cancer, breast cancer, colorectal cancer and healthy control samples, as well as from prostate cancer cell lines, clinical prostate tumours and xenografts. The following *ARLTS1* primers were used for sequencing: TGCACCAGGATGGATTTGTC (forward 5'->3') and GATTTCAGGAGGCTCCACAG (reverse 5' -> 3'), which

enable the investigation of all *ARLTS1* variants in a single amplicon. PCR products were purified in 96-well Acro Prep Filter Plates (Pall Life Sciences, Ann Arbor, MI) using the Perfect Vac Manifold vacuum machine (Eppendorf AG, Hamburg, Germany) or rAPid Alkaline Phosphatase (Roche, Mannheim, Germany) and Exonuclease I (Fermentas, Vilnius, Lithuania).

Sequencing was performed using the ABI PRISM® BigDye™ Terminator Cycle Sequencing Ready Reaction Kit combined with the Applied Biosystems 3130xl Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer's instructions. Sequences were analysed with different versions of Sequencher software (GeneCodes Corporation, Ann Arbor, MI).

4. Genotyping (II, III)

To further investigate the role of the *ARLTS1* SNP T442C in prostate cancer, the genotyping in Study II was performed with the Custom TaqMan® SNP Genotyping Assay using the ABI Prism 7900HT sequence detection system (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer's instructions. Primers and probes for the T442C SNP (rs3803185) were supplied by Applied Biosystems using File Builder 3.1 software.

The genome-wide SNP genotyping in Study III was performed using the HumanOmniExpress BeadChip microarray (Illumina, Inc., San Diego, CA, USA) by the Technology Centre, Institute for Molecular Medicine Finland (FIMM), University of Helsinki.

5. Quantitative real-time reverse transcription-PCR (II)

To analyse the mRNA expression status of *ARLTS1*, total RNA was isolated from lymphoblastoid cell lines, prostate cancer cell lines, BPH samples, clinical tumour samples and xenografts using the Tri-Pure reagent and protocol (Roche, Mannheim, Germany), Qiagen RNeasy MiniKit (Qiagen, Inc., Valencia, CA) or TRIzol reagent (Invitrogen Life Technologies, Carslbad, CA, USA). RNA from cell lines was

reverse transcribed into first-strand cDNA using the SuperScript™ III First-Strand Synthesis SuperMix kit (Invitrogen, Carlsbad, CA, USA) and random hexamers. RNA from clinical tumour samples and xenografts was reverse transcribed as described previously (Linja *et al.* 2001, Saramaki *et al.* 2006a). For the control samples, normal human prostate RNA was obtained from Ambion (Cambridgeshire, United Kingdom). Real-time expression analyses were performed using LightCycler® (Roche, Mannheim, Germany) and the Bio-Rad CFX96™ Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). For the LightCycler® analyses, primers and probe sets were obtained from TIB MolBiol (Berlin, Germany). The PCR reactions were performed with a LightCycler® FastStart DNA Master^{PLUS} HybProbe kit (Roche, Mannheim, Germany). LightCycler® software (Roche, Mannheim, Germany) was used for data analysis. Additionally, the TaqMan assay for *ARLTS1* expression was used. The primers and probes were obtained from TIB MolBiol (Berlin, Germany). Bio-Rad's iQ Supermix was used for the PCR reactions, and CFX Manager Software™ version 1.6 was used for data analysis. *ARLTS1* expression levels were normalised against the housekeeping gene, *TBP* (TATA box binding protein), or RNA levels.

6. Western blotting (II)

ARLTS1 protein expression was analysed by western blotting. Anti -ARL11 (Aviva Systems Biology, San Diego, CA, USA) and anti-actin (pan, clone ACTN05, Neomarkers, Fremont, CA, USA) antibodies were used in this study. Cells were lysed in a buffer containing 50mM Tris-HCl pH 7.5, 150mM NaCl, 0.5% Triton-X-100, 1mM DTT, 1mM PMSF and 1x complete protease inhibitor cocktail (Roche, Mannheim, Germany). The lysates were sonicated 4x30s with a Bioruptor instrument (Diagenode, Liege, Belgium), and the cellular debris was removed by centrifugation. Proteins were separated by polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membranes (Immobilon-P; Millipore, Billerica, MA, USA). Antibodies were detected using HRP-conjugated secondary antibodies (DAKO, Denmark) and Western blotting luminol reagent (Santa Cruz Biotechnologies, CA, USA) by autoradiography.

7. Array comparative genomic hybridisation (II)

The chromosomal aberration status of prostate cancer cell lines and xenografts was analysed by array comparative genomic hybridisation (aCGH) using the Human Genome CGH Microarray Kit 244A (Agilent, Santa Clara, CA, USA), as described by Saramäki OR *et al.* (2006a).

8. Expression quantitative trait loci (eQTL) analysis (III)

Total RNA was extracted from whole blood samples from 84 prostate cancer cases and 15 healthy male relatives from the same families. The samples were collected in PAXgene® Blood RNA Tubes (PreAnalytiX GmbH, Switzerland / Qiagen/BD) using the MagMAX™ for Stabilised Blood Tubes RNA Isolation Kit (Ambion®/Life Technologies, Carlsbad, CA, USA) and the PAXgene Blood miRNA Kit (PreAnalytiX GmbH, Switzerland/Qiagen/BD). The RNA quality was assessed using the Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA). Library preparation, target enrichment and massive parallel paired-end sequencing of the expressed transcripts was performed by the Beijing Genomics Institute (BGI Hong Kong Co., Ltd., Tai Po, Hong Kong) using Illumina HiSeq2000 technology (Illumina Inc., San Diego, CA, USA).

To obtain *ARLTS1* expression values, the RNA sequencing reads were aligned with tophat2 using hg19 as the reference genome (Trapnell, Pachter & Salzberg 2009). The raw read count for *ARLTS1* was calculated using HTseq, and the read counts were transformed to normalised expression values using DESeq (www-huber.embl.de/users/anders/HTSeq/, (Anders, Huber 2010). To investigate the expression quantitative trait loci (eQTL), the linear regression model implemented in PLINK was used to detect transcript specific variants. Associations in *cis* were delineated by a 1 Mb window upstream or downstream of the *ARLTS1* SNP rs9526582. In addition, a new method based on probabilistic indices was used to test for eQTL. This new method is a non-parametric directional test (similar to the well-known Jonckheere-Terpstra test), implemented in our R-Package GeneticTools

which is available on the Comprehensive R Archive Network (http://cran.r-project.org/package=GeneticTools), and a package description is under development (unpublished data). P-values were calculated using permutation tests and adjusted for multiple testing by applying the Benjamini-Hochberg correction.

9. Bioinformatics (II, III)

Co-expression analysis. The *ARLTS1* gene was also studied via *in silico* analysis. An *ARLTS1* co-expression signature was generated from the GeneSapiens (Kilpinen *et al.* 2008) database of mRNA expression. *ARLTS1* expression was studied in 1,587 tumour and 497 normal samples. A Pearson correlation was used to identify genes that positively and negatively correlated with *ARLTS1* among normal samples and tumour samples. For *in silico* analysis of functional *ARLTS1*, gene sets were explored to identify enrichment of functional annotation categories, such as gene ontology (GO), using tools within the EASE program (http://david.abcc.ncifcrf.gov/) (Huang da, Sherman & Lempicki 2009b). The expression levels of the most relevant genes that correlated with *ARLTS1* were checked for enrichment using Gene Ontology. In Study III, co-expression analysis consisted of meta cell line data (n=1,445) originating from over 48 human anatomical parts. A correlation value of >0.30 and a p-value <0.05 were used to determine statistically significant associations. Multiple test corrections were performed using the Benjamini Hochberg and Bonferroni methods.

***In silico functionality prediction*.** In Study III, the RegulomeDB (http://regulome.stanford.edu/) (Boyle *et al.* 2012) and HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php)(Ward, Kellis 2012, ENCODE Project Consortium 2011) databases were used to further elucidate the role of eQTLs in gene regulation.

MDR analysis. The multifactor dimensionality reduction (MDR) program (version 2.0_beta_8.4) was used to examine gene-gene interactions. The MDR analysis program is publicly available from the internet (www.epistasis.org). MDR is a non-parametric, model-free data mining approach and constructive induction algorithm

(Hahn, Ritchie & Moore 2003). MDR selects one genetic model (a one, two, three or four stage locus) that most successfully predicts the phenotype or disease status (e.g., cancer). The *ARLTS1* genotypes from 102 prostate cancer cases and 33 controls were combined with the Illumina Human OmniExpress GWAS database of 700,000 SNPs. Genes functioning in immune-system processes that were included in the MDR analysis were selected from a previously published comprehensive collection made by Loza MJ *et al.* (Loza *et al.* 2007). In short, SNPs involved in apoptosis, cytokine signalling, Toll-like receptor signalling, leukocyte signalling, complement, adhesion and natural killer cell signalling were included in the analysis. The total amount of SNPs was 12,011, of which 4,764 were found in our Illumina Human OmniExpress GWAS data.

10. Statistical analysis (I, II)

After *ARLTS1* mutation screening in breast, prostate and colon cancer cases, the χ^2-test or Fisher's exact test were used to examine a possible association between *ARLTS1* genetic alterations and cancer risk. Each variant within the following groups was analysed using the Haploview program (www.broad.mit.edu/mpg/haploview): prostate cancer cases and controls, breast cancer cases and controls, and breast cancer and prostate cancer cases together versus controls. The case/control data were also analysed using STATA by conducting Pearson's w2-test and Fisher's exact test. Multinomial logistic regression was used to further analyse the prostate cancer data, incorporating WHO (World Health Organization) clinical scores. In addition, logistic regression was used to calculate odd ratios (ORs) and the corresponding 95% confidence interval.

The association of the *ARLTS1* T442C variant with prostate cancer risk in Study II was tested by logistic regression analysis using the SPSS statistical software package (SPSS 15.0; SPSS Inc, Chicago, IL, USA). Associations with the clinical and pathological features of the disease (age at onset, PSA value at diagnosis, T, N and M-stage, WHO grade and Gleason score) were tested among familial and unselected prostate cancer cases using the Kruskal–Wallis test, Fisher's exact test, and t-test included in the R "Stats" package. Prostate cancer patients were classified as having aggressive disease if they had any of the following characteristics: a

locally advanced or metastatic tumour (stage T3, T4, N1 or M1, based on pathology if radical prostatectomy was performed; otherwise, clinical stage), tumour Gleason grade at diagnosis ≥7, poorly differentiated grade (WHO grade III, if no Gleason grade available) or pretreatment PSA at diagnosis ≥20 ng/ml.

RESULTS

1. The contribution of *ARLTS1* variants to familial cancer risk (I)

ARLTS1 was analysed as a risk factor for familial cancers in 2005 (Calin *et al.* 2005). Calin GA *et al.* (2005) found that the nonsense mutation, G446A, was associated with the risk of different familial cancers. To determine the role of *ARLTS1* in Finnish familial and unselected cancer cohorts, we performed mutation screening of the coding region of the *ARLTS1* gene in breast, prostate and colorectal cancer sample sets. Altogether, five sequence variants were found: the previously reported nonsense variant, Trp149Stop (G446A), a synonymous variant, Ser99Ser, two missense variants, Pro131Leu (C392T) and Cys148Arg (T442C), and a novel missense variant, Gly65Val (G194T). The genotype frequencies of the three most important variants in the prostate cancer samples are summarised in Table 5.

Table 5. Association of three *ARLTS1* variants with prostate cancer cases

Variant and sample	**Carrier freq.**	**OR**	**95% CI**	***p***
G446A (Trp149Stop) GA+AA				
Population controls	13/809 (1.6%)	1.00		
Prostate cancer cases, familial	3/164 (1.8%)	1.14	0.32 - 4.05	0.340
Prostate cancer cases, unselected	2/377 (0.5%)	0.33	0.07 - 1.45	0.165
Prostate cancer cases, all	5/541 (0.9%)	0.57	0.20 - 1.61	0.340
T442C (Cys148Arg) CC				
Population controls homozygotes	121/809 (15.0%)	1.00		
Prostate cancer cases, familial	37/164 (22.6%)	1.66	1.09 - 2.51	0.020
Prostate cancer cases, unselected	76/377 (20.2%)	1.44	1.05 - 1.97	0.029
Prostate cancer cases, all	113/541 (20.9%)	1.50	1.13 - 1.99	0.005
G194T (Gly65Val) GT+TT				
Population controls	13/809 (1.6%)	1.00		
Prostate cancer cases, familial	8/164 (4.9%)	3.14	1.28 - 7.70	0.016
Prostate cancer cases, unselected	11/377 (2.9%)	1.84	0.82 - 4.15	0.182
Prostate cancer cases, all	19/541 (3.5%)	2.23	1.09 - 4.55	0.028

Additionally, breast and colorectal cancer samples were analysed. The CC genotype of the T442C variant was slightly more common in both familial and sporadic breast cancer patients compared to controls, but it was similar in prostate cancer cases. However, after adjustments for multiple testing, none of the results remained significant. Additionally, no associations were found with any of the variants and colorectal cancer risk.

Haplotype association with one haploblock including the C392T and T442C variants was detected with Haploview. The haplotype consisted of the common C allele at C392T and the rare C allele at T442C, and it was significantly associated with breast (p=0.032) and prostate cancer (p=0.037). When breast and prostate cancer cases were considered as a combined data set, the CC haplotype showed a nominally significant association with familial (p=0.040), sporadic (p=0.034) and all cancer cases (p=0.017).

2. Contribution of the *ARLTS1* T442C variant to prostate cancer susceptibility (II)

By utilising the previous data from Study I and genotyping more familial and unselected prostate cancer cases, a statistically significant association between the *ARLTS1* T442C variant and prostate cancer risk was validated (Table 6). Significant association was found when comparing the frequency of the homozygous risk allele C to the frequency of the homozygous wild-type allele T. No association was found between the frequency of *ARLTS1* T442C CC and the 375 BPH samples genotyped.

Table 6. Association of the *ARLTS1* T442C variant CC homozygotes with prostate cancer specimens

Sample	Carrier freq.	OR	95% CI	*p*
Population controls	115/760 (15.1%)	1.00		
Prostate cancer cases, familial	50/212 (23.6%)	1.67	1.08-2.56	0.019
Prostate cancer cases, unselected	387/1848 (20.9%)	1.52	1.18-1.97	0.001
Prostate cancer cases, all	437/2060 (21.2%)	1.54	1.20-1.98	0.0007
Clinical prostate tumours	15/53 (28.3%)	-	-	-
Xenografts	8/19 (42.1%)	-	-	-

Direct sequencing was used to examine the frequencies of *ARLTS1* variants in clinical prostate tumours. In the clinical prostate carcinomas, we found variants at four sites: Gly65Val (G194T), Pro131Leu (C392T), Cys148Arg (T442C) and Trp149Stop (G446A). At the T442C site, the frequency of the cancer associated risk genotype CC was relatively high compared to the combined frequency of familial and unselected cases (Table 6). The highest frequency of the *ARLTS1* T442C CC genotype was detected in the xenograft samples (Table 6).

3. Clinical significance of the *ARLTS1* variants (I, II)

When WHO, Gleason score, PSA and T-stage were included in the initial analyses of the prostate cancer data in Study I, the CC genotype at the T442C locus was nominally associated with a high Gleason score (p=0.01) and high PSA at diagnosis (p=0.05) among the unselected prostate cancer cases. When clinical characteristics (age at diagnosis, PSA value at diagnosis, T, N, and M-stage, WHO grade and Gleason score) were included in the analyses of additional germline genotype data in Study II, the CC genotype was significantly associated with a younger age at diagnosis and higher PSA value (p=0.05 and p=0.03, respectively) among the unselected cases. A statistically significant association (p=0.02) was detected when the correlation between aggressive disease status and the occurrence of the *ARLTS1* T442C CC genotype was examined among unselected prostate cancer cases. Within familial prostate cancer cases, no association was found between the CC genotype at the T442C locus and any of the clinical variables.

4. Segregation of *ARLTS1* variants (I, II)

For segregation analysis, additional samples were genotyped from prostate cancer families carrying the three *ARLTS1* variants G446A, T442C and G194T. Three prostate cancer families carrying the G446A variant were analysed, revealing currently unaffected elderly relatives who were hetero- or homozygous for the protein truncating alteration, indicating that the Trp149Stop mutant allele did not co-segregate with disease status (Figure 3A).

Analysis of seven prostate cancer families carrying the G194T variant revealed no clear co-segregation of the variant with prostate cancer phenotype in families that carried the variant (Figure 3B and 3C).

When studying the segregation of the T442C CC genotype in 86 families in which the index was detected as a carrier, complete segregation of the variant CC with the cancer phenotype was seen in one family (Figure 3D). In the rest of the families, segregation was incomplete (e.g., Family 408 in Figure 3C), indicating a clear association between the *ARLTS1* T442C genotype CC or TC and prostate cancer.

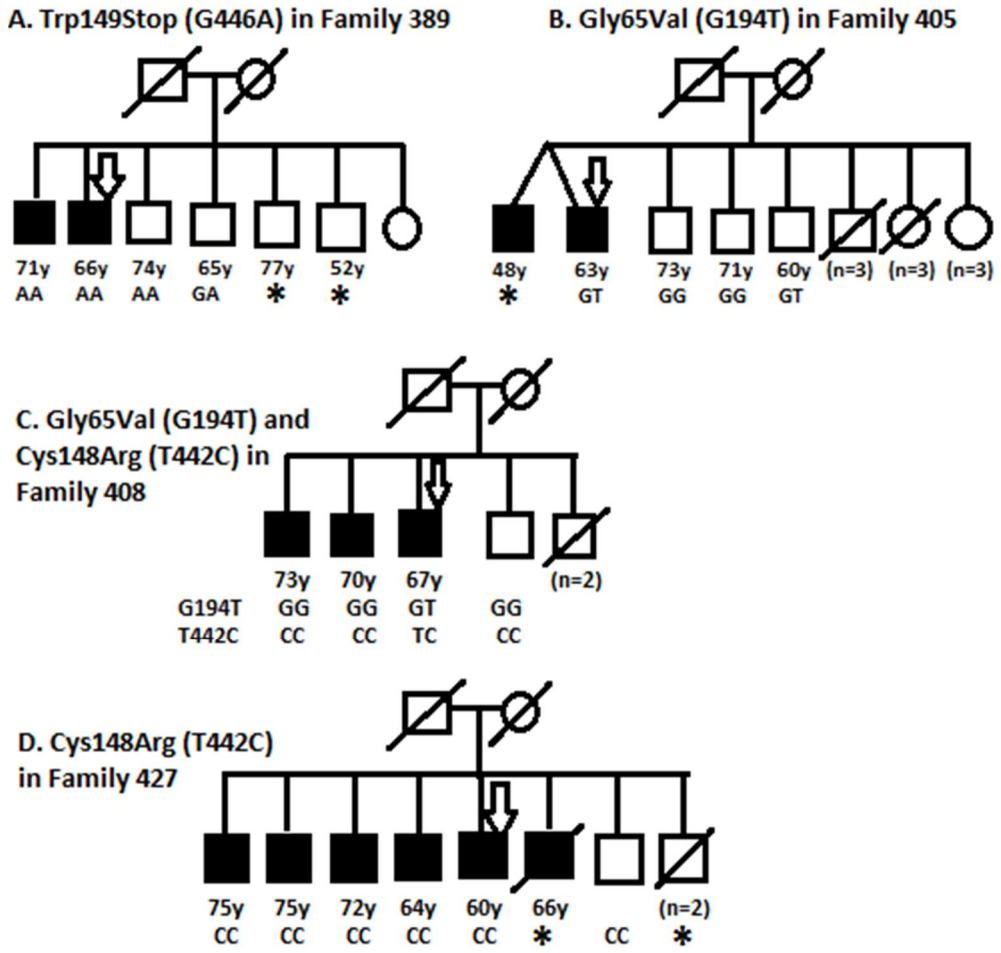


Figure 3. Example of the segregation patterns in one *ARLTS1* G446A variant positive (**A**), two G194T positive (**B** and **C**) and two T442C positive (**C** and **D**) families with prostate cancer. Squares represent males; circles represent females. Open symbols indicate no neoplasm, and filled symbols denote prostate cancer cases. An arrow indicates the individual initially screened for *ARLTS1* sequence variants. Age at diagnosis or current age of the unaffected members for prostate cancer patients (in years) is indicated below the symbol for each family member. An asterisk (*) denotes persons with no available sample.

5. *ARLTS1* RNA expression patterns (II, III)

To investigate the role of the T442C genotype in *ARLTS1* expression, we selected 24 familial patients, including eight from each genotype group, TT, CT and CC. Q-RT-PCR expression analyses were carried out using RNA extracted from lymphoblastoid cell lines obtained from prostate cancer patients. *ARLTS1* mRNA was differentially expressed between the three genotype groups, and it was significantly reduced among patients carrying the CC genotype (CC vs. TT, p=0.02).

In addition, *ARLTS1* RNA expression levels were analysed in 16 LuCaP xenograft samples, six prostate cancer cell lines, two prostate epithelial cell lines and RNA from a normal prostate sample. Among the analysed xenograft samples, two of the samples (LuCaP77 and LuCaP23.12) (13%) displayed a significant reduction in *ARLTS1* expression compared to the normal prostate sample. Total loss of expression was seen in 12 (75%) of the xenograft samples. Two of the xenograft samples, LuCaP23.1 and LuCaP49, had higher *ARLTS1* expression levels compared to normal tissue. Six of the fourteen (43%) *ARLTS1* down-regulated xenograft samples harboured the genotype CC for the T442C variant. Furthermore, loss of the chromosomal region 13q14.3 was detected in 12/15 (80%) xenograft samples, as determined by array comparative genomic hybridisation (aCGH).

Decreased or lost expression of *ARLTS1* was observed in 5/6 (83%) prostate cancer cell lines compared to normal prostate RNA. An *ARLTS1* locus deletion at 13q14.3 was found in PC-3, VCaP, LNCaP and DU-145 cell lines. Among these, the latter also displayed the CC genotype for T442C. Interestingly, the T442C variant was the only *ARLTS1* variant found in any of the prostate cancer cell lines, and expression was lost in the prostate epithelial cell line EP156T. These results indicate that *ARLTS1* expression is low in prostate cancer epithelial cells.

Additionally, *ARLTS1* expression was assessed by Q-RT-PCR in 14 BPH tissue samples, 14 primary and six castration resistant tumour samples. *ARLTS1* expression was significantly lower (p=0.01) in clinical tumours compared to BPH samples.

To explore germline expression patterns, *ARLTS1* RNA expression levels were analysed in total RNA from the whole blood of 84 prostate cancer cases and 15

controls (healthy relatives from the same families). A significant decrease in *ARLTS1* expression was detected in prostate cancer cases (p=0.0037), correlating with the results from the cell line, benign prostatic hyperplasia (BPH) and tumour specimen RNA expression data.

6. ARLTS1 protein expression (II)

ARLTS1 protein expression was determined by western blotting. Cell lysates were obtained from six prostate cancer cell lines (22Rv1, LAPC-4, PC-3, VCaP, LnCaP, DU-145) and the normal prostate epithelial cell line EP156T. ARLTS1 protein expression was highly convergent with *ARLTS1* mRNA status. As expected, based on mRNA expression, LnCaP, DU-145 and EP156T cells showed negative ARLTS1 protein expression. In the prostate cancer cell lines 22Rv1 and PC-3, ARLTS1 protein expression corresponded with mRNA expression values whereas the protein level in LAPC4 cells was lower than the observed *ARLTS1* mRNA expression level. The prostate cancer cell line VCaP showed the highest ARLTS1 protein expression, which did not correspond to the negative RNA expression status observed by Q-RT-PCR analysis.

7. Loss of heterozygosity in clinical prostate tumours (II)

Next, we studied loss of heterozygocity in clinical prostate tumours. Germline DNA from prostate cancer patients carrying the *ARLTS1* T442C variant was analysed by direct sequencing. Sequencing results showed that in two cases, the T allele in whole blood samples was lost in prostate cancer tissue, suggesting a role for *ARLTS1* in tumour suppression (i.e., inactivation of both alleles in carcinogenesis). When clinical parameters (WHO grade, Gleason and T-scores) were examined, one of the two variant-carrying “LOH-patients” had a Gleason score of 9 and a WHO grade of 3, indicating an aggressive form of the disease. The other case had a WHO grade of 3, but no Gleason score was available.

8. eQTL analysis (III)

To identify potential *cis*-acting genetic variants associated with *ARLTS1* transcript levels, eQTL analysis was performed within a selected area of the 13q14 region. Using a linear regression model (PLINK), we were able to detect five eSNPs that affected *ARLTS1* expression. When the calculation window was diminished from 1 Mb to 200 kb, only one SNP, rs7997737, remained. Applying a directional test performed with the R-Package analysis tool, 11 statistically significant eSNPs were identified, and two correlated with the PLINK linear regression model results. The eSNPs found by directional testing were located mainly in non-coding regions. The genomic locations of the eSNPs found in the 13q14 region are displayed in Figure 4. Altogether, 468 genomic variants within the 1 Mb region originating from *ARLTS1* SNP rs9526582 were tested by a linear regression model and a directional test. A statistically significant association in allele specific expression was observed between five SNPs and *ARLTS1* transcript levels, with p-values of 0.011 (rs2075610), 0.004 (rs7997737), 0.008 (rs1543513), 0.001 (rs9568232) and 0.000 (rs9568354). After adjusting for multiple testing, only one eSNP from the directional test remained significant (rs9568354).

The functionality and potential transcriptional regulatory effects of the 14 eSNPs found by eQTL were evaluated using ENCODE data in the RegulomeDB and HaploReg databases. Altogether, nine of the fourteen eQTLs were included in the RegulomeDB database. There was substantial evidence for the regulation of *ARLTS1* for most of the eSNPs as they were located in regulatory regions, transcription factor binding sites, adapted with histone marks or regions shoving open chromatin states with enhancer and regulatory characteristics. Thus, the *ARLTS1* eQTLs are located within regulatory areas of the 13q14 region.

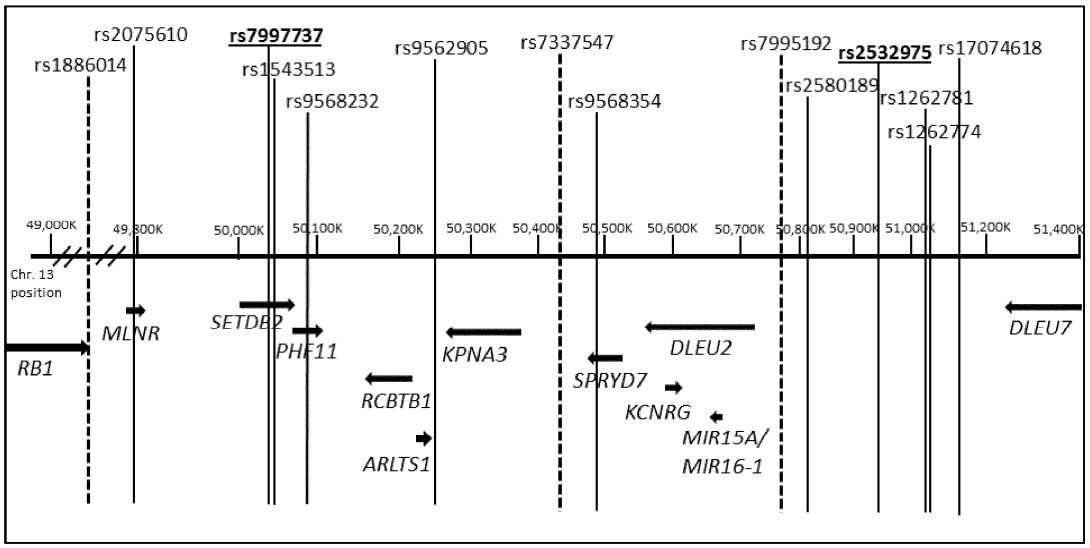


Figure 4. Schematic diagram showing the genomic locations of the eSNPs gathered by eQTL analysis in familial prostate cancer patients, in relation to *ARLTS1*. The gene symbols are as follows: *RB1* (*retinoblastoma 1*), *MLNR* (*motilin receptor*), *SETDB2* (*SET domain, bifurcated 2*/*CLLD8*), *PHF11* (*PHD finger protein 11/NY-REN-34 antigen*), *RCBTB1* (*regulator of chromosome condensation [RCC1] and BTB [POZ] domain containing protein 1*/*CLLD7*), *ARLTS1* (/*ARL11, ADP-ribosylation factor-like 11*), *KPNA3* (*karyopherin alpha 3, importin alpha 4*), *SPRYD7* (*SPRY domain containing 7*/*C13orf1, chromosome open reading frame 1*), *KCNRG* (*potassium channel regulator*), *DLEU2* (*deleted in lymphocytic leukaemia 2*), *MIR-15A* and *MIR16-1* (*microRNA genes 15a and 16-1*) and *DLEU7* (*deleted in lymphocytic leukaemia 7*). Dashed line, results from linear regression model; solid line, results from directional test. SNPs overlapping in both tests are denoted by bold font and are underlined.

9. *ARLTS1* co-expression studies (II, III)

In Study II, *ARLTS1* co-expression signatures were generated from prostate cancer specimens (n=1,587), normal samples (n=497) and human cell lines (n=1,445; from 48 cancer subtypes). The results from co-expression analyses within tumour specimens illustrate a strong association with immune system processes, revealing an enrichment score for the cluster of 6.67 and a p-value of 2.1e-7. These results were generated from the top 100 genes, with a correlation value greater than 0.3

among normal samples. Among cancer samples, immune system processes had the highest enrichment score, 14.89, with a p-value of 5.3e-21 and an adjusted Benjamin score of 2.8e-17. These findings were confirmed using a separate dataset from expo (IGC: Expression Project for Oncology, 2008 [http://www.intgen.org/expo.cfm]). Our results illustrate that 56% of the genes were commonly identified to positively correlate with *ARLTS1* in either dataset composed of cancer samples. Genes identified to intersect with a positive correlation above 0.3 yielded a very strong gene ontology enrichment score of 18.12 for immune system processes with a p-value of 2.1e-24. Gene expression among prostate cancer specimens was very low for *ARLTS1*.

In the cell line cohort (n=1,445, Study III) from the GeneSapiens database, 1,381 genes with correlation value >0.30 and a p-value <0.05 were found to be positively correlated with the *ARLTS1* gene. DAVID GO functional clustering (http://david.abcc.ncifcrf.gov/tools.jsp) (Huang da, Sherman & Lempicki 2009b, Huang da, Sherman & Lempicki 2009a) of the prostate cancer cell line cohort revealed a strong association with nuclear and zinc-finger protein processes when all the genes positively correlating with *ARLTS1* were taken into account (with a more stringent correlation value >0.50). Additionally, nuclear processes (nucleus, intracellular organelles, transcription and DNA-binding) were enriched. The enrichment score was 13.49, with a p-value of 1.1e-32. The adjusted Benjamin score was 5.1e-30, and the cluster of zinc-finger binding proteins displayed a p-value of 9.0e-22. The same phenomenon was observed within the entire set of cell line data (meta cohort), with genes showing a correlation value >0.30. *ARLTS1* co-expression genes (correlation value >0.50) from the meta cell line data revealed that immune system processes (B-/T-cell activation, leukocyte/lymphocyte differentiation and activation) displayed an enrichment score of 2.94 (p-value 5.57e-7, adjusted Benjamin score 2.9e-5).

ARLTS1 co-expression data regarding genes negatively correlated with *ARLTS1* (n=2,722) with correlation value <–0.50 revealed strong gene ontology with glycoproteins (enrichment score 52.13, p-value 3.4e-72) and plasma membrane proteins (enrichment score 38.35, p-value 7.9e-32) in prostate cancer cell lines. Within the negatively correlated genes, a cluster of immunoglobulin domain containing proteins harboured an enrichment score of 12.23, with a p-value of 5.4e-

24. The GO term “cytokine activity” revealed an enrichment score of 10.43, with a p-value of 6.5e-10 within negatively correlating genes.

10. MDR analysis (III)

The 102 familial prostate cancer cases and 33 healthy control individuals included in the Illumina genotyping were sequenced, and altogether, six *ARLTS1* variants were detected at the same amplicon. All the variants were previously reported (rs117251022, rs3803186, rs147120792, rs3803185, rs138452698 and G446A [Trp149Stop]). To investigate genotypic *ARLTS1* interactions, the multifactor dimensionality reduction method was used. Gene-gene interaction status between the *ARLTS1* T442C variant and genes functioning in immune system processes (Loza MJ *et al.* 2007) was calculated, but no statistically significant *ARLTS1* interactions were found in this study cohort.

