

Institute for Human Genetics, University of Ulm
Director: Prof. Dr. med. Christian Kubisch

On the impact
of risk variants in the c-*MYC* gene region
on prostate cancer development

Dissertation for the obtainment of the doctoral degree
in Human Biology (Dr. hum. biol.)
at the Faculty of Medicine, University of Ulm

Submitted by Silvia Kastler,
born in Günzburg

Ulm, 2011

Acting decane: Prof. Dr. Thomas Wirth

First correspondent: Prof. Dr. Günter Assum

Second correspondent: PD. Dr. Marcus Cronauer

Day of Defence: 06/05/2011

Table of Contents

List of abbreviations.....	III
1. INTRODUCTION.....	1
1.1 PROSTATE CANCER	1
1.1.1 DIAGNOSIS, GRADING, STAGING AND THERAPY	1
1.1.2 EPIDEMIOLOGY OF PROSTATE CANCER	3
1.2 SEARCH FOR PROSTATE CANCER SUSCEPTIBILITY GENES	4
1.3 THE 8Q24 PROSTATE CANCER RISK REGION	7
1.3.1 PUTATIVE PROSTATE CANCER RISK FACTORS ON 8Q24.21	9
1.4 AIM OF THIS WORK.....	11
2. MATERIAL AND METHODS	12
2.1. MATERIAL.....	12
2.1.1. CLINICAL SPECIMENS, TISSUE AND CELL LINES	12
2.1.2. CHEMICALS, SOLUTIONS AND BUFFERS.....	14
2.1.3. ENZYMES AND REAGENT KITS	19
2.1.4. ANTIBODIES, LADDERS AND OLIGONUCLEOTIDES	20
2.1.5. TECHNICAL EQUIPMENT, FURTHER LABORATORY MATERIAL AND SOFTWARE	23
2.2. METHODS	25
2.2.1. ISOLATION OF RNA AND DNA FROM CULTIVATED CELLS AND FROZEN TISSUE	25
2.2.2. PCR, SEQUENCING, SNAPSHOT ANALYSIS, MLPA AND RACE.....	27
2.2.3. SPECIFIC PCR FOR <i>POU5F1P1</i> AND <i>OCT4</i>	33
2.2.4. SEMIQUANTITATIVE EXPRESSION ANALYSIS.....	33
2.2.5. QUANTITATIVE EXPRESSION ANALYSIS WITH TAQMAN ANALYSIS	34
2.2.6. ANALYSIS OF CORRELATION BETWEEN 8Q24.21 TRANSCRIPTION LEVELS, RISK VARIANTS AND CLINICAL ASPECTS	38
2.2.7. SEQUENCING OF THE RETROGENE <i>POU5F1P1</i> ON CHROMOSOME 8Q24.21	38
2.2.8. MEASUREMENT OF AN ALLELIC SHIFT IN EXPRESSION OF <i>POU5F1P1</i> TRANSCRIPTS ..	39
2.2.9. ASSOCIATION STUDY.....	42
2.2.9.1. Examination of single <i>POU5F1P1</i> variants for association with prostate cancer	45
2.2.9.2. Estimation of <i>POU5F1P1</i> haplotypes and association with prostate cancer	46
2.2.10. EXAMINATION OF THE <i>POU5F1P1</i> PROTEIN.....	46
2.2.10.1. Immunohistochemistry.....	46
2.2.10.2. Western Blot.....	48

3. RESULTS	54
3.1. ANALYSIS OF THE PSEUDOGENE <i>POU5F1P1</i> IN PROSTATIC TISSUE	54
3.1.1. EXAMINATION OF <i>POU5F1P1</i> CONSERVATION IN PRIMATES	54
3.1.2. <i>POU5F1P1</i> -RELATED SEQUENCE HOMOLOGY WITHIN THE HUMAN GENOME	56
3.1.3. SPECIFIC PCRS ON <i>OCT4</i> AND <i>POU5F1P1</i>	56
3.1.4. QUANTITATIVE EXPRESSION ANALYSIS OF <i>POU5F1P1</i>	58
3.2. EXAMINATION OF CAUSATIVE EFFECTS FOR <i>POU5F1P1</i> OVEREXPRESSION IN PROSTATIC CARCINOMA	62
3.2.1. INVESTIGATION OF COPY NUMBER VARIATION OF 8Q	62
3.2.2. EXAMINATION OF TRANSCRIPTIONAL START SITES OF <i>POU5F1P1</i>	65
3.2.3. EXPRESSION OF OTHER RNAs ON CHROMOSOME 8Q24.21 IN PROSTATIC CARCINOMA TISSUE	68
3.2.4. THE PUTATIVE IMPACT OF PROSTATE CANCER RISK VARIANTS ON THE EXPRESSION LEVEL OF 8Q24.21 TRANSCRIPTS	73
3.3. ASSOCIATION STUDY WITH CODING VARIANTS IN <i>POU5F1P1</i>	79
3.3.1. ASSOCIATION OF SINGLE VARIANTS IN <i>POU5F1P1</i> WITH PROSTATE CANCER	81
3.3.2. ASSOCIATION OF <i>POU5F1P1</i> HAPLOTYPES WITH PROSTATE CANCER	84
3.4. INVESTIGATION OF THE <i>POU5F1P1</i> ENCODED PROTEIN	86
3.4.1. IMMUNOHISTOCHEMISTRY	87
3.4.2. WESTERN BLOT	90
4. DISCUSSION	93
4.1. OVEREXPRESSION OF <i>POU5F1P1</i> IN PROSTATIC CARCINOMA TISSUE	94
4.2. ON THE IMPACT OF 8Q AMPLIFICATION OR PROSTATE CANCER RISK VARIANTS ON <i>POU5F1P1</i> OVEREXPRESSION IN PROSTATIC CARCINOMA TISSUE	95
4.3. ON THE IMPACT OF VARIANTS IN <i>POU5F1P1</i> ON PROSTATE CANCER DEVELOPMENT	98
4.4. EXAMINATION OF THE <i>POU5F1P1</i> ENCODED PROTEIN IN PROSTATIC TISSUE	101
4.5. CONCLUSION OF THIS WORK	103
4.6. ALTERNATIVE TARGETS OF PROSTATE CANCER RISK VARIANTS ON 8Q24.21	103
5. SUMMARY	106
6. REFERENCES	108
APPENDIX	119

List of abbreviations

AJCC	American Joint Committee On Cancer
APS	ammoniumpersulphate
bp	base pairs
BLAST	Basic Local Alignment Search Tool
BPH	benign prostate hyperplasia
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CGH	comparative genome hybridisation
Chr pos	chromosomal position
CI	confidence interval
CIP	calf intestine phosphatase
C _T	threshold cycle
°C	degrees Celsius
DAB	diaminobenzidine
ddNTPs	dideoxyribonucleotide
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide
DTE	dithiothreitol
DTT	dithioerythritol
demin	demineralised
ECL	enhanced chemoluminescence
<i>E.coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EST	expressed sequence tag
EtOH	ethanol
ExpPASy	Expert Proteom Analysis System
FCS	fetal calf serum
FISH	Fluorescence in-situ Hybridisation
FRET	Fluorescence Resonance Energy Transfer
GCG	Genetic Computer Group
h	hour
HCl	hydrochloric acid
H & E	haematoxylin & eosin

HPC	hereditary prostate cancer
HRP	horseradish peroxidase
ICPCG	International Consortium for Prostate Cancer Genetics
kb	kilo bases
LD	linkage disequilibrium
Ig	immunoglobulin
Linc RNA	large intergenic non-coding ribonucleic acid
MAF	minor allele frequency
min	minute
MLPA	Multiplex Ligation-dependent Probe Amplification
mRNA	messenger ribonucleic acid
µg	microgram
µm	micrometer
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Center for Biotechnology Information
ng	nanogram
nm	nanometer
OR	odds ratio
PAGE	polyacrylamid gel electrophoresis
PBS	phosphate buffered saline
PCa	prostate cancer
PCAP	Predisposing for Prostate Cancer
PCR	polymerase chain reaction
PIN	prostatic intratubular neoplasia
PSA	prostate-specific antigen
RACE	rapid amplification of 5'dDNA ends
RAF	risk allele frequency
RefSeq	Reference Sequence
RLM	RNA ligase-mediated
RT-PCR	reverse transcription polymerase chain reaction
SAP	shrimp alkaline phosphatase
SDS	sodium dodecyl sulphate
SNP	single nucleotide polymorphism

TAP	tobacco acid pyrophosphatase
TBST	Tris Buffered Saline Tween [®] 20
TE	Tris EDTA
TEMED	tetramethylethylenediamine
TIN	intratubular germ cell neoplasia of the testis
T,N,M	primary tumour, lymph nodes, metastases
u	unit
UICC	Union for International Cancer Control
V	Volt

1. Introduction

1.1 Prostate cancer

The prostate is an exocrine gland, which is part of the male internal sexual organs. It is made of the glandular epithelium and fibromuscular stroma, and it expels a secretion which makes up to 30 to 40 % of the ejaculate, which further consists of the sperm produced by the testis and the secretion of the seminal vesicle. The prostate is located right under the bladder, ventral of the rectum and is embedding the urethra. Its size is approximately that of a walnut and its weight is around 20 g. A frequent non-cancerous neoplasia of the prostatic gland is the benign prostatic hyperplasia (= BPH), which possibly originates at the age of 30 or later, but it does not invade other tissues. It consists of excess glands and stroma in the transition zone of the prostate close to the urethra. Symptoms are problems with urinating and with sexual functions. It can be treated by transurethral resection. A lesion of the prostate harbouring neoplastic cell growth, that line pre-existing acini and ducts, is the prostatic intratubular neoplasia (PIN), which occurs in the peripheral zone of the prostate. Approximately seventy to eighty percent of prostatic carcinoma originates here, and high-grade PIN is supposed to be the precursor of prostate cancer. It is estimated that one of six men in the U.S.A. will become diagnosed with prostate cancer during lifetime, and one of 36 will die of this disease. In the year 2004, prostate cancer was responsible for around 10 % of all cancer-related deaths in men. Most prostate cancer patients die of metastases, which predominantly occur in the bone or other organs.

1.1.1 Diagnosis, grading, staging and therapy

An early detection and treatment is important for successful treatment of prostate cancer. Digital rectal examination is recommended as early recognition programme for men older than 45. Prostate-specific antigen (= PSA) is produced in the epithelial cells of the prostate, and measuring PSA values in the blood serum is another common procedure for the screening for prostate cancer (Schroder et al., 2009). Values of 3.0 ng / ml PSA or more are indications for punch biopsy and further examination. After punch biopsy, tissue is examined for carcinoma tissue and subsequently graded for prognosis and therapy.

There are different types of grading of prostatic carcinoma. The standard procedure for tumour grading follows Gleason and Mellinger, and is based on the histological arrangement of the tumour (Figure 1). There are five different grades (grade 1 to 5),

representing different degrees of cell differentiation. The sum of the two most predominant grades, observed in different parts of the tumour, corresponds to the Gleason score reaching values between 2 and 10. Figure 1 shows the Gleason grades in a copy of the original drawing of Dr. Gleason, 1974 (Gleason and Mellinger, 1974).

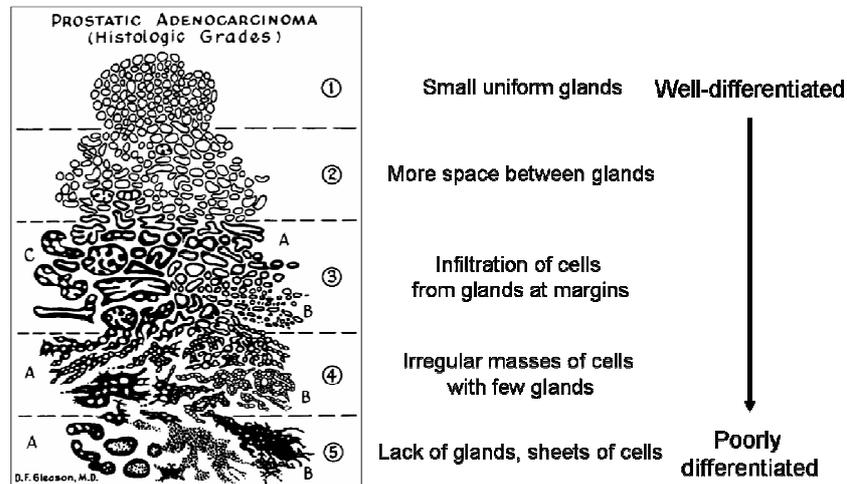


Figure 1: Gleason grades: standard drawing and the corresponding characteristics, copy of the original painting of Dr. Gleason (Gleason and Mellinger, 1974)

Nine different growth patterns were classified into five different Gleason grades. For tumour grading, the sum of the two most frequent grades, observed in different parts of the tumour, is calculated, thus, reaching values from 2 to 10.

The American Joint Committee On Cancer (= AJCC) recommends a grading into three categories. This grading is similar to the grading according to Dhom, which formerly was used in Germany and follows a grading scheme that differentiates between grade I (well-differentiated), grade II (moderately-differentiated) and grade III (poorly-differentiated) (AJCC, 2010)

Staging of prostate cancer is done according to the TNM classification following the Union for International Cancer Control (= UICC), which is based on the spread of malignant tissue and examines the primary tumour (= T), lymph nodes (= N) and metastases (= M). The T category examines the size of the primary tumour. The N category examines if there are tumour cells in the lymph nodes. Finally, the M category determines if metastases are present in distinct organs (UICC, 2009). Depending on the tumour grade and stage, various therapeutical schemes are recommended. Curative forms of prostate cancer therapy are prostatectomy and irradiation. Forms of palliative therapy are androgen ablation or application of cytostatica. In some cases, a combination of these therapies is indicated. Since there are frequent side effects, especially of prostatectomy, like incontinence and impotence, in older men with small carcinoma, a strategy of “watchful waiting” is recommended.