DISCUSSION

The aim of prostate cancer genetics is to identify genetic changes linked to the risk of developing the disease. In recent years, extensive research on prostate cancer aetiology and genome-wide association studies revealed several common low penetrance genetic alterations. The association of these variants with the clinicopathologic features of prostate cancer and patient prognosis remains unclear, and in general, the results lack clinical implications.

There is major evidence of a hereditary component contributing to prostate cancer. The tumour suppressor gene *ARLTS1* is associated with the risk of familial cancer, and the mechanisms contributing to cancer susceptibility and altered *ARLTS1* expression may be through apoptosis and methylation, respectively (Yendamuri *et al.* 2007, Petrocca *et al.* 2006, Yang, Yu & Peng 2011). In this thesis, the contribution of *ARLTS1* variants to Finnish familial cancers were examined, and the role of the *ARLTS1* T442C substitution in prostate cancer susceptibility was investigated. Additionally, *ARLTS1* expression and regulation patterns were explored.

1. The contribution of *ARLTS1* variants to prostate cancer susceptibility (I, II)

Trp149Stop. It was determined that the *ARLTS1* nonsense variant G446A contributes to familial, but not sporadic cancer (Calin *et al.* 2005). Additionally, a positive association between the G446A variant and breast cancer was found in three distinct studies (Frank *et al.* 2006b, Frank, Klaes & Burwinkel 2005, Akisik, Yazici & Dalay 2011) (Table 2). However, the association between G446A and prostate cancer risk was not confirmed in this thesis study or in studies of colorectal cancer, melanoma, chronic lymphocytic leukaemia or basal cell carcinoma of the skin (Castellvi-Bel *et al.* 2007, Frank *et al.* 2006a, Sellick, Catovsky & Houlston 2006b, Ng *et al.* 2007, Li *et al.* 2007, Frank *et al.* 2006, Masojc *et al.* 2007) (Table 2). The study by Masojc *et al.* (2006) reported no differences in the frequencies of

the G446A variant in breast, prostate, laryngeal, thyroid papillary cancer or melanoma within cases and controls in Poland. Furthermore, the G446A polymorphism was not significantly more frequent in cancer familial aggregation cases, except for families in which the proband had a melanoma. Interestingly, the G446A variant was completely absent from both the clinical tumour samples and xenograft samples included in this thesis (Study II), strengthening our claim that it is not important for the tumorigenesis of prostate cancer.

Cys148Arg. Previously, the *ARLTS1* Cys148Arg (T442C) variant was found to be significantly associated with familial and high-risk familial breast cancer (Frank *et al.* 2006b), both sporadic and familial colorectal cancer risk (Castellvi-Bel *et al.* 2007), familial ovarian cancer risk (Yang *et al.* 2009) and melanoma (Frank *et al.* 2006) (Table 3). In the ovarian cancer study, investigators also found higher rates of advanced stage tumours in *ARLTS1* T442C carriers. The *ARLTS1* T442C variant was also associated with increased risk of early-onset ovarian cancer (Yang, Yu & Peng 2009). However, there was no association with this risk allele and bilateral BRCA (Frank *et al.* 2006b), CLL (Sellick, Catovsky & Houlston 2006b, Ng *et al.* 2007), BCC (Li *et al.* 2007) or colorectal cancer (Frank *et al.* 2006a) (Table 3). Additionally, in the original study by Calin GA *et al.* (2005), no association was found between this allele and familial or sporadic cancer.

In Study I of this thesis, the role of the Cys148Arg substitution in familial, sporadic and pooled breast, prostate and colorectal cancer risk was investigated. When the C allele of the T442C variant was assumed to be recessive, a significant association was observed between the CC genotype and prostate cancer risk (OR 1.50, 95% CI 1.13-1.99, p=0.005) for all data (familial and unselected) (Table 5). In the follow-up study (Study II), more samples were genotyped for *ARLTS1* T442C, and we observed a statistically significant association with prostate cancer risk among both familial (OR 1.67, 95% CI 1.08-2.56, p=0.019) and unselected prostate cancer patients (OR 1.52, 95% CI 1.18-1.97, p=0.001) (Table 6). In the combined data set encompassing both familial and unselected cases, the overall risk was increased further (OR 1.54, 95% CI 1.20-1.98, p=0.0007), confirming the role of *ARLTS1* T442C as a prostate cancer predisposing variant. In addition to the improved statistical results, increasing the amount of samples increased the power to detect an association from approximately 94% to 100%. Further follow-up with an

even larger sample size would likely make the association stronger, and it would also be necessary to validate the clinical importance of the marker.

Interestingly, only the homozygous form of the C allele of the T442C variant was associated with prostate cancer risk compared to the wildtype genotype. These data support the results of previous studies on breast cancer (Frank *et al.* 2006b), which showed that C is the risk allele, whereas the T allele has a protective or neutral effect on carcinogenesis. Furthermore, T442C was the only *ARLTS1* variant that was present at a greater frequency among prostate cancer tissue samples and xenografts, with frequencies of 28.3% and 42.1%, respectively. These data indicate a role for *ARLTS1* T442C in prostate carcinogenesis. No association was observed between this variant and BPH, supporting the theory that BPH and prostate cancer are independent events or at least that *ARLTS1* is not consequential for BPH transformation to prostate cancer. Thus, our results strongly suggest that the homozygous CC genotype of the *ARLTS1* T442C variant contributes to prostate cancer susceptibility.

Gly65Val. The novel *ARLTS1* variant, the Gly65Val (G194T) substitution, may have an effect on prostate cancer risk. The G194T alteration was seen at higher frequency in familial prostate cancer patients compared to healthy controls (OR 3.14, 95% CI 1.28-7.70, p=0.016) (Table 5). Controversially, no risk effect was seen in unselected cases. Further studies are necessary to validate the significance of this result. This residue is of interest because it is highly conserved both among species and among the gene family (www.ensembl.org). In addition, functional analyses indicate that it may be important for the communication of the N-terminus of the protein with the nucleotide binding site (Pasqualato, Renault & Cherfils 2002). Taken together, these results suggest that the *ARLTS1* G194T variant increases familial prostate cancer risk in the Finnish population, although it would be interesting to see whether its importance is increased in a larger cohort of Finnish familial prostate cancer patients and other populations.

Overall, the studies performed to date have reported controversial results regarding *ARLTS1* variants and cancer risk. This is likely due to a low number of familial patients with mixed ethnic backgrounds (population stratification) from different cancers, especially in the original study in year 2005 (Calin *et al.* 2005). In

that study, five heterogeneous populations were pooled, possibly inflating the Type I error rate.

Population stratification indicates the presence of allele frequency differences between subpopulations within a population due to different ancestry (Freedman *et al.* 2004); the cases and controls are not matched for their genetic background, masking the actual associations. In this thesis study, we focused on identifying genetic risks for prostate cancer in the Finnish population. It is known that Finns are an isolated, homogenous founder population that is suitable for genetic studies (Kristiansson, Naukkarinen & Peltonen 2008), and these types of genetic isolates are beneficial to clarify the missing heritability beyond complex diseases (Zuk *et al.* 2012). Additionally, data obtained from the diseased individuals were collected from well-defined and accurate national registries. However, the population controls used in sequencing studies were from a non-characterised population of anonymous male blood donors. Blood donors indicated that they were healthy, but we cannot guarantee their health status. Thus, this may affect the results of this thesis study.

Although functional studies provide evidence that impairment or loss of *ARLTS1* function is important for cancer progression (Calin *et al.* 2005, Petrocca *et al.* 2006, Yendamuri *et al.* 2007), the role of *ARLTS1* mutations in an inherited susceptibility to familial cancer at the functional and phenotypic levels is still controversial. In this study, one aim was to conduct further functional studies on the T442C variant. We attempted to clone full-length *ARLTS1* with the TT wildtype into a lentiviral vector for subsequent transfection into the DU145 cell line, which has the CC genotype (harbouring also 13q14 deletion) to monitor cell growth, differentiation and viability. However, the prolonged, diverse cloning study ended with no outcome from the cultivation of the bacteria (with incorrect products), suggesting a possible toxicity or interference due to a bacterial immune response against *ARLTS1*.

2. *ARLTS1* expression status and prostate cancer (II, III)

ARLTS1 expression was reduced in most of the clinical prostate cancer samples, xenografts and cell lines compared to BPH samples, as determined by quantitative reverse-transcriptase-polymerase-chain-reaction assay (Q-RT-PCR). By studying the whole blood-derived RNA of familial prostate cancer patients, we verified that

ARLTS1 expression was reduced, demonstrating an effect of germline alteration in the *ARLTS1* expression levels for the first time. This result is consistent with previous reports that *ARLTS1* expression is very low or absent in lung carcinomas, CLL cells (Calin *et al.* 2005, Yendamuri *et al.* 2007), ovarian primary tumours and cell lines (Petrocca *et al.* 2006) compared with their normal counterparts. A tumour suppressor function of the *ARLTS1* gene was demonstrated by Calin GA *et al.* (2005), who found that the transduction of full-length *ARLTS1* into A549 cells in Nu/Nu mice decreased tumour growth compared to the empty vector. The ability of *ARLTS1* to suppress tumour formation in preclinical models was also observed in ovarian (Petrocca *et al.* 2006) and lung cancer cells (Yendamuri *et al.* 2007). However, the previous results are based on somatic mutations in cancerous cell lines, whereas our whole-blood data reveals a novel expression difference at the germline level. In the future, these findings could be used to enable screening and detection of at-risk patients, even before clinical diagnosis. However, it is of note that in this thesis, the amount of whole blood-derived RNA samples was low, and the result should be validated in a larger sample cohort. Additionally, the samples were not collected in a controlled manner (e.g., with restrictions on diet and/or medication), and there might be expression differences in the blood at different time points.

Analysis of ARLTS1 protein levels in prostate cancer cell lines was, for the most part, convergent with *ARLTS1* mRNA levels. However, in one sample with the 13q14.3 deletion (the VCaP cell line), high protein expression was detected. This divergent result may originate for several reasons, and it warrants further investigation. The *ARLTS1* tumour suppressor gene may have been duplicated, or it may have pseudogenes. This inconsistent result could also arise from technical reasons due to deficiencies in our probe-based Q-RT-PCR method. Nevertheless, the VCaP deletion has not been confirmed as heterozygous, and these protein expression results further indicate that *ARLTS1* expression is low in prostate cancer epithelial cells.

There are various potential mechanisms underlying the reduced or absent *ARLTS1* expression, including promoter hypermethylation or deletions in the gene region. In lung cancer cell lines and ovarian carcinomas, *ARLTS1* is down-regulated due to DNA methylation in its promoter region (Yendamuri *et al*. 2007, Petrocca *et al.* 2006). Additionally, *ARLTS1* is down-regulated due to promoter

hypermethylation in sporadic cases of B-CLL and lung cancers (Calin *et al.* 2005). Petrocca F *et al.* (2006) showed that *ARLTS1* is frequently down-regulated in ovarian tumours and cell lines, and restoring *ARLTS1* expression by expression of adenoviral ARLTS1 or treatment with the demethylating agent 5-AZA-deoxycytidine (5-AZA) effectively induced apoptosis in vitro and suppressed ovarian cancer tumorigenity in nude mice (Petrocca *et al.* 2006). In this thesis, we confirmed the reduction or loss of *ARLTS1* expression in prostate cancer specimens (including whole-blood RNA, prostate cancer cell lines, xenografts and tumours). *ARLTS1* expression was also significantly reduced in lymphoblastoid cell line samples from familial prostate cancer patients carrying the CC genotype. In fact, during the course of this thesis, bisulphite sequencing was used to examine whether promoter methylation is responsible for the down-regulation of *ARLTS1* in prostate cancer. First, Q-RT-PCR was used to examine the expression status of *ARLTS1* in eight previously 5-AZA/TSA treated and demethylated prostate cancer cell lines. Preliminary studies suggested that TSA alone was able to restore *ARLTS1* expression in the LnCaP prostate cancer cells and the normal prostate epithelial cell line PrEC. TSA and 5-AZA were both required to restore *ARLTS1* expression in PC3 and LAPC4 prostate cancer cell lines, revealing that both DNA hypermethylation and histone modifications may regulate *ARLTS1* expression. However, bisulphite sequencing of the CpG islands of the promoter region was not successful. This result was supported by the lack of official CpG islands in the promoter region of *ARLTS1* in different CpG databases, which is contrary to previous reports in lung and ovarian cancers, as detailed above. However, DNA methylation and histone modifications may regulate *ARLTS1* expression in prostate cancer, but additional studies are required to confirm this suspicion.

We demonstrated that LOH exists in prostate cancer samples, indicating that heterozygous, somatic deletions may affect *ARLTS1* expression levels. *ARLTS1* resides at 13q14.3, a region that was reportedly deleted in a variety of hematopoietic and solid tumours (Calin *et al.* 2005), including prostate cancer.

Gene expression can also be regulated by nearby variants. eQTL analysis of whole blood-derived RNA from familial prostate cancer patients revealed that 14 eSNPs, located within the 13q14 region, significantly influence *ARLTS1* transcript levels. Lymphoblastoid cell lines (LCLs) are widely used in eQTL studies because they are an easily accessible source of patient samples of a single cell type.

Particularly for prostate cancer, large numbers of tumour specimens from multiple foci are difficult to collect. Using prostate cancer specimens, it is possible to find tissue-specific genetic effects, but the use of lymphoblastoid cell lines or whole blood from familial prostate cancer patients enables the detection of possible heritable germline differences that may contribute to prostate cancer susceptibility. It has been argued that SNP-transcript approaches using LCLs from small sample sets are underpowered (Min *et al.* 2011), but many studies have identified overlapping eQTLs in cell lines and primary tissues (Bullaughey *et al.* 2009, Ding *et al.* 2010, Nica *et al.* 2011). Additionally, data from the ENCODE consortium indicated that the lymphoblastoid cell line GM12878 is one of three cell lines in the higher priority Tier 1 cohort (ENCODE Project Consortium *et al.* 2012), which assists with interpreting the results collected from LCLs.

The eSNPs gathered by eQTL analysis were mainly located in intergenic, non-coding genomic regions that have the characteristics of regulatory elements and transcription factor activity. Importantly, most of the reported GWAS SNPs significantly affecting prostate cancer risk resided in intergenic or intronic regions upstream and downstream of genes or in regulatory regions and splice sites (Catalog of Published Genome-Wide Association Studies). Thus, our results are in agreement with existing evidence that the noncoding regions of the human genome are important for prostate cancer progression and susceptibility. However, the eQTL results must be confirmed with functional studies in prostate cancer cell lines.

3. *ARLTS1* and immune system processes (II, III)

Chronic or recurrent inflammation have been implicated in the initiation and development of several human cancers, including those of the stomach, liver, colon, and urinary bladder, and a role for chronic inflammation in the aetiology of prostate cancer has been proposed (De Marzo *et al.* 2007). Previously, it was shown that *ARLTS1* induces apoptosis. In the case of prostatic inflammation and antigen engagement, the need for apoptosis increases, and the processes of the immune system, together with endogenous inflammatory cells (i.e., T and B lymphocytes), become activated. Here, *ARLTS1* expression was significantly decreased among CC variant carriers (compared to the wild-type allele T carriers) in the lymphoblastoid

cell lines of familial prostate cancer patients. These data suggest that the function of *ARLTS1* is altered due to the risk genotype, leading to decreased apoptosis. Thus, the *ARLTS1* T442C CC genotype may contribute to the immune response by diminishing apoptosis, decreasing defence mechanisms and cancer progression. Lymphoblastoid cell lines represent an essential component of the immune response, so our findings indicate that *ARLTS1* functions in immune system processes. Indeed, we observed a strong correlation between *ARLTS1* expression and immune system processes using bioinformatic *in silico* –procedures, so functional studies in the laboratory are necessary to confirm this connection.

To further investigate whether the *ARLTS1* gene is involved in gene-gene interaction networks with genes involved in immunological procedures, we performed MDR analysis. However, no genotypic markers significantly interacted with *ARLTS1* within the 4,764 SNPs analysed. One possible reason for this negative result may be the relatively limited sample set, consisting of only 135 cases and controls.

Like all other organs, the normal prostate contains endogenous inflammatory cells and immune response mechanisms. Consequently, *ARLTS1* may not function exclusively as a tumour suppressor but also on immune response functions that occur either locally in the prostate or in peripheral tissues.

Next, we analysed expression quantitative trait loci results from the entire 13q14 region, encompassing 1 Mb up- and downstream of the *ARLTS1* gene. Interestingly, five of the eSNPs found by eQTL analysis are located within the protein-coding genes *SETDB2, PHF11, SPRYD7* and *MLNR*. Two eQTLs were positioned in the *SETDB2* (*SET domain, bifurcated 2*) gene, also known as *CLLD8* (*chronic lymphocytic leukaemia deletion region gene 8 protein*), which functions mainly in epigenetic regulation (Zhang *et al.* 2003). The *PHF11* gene is a positive regulator of Th1-type cytokine gene expression through nuclear factor kappa B (NF-kB) (Clarke *et al.* 2008), and it is highly expressed in T-cells (especially in Th1- compared to Th2-cells), B-cells, natural killer cells and mature dendritic cells (Clarke *et al.* 2008). A splice variant utilising an alternative exon 1 of the *PHF11* gene was expressed in immune-related tissues and cells, and several alleles of the gene were associated with serum IgE levels and asthma (Zhang *et al.* 2003). The *SPRYD7* gene contains a SPRY/B30.2 protein domain, which is present in a variety of cellular proteins, mediates protein-protein interactions and negatively regulates cytokines

(Nicholson, Hilton 1998, Woo *et al.* 2006). The identified eSNPs did not affect the expression of the host gene, suggesting that these eSNPs act as specific regulators of *ARLTS1* expression. However, no explicit conclusions about the actual causal interaction between these variants and *ARLTS1* can be made without further functional validation. The eQTL result presented in this study was from a small sample set, so further functional validation would be warranted only after eQTL analysis was performed on a much larger sample set of prostate cancer patients. Inflammation remains a complicated phenomenon that displays convergent results in regard to prostate cancer susceptibility, indicating the need for future studies to further elucidate the connection between the two.

4. The susceptibility locus 13q14 and prostate cancer

Loss of heterozygosity (LOH) is a frequent event in breast, prostate and colorectal cancer. The 13q14 locus, where *ARLTS1* resides, is among the most frequently deleted chromosomal regions, especially in prostate cancer, as the LOH of 13q14 has been detected in both sporadic (Latil *et al.* 2003, Valeri *et al.* 2005) and hereditary cancer (Rokman *et al.* 2001). Interestingly, LOH in the same area was also reported in a recent multi-centre genome-wide linkage search for new prostate cancer susceptibility genes in families with at least five affected members, suggesting that this locus may be linked to high penetrant predisposition (Xu *et al.* 2005a). Additionally, a region close to 13q14, chromosomal locus 13q13.3, was reported for the first time to have a strong linkage to prostate cancer in the Finnish population (Cropp *et al.* 2011). These data and the presence of putative tumour suppressor genes in the region suggest the importance of this locus to prostate cancer susceptibility.

However, there has been no explicit candidate gene or variants discovered in this region, and the genes and alleles that have been identified exhibit only low penetrant activity. There may be a joint effect between 13q14 genes or genes in the region may produce a larger transcript variant consisting of more than one gene, as has been proposed with *SETDB2* and *PHF11* (Zhang *et al.* 2003). Because the 13q14 region is deleted in many cancers, including prostate cancer, and harbours several tumour suppressor genes, it is implicated as a prostate cancer susceptibility locus. Therefore, there is need for more detailed analysis with larger, more comprehensive sets of prostate cancer patients to further characterise the 13q14 region. Nevertheless, other chromosomal regions, including 8q24, 2q and 17q, have been studied more intensively, and there is strong evidence that these regions are associated with prostate cancer susceptibility.

ARLTS1 is a cancer-predisposing gene with proven tumour suppressor properties. However, little is known about the precise functional properties of the *ARLTS1* gene, especially on pathways, interactions in the 13q14 region and the conformation of the protein. The association between the *ARLTS1* T442C variant and prostate cancer risk found in this thesis may be due to an interaction network of different SNPs in the 13q14 region. Some polymorphisms have a direct functional effect, but SNPs may also be associated with susceptibility through linkage

disequilibrium with another functional variant. There might also be other eQTLs that affect the expression of genes in the 13q14 region. Thus, it is crucial to further elucidate the functional consequences of the variants and loci associated with prostate cancer susceptibility. The *ARLTS1* variant, T442C, may be included in a haploblock or there may still be a missing variant. Several low-penetrant genes (including *ARLTS1*) and variants with minor effects operate in concert with environmental and nongenetic factors. These variants may be individually modest, but they are likely to contribute substantially in combination to promote the overall cancer incidence. Further studies are needed to elucidate the actual role of *ARLTS1*, the interacting genes and proteins and the 13q14 region in prostate cancer susceptibility.

5. Future prospects

Prostate cancer is a heterogeneous and complex disease for which several susceptibility loci have been implicated through genome-wide linkage and association studies. Aging and improved diagnostics are likely responsible for the increase in the number of new cases, but the incidence is also influenced by unknown factors. The growing number of new prostate cancer cases puts pressure on the health care system, and new tools for prostate cancer diagnostics, prognostics and treatment are needed. In particular, new tools are needed to avoid over treatment and unnecessary biopsies as well as to distinguish aggressive cancers from indolent ones. However, despite the comprehensive research conducted over the past decades, the aetiological risk factors and genes that cause genetic susceptibility to prostate cancer remain largely unknown. This lack of knowledge has hampered effective cancer prevention and the development of better treatment options.

The inclusive aim of this study was to provide new information on prostate cancer genetics and the role of inheritance in prostate cancer progression. We achieved this goal by studying the cancer susceptibility gene *ARLTS1* and a prostate cancer locus, 13q14, leading to the identification of putative, clinically important biomarkers to be exploited in the diagnosis, prevention and monitoring of prostate cancer.

The identification of low-penetrant cancer predisposing genes (such as *ARLTS1*) explains only a fraction of the disease risk. In the future, it will be important to explore the interactions and shared mechanisms between these genes by taking advantage of more detailed cohorts, international collaborations and new integrative approaches. In addition, further advances are needed to facilitate personalised prostate cancer prevention, diagnosis and treatment. New translational applications are being harnessed, including the "omics" approach, which incorporates proteomics and metabolomics data, molecular system biology, large epidemiological data, molecular genetics, epigenetics and genomics using well-defined cohorts (Shen, Abate-Shen 2010, Vucic *et al.* 2012). However, computational biology approaches still pose validation challenges for many diseases (Freedman *et al.* 2011). Additional challenges include housing the growing amount of data due to storage constraints, data security and confidentiality reasons (Kilpivaara, Aaltonen 2013). In the future, further collaboration and collective thinking at the multi-group level or within large international consortia will be needed to enable better genetic and clinical outcomes for genetic-epidemiological questions (Goh *et al.* 2012, Freedman *et al.* 2011). National and international biobanks are being created or are already completed that offer well characterised cohorts of samples and controls. Together with the great advances in our understanding of cancer biology, there is also an immense need for international standards and the education of health care professionals (Kilpivaara, Aaltonen 2013).

CONCLUSIONS

The present study was conducted to provide new information regarding prostate cancer genetics and the role of inheritance in prostate cancer progression by investigating the tumour suppressor gene, *ARLTS1*, which is located within the prostate cancer susceptibility locus, 13q14.

The major findings of this thesis were the following:

1. The *ARLTS1* variant G446A (Trp149Stop) is not a cancer predisposing allele in the Finnish population.

2. The *ARLTS1* variant T442C (Cys148Arg) is a prostate cancer predisposing allele that contributes to both familial and sporadic cancer risk and the G194T (Gly65Val) variant may be associated with prostate cancer risk in the Finnish population.

3. *ARLTS1* expression is significantly reduced in somatic prostate cancer specimens and at the germline level, indicating a role in tumour suppression.

4. *ARLTS1* expression and function seem to be connected with immune system processes.

5. The results of the expression quantitative trait loci analysis propose that *ARLTS1* expression is regulated by eSNPs that are located in regulatory and transcriptionally active sites in the 13q14 region.

ACKNOWLEDGEMENTS

This study was carried out in the Laboratory of Cancer Genetics, Institute of Biomedical Technology, University of Tampere and Tampere University Hospital during the years 2005-2013. I wish to thank the former director of IBT, Professor Olli Silvennoinen, M.D., Ph.D., current director of the BioMediTech, Dr. Hannu Hanhijärvi, DDS, Ph.D., and Docent Erkki Seppälä, M.D., Ph.D., Medical Director of the Fimlab Laboratories, for providing excellent research facilities. Tampere Graduate Program in Biomedicine and Biotechnology (TGPBB) is acknowledged for providing me the graduate school position and for its excellent courses and travel grants.

I express my sincere gratitude to my supervisors, Professor Johanna Schleutker, Ph.D., and Docent Tiina Wahlfors, Ph.D. Johanna has provided me the opportunity to join her research group and introduced me into the challenging world of prostate cancer genetics; thank you for your support. Tiina's scientific expertise and guidance helped me to go on and finish this thesis.

My thesis committee members, Professor Teuvo L.J. Tammela, M.D., Ph.D., and Professor Olli Kallioniemi, M.D., Ph.D., are acknowledged for their valuable comments during the years.

The official reviewers of the thesis manuscript, Docent Auli Karhu, Ph.D., and Docent Antti Rannikko, M.D., Ph.D., are warmly thanked for their constructive criticism and feedback.

I am deeply grateful to all the co-authors for their valuable contribution and professional help with my publications, considering the lab work and the writing process. Especially, I want to thank Docent Pia Vahteristo, Ph.D., due to our *ARLTS1* collaboration project. I warmly thank Professor Tapio Visakorpi, M.D., Ph.D., together with his group members, Outi Saramäki, Ph.D., and Leena Latonen, Ph.D., for your help during the course of the second thesis study. The statistical fellows of our group, Martin Schindler, Ph.D., Daniel Fischer, M.Sc., and Tommi Rantapero, M.Sc., are thanked for their priceless efforts. John Patrick Mpindi, M.Sc., is thanked for the co-expression analyses which have strengthened my publications. Additionally, our collaborators at the National Institutes of Health,

USA, especially Joan Bailey-Wilson, Ph.D., together with her students, are thanked for their contribution to my first article.

All the members of our genetics lab are warmly thanked. During these years there has been a quite large cohort of people. Especially, I want to thank the present members of our group; for your help, friendship, our endless conversations about work stuff and personal life and the great company at and out of work: Virpi Laitinen, M.Sc., CMG, Kirsi Kuusisto, M.Sc., Riikka Nurminen, M.Sc., and Mrs. Riina Liikala, together with present TGPBB Coordinator Henna Mattila, Ph.D. Our former post-doc Tarja Ikonen, Ph.D., is thanked for giving me lots of guidance and support in the early phase of my graduate studies. I warmly thank Mrs. Riitta Vaalavuo for helping in the sample issues. I owe my gratitude also to all the former members of our lab: Ms. Linda Enroth, Mrs. Minna Sjöblom, Eija Seppälä, Ph.D., Nina Mononen, Ph.D., Docent Mika Matikainen, M.D., Ph.D., Sanna Pakkanen, M.D., Ph.D., Kimmo Ivori, M.Sc., Martin Schindler, Ph.D., Jarkko Isotalo, Ph.D., Ha Nati, M.Sc., Mimmi Patrikainen, M.Sc., Elisa Vuorinen, M.Sc., and all the project workers. Also all the other co-workers and friends at IBT are warmly thanked, especially the members of Anne Kallioniemi's group, for their company and help in any work-related issues I have had.

I want to thank my family, mother and father, my siblings with their spouses, father- and mother-in-law and all my friends for your support. You have reminded me what’s really important in life and where should I aim for. We have had and always will have great moments together!

Most importantly, I owe my dearest thanks to my closest ones. With the love and patience of my family I managed to complete this sometimes anguished project. My husband and best friend Janne, thanks for always bringing lots of laugh, love and humour to our everyday life. You are also the guy with common sense and you have stood by me all these years through the good times and bad. My son Niila and daughter Vilma, my sunshines, you bring happiness, give endless love and have taught me the dimensions of life. You are everything for me, because eventually; nothing else matters.

I also sincerely thank all the cancer patients and their family members for participating in this study.

This study was financially supported by the Tampere Graduate Program in Biomedicine and Biotechnology (TGPBB), the University of Tampere, the Academy of Finland, the Finnish Cancer Society, the Sigrid Juselius Foundation, Foundation of the Finnish Cancer Institute, the Medical Research Fund of Tampere University Hospital, The Helsinki University Hospital Research Fund, The Reino Lahtikari Foundation, The Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health (USA), The Competitive Research Funding of the Pirkanmaa Hospital District, Finnish Cancer Organizations, Tampere University Hospital, the Orion-Farmos Research Foundation and the Scientific Foundation of the City of Tampere.

Tampere, August 2013

Sanna Siltanen

REFERENCES

1000 Genomes Project Consortium, Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T. & McVean, G.A. 2012, "An integrated map of genetic variation from 1,092 human genomes", *Nature*, vol. 491, no. 7422, pp. 56-65.

Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R.B., Kaariainen, H., Eskelinen, M., Jarvinen, H., Mecklin, J.P. & de la Chapelle, A. 1998, "Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease", *The New England journal of medicine*, vol. 338, no. 21, pp. 1481-1487.

Akamatsu, S., Takata, R., Haiman, C.A., Takahashi, A., Inoue, T., Kubo, M., Furihata, M., Kamatani, N., Inazawa, J., Chen, G.K., Le Marchand, L., Kolonel, L.N., Katoh, T., Yamano, Y., Yamakado, M., Takahashi, H., Yamada, H., Egawa, S., Fujioka, T., Henderson, B.E., Habuchi, T., Ogawa, O., Nakamura, Y. & Nakagawa, H. 2012, "Common variants at 11q12, 10q26 and 3p11.2 are associated with prostate cancer susceptibility in Japanese", *Nature genetics*, vol. 44, no. 4, pp. 426-9, S1.

Akisik, E., Yazici, H. & Dalay, N. 2011, "ARLTS1, MDM2 and RAD51 gene variations are associated with familial breast cancer", *Molecular biology reports*, vol. 38, no. 1, pp. 343-348.

Al Olama, A.A., Kote-Jarai, Z., Giles, G.G., Guy, M., Morrison, J., Severi, G., Leongamornlert, D.A., Tymrakiewicz, M., Jhavar, S., Saunders, E., Hopper, J.L., Southey, M.C., Muir, K.R., English, D.R., Dearnaley, D.P., Ardern-Jones, A.T., Hall, A.L., O'Brien, L.T., Wilkinson, R.A., Sawyer, E., Lophatananon, A., UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK Prostate testing for cancer and Treatment study (ProtecT Study) Collaborators, Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Cooper, C., Donovan, J.L., Hamdy, F.C., Neal, D.E., Eeles, R.A. & Easton, D.F. 2009, "Multiple loci on 8q24 associated with prostate cancer susceptibility", *Nature genetics*, vol. 41, no. 10, pp. 1058-1060.

Alexander, D.D., Mink, P.J., Cushing, C.A. & Sceurman, B. 2010, "A review and meta-analysis of prospective studies of red and processed meat intake and prostate cancer", *Nutrition journal*, vol. 9, pp. 50-2891-9-50.

American Cancer Society 2012, "Cancer Facts and Figures 2012".

American Cancer Society 2013, "Cancer Facts & Figures 2013".

Amundadottir, L.T., Sulem, P., Gudmundsson, J., Helgason, A., Baker, A., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Cazier, J.B., Sainz, J., Jakobsdottir, M., Kostic, J., Magnusdottir, D.N., Ghosh, S., Agnarsson, K., Birgisdottir, B., Le Roux, L., Olafsdottir, A., Blondal, T., Andresdottir, M., Gretarsdottir, O.S., Bergthorsson, J.T., Gudbjartsson, D., Gylfason, A., Thorleifsson, G., Manolescu, A., Kristjansson, K., Geirsson, G., Isaksson, H., Douglas, J., Johansson, J.E., Balter, K., Wiklund, F., Montie, J.E., Yu, X., Suarez, B.K., Ober, C., Cooney, K.A., Gronberg, H., Catalona, W.J., Einarsson, G.V., Barkardottir, R.B., Gulcher, J.R., Kong, A., Thorsteinsdottir, U. & Stefansson, K. 2006, "A common variant associated with prostate cancer in European and African populations", *Nature genetics*, vol. 38, no. 6, pp. 652-658.

Anders, S. & Huber, W. 2010, "Differential expression analysis for sequence count data", *Genome biology*, vol. 11, no. 10, pp. R106-2010-11-10-r106. Epub 2010 Oct 27.

Aqeilan, R.I., Calin, G.A. & Croce, C.M. 2010, "miR-15a and miR-16-1 in cancer: discovery, function and future perspectives", *Cell death and differentiation*, vol. 17, no. 2, pp. 215-220.

Barbieri, C.E., Demichelis, F. & Rubin, M.A. 2012, "Molecular genetics of prostate cancer: emerging appreciation of genetic complexity", *Histopathology*, vol. 60, no. 1, pp. 187-198.

Bartel, D.P. 2004, "MicroRNAs: genomics, biogenesis, mechanism, and function", *Cell*, vol. 116, no. 2, pp. 281-297.

Beltran, H., Yelensky, R., Frampton, G.M., Park, K., Downing, S.R., Macdonald, T.Y., Jarosz, M., Lipson, D., Tagawa, S.T., Nanus, D.M., Stephens, P.J., Mosquera, J.M., Cronin, M.T. & Rubin, M.A. 2013, "Targeted Next-generation Sequencing of Advanced Prostate Cancer Identifies Potential Therapeutic Targets and Disease Heterogeneity", *European urology*, vol. 63, no. 5, pp. 920-926.

Bernards, A. 2003, "GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and Drosophila", *Biochimica et biophysica acta*, vol. 1603, no. 2, pp. 47-82.

Bird, A. 2007, "Perceptions of epigenetics", *Nature*, vol. 447, no. 7143, pp. 396-398.

Bonci, D., Coppola, V., Musumeci, M., Addario, A., Giuffrida, R., Memeo, L., D'Urso, L., Pagliuca, A., Biffoni, M., Labbaye, C., Bartucci, M., Muto, G., Peschle, C. & De Maria, R. 2008, "The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities", *Nature medicine*, vol. 14, no. 11, pp. 1271-1277.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., Cherry, J.M. & Snyder, M.

2012, "Annotation of functional variation in personal genomes using RegulomeDB", *Genome research,* vol. 22, no. 9, pp. 1790-1797.

Bratt, O., Garmo, H., Adolfsson, J., Bill-Axelson, A., Holmberg, L., Lambe, M. & Stattin, P. 2010, "Effects of prostate-specific antigen testing on familial prostate cancer risk estimates", *Journal of the National Cancer Institute,* vol. 102, no. 17, pp. 1336-1343.

Brookman-Amissah, N., Nariculam, J., Freeman, A., Willamson, M., Kirby, R.S., Masters, J.R. & Feneley, M.R. 2007, "Allelic imbalance at 13q14.2 approximately q14.3 in localized prostate cancer is associated with early biochemical relapse", *Cancer genetics and cytogenetics,* vol. 179, no. 2, pp. 118-126.

Bullaughey, K., Chavarria, C.I., Coop, G. & Gilad, Y. 2009, "Expression quantitative trait loci detected in cell lines are often present in primary tissues", *Human molecular genetics,* vol. 18, no. 22, pp. 4296-4303.

Bullrich, F., Fujii, H., Calin, G., Mabuchi, H., Negrini, M., Pekarsky, Y., Rassenti, L., Alder, H., Reed, J.C., Keating, M.J., Kipps, T.J. & Croce, C.M. 2001, "Characterization of the 13q14 tumor suppressor locus in CLL: identification of ALT1, an alternative splice variant of the LEU2 gene", *Cancer research,* vol. 61, no. 18, pp. 6640-6648.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F. & Croce, C.M. 2002, "Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 99, no. 24, pp. 15524-15529.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M. & Croce, C.M. 2004, "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 101, no. 9, pp. 2999-3004.

Calin, G.A., Trapasso, F., Shimizu, M., Dumitru, C.D., Yendamuri, S., Godwin, A.K., Ferracin, M., Bernardi, G., Chatterjee, D., Baldassarre, G., Rattan, S., Alder, H., Mabuchi, H., Shiraishi, T., Hansen, L.L., Overgaard, J., Herlea, V., Mauro, F.R., Dighiero, G., Movsas, B., Rassenti, L., Kipps, T., Baffa, R., Fusco, A., Mori, M., Russo, G., Liu, C.G., Neuberg, D., Bullrich, F., Negrini, M. & Croce, C.M. 2005, "Familial cancer associated with a polymorphism in ARLTS1", *The New England journal of medicine,* vol. 352, no. 16, pp. 1667-1676.

Carpten, J., Nupponen, N., Isaacs, S., Sood, R., Robbins, C., Xu, J., Faruque, M., Moses, T., Ewing, C., Gillanders, E., Hu, P., Bujnovszky, P., Makalowska, I., Baffoe-Bonnie, A., Faith, D., Smith, J., Stephan, D., Wiley, K., Brownstein, M., Gildea, D., Kelly, B., Jenkins, R., Hostetter, G., Matikainen, M.,

Schleutker, J., Klinger, K., Connors, T., Xiang, Y., Wang, Z., De Marzo, A., Papadopoulos, N., Kallioniemi, O.P., Burk, R., Meyers, D., Gronberg, H., Meltzer, P., Silverman, R., Bailey-Wilson, J., Walsh, P., Isaacs, W. & Trent, J. 2002, "Germline mutations in the ribonuclease L gene in families showing linkage with HPC1", *Nature genetics,* vol. 30, no. 2, pp. 181-184.

Carter, B.S., Beaty, T.H., Steinberg, G.D., Childs, B. & Walsh, P.C. 1992, "Mendelian inheritance of familial prostate cancer", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 89, no. 8, pp. 3367-3371.

Castellvi-Bel, S., Castells, A., de Cid, R., Munoz, J., Balaguer, F., Gonzalo, V., Ruiz-Ponte, C., Andreu, M., Llor, X., Jover, R., Bessa, X., Xicola, R.M., Pons, E., Alenda, C., Paya, A., Carracedo, A., Pique, J.M. & Gastrointestinal Oncology Group of the Spanish Gastroenterological Association 2007, "Association of the ARLTS1 Cys148Arg variant with sporadic and familial colorectal cancer", *Carcinogenesis,* vol. 28, no. 8, pp. 1687-1691.

Catalog of Published Genome-Wide Association Studies 7.4.2013-last update, *Catalog of Published Genome-Wide Association Studies; as of 7th of April 2013, the catalog includes 1556 publications and 9811 SNPs.* [Homepage of National Human Genome Research Institute], [Online]. Available: http://www.genome.gov/gwastudies/

Center, M.M., Jemal, A., Lortet-Tieulent, J., Ward, E., Ferlay, J., Brawley, O. & Bray, F. 2012, "International variation in prostate cancer incidence and mortality rates", *European urology,* vol. 61, no. 6, pp. 1079-1092.

Chavrier, P. & Goud, B. 1999, "The role of ARF and Rab GTPases in membrane transport", *Current opinion in cell biology,* vol. 11, no. 4, pp. 466-475.

Chen, C., Frierson, H.F.,Jr, Haggerty, P.F., Theodorescu, D., Gregory, C.W. & Dong, J.T. 2001, "An 800-kb region of deletion at 13q14 in human prostate and other carcinomas", *Genomics,* vol. 77, no. 3, pp. 135-144.

Chung, C.C., Magalhaes, W.C., Gonzalez-Bosquet, J. & Chanock, S.J. 2010, "Genome-wide association studies in cancer--current and future directions", *Carcinogenesis,* vol. 31, no. 1, pp. 111-120.

Ciampa, J., Yeager, M., Amundadottir, L., Jacobs, K., Kraft, P., Chung, C., Wacholder, S., Yu, K., Wheeler, W., Thun, M.J., Divers, W.R., Gapstur, S., Albanes, D., Virtamo, J., Weinstein, S., Giovannucci, E., Willett, W.C., Cancel-Tassin, G., Cussenot, O., Valeri, A., Hunter, D., Hoover, R., Thomas, G., Chanock, S. & Chatterjee, N. 2011, "Large-scale exploration of gene-gene interactions in prostate cancer using a multistage genome-wide association study", *Cancer research,* vol. 71, no. 9, pp. 3287-3295.

Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M. & Croce, C.M. 2005, "miR-15

and miR-16 induce apoptosis by targeting BCL2", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 102, no. 39, pp. 13944-13949.

Clarke, E., Rahman, N., Page, N., Rolph, M.S., Stewart, G.J. & Jones, G.J. 2008, "Functional characterization of the atopy-associated gene PHF11", *The Journal of allergy and clinical immunology,* vol. 121, no. 5, pp. 1148-1154.e3.

Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. 2009, "Mapping complex disease traits with global gene expression", *Nature reviews.Genetics,* vol. 10, no. 3, pp. 184-194.

Cordell, H.J. 2009, "Detecting gene-gene interactions that underlie human diseases", *Nature reviews.Genetics,* vol. 10, no. 6, pp. 392-404.

Crawford, E.D. 2003, "Epidemiology of prostate cancer", *Urology,* vol. 62, no. 6 Suppl 1, pp. 3-12.

Cropp, C.D., Simpson, C.L., Wahlfors, T., Ha, N., George, A., Jones, M.S., Harper, U., Ponciano-Jackson, D., Green, T.A., Tammela, T.L., Bailey-Wilson, J. & Schleutker, J. 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", *International journal of cancer.Journal international du cancer,* vol. 129, no. 10, pp. 2400-2407.

Dawn Teare, M. & Barrett, J.H. 2005, "Genetic linkage studies", *Lancet,* vol. 366, no. 9490, pp. 1036-1044.

De Marzo, A.M., Marchi, V.L., Epstein, J.I. & Nelson, W.G. 1999, "Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis", *The American journal of pathology,* vol. 155, no. 6, pp. 1985-1992.

De Marzo, A.M., Meeker, A.K., Zha, S., Luo, J., Nakayama, M., Platz, E.A., Isaacs, W.B. & Nelson, W.G. 2003, "Human prostate cancer precursors and pathobiology", *Urology,* vol. 62, no. 5 Suppl 1, pp. 55-62.

De Marzo, A.M., Platz, E.A., Sutcliffe, S., Xu, J., Gronberg, H., Drake, C.G., Nakai, Y., Isaacs, W.B. & Nelson, W.G. 2007, "Inflammation in prostate carcinogenesis", *Nature reviews.Cancer,* vol. 7, no. 4, pp. 256-269.

Ding, J., Gudjonsson, J.E., Liang, L., Stuart, P.E., Li, Y., Chen, W., Weichenthal, M., Ellinghaus, E., Franke, A., Cookson, W., Nair, R.P., Elder, J.T. & Abecasis, G.R. 2010, "Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals", *American Journal of Human Genetics,* vol. 87, no. 6, pp. 779-789.

Donaldson, J.G. & Jackson, C.L. 2000, "Regulators and effectors of the ARF GTPases", *Current opinion in cell biology,* vol. 12, no. 4, pp. 475-482.

Dong, H. & Wang, S. 2012, "Exploring the cancer genome in the era of next-generation sequencing", *Frontiers of medicine,* vol. 6, no. 1, pp. 48-55.

Dong, J.T. 2001, "Chromosomal deletions and tumor suppressor genes in prostate cancer", *Cancer metastasis reviews,* vol. 20, no. 3-4, pp. 173-193.

Dong, J.T., Boyd, J.C. & Frierson, H.F.,Jr 2001, "Loss of heterozygosity at 13q14 and 13q21 in high grade, high stage prostate cancer", *The Prostate,* vol. 49, no. 3, pp. 166-171.

Dong, X.Y., Chen, C., Sun, X., Guo, P., Vessella, R.L., Wang, R.X., Chung, L.W., Zhou, W. & Dong, J.T. 2006, "FOXO1A is a candidate for the 13q14 tumor suppressor gene inhibiting androgen receptor signaling in prostate cancer", *Cancer research,* vol. 66, no. 14, pp. 6998-7006.

Eeles, R.A., Kote-Jarai, Z., Al Olama, A.A., Giles, G.G., Guy, M., Severi, G., Muir, K., Hopper, J.L., Henderson, B.E., Haiman, C.A., Schleutker, J., Hamdy, F.C., Neal, D.E., Donovan, J.L., Stanford, J.L., Ostrander, E.A., Ingles, S.A., John, E.M., Thibodeau, S.N., Schaid, D., Park, J.Y., Spurdle, A., Clements, J., Dickinson, J.L., Maier, C., Vogel, W., Dork, T., Rebbeck, T.R., Cooney, K.A., Cannon-Albright, L., Chappuis, P.O., Hutter, P., Zeegers, M., Kaneva, R., Zhang, H.W., Lu, Y.J., Foulkes, W.D., English, D.R., Leongamornlert, D.A., Tymrakiewicz, M., Morrison, J., Ardern-Jones, A.T., Hall, A.L., O'Brien, L.T., Wilkinson, R.A., Saunders, E.J., Page, E.C., Sawyer, E.J., Edwards, S.M., Dearnaley, D.P., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Van As, N., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Cooper, C.S., Southey, M.C., Lophatananon, A., Liu, J.F., Kolonel, L.N., Le Marchand, L., Wahlfors, T., Tammela, T.L., Auvinen, A., Lewis, S.J., Cox, A., FitzGerald, L.M., Koopmeiners, J.S., Karyadi, D.M., Kwon, E.M., Stern, M.C., Corral, R., Joshi, A.D., Shahabi, A., McDonnell, S.K., Sellers, T.A., Pow-Sang, J., Chambers, S., Aitken, J., Gardiner, R.A., Batra, J., Kedda, M.A., Lose, F., Polanowski, A., Patterson, B., Serth, J., Meyer, A., Luedeke, M., Stefflova, K., Ray, A.M., Lange, E.M., Farnham, J., Khan, H., Slavov, C., Mitkova, A., Cao, G., UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, PRACTICAL Consortium & Easton, D.F. 2009, "Identification of seven new prostate cancer susceptibility loci through a genome-wide association study", *Nature genetics,* vol. 41, no. 10, pp. 1116-1121.

Eeles, R.A., Kote-Jarai, Z., Giles, G.G., Olama, A.A., Guy, M., Jugurnauth, S.K., Mulholland, S., Leongamornlert, D.A., Edwards, S.M., Morrison, J., Field, H.I., Southey, M.C., Severi, G., Donovan, J.L., Hamdy, F.C., Dearnaley, D.P., Muir, K.R., Smith, C., Bagnato, M., Ardern-Jones, A.T., Hall, A.L., O'Brien, L.T., Gehr-Swain, B.N., Wilkinson, R.A., Cox, A., Lewis, S., Brown, P.M., Jhavar, S.G., Tymrakiewicz, M., Lophatananon, A., Bryant, S.L., UK Genetic Prostate Cancer Study Collaborators, British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Fisher, C., Jamieson, C., Cooper, C.S., English, D.R., Hopper, J.L., Neal, D.E. & Easton, D.F. 2008, "Multiple newly identified loci associated

with prostate cancer susceptibility", *Nature genetics*, vol. 40, no. 3, pp. 316-321.

Eeles, R.A., Olama, A.A., Benlloch, S., Saunders, E.J., Leongamornlert, D.A., Tymrakiewicz, M., Ghoussaini, M., Luccarini, C., Dennis, J., Jugurnauth-Little, S., Dadaev, T., Neal, D.E., Hamdy, F.C., Donovan, J.L., Muir, K., Giles, G.G., Severi, G., Wiklund, F., Gronberg, H., Haiman, C.A., Schumacher, F., Henderson, B.E., Le Marchand, L., Lindstrom, S., Kraft, P., Hunter, D.J., Gapstur, S., Chanock, S.J., Berndt, S.I., Albanes, D., Andriole, G., Schleutker, J., Weischer, M., Canzian, F., Riboli, E., Key, T.J., Travis, R.C., Campa, D., Ingles, S.A., John, E.M., Hayes, R.B., Pharoah, P.D., Pashayan, N., Khaw, K.T., Stanford, J.L., Ostrander, E.A., Signorello, L.B., Thibodeau, S.N., Schaid, D., Maier, C., Vogel, W., Kibel, A.S., Cybulski, C., Lubinski, J., Cannon-Albright, L., Brenner, H., Park, J.Y., Kaneva, R., Batra, J., Spurdle, A.B., Clements, J.A., Teixeira, M.R., Dicks, E., Lee, A., Dunning, A.M., Baynes, C., Conroy, D., Maranian, M.J., Ahmed, S., Govindasami, K., Guy, M., Wilkinson, R.A., Sawyer, E.J., Morgan, A., Dearnaley, D.P., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Van As, N.J., Woodhouse, C.J., Thompson, A., Dudderidge, T., Ogden, C., Cooper, C.S., Lophatananon, A., Cox, A., Southey, M.C., Hopper, J.L., English, D.R., Aly, M., Adolfsson, J., Xu, J., Zheng, S.L., Yeager, M., Kaaks, R., Diver, W.R., Gaudet, M.M., Stern, M.C., Corral, R., Joshi, A.D., Shahabi, A., Wahlfors, T., Tammela, T.L., Auvinen, A., Virtamo, J., Klarskov, P., Nordestgaard, B.G., Roder, M.A., Nielsen, S.F., Bojesen, S.E., Siddiq, A., Fitzgerald, L.M., Kolb, S., Kwon, E.M., Karyadi, D.M., Blot, W.J., Zheng, W., Cai, Q., McDonnell, S.K., Rinckleb, A.E., Drake, B., Colditz, G., Wokolorczyk, D., Stephenson, R.A., Teerlink, C., Muller, H., Rothenbacher, D., Sellers, T.A., Lin, H.Y., Slavov, C., Mitev, V., Lose, F., Srinivasan, S., Maia, S., Paulo, P., Lange, E., Cooney, K.A., Antoniou, A.C., Vincent, D., Bacot, F., Tessier, D.C., COGS-Cancer Research UK GWAS-ELLIPSE (part of GAME-ON) Initiative, stralian Prostate Cancer Bioresource, UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT (Prostate testing for cancer and Treatment) Study Collaborators, PRACTICAL (Prostate Cancer Association Group to Investigate Cancer-Associated Alterations in the Genome) Consortium, Kote-Jarai, Z. & Easton, D.F. 2013, "Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array", *Nature genetics*, vol. 45, no. 4, pp. 385-391.

Eerola, H., Blomqvist, C., Pukkala, E., Pyrhonen, S. & Nevanlinna, H. 2000, "Familial breast cancer in southern Finland: how prevalent are breast cancer families and can we trust the family history reported by patients?", *European journal of cancer (Oxford, England : 1990)*, vol. 36, no. 9, pp. 1143-1148.

ENCODE Project Consortium 2011, "A user's guide to the encyclopedia of DNA elements (ENCODE)", *PLoS biology*, vol. 9, no. 4, pp. e1001046.

ENCODE Project Consortium, Dunham, I., Kundaje, A., Aldred, S.F., Collins, P.J., Davis, C.A., Doyle, F., Epstein, C.B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B.R., Landt, S.G., Lee, B.K., Pauli, F., Rosenbloom, K.R., Sabo, P., Safi, A., Sanyal, A., Shoresh, N., Simon, J.M., Song, L., Trinklein, N.D.,

Altshuler, R.C., Birney, E., Brown, J.B., Cheng, C., Djebali, S., Dong, X., Dunham, I., Ernst, J., Furey, T.S., Gerstein, M., Giardine, B., Greven, M., Hardison, R.C., Harris, R.S., Herrero, J., Hoffman, M.M., Iyer, S., Kelllis, M., Khatun, J., Kheradpour, P., Kundaje, A., Lassman, T., Li, Q., Lin, X., Marinov, G.K., Merkel, A., Mortazavi, A., Parker, S.C., Reddy, T.E., Rozowsky, J., Schlesinger, F., Thurman, R.E., Wang, J., Ward, L.D., Whitfield, T.W., Wilder, S.P., Wu, W., Xi, H.S., Yip, K.Y., Zhuang, J., Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., Pazin, M.J., Lowdon, R.F., Dillon, L.A., Adams, L.B., Kelly, C.J., Zhang, J., Wexler, J.R., Green, E.D., Good, P.J., Feingold, E.A., Bernstein, B.E., Birney, E., Crawford, G.E., Dekker, J., Elinitski, L., Farnham, P.J., Gerstein, M., Giddings, M.C., Gingeras, T.R., Green, E.D., Guigo, R., Hardison, R.C., Hubbard, T.J., Kellis, M., Kent, W.J., Lieb, J.D., Margulies, E.H., Myers, R.M., Snyder, M., Starnatoyannopoulos, J.A., Tennebaum, S.A., Weng, Z., White, K.P., Wold, B., Khatun, J., Yu, Y., Wrobel, J., Risk, B.A., Gunawardena, H.P., Kuiper, H.C., Maier, C.W., Xie, L., Chen, X., Giddings, M.C., Bernstein, B.E., Epstein, C.B., Shoresh, N., Ernst, J., Kheradpour, P., Mikkelsen, T.S., Gillespie, S., Goren, A., Ram, O., Zhang, X., Wang, L., Issner, R., Coyne, M.J., Durham, T., Ku, M., Truong, T., Ward, L.D., Altshuler, R.C., Eaton, M.L., Kellis, M., Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Roder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Batut, P., Bell, I., Bell, K., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H.P., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Li, G., Luo, O.J., Park, E., Preall, J.B., Presaud, K., Ribeca, P., Risk, B.A., Robyr, D., Ruan, X., Sammeth, M., Sandu, K.S., Schaeffer, L., See, L.H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T.J., Reymond, A., Antonarakis, S.E., Hannon, G.J., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigo, R., Gingeras, T.R., Rosenbloom, K.R., Sloan, C.A., Learned, K., Malladi, V.S., Wong, M.C., Barber, G.P., Cline, M.S., Dreszer, T.R., Heitner, S.G., Karolchik, D., Kent, W.J., Kirkup, V.M., Meyer, L.R., Long, J.C., Maddren, M., Raney, B.J., Furey, T.S., Song, L., Grasfeder, L.L., Giresi, P.G., Lee, B.K., Battenhouse, A., Sheffield, N.C., Simon, J.M., Showers, K.A., Safi, A., London, D., Bhinge, A.A., Shestak, C., Schaner, M.R., Kim, S.K., Zhang, Z.Z., Mieczkowski, P.A., Mieczkowska, J.O., Liu, Z., McDaniell, R.M., Ni, Y., Rashid, N.U., Kim, M.J., Adar, S., Zhang, Z., Wang, T., Winter, D., Keefe, D., Birney, E., Iyer, V.R., Lieb, J.D., Crawford, G.E., Li, G., Sandhu, K.S., Zheng, M., Wang, P., Luo, O.J., Shahab, A., Fullwood, M.J., Ruan, X., Ruan, Y., Myers, R.M., Pauli, F., Williams, B.A., Gertz, J., Marinov, G.K., Reddy, T.E., Vielmetter, J., Partridge, E.C., Trout, D., Varley, K.E., Gasper, C., Bansal, A., Pepke, S., Jain, P., Amrhein, H., Bowling, K.M., Anaya, M., Cross, M.K., King, B., Muratet, M.A., Antoshechkin, I., Newberry, K.M., McCue, K., Nesmith, A.S., Fisher-Aylor, K.I., Pusey, B., DeSalvo, G., Parker, S.L., Balasubramanian, S., Davis, N.S., Meadows, S.K., Eggleston, T., Gunter, C., Newberry, J.S., Levy, S.E., Absher, D.M., Mortazavi, A., Wong, W.H., Wold, B., Blow, M.J., Visel, A.,

Pennachio, L.A., Elnitski, L., Margulies, E.H., Parker, S.C., Petrykowska, H.M., Abyzov, A., Aken, B., Barrell, D., Barson, G., Berry, A., Bignell, A., Boychenko, V., Bussotti, G., Chrast, J., Davidson, C., Derrien, T., Despacio-Reyes, G., Diekhans, M., Ezkurdia, I., Frankish, A., Gilbert, J., Gonzalez, J.M., Griffiths, E., Harte, R., Hendrix, D.A., Howald, C., Hunt, T., Jungreis, I., Kay, M., Khurana, E., Kokocinski, F., Leng, J., Lin, M.F., Loveland, J., Lu, Z., Manthravadi, D., Mariotti, M., Mudge, J., Mukherjee, G., Notredame, C., Pei, B., Rodriguez, J.M., Saunders, G., Sboner, A., Searle, S., Sisu, C., Snow, C., Steward, C., Tanzer, A., Tapanan, E., Tress, M.L., van Baren, M.J., Walters, N., Washieti, S., Wilming, L., Zadissa, A., Zhengdong, Z., Brent, M., Haussler, D., Kellis, M., Valencia, A., Gerstein, M., Raymond, A., Guigo, R., Harrow, J., Hubbard, T.J., Landt, S.G., Frietze, S., Abyzov, A., Addleman, N., Alexander, R.P., Auerbach, R.K., Balasubramanian, S., Bettinger, K., Bhardwaj, N., Boyle, A.P., Cao, A.R., Cayting, P., Charos, A., Cheng, Y., Cheng, C., Eastman, C., Euskirchen, G., Fleming, J.D., Grubert, F., Habegger, L., Hariharan, M., Harmanci, A., Iyenger, S., Jin, V.X., Karczewski, K.J., Kasowski, M., Lacroute, P., Lam, H., Larnarre-Vincent, N., Leng, J., Lian, J., Lindahl-Allen, M., Min, R., Miotto, B., Monahan, H., Moqtaderi, Z., Mu, X.J., O'Geen, H., Ouyang, Z., Patacsil, D., Pei, B., Raha, D., Ramirez, L., Reed, B., Rozowsky, J., Sboner, A., Shi, M., Sisu, C., Slifer, T., Witt, H., Wu, L., Xu, X., Yan, K.K., Yang, X., Yip, K.Y., Zhang, Z., Struhl, K., Weissman, S.M., Gerstein, M., Farnham, P.J., Snyder, M., Tenebaum, S.A., Penalva, L.O., Doyle, F., Karmakar, S., Landt, S.G., Bhanvadia, R.R., Choudhury, A., Domanus, M., Ma, L., Moran, J., Patacsil, D., Slifer, T., Victorsen, A., Yang, X., Snyder, M., White, K.P., Auer, T., Centarin, L., Eichenlaub, M., Gruhl, F., Heerman, S., Hoeckendorf, B., Inoue, D., Kellner, T., Kirchmaier, S., Mueller, C., Reinhardt, R., Schertel, L., Schneider, S., Sinn, R., Wittbrodt, B., Wittbrodt, J., Weng, Z., Whitfield, T.W., Wang, J., Collins, P.J., Aldred, S.F., Trinklein, N.D., Partridge, E.C., Myers, R.M., Dekker, J., Jain, G., Lajoie, B.R., Sanyal, A., Balasundaram, G., Bates, D.L., Byron, R., Canfield, T.K., Diegel, M.J., Dunn, D., Ebersol, A.K., Ebersol, A.K., Frum, T., Garg, K., Gist, E., Hansen, R.S., Boatman, L., Haugen, E., Humbert, R., Jain, G., Johnson, A.K., Johnson, E.M., Kutyavin, T.M., Lajoie, B.R., Lee, K., Lotakis, D., Maurano, M.T., Neph, S.J., Neri, F.V., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Rynes, E., Sabo, P., Sanchez, M.E., Sandstrom, R.S., Sanyal, A., Shafer, A.O., Stergachis, A.B., Thomas, S., Thurman, R.E., Vernot, B., Vierstra, J., Vong, S., Wang, H., Weaver, M.A., Yan, Y., Zhang, M., Akey, J.A., Bender, M., Dorschner, M.O., Groudine, M., MacCoss, M.J., Navas, P., Stamatoyannopoulos, G., Kaul, R., Dekker, J., Stamatoyannopoulos, J.A., Dunham, I., Beal, K., Brazma, A., Flicek, P., Herrero, J., Johnson, N., Keefe, D., Lukk, M., Luscombe, N.M., Sobral, D., Vaquerizas, J.M., Wilder, S.P., Batzoglou, S., Sidow, A., Hussami, N., Kyriazopoulou-Panagiotopoulou, S., Libbrecht, M.W., Schaub, M.A., Kundaje, A., Hardison, R.C., Miller, W., Giardine, B., Harris, R.S., Wu, W., Bickel, P.J., Banfai, B., Boley, N.P., Brown, J.B., Huang, H., Li, Q., Li, J.J., Noble, W.S., Bilmes, J.A., Buske, O.J., Hoffman, M.M., Sahu, A.O., Kharchenko, P.V., Park, P.J., Baker, D., Taylor, J., Weng, Z., Iyer, S., Dong, X., Greven, M., Lin, X., Wang, J., Xi, H.S., Zhuang, J., Gerstein, M., Alexander, R.P., Balasubramanian, S., Cheng, C., Harmanci, A., Lochovsky, L., Min, R., Mu, X.J., Rozowsky, J., Yan, K.K., Yip, K.Y. & Birney, E. 2012,

"An integrated encyclopedia of DNA elements in the human genome", *Nature,* vol. 489, no. 7414, pp. 57-74.

Epstein, J.I. 2010, "An update of the Gleason grading system", *The Journal of urology,* vol. 183, no. 2, pp. 433-440.

Ewing, C.M., Ray, A.M., Lange, E.M., Zuhlke, K.A., Robbins, C.M., Tembe, W.D., Wiley, K.E., Isaacs, S.D., Johng, D., Wang, Y., Bizon, C., Yan, G., Gielzak, M., Partin, A.W., Shanmugam, V., Izatt, T., Sinari, S., Craig, D.W., Zheng, S.L., Walsh, P.C., Montie, J.E., Xu, J., Carpten, J.D., Isaacs, W.B. & Cooney, K.A. 2012, "Germline mutations in HOXB13 and prostate-cancer risk", *The New England journal of medicine,* vol. 366, no. 2, pp. 141-149.

Finnish Cancer Registry "Finnish Cancer Registry, Cancer Statistics at www.cancerregistry.fi, last updated on 18.3.2013".

Frank, B., Hemminki, K., Brenner, H., Hoffmeister, M., Chang-Claude, J. & Burwinkel, B. 2006a, "ARLTS1 variants and risk of colorectal cancer", *Cancer letters,* vol. 244, no. 2, pp. 172-175.

Frank, B., Hemminki, K., Meindl, A., Wappenschmidt, B., Klaes, R., Schmutzler, R.K., Untch, M., Bugert, P., Bartram, C.R. & Burwinkel, B. 2006b, "Association of the ARLTS1 Cys148Arg variant with familial breast cancer risk", *International journal of cancer.Journal international du cancer,* vol. 118, no. 10, pp. 2505-2508.

Frank, B., Klaes, R. & Burwinkel, B. 2005, "Familial cancer and ARLTS1", *The New England journal of medicine,* vol. 353, no. 3, pp. 313-4; author reply 313-4.

Frank, B., Meyer, P., Boettger, M.B., Hemminki, K., Stapelmann, H., Gast, A., Schmitt, C., Kumar, R., Sergi, C. & Burwinkel, B. 2006, "ARLTS1 variants and melanoma risk", *International journal of cancer.Journal international du cancer,* vol. 119, no. 7, pp. 1736-1737.

Freedman, M.L., Monteiro, A.N., Gayther, S.A., Coetzee, G.A., Risch, A., Plass, C., Casey, G., De Biasi, M., Carlson, C., Duggan, D., James, M., Liu, P., Tichelaar, J.W., Vikis, H.G., You, M. & Mills, I.G. 2011, "Principles for the post-GWAS functional characterization of cancer risk loci", *Nature genetics,* vol. 43, no. 6, pp. 513-518.

Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N., Pato, M.T., Petryshen, T.L., Kolonel, L.N., Lander, E.S., Sklar, P., Henderson, B., Hirschhorn, J.N. & Altshuler, D. 2004, "Assessing the impact of population stratification on genetic association studies", *Nature genetics,* vol. 36, no. 4, pp. 388-393.

Frohling, S. & Dohner, H. 2008, "Chromosomal abnormalities in cancer", *The New England journal of medicine,* vol. 359, no. 7, pp. 722-734.

Garzon, R., Calin, G.A. & Croce, C.M. 2009, "MicroRNAs in Cancer", *Annual Review of Medicine,* vol. 60, pp. 167-179.

Giovannucci, E. 2005, "Tomato products, lycopene, and prostate cancer: a review of the epidemiological literature", *The Journal of nutrition,* vol. 135, no. 8, pp. 2030S-1S.