1.1.2 Epidemiology of prostate cancer

Excluding skin cancer, prostate cancer is the most common type of cancer in men of industrialised countries. It is the second most common cause for cancer death in men, only preceded by lung cancer. Incidence rates are higher in economically developed countries with the highest observed in North America, Australia and northern and central Europe. Lowest incidence rates can be seen in southeastern and south central Asia and northern Africa. African-American men show a higher incidence rate than white men and die more frequently because of this disease. For African-American men, from 2002 to 2006, the incidence rate was 231.9 per 100,000 and year, and the mortality rate was 56.3 per 100,000 and year. In the same time period, the incidence rate of white men for prostate cancer was 146.3 per 100,000 and year, the corresponding mortality rate was 23.6 per 100,000 and year (age-adjusted data to the 2000 U.S. population) (Edwards et al., 2010). Incidence rate in 2002 in Germany was 100 per 100,000, and the mortality rate was estimated to 20 per 100,000 (Robert-Koch Institute, Berlin, Germany, www.rki.de). Amongst ethnicity, two other established risk factors for prostate cancer development are age and familial background (Amundadottir et al., 2004; Crawford, 2003). The relative risk for prostate cancer increases more than twofold when a first-degree relative is affected by the disease and is the highest when affected relatives were diagnosed before age 60 years. It is up to five times higher when there are two affected first-degree relatives (Johns and Houlston, 2003). Twin studies can give information about the heritability and, thus, the overall contribution of inherited genes to the development of a disease (Lesko et al., 1996). Heritability is calculated with the concordance of the disease in monozygous twins that share 100 % of their genes and the concordance in dizygous twins that only share approximately 50 % of their genetic information. A cohort twin study from Sweden, Denmark and Finland identified a heritability of prostate cancer of 0.42, meaning that 42 % of prostate cancer cases can be explained by genetic factors. The heritability of prostate cancer, thus, is the highest observed for all tested types of cancer, including colorectal cancer and breast cancer (Lichtenstein et al., 2000).

Carter et al. defined hereditary prostate cancer to match at least one of the following three criteria: 1. There is a cluster of three or more prostate cancer cases in first-degree relatives; 2. There are three prostate cancer cases in three successive generations of the patient's paternal or maternal lineage; 3. There is a cluster of two first-degree relatives of the patient that diseased of prostate cancer at the age of 55 years or younger (Carter et al., 1993). Amongst familial prostate cancer cases, there are also sporadic prostate cancer cases. No clinical or pathological features allow differentiation between inherited or sporadic forms of prostate cancer. Therefore, familial and sporadic cases may result from the same mechanism (Knudson, Jr., 1971).

1.2 Search for prostate cancer susceptibility genes

Linkage studies

For the identification of chromosomal regions that harbour susceptibility genes, contributing to the development of a heritable disease, linkage analysis is performed. Families with cases are tested for cosegregation of polymorphic markers with the disease. Several linkage studies identified putative prostate cancer risk regions (Schaid, 2004):

Locus	chromosome	gene	marker
<i>HPC1</i>	1	<i>RNASEL</i>	D1S218
<i>PCAP</i>	1	-	D1S304
<i>CAPB</i>	1	-	D1S507
<i>8p22</i>	8	<i>MSR1</i>	D8S1827
<i>HPC2</i>	17	<i>ELAC2</i>	D17S969
<i>HPC20</i>	20	-	D20S839
<i>HPCX</i>	X	-	DXS8073

Genes that were discussed as possible candidate genes are *RNASEL*, *MSR1* and *ELAC2*.

The *RNASEL* gene is coding for the 2'-5'-oligoadenylate (2-5A)-dependent ribonuclease L, involved in the immune defense of viruses. It is located in the candidate region called hereditary prostate cancer 1 (= HPC1) on chromosome 1q25, which was identified by linkage analysis. Two families, one of European and one of African descent, showed cosegregation of inactivating mutations with the disease (Carpten et al., 2002). Further studies investigating the role of this gene in the development of prostate cancer resulted in controversial data, and so could not confirm the impact of the *RNASEL* gene on prostate cancer susceptibility for the German population (Maier et al., 2005). A meta-analysis of ten studies reported a genuine genetic effect of one coding variant in *RNASEL* for only a part of prostate cancer cases (Li and Tai, 2006).

Similar controversial results were received for *ELAC2* (= *elaC* homolog 2 (*E.coli*)), which is located on chromosome 17p11 in the so called linkage region hereditary prostate cancer 2 (= HPC2), and encodes the protein ribonuclease Z 2, which catalyses the removal of the 3'trailer of precursor tRNAs.

Another candidate gene *MSR1* (= macrophage scavenger receptor) is located in the linkage region 8p22, and encodes a receptor which is mostly expressed in macrophages. It binds bacteria and other ligands. After finding *MSR1* mutations more frequently in prostate cancer cases than in controls, follow-up studies revealed inconsistent results (Xu et al., 2002). It was discussed that *MSR1* might not be a major prostate cancer gene, but

rather may contribute to a higher prostate cancer risk in African Americans and may have an influence on disease severity (Rennert et al., 2005). A meta-analysis of eight studies suggested that *MSR1* does not confer major risk for prostate cancer, but may confer a moderate risk for prostate cancer in black men (Sun et al., 2006).

For the region Predisposing for Prostate Cancer (= PCAP) the gene *PCTA-1* (= Prostate Carcinoma Tumour Antigen-1) was discussed as candidate gene. But identified variants in *PCTA-1* were not functional (Maier et al., 2002). Gibbs et al. found linkage of prostate cancer-brain cancer susceptibility locus (= *CAPB*) on chromosome 1p36 in 12 families with prostate and primary brain cancers (Gibbs et al., 1999). Replication studies resulted in controversial data (Berry et al., 2000;Cancel-Tassin et al., 2001;Xu et al., 2001). Linkage region *HPC20* was identified by Berry et al. 2000 and confirmed by Cui et al. 2001 and others, but the first study remains the strongest evidence (Berry et al., 2000; Cui et al., 2001). Four genes in the chromosomal region 20q13 (*CSEIL*, *ZNF217*, *MYBL2* and *STK15*) were shown to be overexpressed in prostate cancer, but no candidate gene was identified in this region. Linkage region *HPCX* identified by linkage studies from four groups could be confirmed by several following studies. But till now, no susceptibility gene could be identified in this region.

Chang et al. reported a significant but incomplete co-segregation of two non-coding variants in *CDKN1B* on chromosome 12p encoding a cell cycle inhibitor protein with prostate cancer (Chang et al., 2004). A heterozygous knockout in mice of *CDKN1B* and *PTEN* encoding a phosphatase and located on chromosome 10q in human possesses 100 % penetrance for prostate cancer development (Trotman, 2003). Deletions or mutations of *PTEN* are observed in 10 % of primary tumours and in 63 % of metastatic prostate cancers (Cairns, 1997; Suzuki, 1998). Nevertheless, those two genes did not result in significant linkage values in genome-wide linkage analysis. Xu et al. postulated an epistatic interaction of both genes in prostate cancer development (Xu et al., 2004). The reason for the problems in confirming linkage regions is genetic heterogeneity, possibly caused by multiple incompletely penetrant prostate cancer susceptibility genes.

A further suggestive linkage signal was located on chromosome 8q24.21 (Amundadottir et al., 2006). A genome-wide linkage scan in 871 Icelandic men grouped into 323 families resulted in a suggestive linkage signal on chromosome 8q24.21 with the maximum lod score of 2.11 for the microsatellite D8S529 (Amundadottir et al., 2006).

To increase statistical power of linkage analysis, the International Consortium for Prostate Cancer Genetics (= ICPCG) performed a combined genome-wide linkage study with 1,233 families. They identified several suggestive linkage regions and one significant at 22q12 in 269 families with at least five affected members (Xu et al., 2005). The authors discussed possible false-positive results of the combined analysis because of multiple testing. But

they concluded that if there are susceptibility genes for prostate cancer, they most likely are located in regions showing significant or suggestive linkage in their study.

In summary, results of linkage studies did not identify cancer risk genes with dominant effect for prostate cancer development. In contrast to that, for the identification of common genetic variants, that confer risk for cancer development with moderate effect and low penetrance, case-control studies are performed.

Association studies

Association studies determine the frequency of single nucleotide polymorphisms (= SNPs) in unrelated (familial as well as sporadic) cases compared to controls for examination of the disease risk for variant carriers. More than 500,000 SNPs scattered over the whole genome are tested for association with a disease in genome-wide association studies, which are a powerful approach to identify common, low-penetrance risk loci. Genome-wide association studies for prostate cancer in Iceland, USA and in England identified several risk loci. An overview of loci, associated with a higher risk for prostate cancer development, is shown in Table 1 (Amundadottir et al., 2006;Eeles et al., 2008;Eeles et al., 2009;Gudmundsson et al., 2007a;Gudmundsson et al., 2007b;Gudmundsson et al., 2008;Haiman et al., 2007;Sun et al., 2008;Thomas et al., 2008;Yeager et al., 2007). On chromosome 8q24.21, three independent prostate cancer risk regions were identified.

Table 1: SNPs that showed association with prostate cancer development in genome wide association studies

a) ^a reported risk allele frequency (= RAF) in Europeans, ^b estimated per allele odds ratio from the largest available study to this time point, ^c *P* value for Armitage's trend test, from the first study reporting the replication (not necessarily the current combined evidence); adapted from (Easton and Eeles, 2008). DG8S737 correspond to a dinucleotide AC repeat; hapC consists of 14 SNP alleles; **b)** SNPs in the lower part of the table list SNPs recently identified to be associated with prostate cancer risk; MAF = minor allele frequency (Eeles et al. 2009).

a) Locus	Chromosome	SNP(s)	RAF^a	per allele OR^b	<i>P</i> value^c
2p15	2	rs721048	0.19	1.15	8 x 10 ⁻⁹
3p12	3	rs2660753	0.11	1.18	3 x 10 ⁻⁸
6q25	6	rs9364554	0.29	1.17	6 x 10 ⁻¹⁰
7q21	7	rs6465657	0.46	1.12	10 ⁻⁹
<i>JAZF1</i>	7	rs10486567	0.77	1.12	10 ⁻⁷
8q24 (region1)	8	rs1447295, DG8S737	0.10	1.62	3 x 10 ⁻¹¹
8q24 (region 3)	8	rs6983267	0.50	1.26	9 x 10 ⁻¹³
8q24 (region 2)	8	rs16901979, hapC	0.03	2.1	3 x 10 ⁻¹⁵
<i>HNF1B</i>	17	rs4430796	0.49	1.24	10 ⁻¹¹
<i>HNF1B</i>	17	rs11649743	0.80	1.28	2 x 10 ⁻⁹
17q	17	rs1859962	0.46	1.25	3 x 10 ⁻¹⁰
<i>MSMB</i>	10	rs10993994	0.40	1.25	9 x 10 ⁻²⁹
<i>CTBP2</i>	10	rs4962416	0.27	1.17	3 x 10 ⁻⁸
11q13	11	rs7931342	0.51	1.19	2 x 10 ⁻¹²
<i>KLK2/KLK3</i>	19	rs2735839	0.85	1.20	2 x 10 ⁻¹⁸
Xp11	X	rs5945619	0.36	1.19	2 x 10 ⁻⁹

Continue next page

Continued Table 1

b) Locus	Chromosome	SNP	MAF	per allele OR	P value
2p21	2	rs1465618	0.23	1.08	1.6×10^{-8}
2q31	2	rs12621278	0.06	0.75	8.7×10^{-23}
4q22	4	rs17021918	0.34	0.90	4.2×10^{-15}
4q24	4	rs12500426	0.46	1.08	1.3×10^{-11}
8p21	8	rs7679673	0.45	0.91	2.6×10^{-14}
11p15	11	rs12155172	0.20	1.05	8.8×10^{-6}
22q13	22	rs2928679	0.42	1.05	7.1×10^{-8}

1.3 The 8q24 prostate cancer risk region

The risk region on chromosome 8q24.21 was first described in linkage analysis mentioned above. For identification of the cause for this linkage signal, microsatellites and indel marker in the neighbored genomic region of chromosome 8 between positions 125 and 135 mb were genotyped in prostate cancer cases and controls resulting in the strongest association of the microsatellite DG8S737 (also mentioned in Table 1). A combined Icelandic case control study resulted in a population attributable risk (= PAR) of 11 % for this marker. DG8S737 is located in a 92 kb spanning region of linkage disequilibrium. Further investigation of the neighbored 600 kb chromosomal region resulted in significant association with prostate cancer of 37 SNPs. The risk allele of SNP rs1447295 showed the strongest association with an odds ratio (= OR) of 1.72. After significant results in genome-wide association studies in different ethnic groups, the association of several independent risk loci on chromosome 8q24, that contribute to the development of prostate cancer, were confirmed in several follow-up studies (Amundadottir et al., 2006; Freedman et al., 2006; Gudmundsson et al., 2009; Haiman et al., 2007; Schumacher et al., 2007; Severi et al., 2007; Suuriniemi et al., 2007; Yeager et al., 2009). Yeager et al. estimated a population attributable risk (= PAR) for the risk allele of rs1447295 of 9 %, and a PAR for the risk allele of rs6983267 of 21 % (Yeager et al., 2007). For the SNP rs6983267, association was also shown for other types of cancer like breast cancer, colorectal cancer, ovary cancer, bladder cancer, cancer of the kidney, thyroid cancer, cancer of the larynx and lung cancer (Easton and Eeles, 2008; Kiemeny et al., 2008; Park et al., 2008; Wokolorczyk et al., 2008). Elucidating functional variants on 8q24.21 and their functional impact on carcinogenesis, therefore, would not only contribute to a better understanding of prostate cancer development but also of other types of cancer. Al Olama et al. examined 322 SNPs in the 8q24.21 risk region for association with prostate cancer. They could show an independent association with prostate cancer risk of eight SNPs allocated to 5 blocks on 8q24.21 (see Table 2).

Table 2: Summarised results of 8 SNPs (single nucleotide polymorphisms) on 8q24.21 that were shown to be independently associated with prostate cancer risk

Data were adopted from Al Olama et al., 2009; ^a dbSNP rs number, ^b NCBI Build 36 position, ^c defined by Al Olama et al. 2009, ^d defined by Amundadottir et al., Gudmundsson et al. and Yeager et al., ^e major/minor allele, ^f minor allele frequency, ^g odds ratio, ^h confidence interval, ⁱ Cochran Armitage test for trend.

SNP rs6983561 was closely correlated with rs16901979 (see region 2 in Table 1).

SNP ^a	Location ^b	block ^c	region ^d	alleles ^e	stage	MAF ^f cases	MAF ^f controls	Per-allele OR ^g (95 % CI ^h)	P ⁱ value
rs12543663		1	-	A/C	1	0.35	0.29	1.28 (1.16-1.41)	8.8x10 ⁻⁷
127,993,841					2	0.33	0.31	1.08 (1.00-1.16)	0.037
rs10086908		1	-	T/C	1	0.26	0.30	0.80 (0.73-0.89)	3.1x10 ⁻⁵
128,081,119					2	0.28	0.30	0.87 (0.81-0.94)	0.0003
rs1016343		2	2	C/T	1	0.24	0.18	1.37 (1.23-1.53)	1.5x10 ⁻⁸
128,162,479					2	0.23	0.20	1.21 (1.12-1.31)	3.2x10 ⁻⁶
rs13252298		2	2	A/G	1	0.26	0.31	0.80 (0.73-0.89)	2.5x10 ⁻⁵
128,164,338					2	0.27	0.30	0.84 (0.78-0.90)	2.9x10 ⁻⁶
rs6983561		2	2	A/C	1	0.05	0.03	2.11 (1.65-2.71)	1.4x10 ⁻⁹
128,176,062					2	0.05	0.03	1.47 (1.21-1.79)	8.9x10 ⁻⁵
rs620861		3	3	C/T	1	0.33	0.39	0.78 (0.70-0.86)	4.7x10 ⁻⁷
128,404,855					2	0.35	0.37	0.90 (0.84-0.96)	0.002
rs6983267		4	3	T/G	1	0.58	0.49	1.42 (1.30-1.56)	1.2x10 ⁻¹³
128,482,487					2	0.57	0.51	1.26 (1.18-1.35)	1.0x10 ⁻¹¹
rs10090154		5	1	C/T	1	0.14	0.08	1.86 (1.60-2.16)	2.2x10 ⁻¹⁶
128,601,319					2	0.14	0.09	1.47 (1.31-1.66)	1.0x10 ⁻¹⁰

SNP rs6983267 in the previously described region 3 is one of these independently associated SNPs. SNP rs6983561 in region 2 was closely correlated with the previously described rs16901979 ($r^2 = 0.84$, $D' = 0.93$). By multiple regression analysis, Al Olama et al. examined that the SNP rs10090154 in region 1 was sufficient to explain all other associated variants in this region as well as rs1447295 (Bergerat and Ceraline, 2009). A linkage disequilibrium plot of 8q24.21 with prostate cancer regions 1, 2 and 3 as well as blocks 1 to 5 defined by Al Olama et al. with eight risk variants are depicted in Figure 2.

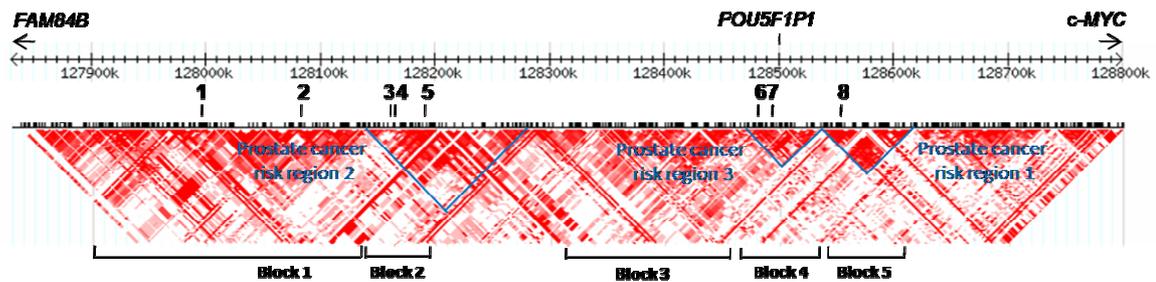


Figure 2: Linkage disequilibrium (LD) plot, based on D' , of regions or blocks on 8q24.21 harbouring SNPs associated with prostate cancer

The LD plot is based on data from CEU HAPMAP (= Haplotype Map of the Human Genome of Utah residents with Northern and Western European ancestry from the CEPH collection, CEPH = Centre d'Etude du Polymorphisme Humain), NCBI (National Center of Biotechnology Information), Build 36. The chart above indicates the position on chromosome 8q. Numeric characters on top correspond to the location of eight SNPs (single nucleotide polymorphisms on 8q24.21), 1: rs12543663, 2: rs10086908, 3: rs1016343, 4: rs13252298, 5: rs16901979, 6: rs6983267 (tagging SNP of region 3 on 8q24.21), 7: rs7837328, 8: rs1447295 (tagging SNP of region 1 on 8q24.21). Blue indicated prostate cancer risk regions were defined to be associated with prostate cancer ((Freedman et al., 2006;Haiman et al., 2007). Black indicated blocks were also reported to show association with prostate cancer (Al Olama et al., 2009). The location of the pseudogene *POU5F1P1* and the closest two genes bordering the gene desert on 8q24.21 in centromeric direction (*FAM84B*) and in telomeric direction (*c-MYC*) are indicated.

1.3.1 Putative prostate cancer risk factors on 8q24.21

Somatic amplification of 8q count to the most frequent chromosomal gains in hormone-refractory recurrent prostatic carcinoma and metastases as well as in primary prostate carcinoma (Cher et al., 1996;Nupponen et al., 1998;Visakorpi et al., 1995). 8q amplification is also seen in other types of cancer like breast cancer and bladder cancer (Kallioniemi et al., 1994;Kallioniemi et al., 1995). Especially 8q24 gain was shown to be a marker of prostate cancer progression, and to be associated with lymph node metastasis (Van Den et al., 1995). Since the prostate cancer associated regions on 8q24.21 are located in a 1.2 MB long gene desert bordered by the two genes *FAM84B* and the proto-oncogene *c-MYC*, it is difficult to identify functional variants and the mechanism, leading to the development of prostate cancer. Recent efforts mostly were focussed on *c-MYC*, and presumed putative regulatory domains affected by 8q24.21 risk variants, that might influence the transcription of *c-MYC*. *c-MYC* plays a role in the regulation of cellular proliferation, differentiation as well as apoptosis, and is deregulated in a variety of cancers (Henriksson and Luscher, 1996;Nesbit et al., 1999). Translocation to the immunoglobuline locus, and thus to a chromosomal region that shows strong activity in B cells, is common in Burkitt's lymphoma, resulting in a higher expression of the translocated *c-MYC* allele than the untranslocated *c-MYC* allele in the same cells (ar-Rushdi et al., 1983;Dalla-Favera et al., 1982). These findings led to the hypothesis that variants on 8q24.21 might function as regulators of *c-MYC* expression.

The only expressed sequence with coding capacity in the so-called gene desert on 8q24.21 corresponds to a pseudogene of *POU5F1* (= POU domain, class 5, transcription factor 1). POU represents a homeodomain protein family with the first described members Pit-1, OCT-1/2 and Unc-86. *POU5F1* is also known as *OCT4*, *OCT3*, *OTF3*, *OTF4* and is subsequently called *OCT4* (= octamer- binding transcription factor 4). *OCT4* is located on chromosome 6 and encodes a transcription factor, responsible for the self-renewal and pluripotency of stem cells in the inner cell mass of the blastocyst (Nichols et al., 1998). Recently, it was shown that human fibroblasts can be reprogrammed to induced pluripotent stem cells by introducing *KLF4*, *SOX2*, *c-MYC* and the *OCT4* gene (Takahashi and Yamanaka, 2006; Zaehres and Scholer, 2007). Introducing just *OCT4* is sufficient to induce pluripotent stem cells from adult mouse neural stem cells as well as from human fetal neural stem cells (Kim et al., 2009a; Kim et al., 2009b). *OCT4* bears potency for the maintenance and induction of pluripotency and is further involved in tumorigenesis. A fusion of *OCT4* to *EWSR1*, that activates *OCT4*, was described in hidradenoma of the skin and mucoepidermoid carcinoma of the salivary glands as well as in sarcoma derived from pelvic bone (Moller et al., 2008; Yamaguchi et al., 2005). Beyond that, it is a marker of germ cell tumours (De Jong and Looijenga, 2006). *OCT4* expression is also reported in bladder cancer, lung cancer, retinoblastoma and prostatic carcinoma (Atlasi et al., 2007; Karoubi et al., 2009; Seigel et al., 2007; Sotomayor et al., 2009).

The *OCT4* pseudogene, called *POU5F1P1* (= *POU5F1* pseudogene 1), is located in prostate cancer risk region 3 on 8q24.21 (see Figure 3). High homology (97 %) in the nucleotide sequence of *OCT4* and its pseudogene *POU5F1P1* might cause a bias in PCR amplification ((Liedtke et al., 2007). Thus, it might be possible that not *OCT4*, but its pseudogene *POU5F1P1*, is expressed in prostatic carcinoma tissue. Studies on the encoded protein POU5F1P1, that shows 95 % homology with the *OCT4* protein, revealed that the putative *POU5F1P1* encoded protein can act as transcription factor and activate a reporter gene, thus not as strong as the *OCT4* protein (Panagopoulos et al., 2008). Its location in a prostate cancer risk region and its high homology to a transcription factor, that is linked to the carcinogenesis and reprogramming of induced pluripotent stem cells, renders *POU5F1P1* a susceptibility gene for prostate cancer. Risk variants on 8q24.21 might have an influence on the transcription level of this pseudogene or, alternatively, coding variants in the open reading frame of this pseudogene could represent functional variants on 8q24.21.

Beyond annotated genes, the target of risk variants on 8q24.21 also might be non-coding RNAs. Possibly such RNAs which exert cellular functions might have an impact on the development of prostate cancer (Mattick and Makunin, 2006). To date, no functional RNA in the risk region on 8q24.21 was identified. Non-coding RNAs with unknown function,

corresponding to splice variants derived from the expressed sequence tag (= EST) AW183883 on 8q24.21, were first described by Amundadottir (Amundadottir et al., 2006). The location of all putative exons of those splice variants, the location of the pseudogene *POU5F1P1* as well as prostate cancer risk region 1, 2 and 3 on 8q24.21 are shown in Figure 3.

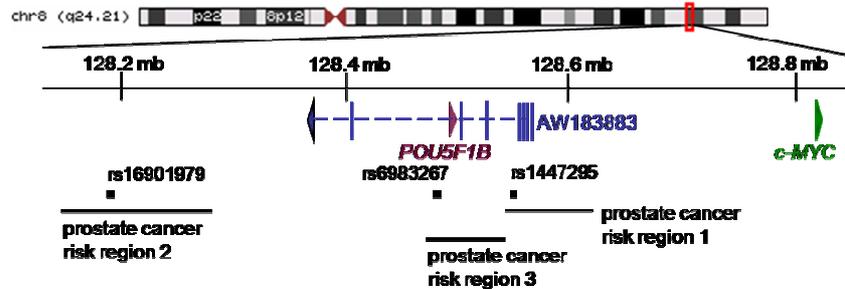


Figure 3: Overview of the location of prostate cancer risk regions as well as the pseudogene *POU5F1P1*, exons of splice variants derived from the Expressed sequence tag (= EST) AW183883 and the *c-MYC* gene on chromosome 8q24.21

Boundaries of the three regions are in correspondence of Haiman et al. according to NCBI, Build 36.3 (Haiman et al., 2007). Dots indicate the location of the prostate cancer risk variants rs16901979 in region 2, rs6983267 in region 3 and rs1447295 in region 1.

1.4 Aim of this work

Several independent association studies confirmed prostate cancer risk regions on 8q24, but to date, there are no convincing data of the functional variants and their mechanism that causes the risk for prostate cancer development.

In this work, the pseudogene *POU5F1P1*, which corresponds to the only sequence with coding capacity in the prostate cancer risk region 3, should be examined as a putative prostate cancer susceptibility gene that might be the target of prostate cancer risk variants on 8q24.21. The pseudogene *POU5F1P1* could exert similar functions as its parent gene, the stem cell factor *OCT4*, when expressed in prostatic tissue cells. It might push a process of reprogramming prostatic cells into prostatic tumour stem cells and so could be involved in carcinogenesis.