Glinsky, G.V., Glinskii, A.B., Stephenson, A.J., Hoffman, R.M. & Gerald, W.L. 2004, "Gene expression profiling predicts clinical outcome of prostate cancer", *The Journal of clinical investigation,* vol. 113, no. 6, pp. 913-923.

Goh, C.L., Schumacher, F.R., Easton, D., Muir, K., Henderson, B., Kote-Jarai, Z. & Eeles, R.A. 2012, "Genetic variants associated with predisposition to prostate cancer and potential clinical implications", *Journal of internal medicine,* vol. 271, no. 4, pp. 353-365.

Gong, Z., Agalliu, I., Lin, D.W., Stanford, J.L. & Kristal, A.R. 2007, "Obesity is associated with increased risks of prostate cancer metastasis and death after initial cancer diagnosis in middle-aged men", *Cancer,* vol. 109, no. 6, pp. 1192-1202.

Graur, D., Zheng, Y., Price, N., Azevedo, R.B., Zufall, R.A. & Elhaik, E. 2013, "On the immortality of television sets: "function" in the human genome according to the evolution-free gospel of ENCODE", *Genome biology and evolution,* vol. 5, no. 3, pp. 578-590.

Gronberg, H., Damber, L., Damber, J.E. & Iselius, L. 1997, "Segregation analysis of prostate cancer in Sweden: support for dominant inheritance", *American Journal of Epidemiology,* vol. 146, no. 7, pp. 552-557.

Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Blondal, T., Gylfason, A., Agnarsson, B.A., Benediktsdottir, K.R., Magnusdottir, D.N., Orlygsdottir, G., Jakobsdottir, M., Stacey, S.N., Sigurdsson, A., Wahlfors, T., Tammela, T., Breyer, J.P., McReynolds, K.M., Bradley, K.M., Saez, B., Godino, J., Navarrete, S., Fuertes, F., Murillo, L., Polo, E., Aben, K.K., van Oort, I.M., Suarez, B.K., Helfand, B.T., Kan, D., Zanon, C., Frigge, M.L., Kristjansson, K., Gulcher, J.R., Einarsson, G.V., Jonsson, E., Catalona, W.J., Mayordomo, J.I., Kiemeney, L.A., Smith, J.R., Schleutker, J., Barkardottir, R.B., Kong, A., Thorsteinsdottir, U., Rafnar, T. & Stefansson, K. 2009, "Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility", *Nature genetics,* vol. 41, no. 10, pp. 1122-1126.

Gudmundsson, J., Sulem, P., Gudbjartsson, D.F., Masson, G., Agnarsson, B.A., Benediktsdottir, K.R., Sigurdsson, A., Magnusson, O.T., Gudjonsson, S.A., Magnusdottir, D.N., Johannsdottir, H., Helgadottir, H.T., Stacey, S.N., Jonasdottir, A., Olafsdottir, S.B., Thorleifsson, G., Jonasson, J.G., Tryggvadottir, L., Navarrete, S., Fuertes, F., Helfand, B.T., Hu, Q., Csiki, I.E., Mates, I.N., Jinga, V., Aben, K.K., van Oort, I.M., Vermeulen, S.H., Donovan, J.L., Hamdy, F.C., Ng, C.F., Chiu, P.K., Lau, K.M., Ng, M.C., Gulcher, J.R., Kong, A., Catalona, W.J., Mayordomo, J.I., Einarsson, G.V., Barkardottir,

R.B., Jonsson, E., Mates, D., Neal, D.E., Kiemeney, L.A., Thorsteinsdottir, U., Rafnar, T. & Stefansson, K. 2012, "A study based on whole-genome sequencing yields a rare variant at 8q24 associated with prostate cancer", *Nature genetics*, vol. 44, no. 12, pp. 1326-1329.

Gudmundsson, J., Sulem, P., Manolescu, A., Amundadottir, L.T., Gudbjartsson, D., Helgason, A., Rafnar, T., Bergthorsson, J.T., Agnarsson, B.A., Baker, A., Sigurdsson, A., Benediktsdottir, K.R., Jakobsdottir, M., Xu, J., Blondal, T., Kostic, J., Sun, J., Ghosh, S., Stacey, S.N., Mouy, M., Saemundsdottir, J., Backman, V.M., Kristjansson, K., Tres, A., Partin, A.W., Albers-Akkers, M.T., Godino-Ivan Marcos, J., Walsh, P.C., Swinkels, D.W., Navarrete, S., Isaacs, S.D., Aben, K.K., Graif, T., Cashy, J., Ruiz-Echarri, M., Wiley, K.E., Suarez, B.K., Witjes, J.A., Frigge, M., Ober, C., Jonsson, E., Einarsson, G.V., Mayordomo, J.I., Kiemeney, L.A., Isaacs, W.B., Catalona, W.J., Barkardottir, R.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. 2007, "Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24", *Nature genetics*, vol. 39, no. 5, pp. 631-637.

Gudmundsson, J., Sulem, P., Rafnar, T., Bergthorsson, J.T., Manolescu, A., Gudbjartsson, D., Agnarsson, B.A., Sigurdsson, A., Benediktsdottir, K.R., Blondal, T., Jakobsdottir, M., Stacey, S.N., Kostic, J., Kristinsson, K.T., Birgisdottir, B., Ghosh, S., Magnusdottir, D.N., Thorlacius, S., Thorleifsson, G., Zheng, S.L., Sun, J., Chang, B.L., Elmore, J.B., Breyer, J.P., McReynolds, K.M., Bradley, K.M., Yaspan, B.L., Wiklund, F., Stattin, P., Lindstrom, S., Adami, H.O., McDonnell, S.K., Schaid, D.J., Cunningham, J.M., Wang, L., Cerhan, J.R., St Sauver, J.L., Isaacs, S.D., Wiley, K.E., Partin, A.W., Walsh, P.C., Polo, S., Ruiz-Echarri, M., Navarrete, S., Fuertes, F., Saez, B., Godino, J., Weijerman, P.C., Swinkels, D.W., Aben, K.K., Witjes, J.A., Suarez, B.K., Helfand, B.T., Frigge, M.L., Kristjansson, K., Ober, C., Jonsson, E., Einarsson, G.V., Xu, J., Gronberg, H., Smith, J.R., Thibodeau, S.N., Isaacs, W.B., Catalona, W.J., Mayordomo, J.I., Kiemeney, L.A., Barkardottir, R.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A. & Stefansson, K. 2008, "Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer", *Nature genetics*, vol. 40, no. 3, pp. 281-283.

Hahn, L.W., Ritchie, M.D. & Moore, J.H. 2003, "Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions", *Bioinformatics (Oxford, England)*, vol. 19, no. 3, pp. 376-382.

Haiman, C.A., Chen, G.K., Blot, W.J., Strom, S.S., Berndt, S.I., Kittles, R.A., Rybicki, B.A., Isaacs, W.B., Ingles, S.A., Stanford, J.L., Diver, W.R., Witte, J.S., Hsing, A.W., Nemesure, B., Rebbeck, T.R., Cooney, K.A., Xu, J., Kibel, A.S., Hu, J.J., John, E.M., Gueye, S.M., Watya, S., Signorello, L.B., Hayes, R.B., Wang, Z., Yeboah, E., Tettey, Y., Cai, Q., Kolb, S., Ostrander, E.A., Zeigler-Johnson, C., Yamamura, Y., Neslund-Dudas, C., Haslag-Minoff, J., Wu, W., Thomas, V., Allen, G.O., Murphy, A., Chang, B.L., Zheng, S.L., Leske, M.C., Wu, S.Y., Ray, A.M., Hennis, A.J., Thun, M.J., Carpten, J., Casey, G., Carter, E.N., Duarte, E.R., Xia, L.Y., Sheng, X., Wan, P., Pooler, L.C., Cheng, I., Monroe, K.R., Schumacher, F., Le Marchand, L., Kolonel, L.N., Chanock, S.J., Van Den Berg, D., Stram, D.O. & Henderson, B.E. 2011,

"Genome-wide association study of prostate cancer in men of African ancestry identifies a susceptibility locus at 17q21", *Nature genetics*, vol. 43, no. 6, pp. 570-573.

Haiman, C.A., Patterson, N., Freedman, M.L., Myers, S.R., Pike, M.C., Waliszewska, A., Neubauer, J., Tandon, A., Schirmer, C., McDonald, G.J., Greenway, S.C., Stram, D.O., Le Marchand, L., Kolonel, L.N., Frasco, M., Wong, D., Pooler, L.C., Ardlie, K., Oakley-Girvan, I., Whittemore, A.S., Cooney, K.A., John, E.M., Ingles, S.A., Altshuler, D., Henderson, B.E. & Reich, D. 2007, "Multiple regions within 8q24 independently affect risk for prostate cancer", *Nature genetics*, vol. 39, no. 5, pp. 638-644.

Hanahan, D. & Weinberg, R.A. 2000, "The hallmarks of cancer", *Cell*, vol. 100, no. 1, pp. 57-70.

Hanahan, D. & Weinberg, R.A. 2011, "Hallmarks of cancer: the next generation", *Cell*, vol. 144, no. 5, pp. 646-674.

Helfand, B.T., Fought, A.J., Loeb, S., Meeks, J.J., Kan, D. & Catalona, W.J. 2010, "Genetic prostate cancer risk assessment: common variants in 9 genomic regions are associated with cumulative risk", *The Journal of urology*, vol. 184, no. 2, pp. 501-505.

Hemminki, K. & Czene, K. 2002, "Age specific and attributable risks of familial prostate carcinoma from the family-cancer database", *Cancer*, vol. 95, no. 6, pp. 1346-1353.

Hollis, B.W., Marshall, D.T., Savage, S.J., Garrett-Mayer, E., Kindy, M.S. & Gattoni-Celli, S. 2012, "Vitamin D(3) supplementation, low-risk prostate cancer, and health disparities", *The Journal of steroid biochemistry and molecular biology*, .

Houlston, R.S. & Peto, J. 2004, "The search for low-penetrance cancer susceptibility alleles", *Oncogene*, vol. 23, no. 38, pp. 6471-6476.

Howlader, N., Noone, A., Krapcho, M., Neyman, N., Aminou, R., Altekruse, S., Kosary, C., Ruhl, J., Tatalovich, Z., Cho, H., Mariotto, A., Eisner, M.L., DR, Chen, H., Feuer, E., Cronin, K. & (eds) 2012, "SEER Cancer Statistics Review, 1975-2009 National Cancer Institute. Bethesda, MD, USA", .

Hsing, A.W., Sakoda, L.C. & Chua, S.,Jr 2007, "Obesity, metabolic syndrome, and prostate cancer", *The American Journal of Clinical Nutrition*, vol. 86, no. 3, pp. s843-57.

Huang da, W., Sherman, B.T. & Lempicki, R.A. 2009a, "Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists", *Nucleic acids research*, vol. 37, no. 1, pp. 1-13.

Huang da, W., Sherman, B.T. & Lempicki, R.A. 2009b, "Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources", *Nature protocols,* vol. 4, no. 1, pp. 44-57.

Hyytinen, E.R., Frierson, H.F.,Jr, Boyd, J.C., Chung, L.W. & Dong, J.T. 1999, "Three distinct regions of allelic loss at 13q14, 13q21-22, and 13q33 in prostate cancer", *Genes, chromosomes & cancer,* vol. 25, no. 2, pp. 108-114.

Isaacs, W. & Kainu, T. 2001, "Oncogenes and tumor suppressor genes in prostate cancer", *Epidemiologic reviews,* vol. 23, no. 1, pp. 36-41.

Ishkanian, A.S., Mallof, C.A., Ho, J., Meng, A., Albert, M., Syed, A., van der Kwast, T., Milosevic, M., Yoshimoto, M., Squire, J.A., Lam, W.L. & Bristow, R.G. 2009, "High-resolution array CGH identifies novel regions of genomic alteration in intermediate-risk prostate cancer", *The Prostate,* vol. 69, no. 10, pp. 1091-1100.

Jacobs, E.J., Rodriguez, C., Bain, E.B., Wang, Y., Thun, M.J. & Calle, E.E. 2007, "Cholesterol-lowering drugs and advanced prostate cancer incidence in a large U.S. cohort", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology,* vol. 16, no. 11, pp. 2213-2217.

Kahn, R.A., Cherfils, J., Elias, M., Lovering, R.C., Munro, S. & Schurmann, A. 2006, "Nomenclature for the human Arf family of GTP-binding proteins: ARF, ARL, and SAR proteins", *The Journal of cell biology,* vol. 172, no. 5, pp. 645-650.

Kilpinen, S., Autio, R., Ojala, K., Iljin, K., Bucher, E., Sara, H., Pisto, T., Saarela, M., Skotheim, R.I., Bjorkman, M., Mpindi, J.P., Haapa-Paananen, S., Vainio, P., Edgren, H., Wolf, M., Astola, J., Nees, M., Hautaniemi, S. & Kallioniemi, O. 2008, "Systematic bioinformatic analysis of expression levels of 17,330 human genes across 9,783 samples from 175 types of healthy and pathological tissues", *Genome biology,* vol. 9, no. 9, pp. R139.

Kilpivaara, O. & Aaltonen, L.A. 2013, "Diagnostic cancer genome sequencing and the contribution of germline variants", *Science (New York, N.Y.),* vol. 339, no. 6127, pp. 1559-1562.

Knudson, A.G.,Jr 1971, "Mutation and cancer: statistical study of retinoblastoma", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 68, no. 4, pp. 820-823.

Kobayashi, M., Ishida, H., Shindo, T., Niwa, S., Kino, M., Kawamura, K., Kamiya, N., Imamoto, T., Suzuki, H., Hirokawa, Y., Shiraishi, T., Tanizawa, T., Nakatani, Y. & Ichikawa, T. 2008, "Molecular analysis of multifocal prostate cancer by comparative genomic hybridization", *The Prostate,* vol. 68, no. 16, pp. 1715-1724.

Kote-Jarai, Z., Olama, A.A., Giles, G.G., Severi, G., Schleutker, J., Weischer, M., Campa, D., Riboli, E., Key, T., Gronberg, H., Hunter, D.J., Kraft, P., Thun, M.J., Ingles, S., Chanock, S., Albanes, D., Hayes, R.B., Neal, D.E., Hamdy, F.C., Donovan, J.L., Pharoah, P., Schumacher, F., Henderson, B.E., Stanford, J.L., Ostrander, E.A., Sorensen, K.D., Dork, T., Andriole, G., Dickinson, J.L., Cybulski, C., Lubinski, J., Spurdle, A., Clements, J.A., Chambers, S., Aitken, J., Gardiner, R.A., Thibodeau, S.N., Schaid, D., John, E.M., Maier, C., Vogel, W., Cooney, K.A., Park, J.Y., Cannon-Albright, L., Brenner, H., Habuchi, T., Zhang, H.W., Lu, Y.J., Kaneva, R., Muir, K., Benlloch, S., Leongamornlert, D.A., Saunders, E.J., Tymrakiewicz, M., Mahmud, N., Guy, M., O'Brien, L.T., Wilkinson, R.A., Hall, A.L., Sawyer, E.J., Dadaev, T., Morrison, J., Dearnaley, D.P., Horwich, A., Huddart, R.A., Khoo, V.S., Parker, C.C., Van As, N., Woodhouse, C.J., Thompson, A., Christmas, T., Ogden, C., Cooper, C.S., Lophatonanon, A., Southey, M.C., Hopper, J.L., English, D.R., Wahlfors, T., Tammela, T.L., Klarskov, P., Nordestgaard, B.G., Roder, M.A., Tybjaerg-Hansen, A., Bojesen, S.E., Travis, R., Canzian, F., Kaaks, R., Wiklund, F., Aly, M., Lindstrom, S., Diver, W.R., Gapstur, S., Stern, M.C., Corral, R., Virtamo, J., Cox, A., Haiman, C.A., Le Marchand, L., Fitzgerald, L., Kolb, S., Kwon, E.M., Karyadi, D.M., Orntoft, T.F., Borre, M., Meyer, A., Serth, J., Yeager, M., Berndt, S.I., Marthick, J.R., Patterson, B., Wokolorczyk, D., Batra, J., Lose, F., McDonnell, S.K., Joshi, A.D., Shahabi, A., Rinckleb, A.E., Ray, A., Sellers, T.A., Lin, H.Y., Stephenson, R.A., Farnham, J., Muller, H., Rothenbacher, D., Tsuchiya, N., Narita, S., Cao, G.W., Slavov, C., Mitev, V., Easton, D.F., Eeles, R.A., UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology, UK ProtecT Study Collaborators, The Australian Prostate Cancer BioResource & PRACTICAL Consortium 2011, "Seven prostate cancer susceptibility loci identified by a multi-stage genome-wide association study", *Nature genetics*, vol. 43, no. 8, pp. 785-791.

Kristiansson, K., Naukkarinen, J. & Peltonen, L. 2008, "Isolated populations and complex disease gene identification", *Genome biology*, vol. 9, no. 8, pp. 109-2008-9-8-109. Epub 2008 Aug 26.

Kundu, J.K. & Surh, Y.J. 2012, "Emerging avenues linking inflammation and cancer", *Free radical biology & medicine*, vol. 52, no. 9, pp. 2013-2037.

Kwon, E.M., Salinas, C.A., Kolb, S., Fu, R., Feng, Z., Stanford, J.L. & Ostrander, E.A. 2011, "Genetic polymorphisms in inflammation pathway genes and prostate cancer risk", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 20, no. 5, pp. 923-933.

Laitinen, S., Karhu, R., Sawyers, C.L., Vessella, R.L. & Visakorpi, T. 2002, "Chromosomal aberrations in prostate cancer xenografts detected by comparative genomic hybridization", *Genes, chromosomes & cancer*, vol. 35, no. 1, pp. 66-73.

Laitinen, V.H., Wahlfors, T., Saaristo, L., Rantapero, T., Pelttari, L.M., Kilpivaara, O., Laasanen, S.L., Kallioniemi, A., Nevanlinna, H., Aaltonen, L., Vessella, R.L., Auvinen, A., Visakorpi, T., Tammela, T.L. & Schleutker, J. 2013,

"HOXB13 G84E Mutation in Finland: Population-Based Analysis of Prostate, Breast, and Colorectal Cancer Risk", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 22, no. 3, pp. 452-460.

Lander, E.S. 2011, "Initial impact of the sequencing of the human genome", *Nature*, vol. 470, no. 7333, pp. 187-197.

Lange, E.M., Gillanders, E.M., Davis, C.C., Brown, W.M., Campbell, J.K., Jones, M., Gildea, D., Riedesel, E., Albertus, J., Freas-Lutz, D., Markey, C., Giri, V., Dimmer, J.B., Montie, J.E., Trent, J.M. & Cooney, K.A. 2003, "Genome-wide scan for prostate cancer susceptibility genes using families from the University of Michigan prostate cancer genetics project finds evidence for linkage on chromosome 17 near BRCA1", *The Prostate*, vol. 57, no. 4, pp. 326-334.

Lange, E.M., Robbins, C.M., Gillanders, E.M., Zheng, S.L., Xu, J., Wang, Y., White, K.A., Chang, B.L., Ho, L.A., Trent, J.M., Carpten, J.D., Isaacs, W.B. & Cooney, K.A. 2007, "Fine-mapping the putative chromosome 17q21-22 prostate cancer susceptibility gene to a 10 cM region based on linkage analysis", *Human genetics*, vol. 121, no. 1, pp. 49-55.

Lange, E.M., Salinas, C.A., Zuhlke, K.A., Ray, A.M., Wang, Y., Lu, Y., Ho, L.A., Luo, J. & Cooney, K.A. 2012, "Early onset prostate cancer has a significant genetic component", *The Prostate*, vol. 72, no. 2, pp. 147-156.

Lapointe, J., Li, C., Giacomini, C.P., Salari, K., Huang, S., Wang, P., Ferrari, M., Hernandez-Boussard, T., Brooks, J.D. & Pollack, J.R. 2007, "Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis", *Cancer research*, vol. 67, no. 18, pp. 8504-8510.

Latil, A., Chene, L., Mangin, P., Fournier, G., Berthon, P. & Cussenot, O. 2003, "Extensive analysis of the 13q14 region in human prostate tumors: DNA analysis and quantitative expression of genes lying in the interval of deletion", *The Prostate*, vol. 57, no. 1, pp. 39-50.

Latil, A., Morant, P., Fournier, G., Mangin, P., Berthon, P. & Cussenot, O. 2002, "CHC1-L, a candidate gene for prostate carcinogenesis at 13q14.2, is frequently affected by loss of heterozygosity and underexpressed in human prostate cancer", *International journal of cancer.Journal international du cancer*, vol. 99, no. 5, pp. 689-696.

Lavender, N.A., Rogers, E.N., Yeyeodu, S., Rudd, J., Hu, T., Zhang, J., Brock, G.N., Kimbro, K.S., Moore, J.H., Hein, D.W. & Kidd, L.C. 2012, "Interaction among apoptosis-associated sequence variants and joint effects on aggressive prostate cancer", *BMC medical genomics*, vol. 5, pp. 11-8794-5-11.

Li, M., Lou, X.Y. & Lu, Q. 2012, "On epistasis: a methodological review for detecting gene-gene interactions underlying various types of phenotypic traits", *Recent patents on biotechnology*, vol. 6, no. 3, pp. 230-236.

Li, X., Gast, A., Rudnai, P., Gurzau, E., Koppova, K., Hemminki, K. & Kumar, R. 2007, "ARLTS1 polymorphisms and basal cell carcinoma of the skin", *Hereditary cancer in clinical practice,* vol. 5, no. 1, pp. 25-29.

Lichtenstein, P., Holm, N.V., Verkasalo, P.K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A. & Hemminki, K. 2000, "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland", *The New England journal of medicine,* vol. 343, no. 2, pp. 78-85.

Lilja, H., Ulmert, D. & Vickers, A.J. 2008, "Prostate-specific antigen and prostate cancer: prediction, detection and monitoring", *Nature reviews.Cancer,* vol. 8, no. 4, pp. 268-278.

Lin, H.Y., Amankwah, E.K., Tseng, T.S., Qu, X., Chen, D.T. & Park, J.Y. 2013, "SNP-SNP Interaction Network in Angiogenesis Genes Associated with Prostate Cancer Aggressiveness", *PloS one,* vol. 8, no. 4, pp. e59688.

Linja, M.J., Savinainen, K.J., Saramaki, O.R., Tammela, T.L., Vessella, R.L. & Visakorpi, T. 2001, "Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer", *Cancer research,* vol. 61, no. 9, pp. 3550-3555.

Loza, M.J., McCall, C.E., Li, L., Isaacs, W.B., Xu, J. & Chang, B.L. 2007, "Assembly of inflammation-related genes for pathway-focused genetic analysis", *PloS one,* vol. 2, no. 10, pp. e1035.

Lu, W., Takahashi, H., Furusato, M., Maekawa, S., Nakano, M., Meng, C., Kikuchi, Y., Sudo, A. & Hano, H. 2006, "Allelotyping analysis at chromosome 13q of high-grade prostatic intraepithelial neoplasia and clinically insignificant and significant prostate cancers", *The Prostate,* vol. 66, no. 4, pp. 405-412.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F., McCarroll, S.A. & Visscher, P.M. 2009, "Finding the missing heritability of complex diseases", *Nature,* vol. 461, no. 7265, pp. 747-753.

Marshall, J.R. 2012, "Diet and prostate cancer prevention", *World journal of urology,* vol. 30, no. 2, pp. 157-165.

Masojc, B., Medrek, K., Debniak, T., Lubinski, J., Wyrwicz, L.S., Koczyk, G., Hoffmann, M. & Rychlewski, L. 2007, "ARLTS1 Trp149Stop mutation and the risk of ovarian cancer", *Cancer research,* vol. 67, no. 9, pp. 4533; author reply 4534.

Masojc, B., Mierzejewski, M., Cybulski, C., van de Wetering, T., Debniak, T., Gorski, B., Jaworowska, E., Tarnowska, C., Lenner, M., Scott, R.J. & Lubinski,

J. 2006, "Cancer Familial Aggregation (CFA) and G446A polymorphism in ARLTS1 gene", *Breast cancer research and treatment,* vol. 99, no. 1, pp. 59-62.

Matlashewski, G., Lamb, P., Pim, D., Peacock, J., Crawford, L. & Benchimol, S. 1984, "Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene", *The EMBO journal,* vol. 3, no. 13, pp. 3257-3262.

Matsuda, T. & Saika, K. 2009, "Comparison of time trends in prostate cancer incidence (1973-2002) in Asia, from cancer incidence in five continents, Vols IV-IX", *Japanese journal of clinical oncology,* vol. 39, no. 7, pp. 468-469.

Michaelson, J.J., Loguercio, S. & Beyer, A. 2009, "Detection and interpretation of expression quantitative trait loci (eQTL)", *Methods (San Diego, Calif.),* vol. 48, no. 3, pp. 265-276.

Min, J.L., Taylor, J.M., Richards, J.B., Watts, T., Pettersson, F.H., Broxholme, J., Ahmadi, K.R., Surdulescu, G.L., Lowy, E., Gieger, C., Newton-Cheh, C., Perola, M., Soranzo, N., Surakka, I., Lindgren, C.M., Ragoussis, J., Morris, A.P., Cardon, L.R., Spector, T.D. & Zondervan, K.T. 2011, "The use of genome-wide eQTL associations in lymphoblastoid cell lines to identify novel genetic pathways involved in complex traits", *PloS one,* vol. 6, no. 7, pp. e22070.

Misumi, T., Yamamoto, Y., Murakami, T., Kawai, Y., Ito, H., Eguchi, S., Yano, S., Nagao, K., Hara, T., Sakano, S., Naito, K., Sasaki, K., Fuellhase, C., Arenas, L., Fichtner, J. & Matsuyama, H. 2010, "Genetic alterations at 13q14 may correlate with differences in the biological behavior of prostate cancer between Japanese and Caucasian men", *Urologia internationalis,* vol. 84, no. 4, pp. 461-466.

Murtola, T.J., Tammela, T.L., Lahtela, J. & Auvinen, A. 2007, "Cholesterol-lowering drugs and prostate cancer risk: a population-based case-control study", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology,* vol. 16, no. 11, pp. 2226-2232.

Nakagawa, H., Akamatsu, S., Takata, R., Takahashi, A., Kubo, M. & Nakamura, Y. 2012, "Prostate cancer genomics, biology, and risk assessment through genome-wide association studies", *Cancer science,* vol. 103, no. 4, pp. 607-613.

Ng, D., Toure, O., Fontaine, L., McMaster, M.L., Goldin, L.R., Caporaso, N. & Toro, J.R. 2007, "No association of ARLTS1 polymorphisms and risk for familial chronic lymphocytic leukaemia", *British journal of haematology,* vol. 137, no. 2, pp. 173-175.

Nica, A.C., Parts, L., Glass, D., Nisbet, J., Barrett, A., Sekowska, M., Travers, M., Potter, S., Grundberg, E., Small, K., Hedman, A.K., Bataille, V., Tzenova Bell, J., Surdulescu, G., Dimas, A.S., Ingle, C., Nestle, F.O., di Meglio, P., Min, J.L.,

Wilk, A., Hammond, C.J., Hassanali, N., Yang, T.P., Montgomery, S.B., O'Rahilly, S., Lindgren, C.M., Zondervan, K.T., Soranzo, N., Barroso, I., Durbin, R., Ahmadi, K., Deloukas, P., McCarthy, M.I., Dermitzakis, E.T., Spector, T.D. & MuTHER Consortium 2011, "The architecture of gene regulatory variation across multiple human tissues: the MuTHER study", *PLoS genetics,* vol. 7, no. 2, pp. e1002003.

Nicholson, S.E. & Hilton, D.J. 1998, "The SOCS proteins: a new family of negative regulators of signal transduction", *Journal of leukocyte biology,* vol. 63, no. 6, pp. 665-668.

Ohori, M., Wheeler, T.M. & Scardino, P.T. 1994, "The New American Joint Committee on Cancer and International Union Against Cancer TNM classification of prostate cancer. Clinicopathologic correlations", *Cancer,* vol. 74, no. 1, pp. 104-114.

Pakkanen, S., Baffoe-Bonnie, A.B., Matikainen, M.P., Koivisto, P.A., Tammela, T.L., Deshmukh, S., Ou, L., Bailey-Wilson, J.E. & Schleutker, J. 2007, "Segregation analysis of 1,546 prostate cancer families in Finland shows recessive inheritance", *Human genetics,* vol. 121, no. 2, pp. 257-267.

Pasqualato, S., Renault, L. & Cherfils, J. 2002, "Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication", *EMBO reports,* vol. 3, no. 11, pp. 1035-1041.

Pelizzola, M. & Ecker, J.R. 2011, "The DNA methylome", *FEBS letters,* vol. 585, no. 13, pp. 1994-2000.

Petrocca, F., Iliopoulos, D., Qin, H.R., Nicoloso, M.S., Yendamuri, S., Wojcik, S.E., Shimizu, M., Di Leva, G., Vecchione, A., Trapasso, F., Godwin, A.K., Negrini, M., Calin, G.A. & Croce, C.M. 2006, "Alterations of the tumor suppressor gene ARLTS1 in ovarian cancer", *Cancer research,* vol. 66, no. 21, pp. 10287-10291.

Platz, E.A., Clinton, S.K. & Giovannucci, E. 2008, "Association between plasma cholesterol and prostate cancer in the PSA era", *International journal of cancer.Journal international du cancer,* vol. 123, no. 7, pp. 1693-1698.

Platz, E.A., Leitzmann, M.F., Visvanathan, K., Rimm, E.B., Stampfer, M.J., Willett, W.C. & Giovannucci, E. 2006, "Statin drugs and risk of advanced prostate cancer", *Journal of the National Cancer Institute,* vol. 98, no. 24, pp. 1819-1825.

Porkka, K.P., Ogg, E.L., Saramaki, O.R., Vessella, R.L., Pukkila, H., Lahdesmaki, H., van Weerden, W.M., Wolf, M., Kallioniemi, O.P., Jenster, G. & Visakorpi, T. 2011, "The miR-15a-miR-16-1 locus is homozygously deleted in a subset of prostate cancers", *Genes, chromosomes & cancer,* vol. 50, no. 7, pp. 499-509.

Rizos, C., Papassava, M., Golias, C. & Charalabopoulos, K. 2010, "Alcohol consumption and prostate cancer: a mini review", *Experimental oncology*, vol. 32, no. 2, pp. 66-70.

Robbins, C.M., Tembe, W.A., Baker, A., Sinari, S., Moses, T.Y., Beckstrom-Sternberg, S., Beckstrom-Sternberg, J., Barrett, M., Long, J., Chinnaiyan, A., Lowey, J., Suh, E., Pearson, J.V., Craig, D.W., Agus, D.B., Pienta, K.J. & Carpten, J.D. 2011, "Copy number and targeted mutational analysis reveals novel somatic events in metastatic prostate tumors", *Genome research*, vol. 21, no. 1, pp. 47-55.

Rokman, A., Koivisto, P.A., Matikainen, M.P., Kuukasjarvi, T., Poutiainen, M., Helin, H.J., Karhu, R., Kallioniemi, O.P. & Schleutker, J. 2001, "Genetic changes in familial prostate cancer by comparative genomic hybridization", *The Prostate*, vol. 46, no. 3, pp. 233-239.

Salovaara, R., Loukola, A., Kristo, P., Kaariainen, H., Ahtola, H., Eskelinen, M., Harkonen, N., Julkunen, R., Kangas, E., Ojala, S., Tulikoura, J., Valkamo, E., Jarvinen, H., Mecklin, J.P., Aaltonen, L.A. & de la Chapelle, A. 2000, "Population-based molecular detection of hereditary nonpolyposis colorectal cancer", *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 18, no. 11, pp. 2193-2200.

Saramaki, O.R., Porkka, K.P., Vessella, R.L. & Visakorpi, T. 2006a, "Genetic aberrations in prostate cancer by microarray analysis", *International journal of cancer.Journal international du cancer*, vol. 119, no. 6, pp. 1322-1329.

Saramaki, O.R., Tammela, T.L., Martikainen, P.M., Vessella, R.L. & Visakorpi, T. 2006b, "The gene for polycomb group protein enhancer of zeste homolog 2 (EZH2) is amplified in late-stage prostate cancer", *Genes, chromosomes & cancer*, vol. 45, no. 7, pp. 639-645.