First of all, specific *POU5F1P1* expression in prostatic carcinoma derived cell lines and primary prostatic carcinoma tissue should be examined. If *POU5F1P1* is the target of prostate cancer risk variants on 8q24.21, causal variants might affect the *POU5F1P1* expression level or, alternatively, might represent coding variants in *POU5F1P1*. Haplotypes of distinct variants in *POU5F1P1* may represent functional units regarding the encoded protein that might mediate the risk for prostate cancer development.

2. Material and Methods

2.1. Material

2.1.1. Clinical specimens, tissue and cell lines

Clinical specimens for sequencing of the open reading frame of *POU5F1P1* and performance of an association study with variants in this pseudogene

DNA was previously isolated from lymphocytes from 535 unrelated prostate cancer patients (346 sporadic cases and 189 familial prostate cancer cases) as well as 213 unrelated controls served as template. Control samples represent male individuals with normal PSA levels. All samples were previously collected within the familial prostate cancer project by the Clinic of Urology and the Institute of Human Genetics. Informed consent was obtained from all probands, according to the Institutional review board of Ulm. Constellation of the patients' collective that was used for data analysis concerning tumour characteristics and age of diagnosis is shown in Table 3.

Table 3: Constellation of the sample collective for the association study concerning tumour characteristics and age of diagnosis

Shown are the parameters tumour stage, tumour grade and age of diagnosis. The character "-" indicates no available information about this parameter. For stratification, parameters were dichotomised in low and high tumour stage or tumour grade as well as early and late age of diagnosis, respectively.

Parameter	Sporadic cases, number (%)	Familial cases, number (%)	All cases, number (%)	Stratification group
Tumour stage				
-	20	9	29	
T1/2 (low)	180 (52)	108 (59)	288 (54)	low
T3/4 (high)	145 (42)	67 (36)	212 (40)	high
Tumour grade				
-	21	10	31	
G1/II (low)	238 (69)	139 (76)	377 (71)	low
GII-III/III (high)	86 (25)	35 (19)	121 (23)	high
age of diagnosis				
-	22	4	26	
< 65 (early)	168 (49)	102 (55)	270 (51)	early
>= 65 (late)	155 (45)	78 (42)	233 (44)	late

Clinical specimens for exploratory research on primary prostatic carcinoma and carcinoma surrounding tissue

Frozen primary prostatic carcinoma tissue and carcinoma surrounding prostatic tissue samples from 35 patients were originally obtained from radical prostatectomies performed at the Department of Urology, University of Ulm, Germany and was gratefully adopted from Dr. Natascha Bachmann. The mean age of diagnosis of these patients was 64.2 years (range: 51 to 73). Patients' characteristics and clinical data are described in detail in Table 4. After prostatectomy tissue was stored at -70 °C. For further processing tissue sections were embedded in Tissue-Tek® and fixed on a cork plate.

Table 4: Clinical parameters of 35 prostate cancer patients

Frozen prostatic carcinoma tissue and carcinoma surrounding tissue of 35 prostate cancer patients were obtained from radical prostatectomy. Listed here are clinical parameters with the corresponding number of cases and percentual fraction of affected patients.

Parameter	Number of cases (%)
T-stage (primary tumour)	
T ₁	1
T ₂	16 (46 %)
T ₃	15 (43 %)
T ₄	3 (9 %)
N-stage (lymph node involvement)	
N ₀	32 (91 %)
N ₁	3 (9 %)
Grade	
GII	24 (69 %)
GII-III	6 (17 %)
GIII	3 (9 %)
Not specified	2
Gleason sum	
5	6 (17 %)
6	6 (17 %)
7	18 (51 %)
8	3 (9 %)
9	1
Not specified	1
Mean age at diagnosis in years (standard deviation)	64.2 (5.5)

Clinical specimen for Immunohistochemistry experiments

Immunohistochemistry experiments were performed on paraffin embedded prostatic tissue from a familial prostate cancer patient. Patient's data were: age of diagnosis: 72 years, pathological T3a, pathological Nx, M0, grade II, Gleason score 7 (4+3).

Further tissue

Frozen TIN (intratubular germ cell neoplasia of the testis) served as positive control for Western Blot experiments and paraffin embedded seminoma (germ cell tumour of the testis) served as positive control for Immunohistochemistry experiments (both gratefully received from Prof. Dr. Möller, Department of Pathology, University of Ulm).

Cell lines

Human cells cultivated for this work are listed in Table 5.

Table 5: Cultivated cells and their origin

Cell line	Cell type origin
DU145 (Prostate cancer cell line)	brain metastasis
PC-3 (Prostate cancer cell line)	bone metastasis
LNCaP (Prostate cancer cell line)	lymph node metastasis
22Rv1 (Prostate cancer cell line)	primary prostate cancer cells, xenograft
PNT1B	immortalised nonmalignant epithelial prostate cells
NTERA-2 clone D1 (Gratefully received from Hans R. Schoeler, Muenster, Germany)	human embryonal carcinoma cell line

DNAs isolated from the following primate lymphoblastoid cells obtained from ECACC (European Collection of Cell Cultures, www.ecacc.org.uk) were gratefully obtained from Dr. Claudia Schmegner:

Macaca mulatta (Rhesus Macaque), Hylobates lar (white-handed gibbon), Pan paniscus (Bonobo), Gorilla Gorilla (Gorilla), Pongo pygmaeus (orangutan), Pan troglodytes (chimpanzee)

2.1.2. Chemicals, solutions and buffers

Chemicals used in this work are listed in Table 6.

Table 6: Chemicals and corresponding manufacturers

Chemical	Manufacturer
Acetic acid	Merck, Darmstadt, Germany
Amido black 10B	AppliChem GmbH, Darmstadt, Germany
Ampuwa water	Fresenius Kabi, Friedberg, Germany
Ammoniumpersulphate (= APS)	Bio-Rad, Hercules, CA, USA
Bromphenol blue	Merck, Darmstadt, Germany
bovine serum albumin (= BSA)	Sigma-Aldrich, St. Louis, USA
Buffer (10x) with EDTA	Applied Biosystems, Warrington, UK
Cellulose acetate membrane	Sartorius Stedium Biotech, Goettingen, Germany
Dulbecco's MEM (= DMEM) (10x)	BioChrom, Berlin, Germany
Ethylenediaminetetraacetic acid (EDTA) 1 % (for cell culture)	BioChrom, Berlin, Germany

Continue next page

Material and Methods

Continued Table 6

Chemical	Manufacturer
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, St. Louis, USA
Deoxyribonucleosidetriphosphate (dNTPs)	MBI Fermentas, St. Leon-Rot
Dithioerythritol (DTE)	Merck, Darmstadt, Germany
Dithiothreitol (DTT)	Merck, Darmstadt, Germany
Ethylenediaminetetraacetic acid (EDTA) 1 % (for cell culture)	BioChrom, Berlin, Germany
Eosin solution	Sigma-Aldrich, St. Louis, USA
Ethylene glycol tetraacetic acid (EGTA)	Sigma-Aldrich, St. Louis, USA
Ethanol	Sigma-Aldrich, St. Louis, USA
Ethidium bromide (10 mg/ml)	BioChrom, Deventer, The Netherlands
Fetal Calf Serum (= FCS)	Biochrom AG, Berlin
Formic acid, 100 %	AppliChem GmbH, Darmstadt, Germany
Glycine	Sigma-Aldrich, St. Louis, USA
Hamilton needle	Hamilton, Bonaduz, Switzerland
Hanks salt solution (10x)	BioChrom, Berlin, Germany
Harris Haematoxylin solution, Papanicolaous solution 1a	Merck, Darmstadt, Germany
HEPES	Roth, Karlsruhe, Germany
HiDi™ Formamide	Applied Biosystems, Warrington, UK
Hydrochloric acid (HCl)	Merck, Darmstadt, Germany
Hydrogen peroxidase, 30 % H ₂ O ₂	Merck, Darmstadt, Germany
Isopropanol	Fluka, Buchs, Switzerland
Kalium chloride	Merck, Darmstadt, Germany
Luminol (= 3-aminophthalhydrazide)	Fluka, Buchs, Switzerland
β-Mercaptoethanol	Serva, Heidelberg, Germany
Modified Eagle Medium (= MEM) – earle (10x)	BioChrom, Berlin, Germany
Methanol	Sigma-Aldrich, St. Louis, USA
Milk powder, blotting grade	Roth, Karlsruhe, Germany
Dinatrium-ethylenediamin-tetraacetate (Na ₂ EDTA)	BioChrom AG, Berlin, Germany
Nitrocellulose membrane (Western Blot)	GE Healthcare, Buckinghamshire, UK
p-Coumaric acid (= 4-hydroxycinnamic acid)	Sigma, St. Louis, USA
3100 POP-6™ Performance Optimized Polymer	Applied Biosystems, Foster City, USA
Proteinase inhibitor	Roche, Indianapolis, USA
SDS ultrapure	Roth, Karlsruhe, Germany
Sodium chloride (NaCl)	AppliChem GmbH, Darmstadt, Germany
Sodium hydroxide	Merck, Darmstadt, Germany
Tissue-Tek®	Sakura Finetek Europe, Zoeterwoude, NL
Trichloroacetic acid	Merck, Darmstadt, Germany
Tris ultrapure	usb corporation, Cleveland, USA
Trypsin (2.5 %, w/v)	BioChrom AG, Berlin, Germany
Tween® 20	Sigma, St. Louis, USA
Ultrapure (Chromanorm) water	VWR, Darmstadt, Germany
Whatman filter paper	Bio-Rad, Hercules, CA, USA
Hyperfilm™ ECL	GE Healthcare, Buckinghamshire, UK
Xylencyanol	Sigma, St. Louis, USA

Solutions and buffers used in this work are listed in Table 7.

Material and Methods

Table 7: Solutions and buffers with corresponding assemblies

If it is not indicated otherwise storage was at room temperature.

Solution with contents	volume / amount
Amido black destaining solution, always prepared directly before usage:	
Methanol	237.5 ml
Acetic acid	25 ml
Aqua demin	ad 500 ml
Amido black stain solution, storage at 4 °C and protected from light:	
Amido black	2.5 g
Methanol	225 ml
Acetic acid	50 ml
Aqua demin	ad 500 ml
10 % Ammoniumpersulphate (APS), storage at 4 °C:	
Ammoniumpersulphate	1 g
Aqua bidest	10 ml
Blotting buffer:	
Tris	3.03 g
Glycine	14.64 g
Methanol	200 ml
SDS	1 g
Bromphenolblue	5 % in aqua bidest
Dissolving solution, always prepared directly before usage:	
Formic acid	80 ml
Acetic acid	10 ml
Trichloroacetic acid	10 g
1x DMEM:	
10x DMEM	50 ml
Fetal bovine serum	50 ml
L-glutamine	5 ml
Gentamycine	2.5 ml
NaHCO ₃ (7.5 %)	14.6 ml
Ultrapure water	ad 500 ml
dNTP stock solution (5 mM): 1.25 mM of each of dATP, dCTP, dGTP, dTTP in ultrapure water	
DTT (2 M), aliquoted to 200 µl it was stored at -20 °C:	
DTT	308.4 mg
Aqua bidest	2 ml
ECL solution 1, storage at 4 °C protected from light (not longer than 2 days):	
Luminal stock solution	250 µl
p-coumaric acid stock solution	110 µl
1 M Tris-HCl (pH value 8.5)	2.5 ml
Aqua demin	ad 25 ml
ECL solution 2, storage at 4 °C protected from light (not longer than 2 days):	
H ₂ O ₂ (30 %)	16 µl
1 M Tris-HCl (pH value 8.5)	2.5 ml
Aqua demin	ad 25 ml
ECL use solution, storage at 4 °C:	
ECL solution 1 and ECL solution 2 were mixed 1:1 and stored protected from light until usage.	
The mixture was not stored longer than 24 hours.	

Continue next page

Material and Methods

Continued Table 7

Solution with contents	volume / amount
EDTA (50 mM):	
EDTA	186.1 mg
Aqua bidest	10 ml
EDTA (0.5 M):	
EDTA	37.2 g EDTA
Aqua bidest	200 ml
NaOH	ad pH value 8.0
EGTA solution (50 mM):	
EGTA	190.2 mg
Aqua bidest	10 ml
10x electrophoresis buffer for Western Blot:	
SDS	25 g
Glycine	360 g
Tris	75 g
Aqua demin	ad 2.5 l
Eosin solution, working solution:	
1 % in aqua bidest, adding one drop of pure acetic acid per 100 ml	
Ethidium bromide:	10 mg/ml in aqua bidest
1x HANKS:	
10x Hanks solution	50 ml
NaHCO ₃ (7.5 %)	2.35 ml
Ultrapure water	ad 500 ml
HEPES solution (1 M), prepared directly before usage:	
HEPES	2.383 g
Aqua bidest	10 ml
KCl (1 M):	
KCl	745.6 mg
Aqua bidest	10 ml
Luminol stock solution, storage at -20°C:	
3-aminophtalhydrazide	0.885 g
DMSO	ad 20m
Loading buffer for DNA and RNA electrophoresis:	
Glycerine (87 %)	1.7 m
Xylencyanol (5 % in aqua bidest)	0.25 ml
Bromphenol blue (5 % in aqua bidest)	0.25 ml
Aqua bidest	2.8 ml
NaCl (5 M):	
NaCl	2.922 g
Aqua bidest	10 ml
NaHCO ₃ (7.5 %):	
NaHCO ₃	37.5 g
Aqua bidest	500 ml
NP40 (10 % (v/v)):	
NP40	1 ml
Aqua bidest	9 ml

Continue next page

Material and Methods

Continued Table 7

Solution with contents	volume / amount
10x PBS:	
PBS	47.75 g
Aqua bidest	ad 500 ml
Sterile filtered	
p-Coumaric acid (= 4-hydroxycinnamic acid) stock solution, storage at -20 °C:	
p-coumaric acid	0.296 g
DMSO	ad 20 ml
Proteinase inhibitor (10x stock solution), storage at 4 °C:	
1.5 complete tablet was solved in 1 ml aqua bidest. Directly before usage it was diluted 1:10 with aqua bidest to get 1x working solution.	
Sample buffer for Western Blot storage at -20 °C:	
Tris	0.908 g
SDS	2.4 g
solved in aqua demin and pH value was set to 6.9.	
Glycerine	15.12 g
DTE	0.06 g
Bromphenol blue (10 mg/ml)	800 µl
Aqua demin	ad 40 ml
RPMI:	
RPMI	500 ml
Fetal bovine serum	50 ml
Gentamycine	3 ml
SE buffer:	
NaCl	75 mM
Na ₂ EDTA	25 mM
pH value 8.0	
DNA ladder (for DNA and RNA electrophoresis):	
DNA ladder	100 µl
Loading buffer	100 µl
Aqua bidest	800 µl
5x TBE buffer:	
Tris	270 g
Boric acid	137.5 g
EDTA	18.6 g
pH value 8.3	
Aqua bidest	ad 5 l
TE buffer:	
Tris	10 mM
EDTA	1 mM
pH value 8.0	
TBS (10x):	
Tris	48.44 g
NaCl	116.88 g
Aqua bidest	ad 2 l

Continue next page

Material and Methods

Continued Table 7

Solution with contents	volume / amount
1x TBS, 0.5 % Tween (v/v):	
TBS (10x)	250 ml
Tween 20	12.5 ml
Aqua bidest	ad 2.5 l
Trypsin (0.15 %) (for cell culture)	
PBS (10x)	43 ml
Trypsin (2.5 %)	30 ml
EDTA (1 %)	40 ml
NaHCO ₃ (7.5 %)	7.5 ml
Ultrapure water	ad 500 ml
Xylencyanol: 5 % in aqua bidest	

2.1.3. Enzymes and reagent kits

Enzymes used in this work are listed in Table 8.

Table 8: Enzymes and corresponding manufacturers

Enzyme	Manufacturer
Calf intestinal alkaline phosphatase (10,000u/ml)	New England BioLabs, Ipswich, MA, USA
DNase I recombinant, Rnase-free	(10 u/μl) Roche, Indianapolis, USA
Exonuclease I	USB corporation, Cleveland, Ohio, USA
Go-Taq [®] Flexi DNA Polymerase	Promega, Madison, USA
illustra [™] Taq DNA Polymerase	GE Healthcare, Buckinghamshire, UK
RNase A, DNase free	Fermentas, Leon-Rot, Germany
Superscript [™] III Reverse Transkriptase (200 u/μl)	Invitrogen, Karlsruhe, Germany
Shrimp alkaline phosphatase (1 u/μl)	Amersham Life Science Inc., Arlington Heights, USA

Reagent kits that were used in this work are listed in Table 9.

Table 9: Reagent kits and corresponding manufacturers

Enzyme	Manufacturer
Big Dye [®] Terminator v3.1 Cycle Sequencing Kit	Applied Biosystems, Warrington, UK
Dako REAL [™] EnVision [™] Detection System Code K5007	Dako, Glostrup, Denmark
GeneRacer [™] Kit	Invitrogen, Karlsruhe, Germany
illustra [™] GFX PCR DNA and Gel Band Purification Kit	GE Healthcare, Buckinghamshire, UK
illustra TempliPhi [™] HT DNA Amplification Kit	GE Healthcare, Buckinghamshire, UK
Montage [™] SEQ ₉₆ Sequencing Reaction Cleanup Kit	Millipore, Bedford, USA
QuantiTect [®] Multiplex RT-PCR Kit	Qiagen, Hilden, Germany
RNeasy Mini Kit	Qiagen, Hilden, Germany
SALSA MLPA kit P014 Chromosome 8, lot 0108	MRC Holland, Amsterdam, The Netherlands
SNaPshot [®] Multiplex Kit	Applied Biosystems, Warrington, UK
SuperScript [™] III First-Strand Synthesis System for RT-PCR	Invitrogen, Karlsruhe, Germany
Taqman Genotyping Mastermix	Applied Biosystems, Warrington, UK
TOPO TA Cloning [®] Kit for Sequencing	Invitrogen, Karlsruhe, Germany

2.1.4. Antibodies, ladders and oligonucleotides

Antibodies used in this work for Immunohistochemistry and Western Blot assays are listed in Table 10.

Table 10: Primary and secondary antibodies with corresponding manufacturers

Antibody	Manufacturer
<u>Primary antibodies:</u>	
sc-8629 Oct-3/4 (C-20) plus blocking peptide sc-8629 P	Santa Cruz Biotechnology, Santa Cruz, CA, USA
sc-8628 Oct-3/4 (N-19) plus blocking peptide sc-8628 P	Santa Cruz Biotechnology, Santa Cruz, CA, USA
sc-5279 Oct-3/4 (C-10)	Santa Cruz Biotechnology, Santa Cruz, CA, USA
ANTI- β -AKTIN, clone AC-74, (A5316), mouse monoclonal (31 mg/ml)	Sigma-Aldrich, Munich, Germany
<u>Secondary antibodies:</u>	
sc-2922, rabbit anti goat IgG-HRP	Santa Cruz Biotechnology, Santa Cruz, CA, USA
goat anti-mouse IgG+IgM (H+L)-HRP	Jackson ImmunoResearch, Westgrove, PA, USA

Ladders for different assays used in this work are listed in Table 11.

Table 11: Different ladders with corresponding manufacturers

Ladder	Manufacturer
<u>For standard gel electrophoresis:</u>	
100 bp DNA ladder (500 μ g/ml)	BioLabs, New England, USA
1 kB DNA ladder (500 μ g/ml)	BioLabs, New England, USA
Ultra low range DNA ladder	Fermentas, Leon-Rot, Germany
<u>For capillary gel electrophoresis with the ABI PRISM 3100 Genetic Analyzer:</u>	
ROX 500 HD	Applied Biosystems, Warrington, UK
GeneScan TM -120 LIZ TM Size Standard	Applied Biosystems, Warrington, UK
<u>For Western Blot assays:</u>	
Protein size standard dual colour (#161 0374)	BioRad, Hercules, CA

All primers were received from biomers.net, the biopolymer factory, Ulm, Germany. A stock solution (100 μ M in ultrapure water) was generated. Working solution was 10 μ M if not indicated otherwise.

Taqman probes were received from Applied Biosystems, Warrington, UK.

Primers for defined transcripts were designed on the following annotated sequences: *c-MYC*: Transcript ID ENST00000377970, a splice variant derived from the EST AW183883: Accession Number DQ515897, *HPRT1*: Transcript ID: ENST00000298556, *OCT4*: Transcript ID: ENST00000383334, *POU5F1P1*: Transcript ID: ENST00000391675. Primer and probe sequences as well as annealing conditions and expected product sizes are listed in Table 12. Primers which are not enlisted here were generated on the RefSeq sequence.

Material and Methods

Table 12: Primers with corresponding sequences, annealing temperature as well as expected PCR product size

Primer nomenclature	Sequence	Annealing, PCR product size
<u>PCR primer for expression of <i>POU5F1P1</i>:</u>		
POU5F1_chr8_hin	5'-GGGAAGGTGTTTCAGCC*T*AA <u>A</u> -3'	55 °C, 261 bp
POU5F1_chr8_rev	5'-CGATGTGGCTGATCTGC <u>A</u> G <u>I</u> -3'	55 °C, 261 bp
<u>The <i>POU5F1P1</i> specific nucleotides are underlined in the oligonucleotide sequences. The forward primer contains one additional mismatching base (marked with asterisks) to further increase the specificity for the pseudogene:</u>		
<u>PCR primer for expression of <i>OCT4</i>:</u>		
OCT4_chr6_ex1h	5'-GGTTGAGTAGTCCCTTCGC-3'	60 °C, 567 bp
OCT4_chr6_ex2r	5'-CACATCGGCCTGTGTATATC-3'	60 °C, 567 bp
<u>Primers for amplification of the genomic <i>POU5F1P1</i> sequence in primates and prostate cancer patients for sequencing and measurement of the allelic shift in expression:</u>		
POU8_gen_hin	5'-CACTACAGCACACTTGTCAC-3'	59 °C, 2 020 bp
POU8_gen_rev	5'-CCATCAAGTAGCACTCTAGGA-3'	59 °C, 2020 bp
<u>Primers for sequencing the genomic <i>POU5F1P1</i> sequence in primates:</u>		
POU8_seq_1h	5'-TATGACACACACAGCCATAC-3'	
POU8_seq_2h	5'-TGGAGACCTCTCAGCCTGA-3'	
POU8_seq_3h	5'-AGAAGTGGGTGGAGGAAGC-3'	
POU8_seq_4r	5'-CTTCCCTCCAACCAAGTTGC-3'	
<u>Primers for sequencing of the genomic <i>POU5F1P1</i> sequence in prostate cancer patients:</u>		
POU_n8_r	5'-GCCTGGTGAATGAGCAATT-3'	
POU8_gen_hin	5'-CACTACAGCACACTTGTCAC-3'	
POU_n1_r	5'-CGAAATCCGAAGCCAGGTGT-3'	
POU_k_hin	5'-GAGTAGTCCCTTCGCAAGC-3'	
POU_n5_r	5'-GCTTTGATGTCCTGGGACTT-3'	
POU5F1_chr8_rev	5'-CGATGTGGCTGATCTGCAGT-3'	
POU5F1_chr8_hin	5'-GGGAAGGTGTTTCAGCCTAAA-3'	
POU_k_rev	5'-GTTTGAATGCATGGGAGAGC-3'	
POU_n4_hin	5'-TATGCAAAGCAGAAACCCTCA-3'	
POU_n7_hin	5'-GACCAGTGTCTTTCTCC-3'	
POU8_gen_rev	5'-CCATCAAGTAGCACTCTAGGA-3'	
<u>Primers for nested PCR for RLM-5' RACE:</u>		
GeneRacer™ 5' Primer	5'-CGACTGGAGCACGAGGACTGA-3'	66 °C, several sizes
GeneRacer™ 5' Nested Primer	5'-GGACTGACATGGACTGAAGGAGTA-3'	66 °C, several sizes
POU_5-RACE_GSP	5'-GCTGAACACCTTCCCAAATAGAACG-3'	66 °C, several sizes
POU_5-RACE_nest1	5'-CCTGGGACTTCTCCGGGTTTGCTC-3'	66 °C, several sizes
<u>Primers for sequencing RACE PCR products:</u>		
A_Ex1_seq1h	5'-ACATGGTGAGCAACTGAGGC-3'	
A1_Ex3_seq1h	5'-CCTGACTCACCCAGGCTA-3'	
B_Ex1_seq1h	5'-GGATAAGGAACATTCACATC-3'	
C_Ex3_seq1h	5'-AGTGGTGGACAGCAGAGGAC-3'	
D_Ex3_seq1h	5'-TGCATGGAACTATTGATGC-3'	
E_Ex3_seq1h	5'-CCCTCCGTGTTTCATCCCAATC-3'	
F_Ex2_seq1h	5'-CTATGTGCAGAGTCTTCATTTTC-3'	
G_Ex2_seq1h	5'-TAAACCTGAGTGCATGGAAAC-3'	
POU5F1_chr8_1rev	5'-GCGAGAAGGCGAAATCCGA-3'	
BX108_ex1_rev	5'-TGCTTGAGTGTCTCACAAC-3'	
M13	5'-CAGGAAACAGCTATGAC-3'	
T7	5'-TAATACGACTCACTATAGGG-3'	