Schaid, D.J. 1998, "Transmission disequilibrium, family controls, and great expectations", *American Journal of Human Genetics*, vol. 63, no. 4, pp. 935-941.

Schaid, D.J. 2004, "The complex genetic epidemiology of prostate cancer", *Human molecular genetics*, vol. 13 Spec No 1, pp. R103-21.

Schleutker, J., Matikainen, M., Smith, J., Koivisto, P., Baffoe-Bonnie, A., Kainu, T., Gillanders, E., Sankila, R., Pukkala, E., Carpten, J., Stephan, D., Tammela, T., Brownstein, M., Bailey-Wilson, J., Trent, J. & Kallioniemi, O.P. 2000, "A genetic epidemiological study of hereditary prostate cancer (HPC) in Finland: frequent HPCX linkage in families with late-onset disease", *Clinical cancer research : an official journal of the American Association for Cancer Research*, vol. 6, no. 12, pp. 4810-4815.

Schultz, C., Meier, M. & Schmid, H.P. 2011, "Nutrition, dietary supplements and adenocarcinoma of the prostate", *Maturitas*, vol. 70, no. 4, pp. 339-342.

Schumacher, F.R., Berndt, S.I., Siddiq, A., Jacobs, K.B., Wang, Z., Lindstrom, S., Stevens, V.L., Chen, C., Mondul, A.M., Travis, R.C., Stram, D.O., Eeles, R.A., Easton, D.F., Giles, G., Hopper, J.L., Neal, D.E., Hamdy, F.C., Donovan, J.L., Muir, K., Al Olama, A.A., Kote-Jarai, Z., Guy, M., Severi, G., Gronberg, H., Isaacs, W.B., Karlsson, R., Wiklund, F., Xu, J., Allen, N.E., Andriole, G.L., Barricarte, A., Boeing, H., Bueno-de-Mesquita, H.B., Crawford, E.D., Diver, W.R., Gonzalez, C.A., Gaziano, J.M., Giovannucci, E.L., Johansson, M., Le Marchand, L., Ma, J., Sieri, S., Stattin, P., Stampfer, M.J., Tjonneland, A., Vineis, P., Virtamo, J., Vogel, U., Weinstein, S.J., Yeager, M., Thun, M.J., Kolonel, L.N., Henderson, B.E., Albanes, D., Hayes, R.B., Feigelson, H.S., Riboli, E., Hunter, D.J., Chanock, S.J., Haiman, C.A. & Kraft, P. 2011, "Genome-wide association study identifies new prostate cancer susceptibility loci", *Human molecular genetics,* vol. 20, no. 19, pp. 3867-3875.

Sellick, G.S., Catovsky, D. & Houlston, R.S. 2006a, "Familial cancer associated with a polymorphism in ARLTS1", *The New England journal of medicine,* vol. 354, no. 11, pp. 1204-5; author reply 1204-5.

Sellick, G.S., Catovsky, D. & Houlston, R.S. 2006b, "Relationship between ARLTS1 polymorphisms and risk of chronic lymphocytic leukemia", *Leukemia research,* vol. 30, no. 12, pp. 1573-1576.

Sevignani, C., Calin, G.A., Siracusa, L.D. & Croce, C.M. 2006, "Mammalian microRNAs: a small world for fine-tuning gene expression", *Mammalian genome : official journal of the International Mammalian Genome Society,* vol. 17, no. 3, pp. 189-202.

Sfanos, K.S. & De Marzo, A.M. 2012, "Prostate cancer and inflammation: the evidence", *Histopathology,* vol. 60, no. 1, pp. 199-215.

Shang, Z., Zhu, S., Zhang, H., Li, L. & Niu, Y. 2013, "Germline Homeobox B13 (HOXB13) G84E Mutation and Prostate Cancer Risk in European Descendants: A Meta-analysis of 24 213 Cases and 73 631 Controls", *European urology,* .

Shen, M.M. & Abate-Shen, C. 2010, "Molecular genetics of prostate cancer: new prospects for old challenges", *Genes & development,* vol. 24, no. 18, pp. 1967-2000.

Siegel, R., DeSantis, C., Virgo, K., Stein, K., Mariotto, A., Smith, T., Cooper, D., Gansler, T., Lerro, C., Fedewa, S., Lin, C., Leach, C., Cannady, R.S., Cho, H., Scoppa, S., Hachey, M., Kirch, R., Jemal, A. & Ward, E. 2012, "Cancer treatment and survivorship statistics, 2012", *CA: a cancer journal for clinicians,* vol. 62, no. 4, pp. 220-241.

Siegel, R., Naishadham, D. & Jemal, A. 2012, "Cancer statistics, 2012", *CA: a cancer journal for clinicians,* vol. 62, no. 1, pp. 10-29.

Sossey-Alaoui, K., Kitamura, E., Head, K. & Cowell, J.K. 2002, "Characterization of FAM10A4, a member of the ST13 tumor suppressor gene family that maps

to the 13q14.3 region associated with B-Cell leukemia, multiple myeloma, and prostate cancer", *Genomics,* vol. 80, no. 1, pp. 5-7.

Stevens, J.B., Abdallah, B.Y., Horne, S.D., Liu, G., Bremer, S.W. & Heng, H.H. 2001, "Genetic and Epigenetic Heterogeneity in Cancer" in *eLS* John Wiley & Sons, Ltd.

Sun, J., Turner, A., Xu, J., Gronberg, H. & Isaacs, W. 2007, "Genetic variability in inflammation pathways and prostate cancer risk", *Urologic oncology,* vol. 25, no. 3, pp. 250-259.

Syrjakoski, K., Vahteristo, P., Eerola, H., Tamminen, A., Kivinummi, K., Sarantaus, L., Holli, K., Blomqvist, C., Kallioniemi, O.P., Kainu, T. & Nevanlinna, H. 2000, "Population-based study of BRCA1 and BRCA2 mutations in 1035 unselected Finnish breast cancer patients", *Journal of the National Cancer Institute,* vol. 92, no. 18, pp. 1529-1531.

Takata, R., Akamatsu, S., Kubo, M., Takahashi, A., Hosono, N., Kawaguchi, T., Tsunoda, T., Inazawa, J., Kamatani, N., Ogawa, O., Fujioka, T., Nakamura, Y. & Nakagawa, H. 2010, "Genome-wide association study identifies five new susceptibility loci for prostate cancer in the Japanese population", *Nature genetics,* vol. 42, no. 9, pp. 751-754.

Tao, S., Feng, J., Webster, T., Jin, G., Hsu, F.C., Chen, S.H., Kim, S.T., Wang, Z., Zhang, Z., Zheng, S.L., Isaacs, W.B., Xu, J. & Sun, J. 2012, "Genome-wide two-locus epistasis scans in prostate cancer using two European populations", *Human genetics,* vol. 131, no. 7, pp. 1225-1234.

Tavtigian, S.V., Simard, J., Teng, D.H., Abtin, V., Baumgard, M., Beck, A., Camp, N.J., Carillo, A.R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J.M., Frank, D., Frye, C., Ghaffari, S., Gupte, J.S., Hu, R., Iliev, D., Janecki, T., Kort, E.N., Laity, K.E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K.T., Reid, J.E., Richards, S., Schroeder, M., Smith, R., Snyder, S.C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A.M., Labrie, F., Skolnick, M.H., Neuhausen, S., Rommens, J. & Cannon-Albright, L.A. 2001, "A candidate prostate cancer susceptibility gene at chromosome 17p", *Nature genetics,* vol. 27, no. 2, pp. 172-180.

Taylor, B.S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B.S., Arora, V.K., Kaushik, P., Cerami, E., Reva, B., Antipin, Y., Mitsiades, N., Landers, T., Dolgalev, I., Major, J.E., Wilson, M., Socci, N.D., Lash, A.E., Heguy, A., Eastham, J.A., Scher, H.I., Reuter, V.E., Scardino, P.T., Sander, C., Sawyers, C.L. & Gerald, W.L. 2010, "Integrative genomic profiling of human prostate cancer", *Cancer cell,* vol. 18, no. 1, pp. 11-22.

Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A., Crenshaw, A., Cancel-Tassin, G., Staats, B.J., Wang, Z., Gonzalez-Bosquet, J., Fang, J., Deng, X., Berndt, S.I., Calle, E.E., Feigelson, H.S., Thun, M.J., Rodriguez, C., Albanes, D., Virtamo,

J., Weinstein, S., Schumacher, F.R., Giovannucci, E., Willett, W.C., Cussenot, O., Valeri, A., Andriole, G.L., Crawford, E.D., Tucker, M., Gerhard, D.S., Fraumeni, J.F.,Jr, Hoover, R., Hayes, R.B., Hunter, D.J. & Chanock, S.J. 2008, "Multiple loci identified in a genome-wide association study of prostate cancer", *Nature genetics,* vol. 40, no. 3, pp. 310-315.

Todd, R. & Wong, D.T. 1999, "Oncogenes", *Anticancer Research,* vol. 19, no. 6A, pp. 4729-4746.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., Lee, C., Montie, J.E., Shah, R.B., Pienta, K.J., Rubin, M.A. & Chinnaiyan, A.M. 2005, "Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer", *Science (New York, N.Y.),* vol. 310, no. 5748, pp. 644-648.

Trapnell, C., Pachter, L. & Salzberg, S.L. 2009, "TopHat: discovering splice junctions with RNA-Seq", *Bioinformatics (Oxford, England),* vol. 25, no. 9, pp. 1105-1111.

Valeri, A., Fromont, G., Sakr, W., Azzouzi, R., Dey, J., Chantrel-Groussard, K., Latil, A., Berthon, P., Cussenot, O., Pontes, J.E. & Cher, M.L. 2005, "High frequency of allelic losses in high-grade prostate cancer is associated with biochemical progression after radical prostatectomy", *Urologic oncology,* vol. 23, no. 2, pp. 87-92.

Varghese, J.S. & Easton, D.F. 2010, "Genome-wide association studies in common cancers--what have we learnt?", *Current opinion in genetics & development,* vol. 20, no. 3, pp. 201-209.

Veyrieras, J.B., Kudaravalli, S., Kim, S.Y., Dermitzakis, E.T., Gilad, Y., Stephens, M. & Pritchard, J.K. 2008, "High-resolution mapping of expression-QTLs yields insight into human gene regulation", *PLoS genetics,* vol. 4, no. 10, pp. e1000214.

Via, M., Gignoux, C. & Burchard, E.G. 2010, "The 1000 Genomes Project: new opportunities for research and social challenges", *Genome medicine,* vol. 2, no. 1, pp. 3.

Visakorpi, T., Kallioniemi, A.H., Syvanen, A.C., Hyytinen, E.R., Karhu, R., Tammela, T., Isola, J.J. & Kallioniemi, O.P. 1995, "Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization", *Cancer research,* vol. 55, no. 2, pp. 342-347.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A.,Jr & Kinzler, K.W. 2013, "Cancer genome landscapes", *Science (New York, N.Y.),* vol. 339, no. 6127, pp. 1546-1558.

Vucic, E.A., Thu, K.L., Robison, K., Rybaczyk, L.A., Chari, R., Alvarez, C.E. & Lam, W.L. 2012, "Translating cancer 'omics' to improved outcomes", *Genome research,* vol. 22, no. 2, pp. 188-195.

Wang, M., Liu, F., Hsing, A.W., Wang, X., Shao, Q., Qi, J., Ye, Y., Wang, Z., Chen, H., Gao, X., Wang, G., Chu, L.W., Ding, Q., OuYang, J., Gao, X., Huang, Y., Chen, Y., Gao, Y.T., Zhang, Z.F., Rao, J., Shi, R., Wu, Q., Zhang, Y., Jiang, H., Zheng, J., Hu, Y., Guo, L., Lin, X., Tao, S., Jin, G., Sun, J., Lu, D., Zheng, S.L., Sun, Y., Mo, Z., Yin, C., Zhang, Z. & Xu, J. 2012, "Replication and cumulative effects of GWAS-identified genetic variations for prostate cancer in Asians: a case-control study in the ChinaPCa consortium", *Carcinogenesis*, vol. 33, no. 2, pp. 356-360.

Ward, L.D. & Kellis, M. 2012, "HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants", *Nucleic acids research*, vol. 40, no. Database issue, pp. D930-4.

Wei, M.Y. & Giovannucci, E.L. 2012, "Lycopene, Tomato Products, and Prostate Cancer Incidence: A Review and Reassessment in the PSA Screening Era", *Journal of oncology*, vol. 2012, pp. 271063.

Wennerberg, K., Rossman, K.L. & Der, C.J. 2005, "The Ras superfamily at a glance", *Journal of cell science*, vol. 118, no. Pt 5, pp. 843-846.

Woo, J.S., Suh, H.Y., Park, S.Y. & Oh, B.H. 2006, "Structural basis for protein recognition by B30.2/SPRY domains", *Molecular cell*, vol. 24, no. 6, pp. 967-976.

Wyatt, A.W., Mo, F., Wang, Y. & Collins, C.C. 2013, "The diverse heterogeneity of molecular alterations in prostate cancer identified through next-generation sequencing", *Asian Journal of Andrology*.

Xin, L. 2012, "Cells of origin for cancer: an updated view from prostate cancer", *Oncogene*.

Xu, J., Dimitrov, L., Chang, B.L., Adams, T.S., Turner, A.R., Meyers, D.A., Eeles, R.A., Easton, D.F., Foulkes, W.D., Simard, J., Giles, G.G., Hopper, J.L., Mahle, L., Moller, P., Bishop, T., Evans, C., Edwards, S., Meitz, J., Bullock, S., Hope, Q., Hsieh, C.L., Halpern, J., Balise, R.N., Oakley-Girvan, I., Whittemore, A.S., Ewing, C.M., Gielzak, M., Isaacs, S.D., Walsh, P.C., Wiley, K.E., Isaacs, W.B., Thibodeau, S.N., McDonnell, S.K., Cunningham, J.M., Zarfas, K.E., Hebbring, S., Schaid, D.J., Friedrichsen, D.M., Deutsch, K., Kolb, S., Badzioch, M., Jarvik, G.P., Janer, M., Hood, L., Ostrander, E.A., Stanford, J.L., Lange, E.M., Beebe-Dimmer, J.L., Mohai, C.E., Cooney, K.A., Ikonen, T., Baffoe-Bonnie, A., Fredriksson, H., Matikainen, M.P., Tammela, T.L., Bailey-Wilson, J., Schleutker, J., Maier, C., Herkommer, K., Hoegel, J.J., Vogel, W., Paiss, T., Wiklund, F., Emanuelsson, M., Stenman, E., Jonsson, B.A., Gronberg, H., Camp, N.J., Farnham, J., Cannon-Albright, L.A., Seminara, D. & The ACTANE Consortium 2005a, "A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics", *American Journal of Human Genetics*, vol. 77, no. 2, pp. 219-229.

Xu, J., Lowey, J., Wiklund, F., Sun, J., Lindmark, F., Hsu, F.C., Dimitrov, L., Chang, B., Turner, A.R., Liu, W., Adami, H.O., Suh, E., Moore, J.H., Zheng, S.L., Isaacs, W.B., Trent, J.M. & Gronberg, H. 2005b, "The interaction of four genes in the inflammation pathway significantly predicts prostate cancer risk", *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*, vol. 14, no. 11 Pt 1, pp. 2563-2568.

Xu, J., Mo, Z., Ye, D., Wang, M., Liu, F., Jin, G., Xu, C., Wang, X., Shao, Q., Chen, Z., Tao, Z., Qi, J., Zhou, F., Wang, Z., Fu, Y., He, D., Wei, Q., Guo, J., Wu, D., Gao, X., Yuan, J., Wang, G., Xu, Y., Wang, G., Yao, H., Dong, P., Jiao, Y., Shen, M., Yang, J., Ou-Yang, J., Jiang, H., Zhu, Y., Ren, S., Zhang, Z., Yin, C., Gao, X., Dai, B., Hu, Z., Yang, Y., Wu, Q., Chen, H., Peng, P., Zheng, Y., Zheng, X., Xiang, Y., Long, J., Gong, J., Na, R., Lin, X., Yu, H., Wang, Z., Tao, S., Feng, J., Sun, J., Liu, W., Hsing, A., Rao, J., Ding, Q., Wiklund, F., Gronberg, H., Shu, X.O., Zheng, W., Shen, H., Jin, L., Shi, R., Lu, D., Zhang, X., Sun, J., Zheng, S.L. & Sun, Y. 2012, "Genome-wide association study in Chinese men identifies two new prostate cancer risk loci at 9q31.2 and 19q13.4", *Nature genetics*, vol. 44, no. 11, pp. 1231-1235.

Xu, J., Zheng, S.L., Komiya, A., Mychaleckyj, J.C., Isaacs, S.D., Hu, J.J., Sterling, D., Lange, E.M., Hawkins, G.A., Turner, A., Ewing, C.M., Faith, D.A., Johnson, J.R., Suzuki, H., Bujnovszky, P., Wiley, K.E., DeMarzo, A.M., Bova, G.S., Chang, B., Hall, M.C., McCullough, D.L., Partin, A.W., Kassabian, V.S., Carpten, J.D., Bailey-Wilson, J.E., Trent, J.M., Ohar, J., Bleecker, E.R., Walsh, P.C., Isaacs, W.B. & Meyers, D.A. 2002, "Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk", *Nature genetics*, vol. 32, no. 2, pp. 321-325.

Yang, X.Y., Yu, H., Lin, W. & Peng, Z.L. 2009, "Effects of ARLTS1 gene on growth and apoptosis of epithelial ovarian cancer SKOV3 cells", *Sichuan da xue xue bao.Yi xue ban = Journal of Sichuan University.Medical science edition*, vol. 40, no. 1, pp. 6-10.

Yang, X.Y., Yu, H. & Peng, Z.L. 2011, "Involvement of ARLTS1 in chemotherapy and apoptosis in ovarian cancer cell line", *Archives of Gynecology and Obstetrics*, vol. 284, no. 5, pp. 1241-1246.

Yang, X.Y., Yu, H. & Peng, Z.L. 2009, "Polymorphism of ARLTS1 gene in early-onset ovarian cancer", *Sichuan da xue xue bao.Yi xue ban = Journal of Sichuan University.Medical science edition*, vol. 40, no. 6, pp. 978-982.

Yang, X.Y., Yu, H., Xi, M.R., Yang, K.X., Pan, X.L., Hu, M. & Peng, Z.L. 2009, "Association of the ARLTS1 variants with familial ovarian cancer risk in China", *International journal of gynecological cancer : official journal of the International Gynecological Cancer Society*, vol. 19, no. 4, pp. 585-590.

Yeager, M., Chatterjee, N., Ciampa, J., Jacobs, K.B., Gonzalez-Bosquet, J., Hayes, R.B., Kraft, P., Wacholder, S., Orr, N., Berndt, S., Yu, K., Hutchinson, A., Wang, Z., Amundadottir, L., Feigelson, H.S., Thun, M.J., Diver, W.R.,

Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F.R., Cancel-Tassin, G., Cussenot, O., Valeri, A., Andriole, G.L., Crawford, E.D., Haiman, C.A., Henderson, B., Kolonel, L., Le Marchand, L., Siddiq, A., Riboli, E., Key, T.J., Kaaks, R., Isaacs, W., Isaacs, S., Wiley, K.E., Gronberg, H., Wiklund, F., Stattin, P., Xu, J., Zheng, S.L., Sun, J., Vatten, L.J., Hveem, K., Kumle, M., Tucker, M., Gerhard, D.S., Hoover, R.N., Fraumeni, J.F.,Jr, Hunter, D.J., Thomas, G. & Chanock, S.J. 2009, "Identification of a new prostate cancer susceptibility locus on chromosome 8q24", *Nature genetics*, vol. 41, no. 10, pp. 1055-1057.

Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N., Wang, Z., Welch, R., Staats, B.J., Calle, E.E., Feigelson, H.S., Thun, M.J., Rodriguez, C., Albanes, D., Virtamo, J., Weinstein, S., Schumacher, F.R., Giovannucci, E., Willett, W.C., Cancel-Tassin, G., Cussenot, O., Valeri, A., Andriole, G.L., Gelmann, E.P., Tucker, M., Gerhard, D.S., Fraumeni, J.F.,Jr, Hoover, R., Hunter, D.J., Chanock, S.J. & Thomas, G. 2007, "Genome-wide association study of prostate cancer identifies a second risk locus at 8q24", *Nature genetics*, vol. 39, no. 5, pp. 645-649.

Yendamuri, S., Trapasso, F. & Calin, G.A. 2008, "ARLTS1 - a novel tumor suppressor gene", *Cancer letters*, vol. 264, no. 1, pp. 11-20.

Yendamuri, S., Trapasso, F., Ferracin, M., Cesari, R., Sevignani, C., Shimizu, M., Rattan, S., Kuroki, T., Dumon, K.R., Bullrich, F., Liu, C.G., Negrini, M., Williams, N.N., Kaiser, L.R., Croce, C.M. & Calin, G.A. 2007, "Tumor suppressor functions of ARLTS1 in lung cancers", *Cancer research*, vol. 67, no. 16, pp. 7738-7745.

Yin, Z., Spitz, M.R., Babaian, R.J., Strom, S.S., Troncoso, P. & Kagan, J. 1999, "Limiting the location of a putative human prostate cancer tumor suppressor gene at chromosome 13q14.3", *Oncogene*, vol. 18, no. 52, pp. 7576-7583.

Zeegers, M.P., Jellema, A. & Ostrer, H. 2003, "Empiric risk of prostate carcinoma for relatives of patients with prostate carcinoma: a meta-analysis", *Cancer*, vol. 97, no. 8, pp. 1894-1903.

Zhang, Y., Leaves, N.I., Anderson, G.G., Ponting, C.P., Broxholme, J., Holt, R., Edser, P., Bhattacharyya, S., Dunham, A., Adcock, I.M., Pulleyn, L., Barnes, P.J., Harper, J.I., Abecasis, G., Cardon, L., White, M., Burton, J., Matthews, L., Mott, R., Ross, M., Cox, R., Moffatt, M.F. & Cookson, W.O. 2003, "Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma", *Nature genetics*, vol. 34, no. 2, pp. 181-186.

Zheng, S.L., Liu, W., Wiklund, F., Dimitrov, L., Balter, K., Sun, J., Adami, H.O., Johansson, J.E., Sun, J., Chang, B., Loza, M., Turner, A.R., Bleecker, E.R., Meyers, D.A., Carpten, J.D., Duggan, D., Isaacs, W.B., Xu, J. & Gronberg, H. 2006, "A comprehensive association study for genes in inflammation pathway

provides support for their roles in prostate cancer risk in the CAPS study", *The Prostate,* vol. 66, no. 14, pp. 1556-1564.

Zheng, S.L., Sun, J., Wiklund, F., Smith, S., Stattin, P., Li, G., Adami, H.O., Hsu, F.C., Zhu, Y., Balter, K., Kader, A.K., Turner, A.R., Liu, W., Bleecker, E.R., Meyers, D.A., Duggan, D., Carpten, J.D., Chang, B.L., Isaacs, W.B., Xu, J. & Gronberg, H. 2008, "Cumulative association of five genetic variants with prostate cancer", *The New England journal of medicine,* vol. 358, no. 9, pp. 910-919.

Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. 2012, "The mystery of missing heritability: Genetic interactions create phantom heritability", *Proceedings of the National Academy of Sciences of the United States of America,* vol. 109, no. 4, pp. 1193-1198.

ORIGINAL COMMUNICATIONS

I

Siltanen S., Syrjäkoski K., Fagerholm R., Ikonen T., Lipman P., Mallott J., Holli K., Tammela T.L., Järvinen H.J., Mecklin J.P., Aittomäki K., Blomqvist C., Bailey-Wilson J.E., Nevanlinna H., Aaltonen L.A., Schleutker J., Vahteristo P. 2008, "*ARLTS1* germline variants and the risk for breast, prostate and colorectal cancer", *Eur J Hum Genet*. vol. 16, no. 8, pp. 983-91.

http://www.nature.com/ejhg/journal/v16/n8/pdf/ejhg200843a.pdf

OPEN ACCESS Freely available online

PLoS one

Contribution of *ARLTS1* Cys148Arg (T442C) Variant with Prostate Cancer Risk and ARLTS1 Function in Prostate Cancer Cells

Sanna Siltanen[1☯], Tiina Wahlfors[1☯], Martin Schindler[1], Outi R. Saramäki[1], John Patrick Mpindi[2], Leena Latonen[1], Robert L. Vessella[3,4], Teuvo L. J. Tammela[5], Olli Kallioniemi[2], Tapio Visakorpi[1], Johanna Schleutker[6]*

1 Institute of Biomedical Technology/BioMediTech, University of Tampere and Centre for Laboratory Medicine, Tampere University Hospital, Tampere, Finland, **2** Institute for Molecular Medicine, University of Helsinki, Helsinki, Finland, **3** Department of Urology, University of Washington, Seattle, Washington, United States of America, **4** Puget Sound VA Medical System, Seattle, Washington, United States of America, **5** Department of Urology, University of Tampere and Tampere University Hospital, Tampere, Finland, **6** Department of Medical Biochemistry and Genetics, University of Turku, Turku, Finland

Abstract

ARLTS1 is a recently characterized tumor suppressor gene at 13q14.3, a region frequently deleted in both sporadic and hereditary prostate cancer (PCa). *ARLTS1* variants, especially Cys148Arg (T442C), increase susceptibility to different cancers, including PCa. In this study the role of Cys148Arg substitution was investigated as a risk factor for PCa using both genetic and functional analysis. Cys148Arg genotypes and expression of the *ARLTS1* were explored in a large set of familial and unselected PCa cases, clinical tumor samples, xenografts, prostate cancer cell lines and benign prostatic hyperplasia (BPH) samples. The frequency of the variant genotype CC was significantly higher in familial (OR = 1.67, 95% CI = 1.08–2.56, $P = 0.019$) and unselected patients (OR = 1.52, 95% CI = 1.18–1.97, $P = 0.001$) and the overall risk was increased (OR = 1.54, 95% CI = 1.20–1.98, $P = 0.0007$). Additional analysis with clinicopathological data revealed an association with an aggressive disease (OR = 1.28, 95% CI = 1.05-∞, $P = 0.02$). The CC genotype of the Cys148Arg variant was also contributing to the lowered *ARLTS1* expression status in lymphoblastoid cells from familial patients. In addition significantly lowered *ARLTS1* expression was observed in clinical tumor samples compared to BPH samples ($P = 0.01$). The *ARLTS1* co-expression signature based on previously published microarray data was generated from 1587 cancer samples confirming the low expression of ARLTS1 in PCa and showed that *ARLTS1* expression was strongly associated with immune processes. This study provides strong confirmation of the important role of *ARLTS1* Cys148Arg variant as a contributor in PCa predisposition and a potential marker for aggressive disease outcome.

Citation: Siltanen S, Wahlfors T, Schindler M, Saramäki OR, Mpindi JP, et al. (2011) Contribution of *ARLTS1* Cys148Arg (T442C) Variant with Prostate Cancer Risk and ARLTS1 Function in Prostate Cancer Cells. PLoS ONE 6(10): e26595. doi:10.1371/journal.pone.0026595

Editor: Surinder K. Batra, University of Nebraska Medical Center, United States of America

Received April 26, 2011; **Accepted** September 29, 2011; **Published** October 20, 2011

Funding: This work was supported by The Competitive Research Funding of the Pirkanmaa Hospital District (grant number 9L091), Reino Lahtikari Foundation, Finnish Cancer Organisations, Sigrid Juselius Foundation and Academy of Finland [grant number 126714 to T.W. and 116437 to J.S.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Johanna.Schleutker@utu.fi

☯ These authors contributed equally to this work.

Introduction

ADP-ribosylation factor-like tumor suppressor gene 1 (*ARLTS1*), also known as ADP-ribosylation factor-like protein 11 (*ARL11*), is a newly characterized gene located at locus 13q14.3. ARLTS1 belongs to the ADP-ribosylation factor (ARF)-ARF-like (ARL) family of the Ras protein superfamily [1]. ARFs are guanine-nucleotide-binding proteins which are critical components of several different eukaryotic vesicle trafficking pathways. As with other members of the Ras superfamily, ARFs function as molecular switches by cycling between inactive GDP- and active GTP-bound conformations [2]. ARLTS1 has been characterized as an intracellular protein having tissue specific expression in the lung and leucocytes. *ARLTS1* variants, such as the nonsense polymorphism Trp149Stop (G446A) and missense polymorphism Cys148Arg (T442C), have been suggested to have a role in different cancers [3–6].

Prostate cancer is the most frequently diagnosed cancer in males in many countries, including Finland. Aging and improved diagnostics most evidently increase the number of new cases, but the incidence is influenced also by some unknown factors. Growing number of new cases create pressure to health care system and new tools for PCa diagnostics, prognostics and treatment are required, especially to avoid over treatment and unnecessary biopsies. During the last several years there has been extensive research in PCa etiology and genome-wide association studies have revealed several common low penetrance genetic alterations. The association of these variants with clinicopathologic features and prognosis remains unclear and results are lacking clinical implications.

We recently showed a significant association withCys148Arg (T442C) variant and the risk of PCa [7]. Further evaluation of this variant is warranted to increase the power of the association and study the functional role of the variant in PCa. More samples are

also needed to evaluate the implication of this variant to clinical outcome and a potential role in predictive biomarker of PCa.

Besides the genetic variants, DNA copy number aberrations are one of the most frequently observed genetic changes in familial and sporadic PCa [8–10]. In most of the cases target genes for the aberrations are not fully identified. Interestingly, allelic imbalance (AI) has been detected at 13q14.2-13q14.3, and it is an important event in the progression of localized PCa [11]. Differences of 13q14 loss of heterozygosity (LOH) in different PCa groups could also be used to distinguish clinically insignificant PCa [12,13].

In this study we analyzed the role of *ARLTS1* in more detail, especially the role of Cys148Arg (T442C) in PCa risk. Chromosomal aberration in 13q14.3 was analyzed with aCGH to evaluate the *ARLTS1* copy number changes in PCa xenografts and cell lines. The expression of *ARLTS1* was studied in clinical tumor samples, BPH samples and also co-expression data form previously published data was analyzed.

Methods

Study population

All samples collected are of Finnish origin. Identification and collection of the Finnish HPC families have been described elsewhere [14]. The familial samples genotyped in this study had at least one affected first or second degree relative. The clinical characteristics of the familial patients can be found in Table S1.

The unselected consecutive prostate cancer patients were diagnosed with PCa between 1999 and 2005 in the Department of Urology at Tampere University Hospital. The hospital is a regional referral center in the area for all patients with PCa, which results in an unselected, population-based collection of patients. The clinical characteristics of the consecutive unselected PCa can be found in Table S1.

A set of benign prostatic hyperplasia (BPH) cases was also used in this study. The diagnosis of this BPH cohort was based on lower-urinary tract symptoms, free uroflowmetry, and evidence of increased prostate size obtained by palpation or transrectal ultrasound. If PSA was elevated or digital rectal examination or transrectal ultrasound showed any abnormality indicative of PCa, the patients underwent biopsies to exclude diagnoses of PCa, high-grade prostate intraepithelial neoplasia (PIN), atypical small acinar cell proliferation (ASAP), or suspicion of malignancy.

Clinical prostate tumors were obtained from Tampere University Hospital (Tampere, Finland) including freshly frozen prostate tumor specimens representing benign prostate hyperplasia (BPH, $n = 14$), androgen-dependent ($n = 14$) and hormone-refractory ($n = 6$) carcinomas. The specimens were histologically examined for the presence of tumor cells using H&E staining. Only samples containing >60% cancerous or hyperplastic epithelial cells were selected for the analyses. The BPH samples were obtained from prostatectomy specimens from cancer patients and were histologically verified not to contain any cancerous cells. Samples from hormone-refractory carcinomas were obtained from transurethral resections of prostate (TURP) from patients experiencing urethral obstruction despite ongoing hormonal therapy. The time from the beginning of hormonal therapy to progression (TURP) varied from 15 to 60 months. The use of clinical tumor material was approved by the Ethical Committee of Tampere University Hospital.

The control samples consisted of DNA samples from anonymous, voluntary, and healthy blood donors obtained from the Blood Center of the Finnish Red Cross in Tampere. The EDTA blood samples used as reference in Q-RT-PCR were from healthy anonymous donors.

Patient information and samples were obtained with full informed consent. The study was performed under appropriate research permissions from the Ethics Committees of the Tampere University Hospital, Finland, as well as the Ministry of Social Affairs and Health in Finland.

Cell lines and xenografts

The PCa cell lines LNCaP, DU145, PC-3, NCI-660 and 22Rv1 were obtained from American Type Culture Collection (Manassas, VA, USA), whereas LAPC-4 was kindly provided by Dr. Charles Sawyers (UCLA, Los Angeles, CA), and the VCaP cell line was provided by Dr. Jack Schalken (Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands). All cell lines were cultured under recommended conditions. PrEC primary cells were obtained from Lonza (Walkersville, MD, USA). EP156T is an h-tert immortalized normal prostate epithelial cell line [15] that was made available by one of the authors (O.K.). DNA was extracted according to standard laboratory protocols and total RNA was extracted with the TRIzol reagent (Invitrogen Life Technologies, Carlsbad, CA, USA).

Nineteen human PCa xenografts of the LuCaP series used in direct sequencing and fifteen of them used in Q-RT-PCR were made available by one of the authors (R.L.V.) and have been described elsewhere [16,17].

The lymphoblastoid cell lines were derived by Epstein-Barr virus transformation of peripheral mononuclear leucocytes from patients. Lymphoblastoid cell lines were grown in RPMI-1640 medium (Lonza, Walkersville, MD, USA) supplemented with 10% fetal bovine serum (Sigma-Aldrich, St. Louis, MO, USA) and antibiotics. Cell pellets were snap-frozen and total RNA was extracted from cells with Trizol® according to the instructions of the manufacturer (Invitrogen, Carlsbad, CA, USA).

Mutation screening

Mutation screening of the genomic DNA was performed by direct sequencing. Sequencing was performed in an Applied Biosystems 3130xl Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA, USA) according to the instructions of the manufacturer. Primers and PCR conditions used in the mutation screening are available upon request.

Genotyping

Genotyping was done with the Custom TaqMan® SNP Genotyping Assay using the ABI Prism 7900HT sequence detection system (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer's instructions. Primers and probes for the Cys148Arg (T442C) SNP rs3803185 were supplied by Applied Biosystems via the File Builder 3.1 software.

Quantitative real-time reverse transcription-PCR

Gene expression analyses were done using LightCycler® (Roche, Mannheim, Germany) and Bio-Rad CFX96™ Real-Time PCR detection system (Bio-Rad Laboratories, Hercules, CA, USA). Total RNA was isolated from cell lines, clinical tumor samples and xenografts as described previously [18,19]. RNA from cell lines was reverse transcribed into first-strand cDNA using SuperScript™ III First-Strand Synthesis SuperMix kit (Invitrogen, Carlsbad, CA, USA) and random hexamers. RNA from clinical tumor samples and xenografts was reverse transcribed as described previously [18,19]. Normal human prostate RNA was obtained from Ambion (Cambridgeshire, United Kingdom). For the LightCycler® analyses, primers and probe sets were obtained from TIB MolBiol (Berlin, Germany). The PCR reactions were

performed with a LightCycler® FastStart DNA Master^{PLUS} HybProbe kit (Roche, Mannheim, Germany). LightCycler® software (Roche, Mannheim, Germany) was used for data analysis. The TaqMan assay for *ARLTS1* was used for Bio-Rad analyses. The primers and probes were obtained from TIB MolBiol (Berlin, Germany). Bio-Rad's iQ Supermix was used for the PCR reactions and CFX Manager Software^{TM} version 1.6 for data analysis. Expression levels of *ARLTS1* were normalized against the housekeeping gene *TBP* (TATA box binding protein) or against RNA levels. *TBP* was chosen as the reference gene, because there are no known retropseudogenes for it, and the expression of *TBP* is lower than that of many commonly used, abundantly expressed reference genes [20].

Western blotting

Cells were lysed in a buffer containing 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5% Triton-X-100, 1 mM DTT, 1 mM PMSF and 1× complete protease inhibitor cocktail (Roche, Mannheim, Germany). Lysates were sonicated 4×30 s with Bioruptor instrument (Diagenode, Liege, Belgium) and the cellular debris was removed by centrifugation. Proteins were separated by polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membrane (Immobilon-P; Millipore, Billerica, MA, USA). Primary antibodies used were anti-ARL11 (Aviva Systems Biology, San Diego, CA, USA) and anti-actin (pan, clone ACTN05, Neomarkers, Fremont, CA, USA), which were detected by HRP-conjugated secondary antibodies (DAKO, Denmark) and Western blotting luminol reagent (Santa Cruz Biotechnologies, CA, USA) by autoradiography.

Array comparative genomic hybridization

Array comparative genomic hybridization (aCGH) was performed on PCa cell lines and xenografts with the Human Genome CGH Microarray Kit 244A (Agilent, Santa Clara, CA, USA), and as described in Saramäki OR et al, 2006 [19].

Statistical analysis

Association of the Cys148Arg (T442C) variant was tested by logistic regression analysis using SPSS statistical software package (SPSS 15.0; SPSS Inc, Chicago, IL, USA). Association with clinical and pathological features of the disease (age at onset, PSA value at diagnosis, T, N, and M-stage, WHO grade and Gleason score) was tested among familial and unselected PCa cases by the Kruskal–Wallis test, Fisher's exact test, and t-test included in R "Stats" package. PCa patients were classified as having an aggressive disease if they had any of the following characteristics: a locally advanced or metastatic tumor (stage T3, T4, N1, or M1, based on pathology if radical prostatectomy was done; otherwise, clinical stage), tumor Gleason grade at diagnosis $\geq$7, poorly differentiated grade (WHO grade III, if no Gleason grade available) or pretreatment PSA at diagnosis $\geq$20 ng/ml.

Co-expression analysis of *ARLTS1* in previously published microarray datasets

The *ARLTS1* co-expression signature was generated from the GeneSapiens [21] database of mRNA expression. *ARLTS1* expression among 1587 tumor samples and among 497 normal samples was determined. We used Pearson correlation to identify genes that positively and negatively correlated with *ARLTS1* among normal samples and tumor samples.

Functional annotation analysis

Gene sets were explored for enrichment of functional annotation categories such as gene ontology (GO), using tools within the EASE program (http://david.abcc.ncifcrf.gov/) [22]. The top 100 genes that were positively correlating with *ARLTS1* among normal samples were checked for enriched Gene Ontology.

Results

The whole coding region of *ARLTS1* was previously screened in a total of 164 familial PCa patients and in a total of 377 unselected PCa patients, as well as in 381 control samples [7]. Here the Cys148Arg (T442C) variant was further genotyped in 48 familial PCa patients, 1471 unselected PCa patients, 375 BPH patients and 379 controls. By utilizing our previous data and genotyping more familial and unselected PCa cases, we validated a statistically significant association between Cys148Arg (T442C) and PCa risk (Table 1). When considering the familial PCa cases, and the association with risk allele C, the homozygous form CC reached a *P* value 0.019 (OR 1.67; 95% CI 1.08–2.56). Within unselected cases, the risk was also significant (OR 1.52, 95% CI 1.18–1.97, $P = 0.001$). When combining the datasets of familial and unselected cases (2060 samples total), the Cys148Arg (T442C) variant was found to be associated significantly with PCa risk (OR 1.54, 95% CI 1.20–1.98, $P = 0.0007$). Significant association was only found when comparing the frequencies of the homozygous form of the risk allele C to the frequencies of the homozygous form of the wild-type allele T. No association was found between the frequency of CC and BPH (Table 1).

When studying the segregation of Cys148Arg (T442C) CC genotype in 86 families where the index was detected as a carrier, complete segregation of the variant CC with cancer phenotype was seen in one family (family 427 in Fig. 1). In the rest of the families segregation was incomplete, yet indicating a clear association with *ARLTS1* genotype CC or TC and cancer phenotype (families 402 and 408 in Fig. 1).

Direct sequencing was used to examine the frequencies of *ARLTS1* variants in clinical prostate tumors. In the clinical prostate carcinomas, we found variants at four sites, Gly65Val (G194T), Pro131Leu (C392T), Cys148Arg (T442C) and Trp149Stop (G446A) (Table 2). At the Cys148Arg (T442C) site, the frequency of the cancer associated risk genotype CC was relatively high (28.3%) compared to the combined frequency of familial and unselected cases 21.2% (Table 1). However, the highest frequency of Cys148Arg (T442C) CC genotype was detected in xenograft samples (42.1%). This is intriguing since LuCaP xenograft samples are considered as a very homogenous representation of PCa.

To investigate the role of Cys148Arg (T442C) genotype in *ARLTS1* expression we selected 24 familial patients, including eight from each genotype group TT, CT and CC. Expression analyses by Q-RT-PCR were carried out by using the RNA extracted from lymphoblastoid cell lines of the patients. *ARLTS1* mRNA was expressed differentially between the three genotype groups (Fig. 2). The expression was significantly reduced among patients carrying the CC genotype ($P = 0.02$).

ARLTS1 expression levels were also analyzed in fifteen LuCaP xenograft samples, six PCa cell lines, two prostate epithelial cell lines and in RNA from normal prostate sample. Among the analyzed xenograft samples, a significant reduction in *ARLTS1* expression was observed in two of the samples (13%) when compared to its expression in normal prostate sample (Fig. 3A). Total loss of expression was seen in 11 (73%) of the xenograft samples (Fig. 3A). Two of the xenograft samples, LuCaP23.1 and LuCaP49, had higher *ARLTS1* expression levels compared to the normal tissue. Six of the fourteen (43%) *ARLTS1* down-regulated

Table 1. Association of the *ARLTS1* Cys148Arg variant with prostate cancer and benign prostatic hyperplasia.

	Cases	Controls			
	n (%)	n (%)	OR*	95% CI	*P*
Prostate cancer patients, all	2060	760			
TT	615 (29.9)	249 (32.8)			
TC	1008 (48.9)	396 (52.1)	1.03	0.85–1.24	0.753
CC	437 (21.2)	115 (15.1)	1.54	1.20–1.98	0.0007
TC+CC	1445 (70.1)	511 (67.2)	1.15	0.96–1.37	0.137
Familial prostate cancer patients	212	760			
TT	65 (30.7)	249 (32.8)			
TC	97 (45.7)	396 (52.1)	0.94	0.66–1.33	0.723
CC	50 (23.6)	115 (15.1)	1.67	1.08–2.56	0.019
TC+CC	147 (69.3)	511 (67.2)	1.1	0.79–1.53	0.563
Unselected prostate cancer patients	1848	760			
TT	550 (29.8)	249 (32.8)			
TC	911 (49.3)	396 (52.1)	1.04	0.86–1.26	0.676
CC	387 (20.9)	115 (15.1)	1.52	1.18–1.97	0.001
TC+CC	1298 (70.2)	511 (67.2)	1.15	0.96–1.38	0.131
Benign prostatic hyperplasia	375	760			
TT	110 (29.3)	249 (32.8)			
TC	203 (54.1)	396 (52.1)	1.16	0.88–1.54	0.299
CC	62 (16.5)	115 (15.1)	1.22	0.83–1.79	0.306
TC+CC	265 (70.7)	511 (67.2)	1.17	0.90–1.54	0.243

**compared to TT homozygotes.*

doi:10.1371/journal.pone.0026595.t001

xenograft samples had a genotype CC for the Cys148Arg (T442C) variant. When array comparative genomic hybridization (aCGH) was performed using these samples, 12/15 xenograft samples showed a loss on the chromosomal region of 13q14.3 (Fig. 3A).

In PCa cell lines, 5/6 (83%) showed a decreased or lost expression of *ARLTS1* when compared to normal prostate RNA (Fig. 3A). An *ARLTS1* locus deletion at 13q14.3 was found in PC-3, VCaP, LNCaP and DU-145. Among these, the latter also had the CC genotype for Cys148Arg (T442C). Interestingly, the Cys148Arg (T442C) variant was the only *ARLTS1* variant found in any of the PCa cell lines. Prostate epithelial cell line EP156T showed loss of expression. These results likely indicate that *ARLTS1* expression is low in the PCa epithelial cells.

ARLTS1 expression was also assessed with the same approach in a panel of fourteen BPH tissue samples, fourteen primary and six hormone-refractory tumor samples (Fig. 3B). *ARLTS1* expression was signinificantly lower ($P = 0.01$) in clinical tumors compared to BPH samples, which further supports its role as a tumor suppressor protein (Fig. 3B).

In clinical prostate tumors, the existence of LOH was studied. Germline DNA from patients carrying the Cys148Arg (T442C) variant was analyzed by direct sequencing. Sequencing results showed that in two cases (samples 261 and 530) the T allele in blood had been replaced with a C allele in cancer tissue (Fig. 3C) which suggests a role for *ARLTS1* in tumor suppression i.e., inactivation of both alleles in carcinogenesis. When clinical parameters (WHO grade, Gleason and T-scores) were examined, one of the two variant-carrying "LOH-patients" had a Gleason score of 9 and WHO grade of 3, indicating an aggressive form of the disease. The other case had a WHO grade of 3 but no Gleason score was available.

ARLTS1 expression was further determined by Western blotting (Fig. 4). Cell lysates were available from six prostate cancer cell lines, (22Rv1, LAPC-4, PC-3, VCaP, LnCaP, DU-145) and from the normal prostate epithelial cell line EP156T. ARLTS1 expression was highly convergent to ARLTS1 mRNA status showed in Fig. 3A. As expected, cell lines LnCaP, DU-145 and EP156T showed negative expression. In prostate cancer cell lines 22Rv1 and PC-3 ARLTS1 expression was corresponding to relative mRNA expression values, whereas in PCA cell line LAPC4 protein expression status was lower than the mRNA expression level. PCA cell line VCaP showed the highest ARLTS1 expression which was not corresponding to negative relative expression status seen in Q-RT-PCR.

When clinical characteristics (age at diagnosis, PSA value at diagnosis, T, N, and M-stage, WHO grade and Gleason score) were included in the analyses of germline genotype data, the CC genotype was significantly associated with a younger age at diagnosis and PSA value ($P = 0.05$ and $P = 0.03$, respectively) among the unselected material. This phenomenon was also observed in our previous study where the CC genotype of Cys148Arg (T442C) was associated with high Gleason scores ($P = 0.01$) and high PSAs (≥ 7) at diagnosis ($P = 0.05$) [7]. When the correlation between aggressive disease status and the occurrence of the CC genotype was examined, we found a statistically significant association ($P = 0.02$) among the unselected cases (Table 3). Within familial PCa cases, no association was found between the CC genotype at the Cys148Arg (T442C) locus and any of the clinical variables or with aggressive PCa.

ARLTS1 expression in the different tissues of the body is shown in Figure S1. The figure illustrates that *ARLTS1* is highly expressed

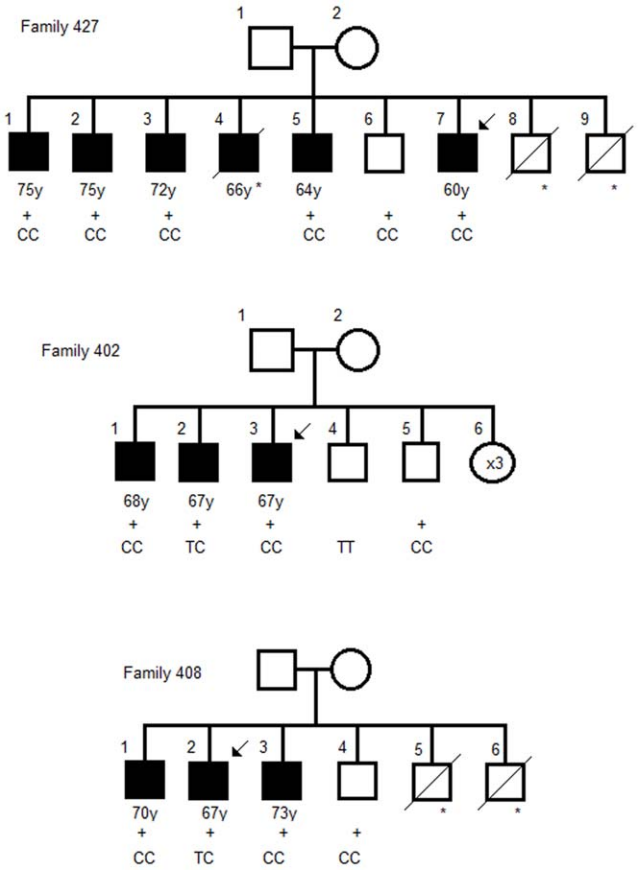


Figure 1. Examples of segregation analysis in three *ARLTS1* Cys148Arg (T442C) variant-positive families with prostate cancer. Squares represent males; circles represent females. Open symbols indicate no neoplasm, and filled symbols denote prostate cancer cases. +indicates the presence of the T442C variant in the DNA sample of the family members, followed by the actual genotype. An arrow indicates the individual initially screened for *ARLTS1* sequence variants. Age at diagnosis for prostate cancer patients (in years) is indicated below the symbol for each family member. An asterisk (*) denotes the persons with no sample available.
doi:10.1371/journal.pone.0026595.g001

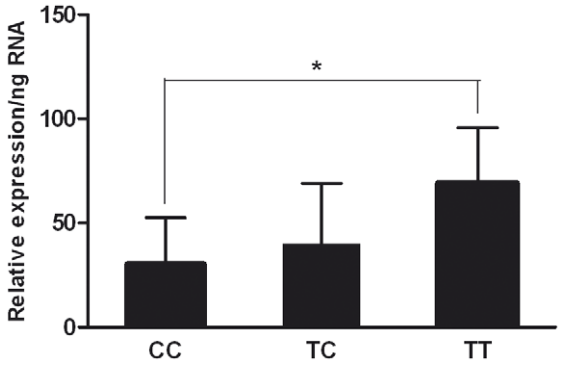


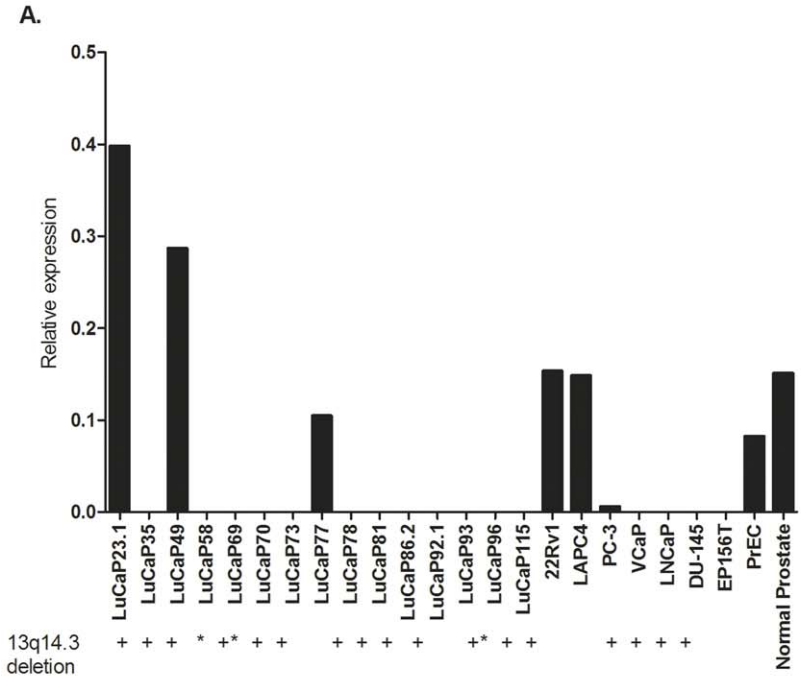
Figure 2. The relative expression of *ARLTS1* in Cys148Arg (T442C) genotyped lymphoblastoid cell lines. Determined by Q-RT-PCR. *, $P<0.05$. Columns, mean of eight individuals; bars, SD.
doi:10.1371/journal.pone.0026595.g002

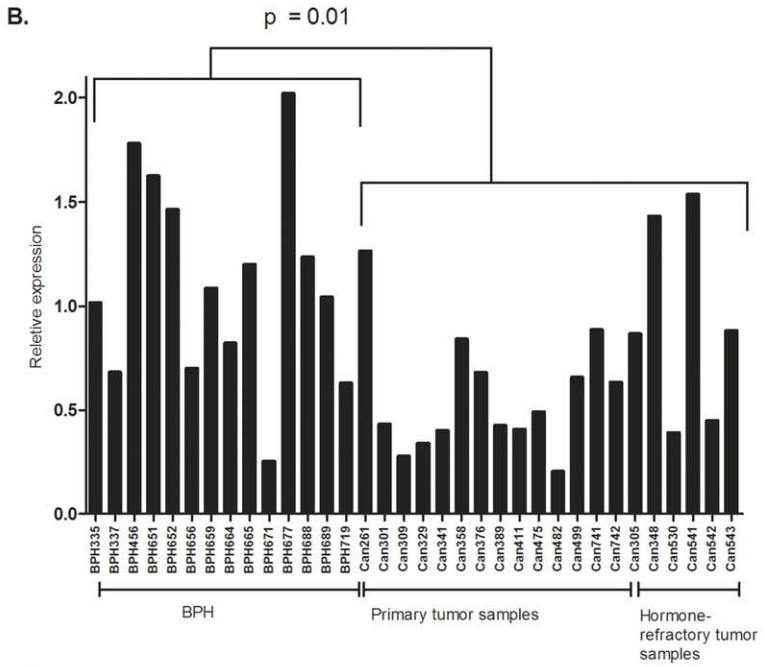
in the hematological and lymphatic system. The results from co-expression analysis illustrate a strong association with immune system processes, revealing an enrichment score for the cluster of 6.67 and a p-value of 2.1e-7. These results were generated from the top 100 genes with a correlation value greater than 0.3 among normal samples. Among cancer samples, the category of immune system processes had the highest enrichment score 14.89 with a p-value of 5.3e-21 and adjusted Benjamin score of 2.8e-17. We did not identify any strong gene ontology associated with the top 100 genes that were negatively correlated with *ARLTS1* among cancer samples. We confirmed our findings using a separate dataset from expo (IGC: Expression Project for Oncology. 2008 [http://www.intgen.org/expo.cfm]). The results illustrate that 56% of the genes were commonly identified to positively correlate with *ARLTS1* in either dataset comprised of cancer samples. The genes with a positive correlation above 0.3 identified in the intersection yielded a very strong gene ontology enrichment score of 18.12 for immune system process with a p-value of 2.1e-24 as shown in Figure S2. Gene expression among PCa specimens is shown to be very low for *ARLTS1*. We could not reliably establish any significant link to PCa for *ARLTS1* using gene expression data due to low expression levels.

Table 2. Observed *ARLTS1* variants and their frequencies in clinical prostate tumors and LuCap xenografts.

Nucleotide change	Amino acid change	Genotype	Clinical prostate tumors n (out of 53, %)	Xenografts n (out of 19, %)
G194T	Gly65Val	GG	52 (98.1)	19 (100)
		GT	1 (1.9)	0 (0)
		TT	0 (0)	0 (0)
C392T	Pro131Leu	CC	43 (81.1)	19 (100)
		CT	10 (18.9)	0 (0)
		TT	0 (0)	0 (0)
T442C	Cys148Arg	TT	23 (43.4)	7 (36.8)
		TC	15 (28.3)	4 (21.1)
		CC	15 (28.3)	8 (42.1)
G446A	Trp149Stop	GG	52 (98.1)	18 (94.7)
		GA	1 (1.9)	1 (5.3)
		AA	0 (0)	0 (0)

doi:10.1371/journal.pone.0026595.t002


A.
Relative expression
0.5
0.4
0.3
0.2
0.1
0.0
LuCaP23.1
LuCaP35
LuCaP49
LuCaP58
LuCaP69
LuCaP70
LuCaP73
LuCaP77
LuCaP78
LuCaP81
LuCaP86.2
LuCaP92.1
LuCaP93
LuCaP96
LuCaP115
22Rv1
LAPC4
PC-3
VCaP
LNCaP
DU-145
EP156T
PrEC
Normal Prostate
13q14.3 deletion
+ + + * +* + + + + + + + +* + + + + + + +


B.
p = 0.01
Reletive expression
2.0
1.5
1.0
0.5
0.0
BPH335
BPH337
BPH456
BPH651
BPH652
BPH656
BPH659
BPH664
BPH665
BPH671
BPH677
BPH688
BPH689
BPH719
Can261
Can301
Can309
Can329
Can341
Can358
Can376
Can389
Can411
Can475
Can482
Can499
Can741
Can742
Can305
Can348
Can530
Can541
Can542
Can543
BPH
Primary tumor samples
Hormone-refractory tumor samples

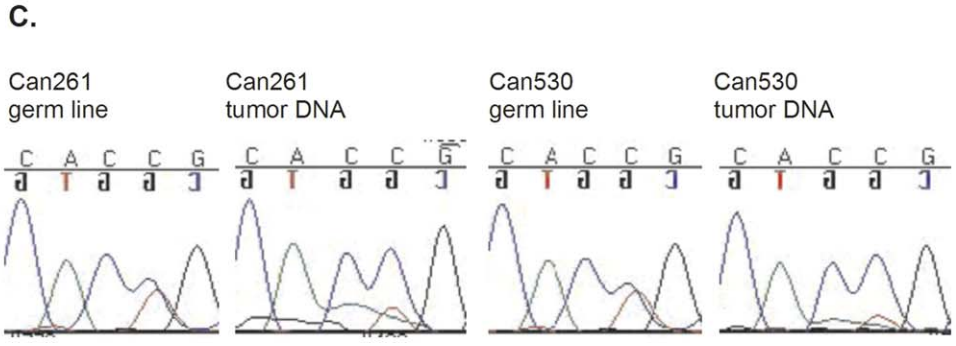

C.
Can261 germ line
Can261 tumor DNA
Can530 germ line
Can530 tumor DNA
C A C C G
C A C C G
C A C C G
C A C C G

Figure 3. Relative expression of *ARLTS1* in A, LuCaP xenografts, prostate cancer cell lines, prostate epithelial cells and in normal prostate and B, in BPH and clinical tumor samples, as determined by Q-RT-PCR. The Cys148Arg (T442C) genotype as well as the choromosomal aberration is shown. **C**, Cys148Arg (T442C) variant in two clinical cancer tissue samples and in two normal blood DNA (sequences are in reverse orientation). * determined as in Saramäki OR et al., 2006.
doi:10.1371/journal.pone.0026595.g003

Discussion

The *ARLTS1* gene and *ARLTS1* polymorphisms have been shown to have a role in the pathogenesis of many cancers. The nonsense polymorphism at the end of the coding region has been revealed to predispose to familial cancer [8]. Functional analyses of the truncated protein have indicated that the Trp149Stop variant might affect apoptosis and tumor suppression. Another *ARLTS1* variant, the missense polymorphism Cys148Arg (T442C), and especially the CC genotype, has been found to be significantly associated with high-risk familial breast cancer [4]. The same variant was also found to be associated with predisposition to melanoma [5]. The roles of *ARLTS1* variants have also been studied with colorectal cancer [23] and chronic lymphocytic leukemia (CLL) [24] but no associations have been found.

In the present study, we showed an association with the *ARLTS1* Cys148Arg (T442C) variant and elevated PCa risk, validating our previous results [7]. Previously, we sequenced 164 familial and 377 unselected PCa samples together with 809 controls, and Cys148Arg (T442C) showed significant association with PCa risk (OR 1.19; 95% CI 1.02–1.39, $P = 0.020$) but the significance did not hold when the data were subdivided into familial and unselected cases. However, when the C allele was assumed to be recessive, all data (familial and unselected) showed significant association between the CC genotype and PCa risk ($P = 0.005$). Here were genotyped more familial and unselected PCa cases and Cys148Arg (T442C) revealed a statistically significant association with PCa risk among both familial (OR = 1.67, 95% CI = 1.08–2.56, $P = 0.019$) and unselected PCa patients (OR = 1.52, 95% CI = 1.18–1.97, $P = 0.001$). In the combined data set of both familial and unselected cases, the overall risk was even more increased (OR = 1.54, 95% CI = 1.20–1.98, $P = 0.0007$). The fact that previous analysis did not hold when divided to subclasses could therefore be explained by the size of the study. The current study has a power of approximately 100% to detect association compared to 94% of the previous study. Further follow-up with larger sample size would likely make the association even stronger. We found no association between this variant and BPH, suggesting that role of the Cys148Arg (T442C) is minor and the cancer predisposing effect emerges only in more advanced cases, especially those with an aggressive PCa status. Alternatively, BPH and PCa are independent events or at least *ARLTS1* is not consequential for BPH transformation to PCa. The longer follow-up of the BPH patients is needed to assess the question whether the patients carrying CC genotype (16%) are the ones developing cancer later on.

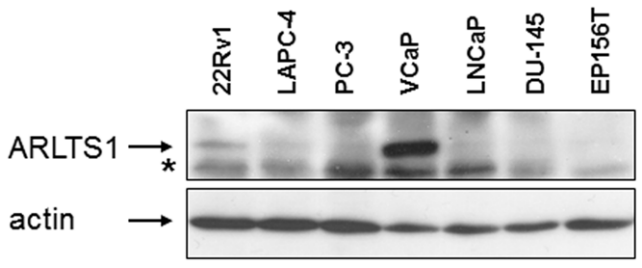


Figure 4. The levels of ARLTS1 protein by Western blotting. Protein expression status in six prostate cancer cell lines and in one normal prostate epithelial cell line. Actin is shown as a loading control. ARL11 expected/observed Mw 21.4 kDa. A non-specific band is marked with an asterisk (*).
doi:10.1371/journal.pone.0026595.g004

Only the homozygous form of the C allele of the Cys148Arg (T442C) variant resulted in an association with PCa risk when compared to the TT genotype. This supports the results from previous studies with breast cancer [4] that showed that C is the risk allele while the T allele has a protective or neutral effect on carcinogenesis. *ARLTS1* Cys148Arg (T442C) in particular has a predisposing effect on sporadic PCa. Further, the Cys148Arg was the only *ARLTS1* variant observed at a greater frequency among PCa tissue samples and xenografts, where the frequencies were 28.3% and over 40 percent, respectively. This also indicates a likely role for *ARLTS1* in prostate carcinogenesis. Interestingly, a total absence of Trp149Stop (G446C) was observed both in clinical tumor samples and in xenografts indicating that this variant has not an important role in PCa development.

The control population in our study was not in Hardy-Weinberg equilibrium (HWE), suggesting that Cys148Arg might be a novel variant. In our data, an excess of heterozygotes was observed which may indicate the presence of over dominant selection or the occurance of outbreeding. The phenomenon of heterozygous advantage allows for natural selection to maintain the polymorphism. The same control population has been used many times in different analyses, and never before has it been discordant from HWE [25]. Interestingly, in the Sub-Saharan African population, the CC genotype does not exist, and in the Asian population it is only present in a small fraction (http://www.ncbi.nlm.nih.gov). Only within the European population does CC exist at a greater proportion. Considering the possible role of *ARLTS1* in the immune system, it may be that the CC genotype has provided a protective effect among Europeans. HW disequilibrium among the controls is not either result of genotyping error because direct sequencing was previously used to genotype 164 of the familial PCa samples and the genotypes matched with the TaqMan assay results. Furthermore, the same assay was used to genotype all the cancer samples and controls.

Quantitative reverse-transcriptase-polymerase-chain-reaction assay (Q-RT-PCR) showed lowered *ARLTS1* expression in most of the clinical PCa samples, compared to BPH samples. This is consistent with the previous findings; *ARLTS1* expression has also been detected to be very low or absent in lung carcinomas and CLL cells [8,26], as well as in ovarian primary tumors and cell lines [27], when compared with the levels of their normal counterparts. Here, using Q-RT-PCR, we used commercially available normal prostate RNA from a 72 year-old male. With respect to PCa disease progression, we could assume that at the age of 72 there might be some changes in the prostatic tissue, and use of control RNA from a younger male might lead to even better results concerning comparisons of *ARLTS1* expression status in normal and diseased samples.

Western analysis of prostate cancer cell lines for levels of ARLTS1 protein was in most part convergent with the ARLTS1 mRNA levels. However, in one sample with 13q14.3 deletion (VCaP), high protein expression was detected. This divergent result may originate from several reasons and needs further investigation. ARLTS1 tumor suppressor gene may have dupli-

Table 3. Association of the Cys148Arg (T442C) variant with clinicopathologic variables of prostate cancer patients.

	Familial cases	Unselected cases
No. of subjects	212	1848
Aggressive disease		
No. of subjects	121	975
TT	40	281
TC	57	471
CC	24	223
OR (one-sided 95% CI), *P* value [CC] *vs.* [TT+TC]	0.62 (0.35 - infinity), 0.95	1.28 (1.05 - infinity), 0.02

doi:10.1371/journal.pone.0026595.t003

cated or activated pseudogenes, or there might also be pseudogenes or polymorphisms within the regulatory miRNAs. However, these protein expression results further confirm that ARLTS1 expression is low in the PCa epithelial cells.