Continue next page

Material and Methods

Continued Table 12

Primer nomenclature	Sequence	Annealing, PCR product size
<u>Primers for semiquantitative expression analysis of c-MYC:</u>		
cmh-6FAM	FAM-5'-GACAGCAGCTCGCCCAA-3'	56 °C, 389 bp
cmr	5'-ACTCTGACACTGTCCAA-3'	56 °C, 389 bp
f_n1_HPRT1-6FAM	FAM-5'-TCCTCCTGAGCAGTCAGC-3'	56 °C, 302 bp
HPRT1_ex3_r	5'-TATTCAGTGCTTTGATGTAATC-3'	56 °C, 302 bp
<u>Primers and MGB-probes for quantitative analysis of splice variants derived from the EST AW183883 and POU5F1P1 expression:</u>		
estchr8_ex6_h3	5'-ATTTTGGACTGACGTTG-3'	60 °C, 102 bp
estchr8_ex8_r2	5'-GGCTAGTTCTGGCTGCTAC-3'	60 °C, 102 bp
p_AW183883	5'-CTCAAGTAGCCTATGGA-3'-TAMRA	60 °C
POU_n9_h	5'-TTCAGTCAACATTTAATGATGCT-3'	60 °C, 121 bp
POU_n8_r	5'-GCCTGGTGAAATGAGCAATT-3'	60 °C, 121 bp
POU5F1P1-probe	VIC-5'-ATTGCTAGTGAGCGTATGA-3'-TAMRA	60 °C
f_n1_HPRT1	5'-TCCTCCTGAGCAGTCAGC-3'	60 °C, 85 bp
r_n1_HPRT1	5'-TGGTCCATCATCACTAATCAGC-3'	60 °C, 85 bp
HPRT1-probe	VIC-5'-CTCCGTTATGGCGACCC-3'-TAMRA	60 °C
<u>Tagman® Genotyping Assays (primer probe mixes commercially available from Applied Biosystems, listed here are only probes with both SNP alleles, indicated underlined, with the respective fluorescence dyes:</u>		
For rs1447295: AGTGCCATTGGGGAGGTATGTA AAA <u>[A/C]</u> GTGCTATGGAAAAAAGCAACAGGA (VIC/FAM)		
For rs6983267: GTCCTTTGAGCTCAGCAGATGAAAG <u>[G/T]</u> CACTGAGAAAAGTACAAAGAATTTT (VIC/FAM)		
For rs16901979: GTGTTAATGATTTAGCATTACTTAT <u>[A/C]</u> TCTGGCAAATGGTATTTTTGAGATA (VIC/FAM)		
<u>Primers for genotyping the SNP rs7837328:</u>		
rs783_hin	5'-TCCAAATTAGGATAGAATTC-3'	53 °C, 200 bp
rs783_rev	5'-AGAAAAGGTATGTCACCATG-3'	53 °C, 200 bp
SNPrs7837328_r	5'-ACATACACTTAGTACTTTCT-3'	
<u>Primer for only genotyping one POU5F1P1 SNP:</u>		
POU8_k_545r	5'-CTCTCCTCAAAGCGGCAGATGGTC-3'	
<u>Primers for genotyping POU5F1P1 SNPs as well as measuring the allelic shift in expression of POU5F1P1:</u>		
POU8_k_527h	5'-ATCCTGGGGTTCTATTTG-3'	
POU8_k_640h	5'-CTCTCTCTCGGTGGAGGAAGCTGACAAC-3'	
POU8_k_684r	5'-CTCTCTCTCTCTCTCTTCGCTTTCTCTTTTCGGGC-3'	
POU8_k_712r	5'-CTCTCTCTCTCTCTCTCAGGTTGCCTCTCACTCGGTTCT-3'	
<u>PCR on mRNA sequence of POU5F1P1 for measurement of an allelic shift in expression:</u>		
POU8_seq_2h	5'-TGGAGACCTCTCAGCCTGA-3'	64 °C, 523 bp
POU5F1_chr8_rev	5'-CGATGTGGCTGATCTGCAGT-3'	64 °C, 523 bp
<u>Primer for only measuring the allelic shift in expression of POU5F1P1:</u>		
POU8_k_545h1	5'-AGGTGTTTCAGCCAAA-3'	

2.1.5. Technical equipment, further laboratory material and software

Technical equipment applied in this work is listed in Table 13.

Table 13: Technical equipment supplied by the Institute of Human Genetics, the Institute of Anatomy and Cell Biology as well as the Department of Pathology

Technical equipment	Manufacturer
ABI PRISM 3100 Genetic Analyzer	Applied Biosystems, Foster City, USA
Axiophot™ microscope	Zeiss, Oberkochen, Germany
150 ml Bottle Top Filter (sterile-filter equipment)	Becton Dickinson Heidelberg, Germany
Cell culture flasks (different sizes)	Nunc, Roskilde, Denmark
Centrifuges	Heraeus, Hanau, Germany; Eppendorf, Hamburg, Germany
Electrophoresis chambers	Central work shop, University of Ulm, Germany
7900HT Fast Real-Time PCR System	Applied Biosystems, Foster City, USA
Gel documentation device	LTT Labortechnik, Wasserburg, Germany
KY-F75U high resolution videocamera	JVC, Tokyo, Japan
Leica CM1900 cryostat	leica Microsystems Nussloch GmbH, Nussloch, Germany
microplate reader Multiskan RC	Labsystems, Helsinki, Finland
<u>Multipipettes:</u>	
Eppendorf (100 µl and 500 µl)	Eppendorf, Hamburg, Germany
Matrix Impact 2® (8 channel, 1-30 µl)	Matrix, Hudson, USA
Matrix Impact 2® (1 channel, 1-30 µl)	Matrix, Hudson, USA
NanoPhotometer	Implen, Munich, Germany
Pellet pestle, disposable for 1.5 ml tubes	Kimble Glass Inc., Vineland, NJ, USA
Pipetman	Gilson, Villiers le Bel, France
Power supply units	BioRad, Hercules, CA, USA
Single-use pipettes (different sizes)	Becton Dickinson Heidelberg, Germany
Sterican® 0.90 mm x 40 mm (needle)	Braun, Kronberg, Germany
BD Plastipak™, 1 ml (syringe)	Becton Dickinson, Heidelberg, Germany
SuperFrost® glass slides	Menzel, Braunschweig, Germany
<u>Thermocycler:</u>	
PTC-100™ thermocycler	MJ Research, Waltham, USA
Eppendorf Mastercycler gradient	Eppendorf, Hamburg, Germany
T-Gradient	Biometra, Göttingen, Germany
Thermomixer 5436	Eppendorf, Hamburg, Germany
Tissue-Tek®	Sakura, Finetek Europe, B.V. Zoeterwoude, The Netherlands
Transilluminator 302 nm	Bachofer, Reutlingen, Germany
Transmitted light microscope	Zeiss, Göttingen, Germany
96 well cell culture plate	greiner bio-one, Frickenhausen, Germany
96 well PCR plate (standard)	peqlab Biotechnologie, Erlangen, Germany
Ultrasonic device Sonoplus	Bandelin electronic, Berlin, Germany
Western blot equipment	BioRad, Hercules, CA, USA

Further laboratory material

Further laboratory material as most pipette tips, reaction tubes, single-use pipettes as well as cell culture plates were used from firms as Eppendorf (Hamburg, Germany), Greiner (Frickenhausen, Germany), Roth (Karlsruhe, Germany) and Biozym (Hameln, Germany).

Software for raw data visualisation and data analysis as well statistical method applications are listed in Table 14.

Table 14: Software with corresponding providers

Software	Provider
Coffalyser V8	MRC Holland, Amsterdam, The Netherlands
Data Collection Software v1.1	Applied Biosystems, Foster City, USA
DNASTAR Lasergene [®] 8 v8.0.2	GATC Biotech AG, Konstanz, Germany
FAMHAP18 (2008)	Tim Becker, IMBIE, Bonn, Germany
GCG 6.0	Genetics Computer Group, University of Wisconsin, Madison, USA
GeneScan Analysis v3.7	Applied Biosystems, Foster City, USA
Genotyper [®] 3.7 NT	Applied Biosystems, Foster City, USA
Haploview 4.1	Broad Institute, Cambridge, MA, USA
Microsoft Office 2003	Microsoft, Redmont, USA
SeqScape v2.5	Applied Biosystems, Foster City, USA
SDS 2.3 and RQ Manager 1.3	Applied Biosystems, Foster City, USA
Sequencing Analysis v3.7	Applied Biosystems, Foster City, USA
Statview	SAS Institute Inc., Cary, NC, USA

2.2. Methods

2.2.1. Isolation of RNA and DNA from cultivated cells and frozen tissue

Cultivating cells

Prostatic carcinoma cell lines (DU145, PC-3, LNCaP, 22Rv1), the prostatic epithelial cell line PNT1B and the human embryonal carcinoma cell line NTERA-2 were cultured in DMEM, containing 10 % FCS and Gentamycin, following standard protocols.

Preparing tissue slices with the cryotome

The chamber of the cryostat was precooled to -20 °C, the specimen head was precooled to -30 °C.

The sample embedded in Tissue-Tek[®] was frozen on the specimen disc with some drops of 0.9 % NaCl solution. It then was fixed on the specimen head and the tissue was cut in 10 µm slices. With a pair of tweezers, five to ten slices were transferred in a 1.5 ml reaction tube. The tube was precooled in case of RNA isolation, or used at room temperature for DNA isolation.

Isolation of RNA and cDNA synthesis

Isolation of RNA from cultivated cells

After washing cultivated cells with cold 1x PBS, they were treated with Trypsin, transferred into a 10 ml reaction tube and centrifuged with 1,000 rpm for 5 min. After washing the cell pellet again in cold 1x PBS, the cells were shredded with syringes (0.90 x 4 mm BL/LB, 20G x 1 ½") and RNA was extracted with the RNeasy[®] Mini Kit, according to manufacturer's instructions. DNase I digestion on column was performed with 1 µl DNase I mixed with 8 µl 10x buffer and 71 µl of RNase-free water. After an incubation time of 15 min at room temperature, columns were centrifuged for 15 sec with 10,000 rpm and subsequently washed with RW1 buffer. For the following procedure, refer to manufacturer's instructions.

Isolation of RNA from frozen tissue

For isolation of RNA from frozen tissue, five to eight tissue slices of 10 µm thickness were used for being shredded with syringes and homogenised with QIAshredder[™] columns. Further procedure was analogous to that described for cultivated cells.

To check the quality of the RNA preparation, it was separated by gel electrophoresis for visualisation of the 28S and 18S rRNA bands.

Additional DNase I digestion following RNA isolation

For the expression analysis of *POU5F1P1*, it was important to digest genomic DNA completely from RNA solution. Therefore, two additional DNase I digestion steps were performed after RNA isolation. 1 µl DNase buffer (10x) and 1 µl DNase I were added to 8 µl RNA. After mixing carefully, the sample was incubated for 15 min at 25 °C. To stop the reaction and inactivate the enzyme, 1 µl EDTA (25 mM stock solution) was added and the sample was incubated for 10 min at 65 °C. Prior to Taqman reaction, DNase I was inactivated without adding EDTA by incubation for 10 min at 95 °C.

cDNA-synthesis

Reverse transcription was performed with the SuperScript™ First Strand Synthesis System for reverse transcription polymerase chain reaction (= RT-PCR) from Invitrogen, refer to manufacturer's instructions for details.

Isolation of DNA

Isolation of DNA from cultivated cells

After pelleting cells, following standard procedure, the pellet was washed with 1 x PBS and centrifuged for 5 min with 10,000 rpm at room temperature. The pellet was resuspended in 5.1 ml SE buffer, 250 µl 20 % SDS and 25 µl proteinase K (10 mg/ml). If there were very little cells, all volumes were reduced to 50 %.

Samples were incubated at 37 °C over night, slightly shaking for denaturing proteins and for lysis of cells. The day after, 1.5 ml NaCl (6 M, sterile, room temperature) were added and the sample was vortexed for 15 sec. Cell fragments were then centrifuged for 15 min with 4,000 rpm. The DNA in the supernatant was transferred into a sterile reaction tube (50 ml).

If the supernatant was not clear, additional centrifugation was performed for 15 min with 4,000 rpm. Subsequently, the supernatant was again transferred into a sterile reaction tube (50 ml), 14 ml absolute ethanol (= EtOH) (corresponding to the twofold volume of the sample) was added and the DNA was precipitated while inverting the tube several times.

The now visible milky DNA precipitate was fished with a plastic applicator and transferred into a 1.5 ml tube with 1.2 ml 70 % EtOH. In this step, salt was washed out of the DNA. The sample was again centrifuged for 10 min with 14,000 rpm. The supernatant was discarded and the DNA pellet was dried upside down for one hour.

Depending on the amount of the DNA precipitate, 100 to 200 μ l 1x TE buffer were added, and the sample was incubated at 37 $^{\circ}$ C over night for dissolving the DNA, completely. DNA samples were kept on 4 $^{\circ}$ C for long-term storage .

Isolation of DNA from tissue

The protocol is similar to that of the isolation of DNA from cell culture, except for starting with a smaller sample volume. For each sample, a sample buffer was prepared containing 2 ml SE buffer, 200 μ l 20 % SDS, 9.8 μ l proteinase K (10 mg/ml) and 40 μ l RNaseA stock solution (10 mg/ml). Depending on the size of the tissue surface, 10 to 20 μ m thick sections of frozen prostatic tissue were used, which should not exceed a volume of 20 mg tissue. The slices were transferred in a 1.5 ml reaction tube with 200 μ l sample buffer, stored at room temperature. For homogenisation of the tissue, the sample was pestled with a pellet pestle and after transferring samples in a sterile 10 ml reaction tube, 2 ml buffer were added. Samples were incubated over night at 37 $^{\circ}$ C for cell lysis.

Further procedure corresponds to that of DNA isolation from cultivated cells. Instead of the volumes indicated there, 0.6 ml NaCl (6 M) and 5.5 ml (corresponding to the twofold volume) absolute EtOH were used.

Checking the DNA quality with gel electrophoresis

For checking the quality of the DNA, 1 μ l DNA and 9 μ l of ultrapure water were added to 2 μ l loading buffer and loaded on a 0.8 % agarose gel. The gel electrophoresis was performed with 290 V for 45 min.

2.2.2. PCR, Sequencing, SNaPshot analysis, MLPA and RACE

PCR

Standard PCR

A standard PCR reaction consisted of the following components:

Volume	Component
2.5 μ l	10x PCR buffer
4 μ l	dTNP mix (5 mM)
1.25 μ l	forward primer (5 μ M)
1.25 μ l	reverses primer (5 μ M)
0.25 μ l	illustra TM Taq DNA Polymerase (5 u/ μ l)
14.75 μ l	ultrapure water
1 μ l	template
25 μl	Total volume

A standard PCR reaction was performed with the following parameters:

Step	Temperature	Time
1	94 °C	3 min
2	94 °C	2 min
3	(annealing temperature)	1 min
4	72 °C	3 min
5	72 °C	5 min

} repeated for 35 cycles

Annealing temperature was optimised for each primer pair, separately. Time of annealing depended on the amplicon size (guideline: 1,000 bp can be synthesized in 1 min). To check the quality and quantity of the amplification, PCR reactions were separated by gel electrophoresis, following standard procedure.