The suggested mechanisms behind the reduced or absent *ARLTS1* expression are variable, including promoter hypermethylation or LOH. It has been confirmed that in lung cancer cell lines and in ovarian carcinomas *ARLTS1* is down-regulated due to DNA methylation in its promoter region [26,27]. *ARLTS1* restoration by adenoviral transduction induced apoptosis. Finally, re-expression of *ARLTS1* suppressed ovarian and lung cancer tumorigenity in nude mice. We were able to show that LOH exists also in PCa samples, meaning that heterozygous deletions may affect *ARLTS1* expression levels. *ARLTS1* resides at 13q14.3, a region that has been reported to be deleted in a variety of hematopoietic and solid tumors [8], including PCa. Future investigations include whether reduced *ARLTS1* expression in PCa may be due to hypermethylation.

Chronic or recurrent inflammation has been implicated in the initiation and development of several human cancers, including those of the stomach, liver, colon, and urinary bladder and a role for chronic inflammation in the etiology of PCa has been proposed [28]. The source for prostatic inflammation is unclear, but it has been suggested that it may be directly related to prostate infecting agents (such as sexually transmitted organisms and viruses), dietary factors and oxidative stress, urine reflux, chemical and physical trauma or a combination of these. Recently experimental evidence was provided by Schlaberg R et al that xenotropic murine leukemia virus-related virus (XMRV) was present in malignant prostatic cells and was significantly associated with more aggressive tumors [29]. Susceptibility genes related to PCa might function in host immune responses and protection against cell and DNA damage caused by oxidative agents. For example, two suggested susceptibility risk genes for hereditary PCa, *RNASEL* and *MSR1*, are both involved with innate immunity. Lymphoblastoid cell lines represent an essential component of the immune response, so this result may indicate that ARLTS1 has a function in immune system processes.

Alterations in important molecular pathways involved in PCa have been implicated in proliferative inflammatory atrophy (PIA) and prostatic intraepithelial neoplasia (PIN) lesions. Tumor suppressor genes NKX3.1 [30], CDKN1B which encodes p27 [28,31] and the phosphatase and tensin homologue (PTEN) [32] are highly expressed in normal prostate epithelium and have shown to be down-regulated or absent in PIN and PCa. Since ARLTS1 acts as a tumor suppressor protein and has a decreased expression in PCa, these recent findings are in line with the results from other suggested PCa tumor suppressor proteins.

Previously it has been shown that *ARLTS1* induces apoptosis in lung cancer cells [8,26] and in ovarian carcinoma [27]. In the case of prostatic inflammation and antigen engagement, the need for apoptosis increases and the processes of the immune system, together with endogenous inflammatory cells (such as T and B lymphocytes) become activated. Here we showed that in the lymphoblastoid cell lines of PCa patients, *ARLTS1* expression was significantly decreased among the CC carrying patients compared to the wild-type allele T carrying patients. The lymphoblastoid cell lines were derived by Epstein-Barr virus transformation of peripheral mononuclear leucocytes from patients, indicating that these cells have encountered antigen stimuli via viral infection. The CC carrying patients had a low *ARLTS1* expression status suggesting that *ARLTS1* function is decreased due to this risk genotype and consequently this leads to decreased apoptosis. This may indicate that the Cys148Arg (T442C) CC genotype contributes to the immune response by diminishing apoptosis rates, decreasing defense mechanisms and finally cancer progression. It has been shown earlier that lung cancer cells carrying the Trp149Stop (G446A) variant and expressing the truncated protein had a reduced capability to induce apoptosis compared to cells expressing the full-length protein [8]. This reduced apoptosis could also be the causal factor in PCa. Our data suggest for the first time that the predisposing effect of the CC genotype of Cys148Arg (T442C) is related to reduced expression in immune system cells (lymphoblasts) rather than in tumor cells where *ARLTS1* expression is naturally very low.

Like all other organs, the normal prostate contains endogenous inflammatory cells and immune response mechanisms. Consequently, the function of *ARLTS1* may not exclusively be based on tumor suppression as suggested before, but also on immune response functions that occur either locally in prostate or in peripheral tissues. This is in line with the findings that PCa is a very heterogeneous disease and different mechanisms of cancer progression may occur.

Taken together, association and expression analyses of the Cys148Arg (T442C) CC genotype and PCa risk suggest that *ARLTS1* Cys148Arg (T442C) variant has a role in PCa predisposition and *ARLTS1* functions via immune system processes.

Supporting Information

Figure S1 Expression of *ARLTS1* in the publicly available GeneSapiens database. The expression of *ARLTS1* is very low in both normal and cancer samples. In the prostate the expression of *ARLTS1* remains very low and no significant associations could be made using gene expression data.
(TIF)

Figure S2 A print from the EASE association analysis.
(TIF)

Table S1 Clinicopatholgic variables of prostate cancer patients.
(DOC)

Acknowledgments

We wish to thank Linda Enroth, Riina Liikala and Riitta Vaalavuo for technical assistance. Cancer patients and their family members are thanked for participating in this study.

References

1. Kahn RA, Der CJ, Bokoch GM (1992) The ras superfamily of GTP-binding proteins: Guidelines on nomenclature. FASEB J 6(8): 2512–2513.
2. Kahn RA, Volpicelli-Daley L, Bowzard B, Shrivastava-Ranjan P, Li Y, et al. (2005) Arf family GTPases: Roles in membrane traffic and microtubule dynamics. Biochem Soc Trans 33(Pt 6): 1269–1272.
3. De Brakeleer S, Teugels E, De Greve J (2005) Familial cancer and ARLTS1. N Engl J Med 353(3): 313–4. author reply 313–4.
4. Frank B, Hemminki K, Meindl A, Wappenschmidt B, Klaes R, et al. (2006) Association of the ARLTS1 Cys148Arg variant with familial breast cancer risk. Int J Cancer 118(10): 2505–2508.
5. Frank B, Meyer P, Boettger MB, Hemminki K, Stapelmann H, et al. (2006) ARLTS1 variants and melanoma risk. Int J Cancer 119(7): 1736–1737.
6. Castellvi-Bel S, Castells A, de Cid R, Munoz J, Balaguer F, et al. (2007) ASSOCIATION OF THE ARLTS1 CYS148ARG VARIANT WITH SPORADIC AND FAMILIAL COLORECTAL CANCER. Carcinogenesis.
7. Siltanen S, Syrjakoski K, Fagerholm R, Ikonen T, Lipman P, et al. (2008) ARLTS1 germline variants and the risk for breast, prostate, and colorectal cancer. Eur J Hum Genet 16(8): 983–991.
8. Calin GA, Trapasso F, Shimizu M, Dumitru CD, Yendamuri S, et al. (2005) Familial cancer associated with a polymorphism in ARLTS1. N Engl J Med 352(16): 1667–1676.
9. Latil A, Chene L, Mangin P, Fournier G, Berthon P, et al. (2003) Extensive analysis of the 13q14 region in human prostate tumors: DNA analysis and quantitative expression of genes lying in the interval of deletion. Prostate 57(1): 39–50.
10. Valeri A, Fromont G, Sakr W, Azzouzi R, Dey J, et al. (2005) High frequency of allelic losses in high-grade prostate cancer is associated with biochemical progression after radical prostatectomy. Urol Oncol 23(2): 87–92.
11. Brookman-Amissah N, Nariculam J, Freeman A, Willamson M, Kirby RS, et al. (2007) Allelic imbalance at 13q14.2 approximately q14.3 in localized prostate cancer is associated with early biochemical relapse. Cancer Genet Cytogenet 179(2): 118–126.
12. Ishkanian AS, Mallof CA, Ho J, Meng A, Albert M, et al. (2009) High-resolution array CGH identifies novel regions of genomic alteration in intermediate-risk prostate cancer. Prostate 69(10): 1091–1100.
13. Nakano M, Takahashi H, Shiraishi T, Lu T, Furusato M, et al. (2008) Prediction of clinically insignificant prostate cancer by detection of allelic imbalance at 6q, 8p and 13q. Pathol Int 58(7): 415–420.
14. Schleutker J, Matikainen M, Smith J, Koivisto P, Baffoe-Bonnie A, et al. (2000) A genetic epidemiological study of hereditary prostate cancer (HPC) in finland: Frequent HPCX linkage in families with late-onset disease. Clin Cancer Res 6(12): 4810–4815.
15. Kogan I, Goldfinger N, Milyavsky M, Cohen M, Shats I, et al. (2006) hTERT-immortalized prostate epithelial and stromal-derived cells: An authentic in vitro model for differentiation and carcinogenesis. Cancer Res 66(7): 3531–3540.
16. Laitinen S, Karhu R, Sawyers CL, Vessella RL, Visakorpi T (2002) Chromosomal aberrations in prostate cancer xenografts detected by comparative genomic hybridization. Genes Chromosomes Cancer 35(1): 66–73.
17. Saramaki OR, Tammela TL, Martikainen PM, Vessella RL, Visakorpi T (2006) The gene for polycomb group protein enhancer of zeste homolog 2 (EZH2) is amplified in late-stage prostate cancer. Genes Chromosomes Cancer 45(7): 639–645.
18. Linja MJ, Savinainen KJ, Saramaki OR, Tammela TL, Vessella RL, et al. (2001) Amplification and overexpression of androgen receptor gene in hormone-refractory prostate cancer. Cancer Res 61(9): 3550–3555.
19. Saramaki OR, Porkka KP, Vessella RL, Visakorpi T (2006) Genetic aberrations in prostate cancer by microarray analysis. Int J Cancer 119(6): 1322–1329.
20. Bieche I, Laurendeau I, Tozlu S, Olivi M, Vidaud D, et al. (1999) Quantitation of MYC gene expression in sporadic breast tumors with a real-time reverse transcription-PCR assay. Cancer Res 59(12): 2759–2765.
21. Kilpinen S, Autio R, Ojala K, Iljin K, Bucher E, et al. (2008) Systematic bioinformatic analysis of expression levels of 17,330 human genes across 9,783 samples from 175 types of healthy and pathological tissues. Genome Biol 9(9): R139.
22. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1): 44–57.
23. Frank B, Hemminki K, Brenner H, Hoffmeister M, Chang-Claude J, et al. (2006) ARLTS1 variants and risk of colorectal cancer. Cancer Lett 244(2): 172–175.
24. Sellick GS, Catovsky D, Houlston RS (2006) Relationship between ARLTS1 polymorphisms and risk of chronic lymphocytic leukemia. Leuk Res 30(12): 1573–1576.
25. Kote-Jarai Z, Easton DF, Stanford JL, Ostrander EA, Schleutker J, et al. (2008) Multiple novel prostate cancer predisposition loci confirmed by an international study: The PRACTICAL consortium. Cancer Epidemiol Biomarkers Prev 17(8): 2052–2061.
26. Yendamuri S, Trapasso F, Ferracin M, Cesari R, Sevignani C, et al. (2007) Tumor suppressor functions of ARLTS1 in lung cancers. Cancer Res 67(16): 7738–7745.
27. Petrocca F, Iliopoulos D, Qin HR, Nicoloso MS, Yendamuri S, et al. (2006) Alterations of the tumor suppressor gene ARLTS1 in ovarian cancer. Cancer Res 66(21): 10287–10291.
28. De Marzo AM, Marchi VL, Epstein JI, Nelson WG (1999) Proliferative inflammatory atrophy of the prostate: Implications for prostatic carcinogenesis. Am J Pathol 155(6): 1985–1992.
29. Schlaberg R, Choe DJ, Brown KR, Thaker HM, Singh IR (2009) XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. Proc Natl Acad Sci U S A 106(38): 16351–16356.
30. Bethel CR, Faith D, Li X, Guan B, Hicks JL, et al. (2006) Decreased NKX3.1 protein expression in focal prostatic atrophy, prostatic intraepithelial neoplasia, and adenocarcinoma: Association with gleason score and chromosome 8p deletion. Cancer Res 66(22): 10683–10690.
31. van Leenders GJ, Gage WR, Hicks JL, van Balken B, Aalders TW, et al. (2003) Intermediate cells in human prostate epithelium are enriched in proliferative inflammatory atrophy. Am J Pathol 162(5): 1529–1537.
32. De Marzo AM, Platz EA, Sutcliffe S, Xu J, Gronberg H, et al. (2007) Inflammation in prostate carcinogenesis. Nat Rev Cancer 7(4): 256–269.

Author Contributions

Conceived and designed the experiments: SS TW JPM LL TV JS. Performed the experiments: SS TW ORS JPM LL. Analyzed the data: SS TW MS LL ORS JPM OK. Contributed reagents/materials/analysis tools: JS RLV TLJT TV OK. Wrote the paper: SS TW TV JS.

OPEN ACCESS Freely available online

PLOS ONE

ARLTS1 and Prostate Cancer Risk - Analysis of Expression and Regulation

Sanna Siltanen[1], Daniel Fischer[2], Tommi Rantapero[1], Virpi Laitinen[1], John Patrick Mpindi[3], Olli Kallioniemi[3], Tiina Wahlfors[1], Johanna Schleutker[1,4]*

1 Institute of Biomedical Technology/BioMediTech, University of Tampere and Fimlab Laboratories, Tampere, Finland, **2** School of Health Sciences, University of Tampere, Tampere, Finland, **3** Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland, **4** Department of Medical Biochemistry and Genetics, Institute of Biomedicine, University of Turku, Turku, Finland

Abstract

Prostate cancer (PCa) is a heterogeneous trait for which several susceptibility loci have been implicated by genome-wide linkage and association studies. The genomic region 13q14 is frequently deleted in tumour tissues of both sporadic and familial PCa patients and is consequently recognised as a possible locus of tumour suppressor gene(s). Deletions of this region have been found in many other cancers. Recently, we showed that homozygous carriers for the T442C variant of the *ARLTS1* gene (ADP-ribosylation factor-like tumour suppressor protein 1 or *ARL11*, located at 13q14) are associated with an increased risk for both unselected and familial PCa. Furthermore, the variant T442C was observed in greater frequency among malignant tissue samples, PCa cell lines and xenografts, supporting its role in PCa tumourigenesis. In this study, 84 PCa cases and 15 controls were analysed for *ARLTS1* expression status in blood-derived RNA. A statistically significant ($p = 0.0037$) decrease of *ARLTS1* expression in PCa cases was detected. Regulation of *ARLTS1* expression was analysed with eQTL (expression quantitative trait loci) methods. Altogether fourteen significant *cis*-eQTLs affecting the *ARLTS1* expression level were found. In addition, epistatic interactions of *ARLTS1* genomic variants with genes involved in immune system processes were predicted with the MDR program. In conclusion, this study further supports the role of *ARLTS1* as a tumour suppressor gene and reveals that the expression is regulated through variants localised in regulatory regions.

Citation: Siltanen S, Fischer D, Rantapero T, Laitinen V, Mpindi JP, et al. (2013) *ARLTS1* and Prostate Cancer Risk - Analysis of Expression and Regulation. PLoS ONE 8(8): e72040. doi:10.1371/journal.pone.0072040

Editor: Amanda Ewart Toland, Ohio State University Medical Center, United States of America

Received April 4, 2013; **Accepted** July 3, 2013; **Published** August 5, 2013

Funding: This work was supported by the Tampere Graduate Program in Biomedicine and Biotechnology salary and Orion-Farmos Research Foundation grant to S.S. and by the Sigrid Juselius Foundation, the Academy of Finland (251074), the Finnish Cancer Organisations, and the Competitive Research Funding of the Pirkanmaa Hospital District (9N069) grants to J.S. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Johanna.Schleutker@utu.fi

Introduction

Prostate cancer (PCa) is a heterogeneous trait, and it is the most common malignancy among men in western countries, including Finland. It is a multifactorial disease, and definitive risk factors include age, ethnic origin and family history. In Finland, the incidence of PCa is 89.4/100,000, and in 2010, 4697 new prostate cancer cases were diagnosed (http://www.cancer.fi/syoparekisteri/en/). Despite extensive research over the last decade, the etiological risk factors and genes that cause genetic susceptibility remain largely unknown. This lack of knowledge has hampered effective cancer prevention and development of better treatments.

Mutations in the known high-penetrance PCa predisposition genes explain only a small fraction of PCa cases. A polygenic model for familial aggregation of cancer has been proposed where several low-penetrance alleles may have a multiplicative and/or modifying effect. One low-penetrant candidate gene is *ARLTS1* (*ARL11*), ADP-ribosylation factor like tumour suppressor protein 1, a putative tumour-suppressor gene on chromosome 13q14, which has been shown to function in many human cancers [1–5]. *ARLTS1* is a member of the ADP-ribosylation factor family that plays a role in apoptotic signalling. The same chromosomal area on 13q has been indicated in a multi-centre genome-wide linkage study in families with at least five affected members [6]. In another recent study, the immediate adjacent region 13q13 showed a suggestive linkage to PCa, with a HLOD>1.9 [7]. In addition, the locus 13q14 is among the most frequently deleted chromosomal regions in somatic tumour tissues in both unselected and hereditary prostate cancers [8,9], suggesting that *ARLTS1* could be a target for both germline and somatic mutations. We previously reported a significant association of *ARLTS1* T442C (rs3803185) homozygote carriers with PCa [10]. This risk genotype was also associated with decreased *ARLTS1* expression in the lymphoblastoid cell line samples of PCa patients, and *ARLTS1* co-expression signatures from data mining revealed that *ARLTS1* expression was strongly associated with immune system processes. These processes are of interest because a link between chronic inflammation and PCa progression has been repeatedly proposed, and multiple genes acting in inflammatory pathways have been linked to PCa susceptibility [11–13].

Complex diseases, such as cancer, are caused by a combination of multiple genetic interactions and environmental factors, which are hard to detect and link to each other. To determine true causal associations using statistical methods and phenotype information, it is advisable to organise individual markers into groups according

to meaningful biological criteria, such as inflammation. By composing variant sets, it is possible to reduce the number of hypotheses being tested, which allows the association between a genomic feature and a phenotype to be more easily detected.

Epistasis in genotype level is defined as the interaction among multiple genes or loci, and this joint genetic effect may be the factor behind "missing heritability", a phenomenon linked to the unexplained portion of hereditary cancer susceptibility, which is observed in PCa. The genome-wide expression quantitative trait loci, eQTL, analysis is a method for studying epistasis in complex traits that is able to detect associations between genotypic and expression data. The eQTL analysis is a widely used approach to gain insight into the role of single nucleotide polymorphisms (SNPs) affecting transcript levels.

In this study, we investigated the mechanisms behind the previously observed association of *ARLTS1* and PCa by focusing on finding gene/expression interactions that dispose patients to PCa, including interactions involving the *ARLTS1* gene. To further investigate the *ARLTS1* expression differences seen previously in tumor samples, prostate cancer and lymphoblastoid cell lines [10], we performed functional eQTL analysis from whole blood derived total RNA from PCa patients. The previously reported *ARLTS1* co-expression with immune system processes [10] was tested by MDR analysis. To our knowledge, this is the first study reporting the findings of *ARLTS1* interacting variants and prostate cancer eQTLs at the 13q14 region.

Materials and Methods

Study population

All the samples were of Finnish origin. The identification and collection of the Finnish HPC families has been described elsewhere [14]. The familial samples analysed in this study had at least two affected first or second degree relatives. Altogether, 102 prostate cancer cases and 33 healthy male family members belonging to 31 families were initially taken into the study population. The clinical characteristics of the familial patients used in RNA sequencing (n = 84) are referred to in Table 1. Average age at diagnosis was 63.0 y.

Patient information and samples were obtained with full written informed consent. The study was performed under appropriate research permissions from the Ethics Committees of the Tampere University Hospital, Finland, as well as the Ministry of Social Affairs and Health in Finland.

Genome-wide SNP Genotyping

Genome-wide SNP genotyping was performed using the HumanOmniExpress BeadChip microarray (Illumina, Inc., San Diego, CA, USA) by the Technology Centre, Institute for Molecular Medicine Finland (FIMM), University of Helsinki. This array covers more than 700,000 markers with an average spacing of 4 kb across the entire genome.

DNA Sequencing

The *ARLTS1* variant status of the PCa patients used in Illumina genotyping was examined by direct sequencing. Sequencing was performed in an Applied Biosystems 3130xl Genetic Analyzer (Life Technologies Corporation, Carlsbad, CA, USA) according to the manufacturer's instructions. Primers and PCR conditions used in the mutation screening are available upon request.

RNA Extraction and sequencing

Total RNA was extracted from 84 PCa cases and 15 healthy male relatives. All subjects belonged to the 31 Finnish HPC families mentioned above. Total RNA was purified from whole blood collected in PAXgene® Blood RNA Tubes (PreAnalytiX GmbH, Switzerland/Qiagen/BD) using the MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit (Ambion®/Life Technologies, Carlsbad, CA, USA) and the PAXgene Blood miRNA Kit (PreAnalytiX GmbH, Switzerland/Qiagen/BD). The RNA quality was assessed using the Agilent 2100 Bioanalyzer and the Agilent RNA 6000 Nano Kit (Agilent Technologies, Santa Clara, CA, USA).

Library preparation, target enrichment and massively parallel paired-end sequencing of expressed transcripts was performed by Beijing Genomics Institute (BGI Hong Kong Co., Ltd., Tai Po, Hong Kong) using Illumina HiSeq2000 technology (Illumina Inc., San Diego, CA, USA).

Table 1. Clinicopathologic findings at diagnosis of the PCa patients used in RNA sequencing (n = 84).

	n (%)
Age at diagnosis	
<65 years	48 (57.1)
≥65 years	35 (41.7)
Stage	
T stage	
T1 (clinically undetectable)	31 (36.9)
T2–T4 (clinically detectable)	48 (57.1)
M stage	
M0 (no evidence of metastasis)	53 (63.1)
M1 (bone metastasis)	0 (0)
MX (bone metastasis cannot be assessed)	27 (32.1)
PSA value	
<20 ng/ml	60 (71.4)
≥20 ng/ml	14 (16.7)
Grade	
Gleason score	
<7	44 (52.4)
7	10 (11.9)
>7	0 (0)

doi:10.1371/journal.pone.0072040.t001

eQTL analysis

To obtain *ARLTS1* expression values, first, the reads from RNA sequencing were aligned with tophat2 using hg19 as the reference genome [15]. The raw read count for *ARLTS1* was calculated using HTseq, and read counts were transformed to normalised expression values using DESeq (www-huber.embl.de/users/anders/HTSeq/, [16]). The linear regression model implemented in PLINK was used to detect transcript specific variants. Associations in *cis* were delineated by a 1 Mb window upstream or downstream of the *ARLTS1* SNP rs9526582, as most of the *cis*-eQTLs are located within or close to the gene of interest [17]. In addition, we applied a new method based on probabilistic indices to test for eQTL. The new method is a non-parametric directional test (similar to the well-known Jonckheere-Terpstra test), implemented in our R-Package GeneticTools that is available on the Comprehensive R Archive Network (http://cran.r-project.org/package = GeneticTools), and a package description is under

development (unpublished data). P-values were calculated using permutation tests and were adjusted for multiple testing by applying the Benjamini-Hochberg correction. We also calculated for each test size α in [0,0.1] the ratio of the amount of expected test rejections and the amount of observed rejections. The α for which the ratio of these two values was maximal, we chose also as an optimal test size.

Gene expression dataset

All microarray gene expression data on cell lines (n = 1445) included in these analyses are publicly available via the Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/, accession numbers; GSE36133, GSE7127, GSE8332, GSE10843, GSE10890, GSE12777, GSE15455, GSE18773, GSE20126, GSE21654 and GSE24795) and GSK Cancer Cell Line Genomic Profiling Data (https://cabig.nci.nih.gov/tools/caArray_GSKdata) (2008) [18] (GlaxoSmithKline). The bulk of the gene expression data was acquired from the Cancer Cell Line Encyclopedia (CCLE) (GSE36133) [19]. We used samples from the most recent and widely cited Affymetrix microarray platform, HGU133_plus 2.0, to perform these analyses.

Gene expression data Normalisation

Gene expression data normalisation was performed from the raw CEL files using the Aroma Affymetrix (Version 1.3.0) R package (http://www.aroma-project.org) based on custom CDF files (version 16) found at http://brainarray.mbni.med.umich.edu [20]. We processed expression for 19,003 distinct genes. All computations were performed in the R statistical environment, employing the BioConductor suite of packages.

Co-expression analysis of *ARLTS1*. The co-expression analysis method was described previously [10]. The meta cell line data (n = 1445) originated from over 48 human anatomical parts. A correlation value >0.30 and a p-value <0.05 were used as determinants for a statistically significant association. We performed multiple test corrections using Benjamini Hochberg and Bonferroni methods. We decided to use the Benjamini Hochberg (BH) method for selecting significantly co-expressed genes because it was moderately strict at calling a gene pair correlation a false positive.

In silico functionality prediction

RegulomeDB (http://regulome.stanford.edu/) and HaploReg (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) [21,22] databases were used to further elucidate the role of eQTLs in gene regulation. RegulomeDB allows he features of DNA and regulatory elements of non-coding regions to be assessed, and HaploReg is a tool for developing mechanistic hypotheses of the impact of candidate regulatory non-coding variants on clinical phenotypes and normal variation.

MDR analysis

The multifactor dimensionality reduction (MDR) program is publicly available from the internet (www.epistasis.org). We used version 2.0_beta_8.4 to examine gene-gene interactions. MDR detects interactions in relatively small sample sizes. MDR is a non-parametric, model-free data mining approach constructive induction algorithm that transforms the high-dimensional data into one-dimensional variables by pooling genotypes into high and low risk groups based on the ratio of cases to controls that have the genotype in question [23]. MDR selects one genetic model (a one, two, three or four stage locus) that most successfully predicts the phenotype or disease status, e.g., cancer. Data are then divided into ten equal parts to perform a 10-fold cross-validation. The model creates a training set (9/10 of the data) and testing set (1/10 data) to evaluate the prediction ability. The procedure repeats this protocol ten times and calculates the cross-validation consistency (CVC). CVC depicts the number of times that particular model is chosen as the best one of those ten intervals. From the MDR results section, testing balanced accuracy (TBA) shows how many instances are correctly classified. The *ARLTS1* genotypes from 102 PCa cases and 33 controls were combined with the GWAS data of 700,000 SNPs.

Selection of genes and SNPs for MDR

The genes functioning in immune system processes and inflammation pathways were selected from a previously published comprehensive collection made by Loza MJ et al. [24] because MDR is not able to run data sets of thousands of SNPs within reasonable time limits. In short, the SNPs selected included those involved in apoptosis, cytokine signalling, Toll-like receptor signalling, leukocyte signalling, complement, adhesion and natural killer cell signalling. We also ran MDR with a subset of genes gathered from our previous *ARLTS1* co-expression studies (e.g., twelve genes within B-cell receptor [BCR] signalling pathway). The total amount of SNPs was 12,011 of which 4,764 were found in our Illumina Human OmniExpress GWAS data.

Results

RNA expression

The *ARLTS1* RNA expression levels were analysed from total RNA of 84 PCa cases and 15 controls. A significant decrease of the *ARLTS1* expression level in PCa cases was detected (p = 0.0037, Fig. 1). This is in concordance with our previous results from cell line, benign prostatic hyperplasia (BPH) and tumour specimen RNA expression data [10].

eQTL analysis

To identify possible *cis*-acting genetic variants associated with *ARLTS1* transcript levels, we performed an eQTL analysis within a special area of the 13q14 region. By a linear regression model (PLINK), we were able to detect 5 eSNPs affecting *ARLTS1* expression (Table 2). When the calculation window was diminished from 1 Mb to 200 kb, only one SNP, rs7997377, remained. With the directional test performed by the R-Package analysis tool, 11 statistically significant eSNPs were found (Table 2), and two were in concordance with the PLINK linear regression model

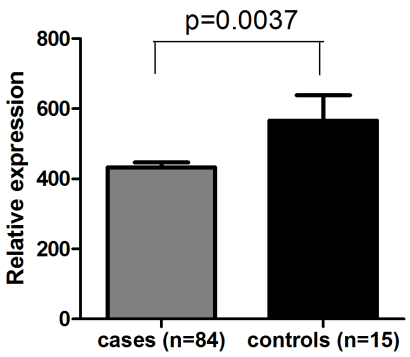


Figure 1. Relative *ARLTS1* RNA expression from PCa patients and healthy controls. Relative *ARLTS1* RNA expression was determined by RNA sequencing analysis. Columns represent means of individuals; bars represent SD.
doi:10.1371/journal.pone.0072040.g001

results. For both test procedures a test size of 0.01 was chosen. The eSNPs found by the directional test were located mainly in non-coding regions (Table 2). The genomic locations of the eSNPs found in the 13q14 region are visualised in Figure 2. Altogether, 468 genomic variants within the 1 Mb region originating from *ARLTS1* SNP rs9526582 were tested by a linear regression model and directional test.

After adjusting for multiple testing, a FDR of 39% in the linear model and approximately 32% in the directional test has to be accepted to keep the significant test results from the marginal p-values. For a more common FDR level of 10% one significant eSNP from the directional test remained significant (rs9568354). When we considered the ratio of expected and observed significant tests, a maximum ratio for $\alpha = 0.011$ in directional test was identified. For that α approximately 2.5 times more significant test results appeared than expected. Hence, we report also the above mentioned non-adjusted p-values for a significance level 0.01. With linear regression model, the amount of observed significant tests matched the amount of expected test rejections under the null hypothesis.

The association of the eSNP genotypes with the *ARLTS1* expression level was calculated. A statistically significant correlation was naturally observed between all the 14 SNPs and *ARLTS1* transcript levels. The eSNP genotype - *ARLTS1* expression association box plots are depicted in Figure 3.

The functionality and possible transcriptional regulatory effect of the 14 eSNPs within genes found by eQTL was evaluated using ENCODE-data in the RegulomeDB and HaploReg databases. Altogether nine of the fourteen eQTLs are reported in RegulomeDB. The findings in the GM12878 lymphoblastoid cell line are emphasized below because this cell line resembles the tissue type from which the RNA sequencing data was retrieved.

The most substantial evidence for the regulation of *ARLTS1* was found for SNP rs2532975. Its regulatory role is supported by its location in a regulatory active region. According to the Chip-Seq data from the GM12878 cell line, rs2532975 resides in a BATF (basic leucine zipper transcription factor) binding site. In addition, using position weight matrix (PWM) matching, a CDC5 (cell cycle serine/threonine-protein kinase) binding motif has been identified that spans the genomic position of this variant. Furthermore, a promyelocytic leukemia zinc finger (PLZF) motif is reported in HaploReg. As reported by HaploReg, CDC5 binding efficiency decreases while PLZF binding increases. Additionally, the chromatin state in the region surrounding rs2532975 might adopt weak enhancer characteristics. This prediction is further strengthened by the presence of two histone marks in this region, H3k4me1 and H3k4me2, identified in the GM12878 cells.