Sequencing

Purifying PCR products prior to sequence reaction

Sequencing reaction is not described in detail here, but follows the principle of the Sanger method. Before preparation of the sequencing reaction, PCR products were treated with exonuclease I and shrimp alkaline phosphatase to destroy PCR primers and to eliminate not incorporated dNTPs. The reaction mix was prepared as follows:

Volume	Component
2.5 µl	exonuclease I (1 u/µl)
2.5 µl	shrimp alkaline phosphatase (1 u/µl)
5 µl	PCR product
10 µl	Total volume

The reaction was incubated at 37 °C for 45 min. Enzymes were heat inactivated by incubation for 15 min at 72 °C.

Standard sequencing reaction consisted of the following components:

Volume	Component
2.5 µl	enzyme treated PCR product
1 µl	5x buffer
1 µl	BigDye reaction mix
0.5 µl	primer (10 µM)
5 µl	Total volume

A standard sequencing reaction was performed with the following parameters:

Step	Temperature	Time
1	96 °C	2 min
2	96 °C	10 sec
3	55 °C	10 sec
4	60 °C	3 min

} repeated for 25 cycles

Sequencing reactions can be stored at 4 °C for several days, until further processing.

Purifying sequence reaction

Before analysing sequence reaction, samples had to be purified from remaining fluorescence-labelled ddNTPs. This was carried out with the Montage™ SEQ₉₆ Sequencing Reaction Cleanup Kit from Millipore or, alternatively, by ethanol precipitation. After precipitation, the pellet was diluted in formamide and denatured at 95 °C for 3 min. Samples were separated with the ABI PRISM 3100 Genetic Analyzer and sequence results were checked with the Sequencing Analysis Software v3.7.

SNaPshot analysis

SNaPshot reaction is suitable for genotyping single nucleotide variants. It is a one-base primer extension reaction where only ddNTPs are provided for elongation. Used ddNTPs are fluorescence-labelled and, thus, the incorporation of a distinct ddNTP to the 3' end of the SNaPshot primer gives information about the present SNP allele. Fluorescence labelling was as follows: ddATP: dR6G (green), ddCTP: dTAMRA™ (black), ddGTP: dR110 (blue), ddTTP: dROX™ (red).

As template for this reaction served a PCR-product, harbouring one or more variants of interest, purified from oligonucleotides and remaining nucleotides from preceding PCR reaction (see purifying PCR products prior to sequence reaction). A standard SNaPshot reaction consisted of the following components:

Volume	Component
1.5 µl	enzyme treated PCR product
2.5 µl	SNaPshot reaction mix
1 µl	SNaPshot primer(s) (2 µM each)
5 µl	Total volume

If more than one variant was genotyped in one reaction, SNaPshot primers, that differed at least 4 bases in length, were chosen.

A standard SNaPshot reaction was performed with the following parameters:

Step	Temperature	Time
1	96 °C	2 min
2	96 °C	10 sec
3	60 °C	30 sec
4	60 °C	3 min

} repeated for 25 cycles

After SNaPshot reaction, 1 µl of calf intestine phosphatase was added to each sample, which was then incubated at 37 °C for 45 min. Enzyme was then heat inactivated at 72 °C for 15 min. 1 µl SNaPshot reaction was added to 20 µl formamide and 0.3 µl LIZ size standard, samples were denatured at 95 °C for 3 min and subsequently separated with the ABI PRISM 3100 Genetic Analyzer. Data were analysed with GeneScan software.

MLPA (Multiplex Ligation-dependent Probe Amplification)

For analysing somatic 8q amplifications in the genomic DNA of the 35 prostatic carcinoma tissues compared to carcinoma surrounding tissue, MLPA experiments were performed using the SALSA MLPA kit P014 Chromosome 8 (lot 0108) from MRC Holland. The kit contained the following probes, encompassing the region of interest on 8q24: one probe harbouring *RNF139* (3MB in centromeric direction of *POU5F1P1*) and two probes harbouring *c-MYC* (300 kB in telomeric direction of *POU5F1P1*). MLPA reaction was performed with 50 to 100 ng of genomic DNA.

For detailed experimental procedure, refer to manufacturer's instructions. In general, MLPA reaction is similar to a multiplex PCR reaction, except that no genomic DNA template is amplified, but probes, that previously hybridised to the target DNA sequence. MLPA reaction consists of four steps (see Figure 12, 3.2.2). After denaturation of genomic DNA strands a probemix is added containing DNA probes complementary to the sequences of interest. The kit used in this work contained 8 probes for chromosome 8p and 24 probes for chromosome 8q as well as 9 control probes, scattered over 6 other chromosomes. Each probe exists of two parts. One part of the probe carries a PCR primer sequence X, and the other one a primer sequence Y. Additionally, there is a stuffer sequence of a distinct length for each probe in order to distinguish between targets. After hybridisation of the probes to the adjacent sequences, the two parts are ligated by a thermostable ligase. Then, all probes are amplified with the universal PCR primer pair, complementary to X and Y. Subsequently, the fragments were separated with the Genetic Analyzer ABI PRISM 3100 and analysed with the GeneScan software from Applied Biosystems. Finally, peak heights from probes of interest and control probes as well as DNA from tumour and tumour surrounding tissue were compared with the Coffalyser V8 software from MRC Holland (using the Coffalyser method for analysis).

RNA ligase-mediated rapid amplification of 5'cDNA ends (RLM-5'RACE)

For the identification of the 5'-ends of the *POU5F1P1* transcripts, the GeneRacer™ Kit from Invitrogen were used. 5'RACE experiments with RNA isolated from different prostatic cell lines (LNCaP, 22rv1, DU145) and from two tumour samples (tumour 6 and tumour 28) were performed, according to the manufacturer's instructions. After isolation of total RNA from prostatic tissue or cell lines, RNA was treated with calf intestine phosphatase (= CIP), which dephosphorylates RNA species without a 5'cap structure, including non-full-length mRNAs. In the following, tobacco acid pyrophosphatase (= TAP) was added to remove the 5'cap structure from intact mRNA, leaving a 5'phosphate. A RNA linker sequence, the GeneRacer™RNA Oligo, was ligated specifically to RNA molecules with this 5' phosphate. Subsequently, reverse transcription with random primers was performed. For PCR amplification of RACE products, a nested PCR was performed, including two separate PCR reactions. For the first PCR, primers complementary to the GeneRacer™RNA Oligo and *OCT4*-related sequences were used. Primers were digested by adding 2 units of exonuclease I to 20 µl PCR reaction and an incubation step for 45 min at 37 °C followed. Enzyme was heat inactivated by incubation for 15 min at 72 °C. The second PCR was performed with nested primers which were located some bases in 3'direction of the first ones. PCR conditions of the first PCR and the nested PCR were as follows:

Volume	Component (first PCR)	Component (second PCR)
5 µl	5x Colorless Flexi Buffer	5x Colorless Flexi Buffer
4 µl	dTNP mix (5 mM)	dTNP mix (5 mM)
1.25 µl	GeneRacer™ 5' Primer	Gene Racer™ 5' Nested Primer
1.25 µl	POU_5-RACE_GSP	POU_5-RACE_nest1
0.25 µl	GoTaq® Flexi DNA Polymerase	GoTaq® Flexi DNA Polymerase
11.25 µl	ultrapure water	ultrapure water
2 µl	cDNA from RLM-5' RACE	exonuclease treated first PCR
25 µl	Total volume	

PCRs were run with the following programme:

Step	Temperature	Time
1	94 °C	3 min
2	93 °C	2 min
3	66 °C	1 min
4	72 °C	3 min
5	72 °C	5 min

} 20 cycles for first PCR;
} 27 cycles for nested PCR

PCR products were separated by gel electrophoresis. PCR bands were excised from the gel with a scalpel and purified using the illustra™ GFX PCR DNA and Gel Band Purification Kit. Experimental procedure followed manufacturer's instructions.

RACE PCR products were cloned with the TOPO TA Cloning® Kit for Sequencing according to the manufacturer's instructions. Vectors encoding for an Ampicillin resistance gene were transformed in *E.coli* TOP10 cells with heat shock. Thus, successfully transformed bacteria were selected by cultivation on LB agar plates containing 0.1 mg/ml Ampicillin. With blue/white selection clones containing a vector without or with the insert can be distinguished (for further details, refer to manufacturer's instructions).

White clones were tested to carry the correct insert by performing a PCR with a single bacteria colony as template. PCR was performed using the primers from the nested PCR and the following reaction conditions:

Volume	Component
2.5 µl	20x PCR buffer
4 µl	dTNP mix (5 mM)
1.25 µl	Gene Racer™ 5' Nested Primer
1.25 µl	POU_5-RACE_nest1
0.4 µl	illustra™ Taq DNA Polymerase
15.6 µl	ultrapure water
a tip of	one bacteria colony
25 µl	Total volume

For plasmid preparation from bacterial template, the illustra TempliPhi™ HT DNA Amplification Kit was used. For experimental procedure, see manufacturer's instructions. The kit is designed to amplify small circular DNA with a high throughput by rolling circle amplification. Primers used for sequencing of the RACE products are listed in Table 12. Samples were analysed with the ABI PRISM 3100 Genetic Analyzer and sequences were visualised with the software Sequencing Analysis v3.7. For sequence alignment, BLAST (Basic Local Alignment Search Tool) function on the homepage of NCBI (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>) was used. The complete sequence for each clone was reproduced and edited with the help of the softwares GCG (Genetics Computer Group) and EditSeq (DNASTAR Lasergene®).

2.2.3. Specific PCR for *POU5F1P1* and *OCT4*

Amplification of the pseudogene *POU5F1P1* on chromosome 8

Primers' 3' ends were designed to match exclusively the sequence of the pseudogene *POU5F1P1* on chromosome 8, but form mismatches with all other known *POU5F1P1*-related sequences. The forward primer contained one additional mismatching base to further increase the specificity for the pseudogene. Primers' sequences are provided in Table 12. For all RT-PCR experiments, DNase I digestion was performed twice to remove all traces of genomic DNA from RNA preparations. This was necessary because the *POU5F1P1* specific primers can also amplify genomic DNA of the intron-less pseudogene. For testing success of DNase I digestion, RNA without reverse transcription was carried along as PCR template for all samples.

PCR products were sequenced to ensure specific amplification.

Amplification of the parent gene *OCT4* on chromosome 6

We used *OCT4* specific primers, the forward one located in exon 1 in the 5' untranslated region which is absent from all pseudogenes and the reverse primer located in exon 2 of the *OCT4* mRNA.

2.2.4. Semiquantitative expression analysis

After RNA isolation and cDNA synthesis of 400 ng RNA from the 35 prostatic carcinoma tissue and corresponding surrounding prostatic tissue, PCR products of *c-MYC* and the housekeeping gene *HPRT1* were amplified in a multiplex reaction. For each tissue sample, triplicates were performed.

PCR conditions were set as follows:

Volume	Component
2.5 µl	10x PCR buffer
4 µl	dTNP mix (5 mM)
1.25 µl	cmh-6FAM (1.5 µM)
1.25 µl	cmr (1.5 µM)
1.25 µl	f_n1_HPRT1-6FAM (5 µM)
1.25 µl	HPRT1_ex3_r (5 µM)
0.25 µl	illustra TM Taq DNA Polymerase (5 u/µl)
11.25 µl	ultrapure water
2 µl	cDNA
25 µl	Total volume

The PCR reaction was done with standard PCR programme conditions. Annealing temperature was set to 56 °C. PCR was performed with 26 cycles. Preceding experiments, done in our working group and not described in detail here, could show that in this cycle, PCR amplification has not yet reached saturation level, ensuring that PCR reaction was stopped in the exponential stage. This issue made it possible to compare different PCR samples with each other.

20 µl Formamide, 0.3 µl ROX size standard and 1 µl PCR product were pipetted into a 96 well plate and samples were denaturated for 2 min at 95 °C. Each of the forward primers was labelled with the fluorescent dye 6-FAM (= 6-Carboxyfluorescein). Finally, the two distinct amplicons of *c-MYC* and *HPRT1* of different sizes were separated with the Genetic Analyzer ABI PRISM 3100. The peak height is proportional to the product amount, allowing the measurement of the amount of *c-MYC* and *HPRT1* in a semi-quantitative manner. Data were visualised with the Gene Scan software and peak heights were extracted to Excel for further analysis.

For data analysis, the ratio of *c-MYC* peak height and *HPRT1* peak height was calculated. The arithmetic mean of these ratios resulting from triplicate measurements were plotted for prostatic carcinoma tissue and prostatic carcinoma surrounding tissue.

2.2.5. Quantitative expression analysis with Taqman analysis

Theoretical background

For semi-quantitative expression analysis, a one-step RT-PCR reaction with real-time analysis was performed. The principle of this procedure is schematically shown in Figure 10 and shall be briefly described here. After cDNA synthesis, a PCR was immediately performed containing alongside with PCR primers a Taqman probe, complementary to the target sequence. The Taqman probe consists of a sequence-specific oligonucleotide carrying a reporter fluorescence dye at the 5'end and a quencher dye at the 3'end which absorbs the fluorescence of the reporter dye by fluorescence resonance energy transfer (= FRET). Taqman probes hybridise to target sequences contemporaneous with primer annealing. During polymerisation, the Taqman probe is displaced from the template strand and degraded by the 5'nuclease activity of AmpliTaq Gold® DNA Polymerase, contained in the reaction mix. Separation of reporter dye and quencher dye results in fluorescence of the reporter dye. Accumulation of the PCR product is directly monitored by the increased fluorescence intensity resulting from augmenting separated reporter dyes. According to PCR amplification, an exponential increase of fluorescence intensity takes place.