Another possible candidate for *ARLTS1* regulation is rs9562905. In a Chip-Seq study, a POLA2 binding site was identified in a human embryonic stem cell line, H1-hESC, in the region surrounding rs9562905. HaploReg predicts active enhancer characteristics in the chromatin surrounding rs9562905 in the lymphoblastoid GM12878 cells, similar to rs2532975. This interpretation is supported by the presence of two enhancer

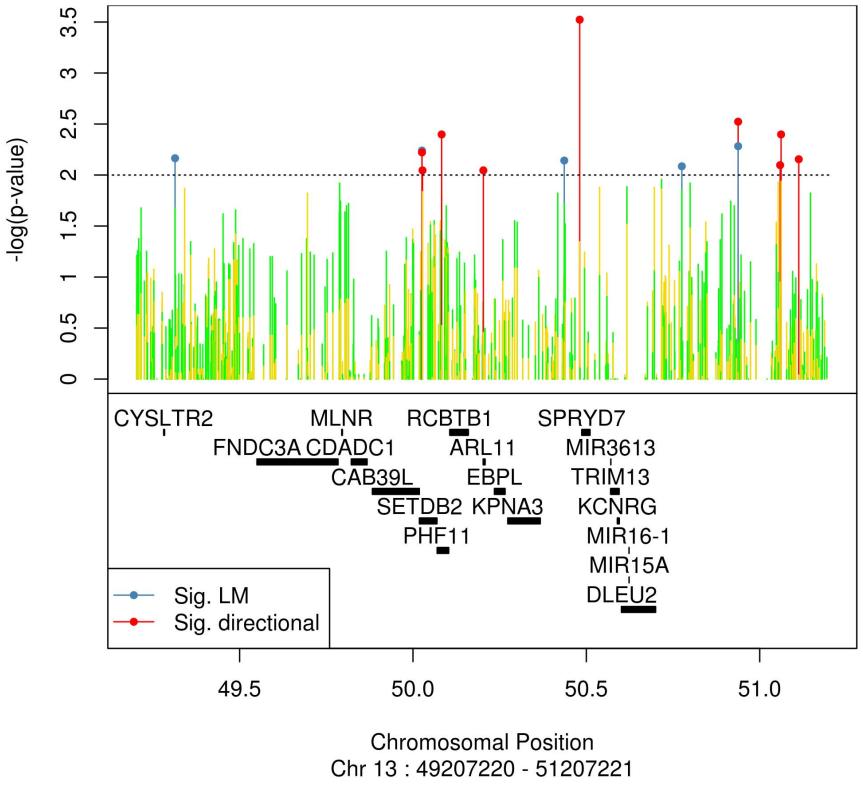


Figure 2. Schematic diagram showing the genomic locations of eSNPs gathered by eQTL analysis in PCa patients. The gene symbols are as follows: *CYSLTR2* (cysteinyl leukotriene receptor 2), *FNCD3A* (fibronectin type III domain containing 3A), *MLNR* (motilin receptor), *CDADC1* (cytidine and dCMP deaminase domain containing 1), *CAB39L* (calcium binding protein 39-like), *SETDB2* (SET domain, bifurcated 2/*CLLD8*), *PHF11* (PHD finger protein 11/NY-REN-34 antigen), *RCBTB1* (regulator of chromosome condensation [RCC1] and BTB [POZ] domain containing protein 1/*CLLD7*), *ARLTS1* (/*ARL11*, ADP-ribosylation factor-like 11), *EBPL* (emopamil binding protein-like), *KPNA3* (karyopherin alpha 3, importin alpha 4), *SPRYD7* (SPRY domain containing *7/C13orf1*, chromosome open reading frame 1), *MIR3613* (microRNA 3613), *TRIM13* (tripartite motif containing 13), *KCNRG* (potassium channel regulator), *MIR-15A* and *MIR16-1* (microRNA genes 15a and 16-1) and *DLEU2* (deleted in lymphocytic leukemia 2).
doi:10.1371/journal.pone.0072040.g002

Table 2. Chromosomal region 13q14 risk variants (eSNPs) associated with differential *ARLTS1* expression.

SNP	Gene	Position	Allele1	Allele2	P-value	Adjusted P-value
Linear regression model						
RS1886014	N/A	49321044	A	G	0,007	0,369
RS7997737	*SETDB2*	50033188	G	A	0,006	0,322
RS7337547	N/A	50443527	C	A	0,008	0,384
RS7995192	N/A	50782599	G	A	0,008	0,331
RS2532975	N/A	50945011	G	A	0,005	0,322
Directional test						
RS2075610	*MLNR*	49795705	G	A	0,010	0,322
RS7997737	*SETDB2*	50033188	G	A	0,008	0,322
RS1543513	*SETDB2*	50034684	A	C	0,007	0,322
RS9568232	*PHF11*	50089844	A	G	0,000	0,322
RS9562905	N/A	50210212	A	C	0,008	0,322
RS9568354	*SPRYD7*	50487993	A	G	0,002	0,000
RS2580189	N/A	50806640	A	G	0,009	0,322
RS2532975	N/A	50945011	G	A	0,001	0,322
RS1262781	N/A	51066171	A	G	0,010	0,322
RS1262774	N/A	51068896	A	G	0,006	0,322
RS17074618	N/A	51153475	A	G	0,006	0,322

doi:10.1371/journal.pone.0072040.t002

associated histone marks, H3k4me1 and H3k4me2. In addition, a FAIRE-sequencing study conducted for the GM12878 cells also implies that this region is in an open chromatin state and is therefore likely to have regulatory activity.

The variants rs1543513, rs7997737 and rs9568354 share similar chromatin structural features in the GM12878 cells. According to HaploReg, the chromatin state associates with the weakly transcribed region. This prediction is confirmed by the elongation of histone mark H3k36me3 in the surrounding regions of the variants rs1543513, rs7997737 and rs9568354. According to RegulomeDB, rs1543513 is located within the Irx3 and Irx6 motifs. However, in HaploReg, the presence of these motifs is not reported. Similarly, the transcription factor binding motif of AIRE (autoimmune regulator) is reported for rs1262781 in RegulomeDB but not in HaploReg. A MZF1 (myeloid zinc finger 1) motif surrounding rs7997737 is reported in both databases. HaploReg reports a negative LOD score difference for rs7997737, which can be interpreted as lowered binding efficiency of MZF. Compared to the three variants mentioned above, rs9568232 shares similar characteristics regarding its chromatin state. According to HaploReg, this chromatin state is related to elongated transcription.

The variants rs2075610 and rs1262774 are located within regions most likely under the control of epigenetic regulation by the polycomb-group proteins in the GM12878 cells. In addition, DNase-Seq studies performed for several cell lines indicate an open chromatin state surrounding rs2075610. However, two repressed state chromatin histone marks, H3k27me3 and H3k9me3, have also been identified. Variant rs1262774 is not reported in the RegulomeDB.

The remaining variants are located in heterochromatin regions, according to HaploReg. For rs7995192, rs2580189 and rs1262781, this prediction is further confirmed by the presence of H3k27me3. In addition, another repressive state associated histone mark, namely H3k9me3, is present in the rs7995192 and rs2580189 regions. Although there is evidence of repressed state chromatin, RegulomeDB reports that transcription factor binding motifs do in fact span the genomic positions of rs2580189 and rs1262781.

Two motifs for XBP-1 (X-box binding protein 1) and ATF6 (activating transcription factor 6) span the region of rs1262781, according to RegulomeDB. HaploReg confirms the presence of XBP-1 and ATF6 motifs and an additional motif for SOX-17 (SRY [sex determining region Y] box 17). HaploReg reports lower predicted binding efficiencies for XBP-1 and ATF6 and a slightly increased binding efficiency for SOX-17.

RegulomeDB reports a ZNF143 (zinc finger protein 143) motif in the region surrounding SNP rs2580189. However, HaploReg does not report the presence of this motif. Taken together, the results gathered from RegulomeDB and HaploReg indicate that the *ARLTS1* eQTLs are located within regulatory areas of the 13q14 region.

Co-expression analysis

The GeneSapiens mRNA expression database data, including *ARLTS1* expression, from 1445 cell lines (48 cancer subtypes) and prostate cancer tumours was available for interaction studies. Altogether 1381 genes with correlation value >0.30 and p-value <0.05 was found to be positively correlating with the *ARLTS1* gene. Using DAVID GO functional clustering, (http://david.abcc.ncifcrf.gov/tools.jsp) [25,26] a strong association with nucleus and zinc-finger protein processes was illustrated when all the genes positively correlating with *ARLTS1* expression ($n = 36$) in the PCa cell line cohort were taken into account (with a more stringent correlation value >0.50). The group of nuclear processes (nucleus, intracellular organelles, transcription and DNA-binding) was enriched when data of PCa cell lines was studied. The enrichment score was 13.49 with a p-value 1.1E-32. The adjusted Benjamin score was 5.1E-30, and the cluster of zinc-finger binding proteins revealed a p-value of 9.0E-22. The same phenomenon was

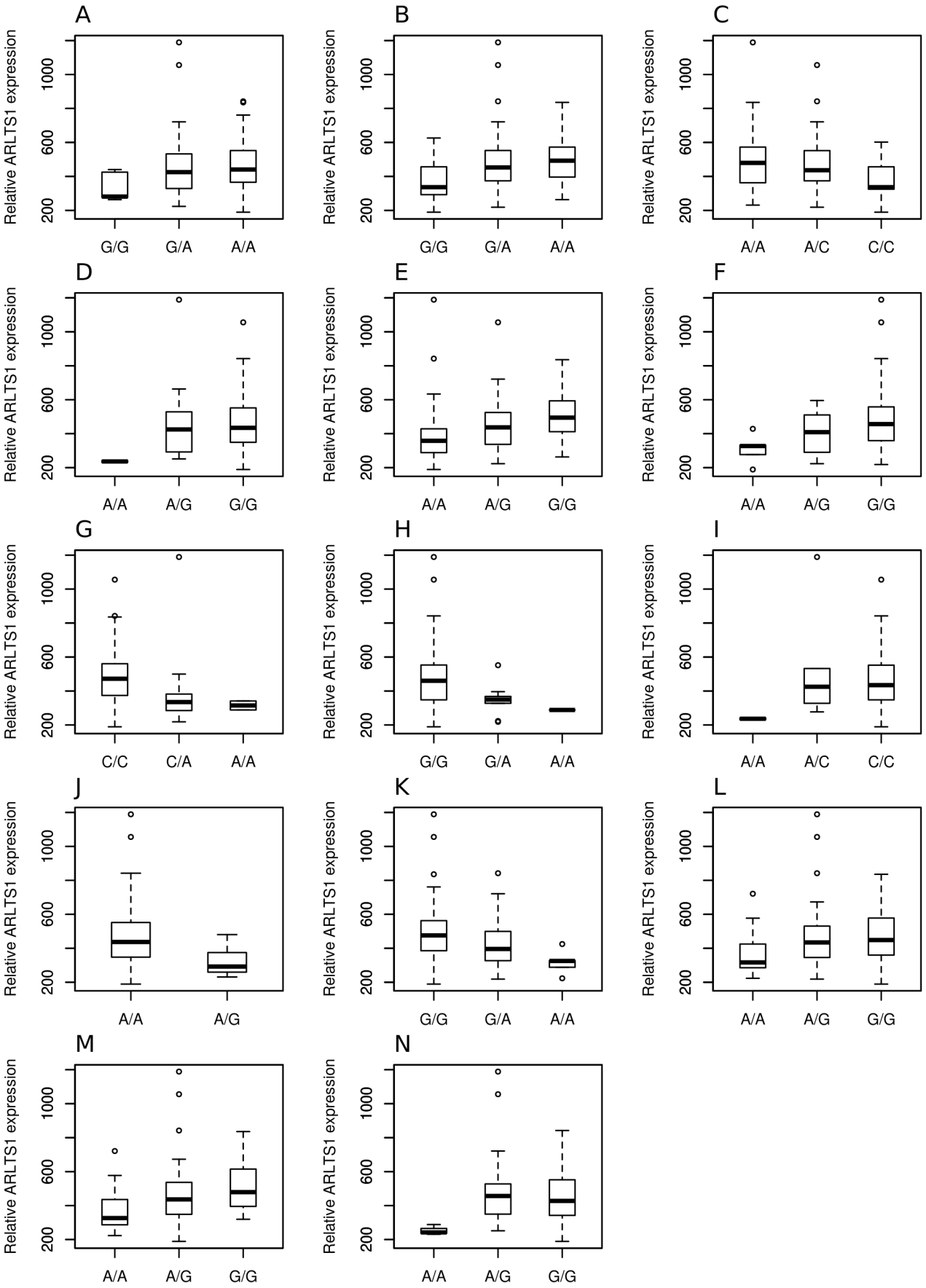

A
Relative ARLTS1 expression
200
600
1000
G/G
G/A
A/A
B
Relative ARLTS1 expression
200
600
1000
G/G
G/A
A/A
C
Relative ARLTS1 expression
200
600
1000
A/A
A/C
C/C
D
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G
E
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G
F
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G
G
Relative ARLTS1 expression
200
600
1000
C/C
C/A
A/A
H
Relative ARLTS1 expression
200
600
1000
G/G
G/A
A/A
I
Relative ARLTS1 expression
200
600
1000
A/A
A/C
C/C
J
Relative ARLTS1 expression
200
600
1000
A/A
A/G
K
Relative ARLTS1 expression
200
600
1000
G/G
G/A
A/A
L
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G
M
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G
N
Relative ARLTS1 expression
200
600
1000
A/A
A/G
G/G

Figure 3. Correlation of the different genotype groups of eSNPs to *ARLTS1* RNA expression levels. A, rs2075610; B, rs7997737; C, rs1543513; D, rs9568232; E, rs9568354; F, rs1886014; G, rs7337547; H, rs7995192; I, rs9562905; J, rs2580189; K, rs2532975; L, rs1262781; M, rs1262774 and N, rs17074618.
doi:10.1371/journal.pone.0072040.g003

observed within the whole data of cell lines (meta cohort) with genes showing a correlation value >0.30. *ARLTS1* co-expression genes (with correlation value >0.50) from the meta cell line data revealed a category of immune system processes (B-/T-cell activation, leukocyte/lymphocyte differentiation and activation), with an enrichment score of 2.94 (p-value 5.57E-7, adjusted Benjamin score 2.9E-5).

ARLTS1 co-expression data of genes negatively correlated to *ARLTS1* identified a strong gene ontology of glycoprotein and plasma membrane protein genes in PCa cell lines (n = 2722, correlation value<−0.50, enrichment score 52.13, p-value 3.4E-72 and 38.35, p-value 7.9E-32, respectively). Within the negatively correlating genes, a cluster of immunoglobulin domain containing proteins harboured an enrichment score of 12.23 with a p-value of 5.4E-24. The GO term "cytokine activity" revealed an enrichment score of 10.43 with a p-value of 6.5E-10 within negatively correlating genes. In addition to the result of *ARLTS1* negatively correlating genes, we also identified clusters of G-protein coupled receptors and cell-cell signalling.

Top five positively and negatively *ARLTS1* correlating genes in cell line data are presented in Table 3 (upper panel). Additionally, results of the *ARLTS1* co-expression with genes (*SETDB2*, *PHF11*, *SPRYD7*, *MLNR*) that harbored eSNPs are presented in the lower part of the Table 3, from the co-expression analysis performed in the meta cell line, prostate cell line data and prostate tumor data.

The expression of *ARLTS1* was positively correlated with the expression of *SETBD2*, in both the whole data of cell lines (meta cell line cohort) (correlation value 0.64, p-value 0.000) and in the

Table 3. *ARLTS1* Co-expression signatures from tumor specimens and cell lines.

Gene	Correlation value	P-value	Samples (n)	pval_corrected*	pval_corrected#
Top five genes					
Co-expression in meta cell line data					
Positive correlation					
BTK	0,69	0	2818	0	0
GPR18	0,66	0	2818	0	0
CXorf21	0,66	0	2818	0	0
P2RY8	0,66	0	2818	0	0
PIK3CG	0,65	0	2818	0	0
Negative correlation					
NCKAP1	−0,62	1,69E-295	2818	3,22E-291	7,70E-295
CDC42BPB	−0,59	1,64E-263	2818	3,11E-259	7,43E-263
GIPC1	−0,57	1,10E-239	2818	2,10E-235	5,01E-239
PTMS	−0,53	2,68E-208	2818	5,09E-204	1,22E-207
KIAA0284	−0,53	1,04E-203	2818	1,97E-199	4,71E-203
Genes harboring eSNPs					
Co-expression in meta cell line data					
MLNR	−0,15	1,35E-16	2818	2,57E-12	2,96E-16
PHF11	**0,44**	**0,000**	**2818**	**0,000**	**0,000**
SETDB2	**0,64**	**0,000**	**2818**	**0,000**	**0,000**
SPRYD7	0,28	0,000	2818	0,000	0,000
Co-expression in PCa cell lines					
MLNR	**−0,35**	**0,0352**	**36**	**1**	**0,081**
PHF11	0,09	0,608	36	1	0,704
SETDB2	**0,61**	**9,01E-05**	**36**	**1**	**0,001**
SPRYD7	0,17	0,336	36	1	0,451
Co-expression in prostate tumor samples					
MLNR	0,03	0,821	75	1	0,934
PHF11	−0,07	0,576	75	1	0,806
SETDB2	−0,07	0,547	75	1	0,788
SPRYD7	**−0,34**	**0,003**	**75**	**1**	**0,045**

*Bonferroni correction.
#Benjamini Hohchberg multiple testing correction.
doi:10.1371/journal.pone.0072040.t003

specific PCa cell lines (correlation value 0.61, p-value 0.00009) (Table 3). Expression of the *PHF11* gene was positively correlated with *ARLTS1* expression in the meta cell line data (correlation value 0.44, p-value 0.000). Expression of *MLNR* and *SPRYD7* was negatively correlated with *ARLTS1* expression. *ARLTS1* and *MLNR* had a correlation value of −0.35 (p-value 0.04) in PCa cell lines, and *ARLTS1* and *SPRYD7* had a correlation value of −0.34 (p-value 0.003) in prostate tumour specimens (Table 3). The strong negative correlation of *ARLTS1* and *SPRYD7* expression levels was also validated in our transcriptome data of 84 PCa cases and 15 controls.

MDR analysis

By direct sequencing, we were able to detect six *ARLTS1* variants at the same amplicon from PCa patients included in the Illumina genotyping (n = 135). All the variants were previously known (rs117251022, rs3803186, rs147120792, rs3803185, rs138452698 and G446A [Trp149Stop]). To elucidate the genotypic *ARLTS1* interactions we used multifactor dimensionality reduction (MDR). Gene-gene interaction status between *ARLTS1* and genes functioning in immune system processes within 102 PCa cases and 33 controls was calculated, but we were not able to find any statistically significant *ARLTS1* interactions in this study cohort (data not shown).

Discussion

ARLTS1 is a cancer-predisposing gene with proven tumour suppressor properties. However, very little evidence on function, especially on pathways, is currently available. In this study, we were able to verify downregulated *ARLTS1* expression in the blood-derived RNA of PCa patient samples, demonstrating for the first time the effect of germline alteration on *ARLTS1* expression levels. Tumour suppressor function of the *ARLTS1* gene has been previously proven by Calin GA et al., who found that transduction of full-length *ARLTS1* to A549 cells in Nu/Nu mice decreased tumour growth when compared to empty vector [1]. The ability of *ARLTS1* to suppress tumour formation in preclinical models has also been observed with ovarian [27] and lung cancer cells [28]. However, these results are based on somatic mutations in cancerous cell lines, whereas our result reveals a novel expression difference at the germline level, supporting the role of *ARLTS1* as a tumour suppressor gene. In the future, these types of findings could be used to enable screening and detection of at-risk patients even before clinical diagnoses.

We have previously genotyped *ARLTS1* variants in prostate, breast and colorectal cancer [29] and produced a prostate cancer follow-up study [10]. In the first study [29], we reported a statistically significant association with *ARLTS1* variants T442C, G194T and prostate cancer risk. However, after adjusting for multiple testing, none of the results were significant. In the follow-up study with larger sample size, we reported a statistically significant association with T442C variant and prostate cancer risk [10]. We reported also a decreased or lost *ARLTS1* RNA or protein expression in clinical prostate tumors, prostate cancer cell lines and xenografts, supporting the role of *ARLTS*1 as a tumor suppressor gene. Thus, there is no conflict between the previous publications and this study that is confirming the tumor suppressor role of *ARLTS1*.

Here, 14 eSNPs located at the 13q14 region were shown to significantly influence *ARLTS1* transcript levels in PCa patients. The eQTL analysis was performed with two methods, a linear model approach and a directional test. One advantage of our directional test method over the commonly used linear model approach is its robustness against outliers and the weaker model assumptions. In the linear model approach, the expression values of the different genotype groups are considered to follow the same normal distribution up to a location shift parameter. If the location shift parameter is not equal to zero, it generates a testing problem. The directional test only assumes that the different expression values follow certain distribution functions, so a testing problem only occurs if the distributions of the genotype groups are stochastically ordered.

Interestingly, 5 of the eSNPs are located within the protein-coding genes *SETDB2*, *PHF11*, *SPRYD7* and *MLNR*. Two eQTLs were positioned in the *SETDB2* (SET domain, bifurcated 2) gene, also known as the *CLLD8* (chronic lymphocytic leukemia deletion region gene 8 protein) gene, which functions mainly in epigenetic regulation [30]. The *PHF11* gene, another hit for the eSNPs, is a positive regulator of Th1-type cytokine gene expression through nuclear factor kappa B, (NF-kB) [30,31]. The third eSNP gene, *SPRYD7*, (SPRY domain containing 7/chromosome 13 open reading frame 1 *[C13or1]*) also known as chronic lymphocytic deletion region gene 6 protein, (*CLLD6*) [32], is proposed to have a tumour suppressor function because it has been detected to be downregulated in B-CLL patients [33]. The SPRY/B30.2 protein domain occurs in a variety of cellular proteins, mediates protein-protein interactions and negatively regulates cytokine activities [34,35]. The fourth eSNP gene, *MLNR/MTLR1*, (G-protein-coupled receptor 38, GPR38) is a member of the G-protein coupled receptor 1 family and is identified as the motilin receptor [36]. The found eSNPs were not affecting the expression of the gene they resided in, suggesting the role of these eSNPs as specific regulators targeting *ARLTS1* expression. However, no explicit conclusions about the actual causal interaction between these variants and *ARLTS1* can be made without further functional validation. The eQTL result gathered in this study was from a relatively small sample set in which analyses encompassed 1 Mb up- and downstream of the *ARLTS1* gene. Thus, further validation in larger sample sets and genomic areas are warranted. However, in our results, one eSNP remained significant even after adjustments for multiple testing. Those adjustments are intricate in the case of eQTL analysis and studies with small sample sizes often suffer from a low detection rate after multiple testing adjustment [37]. One possible strategy to deal with this is to accept a relatively high FDR for the sake of later validation in a larger population. Here, no potential candidate eSNPs were excluded from further studies because the α with largest ratio between observed and rejected tests was considered as optimal test size.

Lymphoblastoid cell lines (LCLs) are widely used in eQTL studies because they are an easily accessible source of patient samples of a single cell type. Particularly with prostate cancer, the tumour specimens of multiple foci are hard to collect in large amounts making it difficult to reliably detect disease-associated traits. By using prostate cancer specimens, it is possible to find tissue-specific genetic effects, but the use of lymphoblastoid cell lines or whole blood of familial prostate cancer patients enables one to detect the possible heritable germline differences which may contribute to prostate cancer susceptibility. It has been argued that SNP-transcript approaches using LCLs of small sample sets are underpowered [38], but many studies have been able to find overlapping eQTLs in cell lines and primary tissues [39–41]. Our finding of the eSNP in the *SPRYD7* gene was endorsed in the co-expression data, in which the *ARLTS1* expression levels were significantly correlated with *SPRYD7* expression, and in the prostate tumour specimens (Table 3). Additionally, in the data from the ENCODE consortium, the significance and usage of LCLs are justified because the lymphoblastoid cell line (GM12878)

is one of the three cell lines in the higher priority Tier 1 cohort [42].

There may be a joint effect between the 13q14 genes. For example, one larger transcript variant that consists of more than one gene has been proposed with *SETDB2* and *PHF11* [30]. Very little is known about the interactions in this area, particularly the genomic collaborators of the *ARLTS1* gene. A recent study identified cellular retinoic acid binding protein 2 (CRABP2) and phosphoglycerate mutase 1 (PGAM1) as novel ARLTS1-binding proteins using the in-frame cDNA library technique [43]. These proteins were not observed in the expression or genotype level of our data, so further studies are needed to elucidate the actual *ARLTS1* interacting genes and proteins.

Considering the interactions between *ARLTS1* and inflammation pathway genes, it was our hypothesis that *ARLTS1* would have interactions with inflammatory genes, as chronic inflammation has been proposed as a prostate cancer risk factor. With MDR analysis, we aimed to investigate the previously observed *ARLTS1* connection with immune system processes [10]. Another aim was to test the hypothesis of genes functioning in inflammation processes affecting prostate cancer risk. However, we failed to substantiate our hypothesis, and statistically significant association between *ARLTS1* T442C SNP and the inflammatory/immune system SNPs was not found within the 700,000 SNPs analysed. For the analysis, we extracted distinct SNP groups from 4,764 markers found from our data of 700,000 variants in total. One possible reason for the negative results may be the relatively limited sample set of 135 cases and controls because we were not able to generalise the results to the test data. We were also concerned about overfitting the data with MDR.

Our previously found association between *ARLTS1* T442C and prostate cancer risk may be caused by the interaction network of different SNPs in the 13q14 region. Variant T442C may be included in a very rare haploblock, or there may still be a missing variant because the region has been connected to other cancers in addition to PCa.

In conclusion, in this study we have shown that the *ARLTS1* expression level is influenced by changes in germline expression levels and that the *ARLTS1* gene is a quantitative trait locus for 14 eSNPs in the 13q14 region. Thus, our results indicate a more complicated network of changes that increase PCa risk than merely one genetic variant in *ARLTS1*.

Acknowledgments

We thank all of the patients who participated in our study and acknowledge the contribution of Riina Liikala and Riitta Vaalavuo for technical assistance.

Author Contributions

Conceived and designed the experiments: SS TW DF JPM JS. Performed the experiments: SS VL DF JPM TR. Analyzed the data: SS DF TR JPM. Contributed reagents/materials/analysis tools: JS OK. Wrote the paper: SS TW JS.

References

1. Calin GA, Trapasso F, Shimizu M, Dumitru CD, Yendamuri S, et al. (2005) Familial cancer associated with a polymorphism in ARLTS1. N Engl J Med 352: 1667–1676.
2. Frank B, Hemminki K, Brenner H, Hoffmeister M, Chang-Claude J, et al. (2006) ARLTS1 variants and risk of colorectal cancer. Cancer Lett 244: 172–175.
3. Frank B, Hemminki K, Meindl A, Wappenschmidt B, Klaes R, et al. (2006) Association of the ARLTS1 Cys148Arg variant with familial breast cancer risk. Int J Cancer 118: 2505–2508.
4. Frank B, Meyer P, Boettger MB, Hemminki K, Stapelmann H, et al. (2006) ARLTS1 variants and melanoma risk. Int J Cancer 119: 1736–1737.
5. Sellick GS, Catovsky D, Houlston RS (2006) Relationship between ARLTS1 polymorphisms and risk of chronic lymphocytic leukemia. Leuk Res 30: 1573–1576.
6. Xu J, Dimitrov L, Chang BL, Adams TS, Turner AR, et al. (2005) A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. Am J Hum Genet 77: 219–229.
7. Cropp CD, Simpson CL, Wahlfors T, Ha N, George A, 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.
8. Visakorpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, et al. (1995) Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. Cancer Res 55: 342–347.
9. Rokman A, Koivisto PA, Matikainen MP, Kuukasjarvi T, Poutiainen M, et al. (2001) Genetic changes in familial prostate cancer by comparative genomic hybridization. Prostate 46: 233–239.
10. Siltanen S, Wahlfors T, Schindler M, Saramaki OR, Mpindi JP, et al. (2011) Contribution of ARLTS1 Cys148Arg (T442C) variant with prostate cancer risk and ARLTS1 function in prostate cancer cells. PLoS One 6: e26595.
11. Sfanos KS, De Marzo AM (2012) Prostate cancer and inflammation: The evidence. Histopathology 60: 199–215.
12. Kwon EM, Salinas CA, Kolb S, Fu R, Feng Z, et al. (2011) Genetic polymorphisms in inflammation pathway genes and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 20: 923–933.
13. De Marzo AM, Platz EA, Sutcliffe S, Xu J, Gronberg H, et al. (2007) Inflammation in prostate carcinogenesis. Nat Rev Cancer 7: 256–269.
14. Schleutker J, Matikainen M, Smith J, Koivisto P, Baffoe-Bonnie A, et al. (2000) A genetic epidemiological study of hereditary prostate cancer (HPC) in finland: Frequent HPCX linkage in families with late-onset disease. Clin Cancer Res 6: 4810–4815.
15. Trapnell C, Pachter L, Salzberg SL. (2009) TopHat: Discovering splice junctions with RNA-seq. Bioinformatics 25: 1105–1111.
16. Anders S, Huber W. (2010) Differential expression analysis for sequence count data. Genome Biol 11: R106-2010-11-10-r106. Epub 2010 Oct 27.
17. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, et al. (2008) High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet 4: e1000214.
18. George RE, Attiyeh EF, Li S, Moreau LA, Neuberg D, et al. (2007) Genome-wide analysis of neuroblastomas using high-density single nucleotide polymorphism arrays. PLoS One 2: e255.
19. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, et al. (2012) The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483: 603–607.
20. Hu R, Qiu X, Glazko G, Klebanov L, Yakovlev A (2009) Detecting intergene correlation changes in microarray analysis: A new approach to gene selection. BMC Bioinformatics 10: 20-2105-10-20.
21. Ward LD, Kellis M (2012) HaploReg: A resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res 40: D930–4.
22. ENCODE Project Consortium. (2011) A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol 9: e1001046.
23. Hahn LW, Ritchie MD, Moore JH (2003) Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics 19: 376–382.
24. Loza MJ, McCall CE, Li L, Isaacs WB, Xu J, et al. (2007) Assembly of inflammation-related genes for pathway-focused genetic analysis. PLoS One 2: e1035.
25. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57.
26. Huang da W, Sherman BT, Lempicki RA. (2009) Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37: 1–13.
27. Petrocca F, Iliopoulos D, Qin HR, Nicoloso MS, Yendamuri S, et al. (2006) Alterations of the tumor suppressor gene ARLTS1 in ovarian cancer. Cancer Res 66: 10287–10291.
28. Yendamuri S, Trapasso F, Ferracin M, Cesari R, Sevignani C, et al. (2007) Tumor suppressor functions of ARLTS1 in lung cancers. Cancer Res 67: 7738–7745.
29. Siltanen S, Syrjakoski K, Fagerholm R, Ikonen T, Lipman P, et al. (2008) ARLTS1 germline variants and the risk for breast, prostate, and colorectal cancer. Eur J Hum Genet 16: 983–991.
30. Zhang Y, Leaves NI, Anderson GG, Ponting CP, Broxholme J, et al. (2003) Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. Nat Genet 34: 181–186.

31. Clarke E, Rahman N, Page N, Rolph MS, Stewart GJ, et al. (2008) Functional characterization of the atopy-associated gene PHF11. J Allergy Clin Immunol 121: 1148–1154.e3.
32. Mabuchi H, Fujii H, Calin G, Alder H, Negrini M, et al. (2001) Cloning and characterization of CLLD6, CLLD7, and CLLD8, novel candidate genes for leukemogenesis at chromosome 13q14, a region commonly deleted in B-cell chronic lymphocytic leukemia. Cancer Res 61: 2870–2877.
33. van Everdink WJ, Baranova A, Lummen C, Tyazhelova T, Looman MW, et al. (2003) RFP2, c13ORF1, and FAM10A4 are the most likely tumor suppressor gene candidates for B-cell chronic lymphocytic leukemia. Cancer Genet Cytogenet 146: 48–57.
34. Nicholson SE, Hilton DJ (1998) The SOCS proteins: A new family of negative regulators of signal transduction. J Leukoc Biol 63: 665–668.
35. Woo JS, Suh HY, Park SY, Oh BH (2006) Structural basis for protein recognition by B30.2/SPRY domains. Mol Cell 24: 967–976.
36. Feighner SD, Tan CP, McKee KK, Palyha OC, Hreniuk DL, et al. (1999) Receptor for motilin identified in the human gastrointestinal system. Science 284: 2184–2188.
37. Loo LW, Cheng I, Tiirikainen M, Lum-Jones A, Seifried A, et al. (2012) Cis-expression QTL analysis of established colorectal cancer risk variants in colon tumors and adjacent normal tissue. PLoS One 7: e30477.
38. Min JL, Taylor JM, Richards JB, Watts T, Pettersson FH, et al. (2011) The use of genome-wide eQTL associations in lymphoblastoid cell lines to identify novel genetic pathways involved in complex traits. PLoS One 6: e22070.
39. Bullaughey K, Chavarria CI, Coop G, Gilad Y (2009) Expression quantitative trait loci detected in cell lines are often present in primary tissues. Hum Mol Genet 18: 4296–4303.
40. Ding J, Gudjonsson JE, Liang L, Stuart PE, Li Y, et al. (2010) Gene expression in skin and lymphoblastoid cells: Refined statistical method reveals extensive overlap in cis-eQTL signals. Am J Hum Genet 87: 779–789.
41. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: The MuTHER study. PLoS Genet 7: e1002003.
42. ENCODE Project Consortium, Dunham I, Kundaje A, Aldred SF, Collins PJ, et al. (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489: 57–74.
43. Lee S, Lee I, Jung Y, McConkey D, Czerniak B (2012) In-frame cDNA library combined with protein complementation assay identifies ARL11-binding partners. PLoS One 7: e52290.