Methodical procedure

In this work, the housekeeping gene *HPRT1*, which was shown to be equally expressed in prostatic carcinoma and carcinoma surrounding tissue, served as an internal control (de Kok et al., 2005). Triplicate reactions were performed for the target gene and for the housekeeping gene. The quantitative expression analysis was performed with the QuantiTect® Multiplex RT-PCR Kit from Qiagen, using the 7900HT Fast Real-Time PCR System from Applied Biosystems. Taqman probes, labelled with VIC at the 5'-end, were obtained from Applied Biosystems.

A standard Taqman reaction consisted of the following components:

Volume	Component
10.0 µl	2x QuantiTect RT-PCR Master Mix
0.2 µl	QuantiTect RT-Mix
0.8 µl	forward primer (10 µM)
0.8 µl	reverse primer (10 µM)
0.4 µl	probe (10 µM)
3.8 µl	RNase free water
4 µl	RNA (25 ng/µl)
20 µl	Total volume

Standard Taqman PCR conditions were:

Step	Temperature	Time	Reaction step
1	50 °C	20 min	reverse transcription
2	95 °C	15 min	<i>Taq</i> activation
3	94 °C	45 sec	denaturation
4	60 °C	45 sec	annealing, synthesis

For standard protocol steps 3 and 4 were repeated for 50 cycles.

Analysis of fluorescent signals was done with the SDS 2.3 software and the RQ Manager 1.2. This software plots the fluorescence signal of the real-time PCR in a log-linear plot versus the PCR cycles.

For analysis for each target sequence, a threshold of fluorescence is defined. Thus, a threshold cycle (= C_T) for each sample can be determined, in which this threshold of fluorescence is reached. To examine RNA amounts, that show stable PCR efficiencies for target and housekeeping gene, standard curves were generated.

Generation of standard curves

For measuring the expression of defined transcripts, first of all, the range of template with similar PCR efficiencies of the target transcript and the housekeeping gene was examined. Different dilutions of RNA from the cell line LNCaP served as template.

After PCR reaction, data were analysed with the SDS 2.3 software and C_T values were determined with the RQ Manager 1.2 software. For the generation of the standard curve, the mean of C_T values were plotted for the logarithm of input RNA amount. The difference of gradients from different PCR assays gives information about the similarity of PCR efficiencies of both assays.

Expression analysis of the pseudogene *POU5F1P1* in prostatic tissue

For the quantification of the *POU5F1P1* transcript and the housekeeping gene *HPRT1*, PCR reaction was performed with 60 cycles. To eliminate genomic DNA, DNase I digestion on column and, additionally, twice after RNA purification was performed. DNase I inactivation was performed for 10 minutes at 95 °C.

The 3'end of the amplicon for the quantification of *POU5F1P1* ended several nucleotides upstream of the *POU5F1P1* coding sequence. Forward primer and probe for *POU5F1P1* are located in chromosome 8 specific sequences.

For the generation of standard curves for *POU5F1P1* and the housekeeping gene *HPRT1*, following RNA dilutions and corresponding RNA amounts were used:

Dilution	Concentration	RNA amount in corresponding reaction
1:3	56 ng/μl	224
1:10	16.8 ng/μl	67.2 ng
1:30	5.6 ng/μl	22.4 ng
1:100	1.7 ng/μl	6.8 ng
1:300	0.56 ng/μl	2.24 ng
1:1,000	0.17 ng/μl	0.68 ng

Triplicate measurements for each RNA dilution were performed.

Fluorescence thresholds were set to 0.09 for *POU5F1P1* and to 0.14 for *HPRT1*, according to manufacturer's recommendation. The mean C_T values of triplicate measurements were plotted for the logarithm of input RNA amount. A linear regression curve was generated to examine the PCR efficiencies for both assays. To examine the correlation of assay efficiencies of both target gene and housekeeping gene, the differences of C_T values for different RNA amounts were plotted for input RNA amount. The gradient of the corresponding calculated linear regression curve shows the difference

in assay efficiencies for these two assays. Gradient deviation should not be greater than 0.1.

Expression analysis of splice variants derived from the EST AW183883 in prostatic tissue

For the generation of standard curves for splice variants derived from the EST AW183883 and the housekeeping gene *HPRT1* following RNA amounts were used:

Dilution	Concentration	RNA amount in corresponding reaction
none	200 ng/μl	
1:10	20 ng/μl	80 ng
1:30	6.67 ng/μl	26.68 ng
1:100	2.0 ng/μl	8.0 ng
1:300	0.67 ng/μl	2.68 ng
1:1,000	0.2 ng/μl	0.8 ng
1:3,000	0.067 ng/μl	0.27 ng
1:10,000	0.02 ng/μl	0.08 ng

Triplicate measurements were performed for each RNA dilution.

For measuring the splice variants derived from the EST AW183883, the fluorescence thresholds were set to 0.18 for the EST AW183883 and to 0.15 for *HPRT1*, according to manufacturer’s recommendation. Data analysis was performed, according to 0. Further processing is described in 0.

Data analysis of real-time PCR performance

After determination of C_T values for the target gene (*POU5F1P1* or splice variants derived from the EST AW183883) and the housekeeping gene (*HPRT1*), mean C_T values of double determinations were calculated. Afterwards, ΔC_T values were calculated as follows:

$$\Delta C_T = \text{mean of } C_T \text{ (target gene)} - \text{mean of } C_T \text{ (housekeeping gene)}$$

Finally, 2^{-ΔCT} values were calculated representing the relative *POU5F1P1* expression compared to the housekeeping gene *HPRT1* (Livak and Schmittgen, 2001). 2^{-ΔCT} values for tumour and tumour surrounding tissue were plotted in a bar diagram and patients were sorted from high to low *POU5F1P1* expression in the tumour tissue. Additionally, relative *POU5F1P1* expression in tumour and tumour surrounding tissue, respectively, were plotted as Box Plots.

2.2.6. Analysis of correlation between 8q24.21 transcription levels, risk variants and clinical aspects

Examination of correlation of eight prostate cancer risk variants on 8q24.21 with the expression of *POU5F1P1* and splice variants derived from the EST AW183883

Correlation of the expressions of *POU5F1P1* or splice variants derived from the EST AW183883 with genotypes of eight variants on 8q24.21 was examined (for location of the variants, please Figure 2, 1.3) with different statistical analysis.

If at least five patients were available for all three possible genotypes (homozygous for the non-risk allele, heterozygous and homozygous for the risk allele), a Kruskal-Wallis test was performed. If the risk allele was very rare and there were only two different groups (homozygous for the non-risk allele and heterozygous), a Mann-Whitney test was performed.

Additionally, coding variants in *POU5F1P1* were tested for correlation with the *POU5F1P1* expression level. Patients were grouped into “homozygous for the risk allele” and “at least one non-risk allele” for the SNP rs6998061, into “at least one risk allele” and “two non-risk alleles” for rs13273814, into “homozygous for the risk allele” and “heterozygous” for rs13274084 and into all three possible genotype groups for rs6998254 and rs7002225. The Mann-Whitney test was performed for two groups and the Kruskal-Wallis test was performed for three groups.

Correlation of expression levels of the pseudogene *POU5F1P1* and splice variants derived from the EST AW183883 was tested with the Spearman Rank correlation test.

Analysis of correlation of 8q24.21 transcript levels with clinical data

Putative correlation of the expression levels of 8q24.21 transcripts with clinical data of prostate cancer patients was examined by different statistical analysis. Correlation of tumour stage (grouped in low = T1/2 and high = T3/4), tumour grade (grouped in low = I/II and high = II-III/III) as well as age at first diagnosis (grouped in early = age of 64 or younger and late = age of 65 or older) with expression data was examined with the Mann-Whitney test. Correlation of Gleason score with the expression data was tested with the Kruskal-Wallis test.

2.2.7. Sequencing of the retrogene *POU5F1P1* on chromosome 8q24.21

Sequencing of *POU5F1P1* in primates

PCR amplification of the genomic *POU5F1P1* sequence was carried out with several DNAs of primates (white-handed gibbon, Bonobo, Gorilla, orangutan and chimpanzee). When the correct product size was amplified, PCR products were sequenced.

Nucleotide sequences were compared to the human *POU5F1P1* sequence by SeqScape analysis and were translated into the potential amino acid sequences with the Translate Tool of the proteomics server ExPASy (= Expert Proteom Analysis System) of the Swiss Institute of Bioinformatics (<http://au.expasy.org/>). Subsequently, primate *POU5F1P1* amino acid sequences were compared to that of the human *POU5F1P1* with protein BLAST (Basic Local Alignment Search Tool) function of NCBI (National Center for Biotechnology Information).

Sequencing of *POU5F1P1* in prostate cancer patients

For the identification of variants in the coding sequence of *POU5F1P1*, the DNAs from 95 unrelated familial prostate cancer patients were sequenced. Primers for amplification and sequencing of the genomic *POU5F1P1* are listed in Table 12.

Calculation of allele frequencies

For the calculation of the allele frequencies of the identified variants in *POU5F1P1*, the different genotypes for each SNP were counted.

n_{11} = number of counted genotypes homozygous for allele 1

n_{12} = number of counted genotypes heterozygous

n_{22} = number of counted genotypes homozygous for allele 2

The allele frequencies p for allele 1 and q for allele 2 were calculated as described in the following.

$$p = \frac{(2 \times n_{11} + n_{12})}{2 \times \text{total number of samples}}$$

$$q = \frac{(2 \times n_{22} + n_{12})}{2 \times \text{total number of samples}}$$

2.2.8. Measurement of an allelic shift in expression of *POU5F1P1* transcripts

Theoretical background

To measure an allelic shift in expression of the *POU5F1P1* transcript, the ratio of transcripts originating from both alleles in mRNA was measured. This ratio was normalised to the ratio of both alleles in genomic DNA which presumably represents a 1:1 ratio. No difference of allelic ratios between mRNA and genomic DNA results in a value of

one meaning that both alleles are transcribed at the same levels. A difference of allelic ratios between mRNA and genomic DNA results in a value different from one indicating that one allele is transcriptional more active than the other.

Methodical procedure

For determination of allelic ratios of the *POU5F1P1* transcript, SNaPshots on the five variants in *POU5F1P1* were performed. First of all, the resolution power of this method was evaluated by generating a standard curve for each of these SNPs.

Generation of standard curves

For measuring different ratios of SNP alleles in an adequate manner, a standard curve with defined SNP allele ratios for each investigated SNP in *POU5F1P1* was generated. For each standard curve, DNAs from control persons homozygous for one of the two SNP alleles were mixed in distinct ratios. For this, the concentrations of genomic DNAs first were measured and adjusted to 10 ng/μl. Nine different DNA compositions from this genomic DNAs, homozygous for allele 1 or allele 2 (subsequently called allele 1 or allele 2, respectively), were generated. Table 15 shows the SNP constellations of the samples in detail.

Table 15: DNA compositions from control samples homozygous for one of the two SNP alleles of the investigated SNPs for the generation of standard curves

Nine different compositions of SNP allele 1 and SNP allele 2 ranging from fractions of 10 % up to 90 % of each allele represented by genomic DNA of control samples homozygous for the respective SNP were generated. The total volume of each sample was 50 μl.

Sample	SNP allele 1 [μl]	SNP allele 1 [%]	SNP allele 2 [μl]	SNP allele 2 [%]
1	5	10	45	90
2	10	20	40	80
3	15	30	35	70
4	20	40	30	60
5	25	50	25	50
6	30	60	20	40
7	35	70	15	30
8	40	80	10	20
9	45	90	5	10

Each DNA composition served as template for three individual PCRs. After exonuclease I and shrimp alkaline phosphatase (= SAP) digestion, each PCR served for SNaPshot experiments, performed according to the standard protocol (described in 2.2.2) using the following primers and mixes. All primers had a concentration of 2 μM.

Primer	Reaction mix	Investigated SNP
POU8_k_527h	a	rs6998061
POU8_k_545h1	b	rs13273814
POU8_k_640h	c	rs13274084
POU8_k_684r	c	rs6998254
POU8_k_712r	c	rs7002225

As SNPs rs6998061 and rs13273814 are located very close, SNaPshot reaction was performed separately (mixes “a” and “b”) to avoid binding blockage. For genotyping rs13274084, rs6998254 and rs7002225, a multiplex SNaPshot reaction was performed (mix c).

After data analysis on the ABI PRISM 3100 Genetic Analyzer, peak heights were determined with the software GeneScan Analysis v3.7. For each DNA composition, the peak height ratio of the ancestral allele and the variant allele was calculated. For linearisation of the data, the natural logarithms of all single values as well as the arithmetic mean of the triplicates were plotted versus the percentage of allele 1 and a regression curve was generated. Standard curves only served for evaluation of resolution power of the SNaPshot method and was not used for calculation of the allelic ratio in the investigated samples.

Measurement of an allelic shift in expression of *POU5F1P1* in heterozygous samples

Patients that showed heterozygosity for at least one SNP in the open reading frame of *POU5F1P1* were selected for this experiment (for detailed genotype data, please see Table 24). For each patient, triplicate PCR reactions on genomic DNA as well as at least six single PCRs on cDNA were performed. Additionally, for evaluation of the results, two different RNA preparations were used for some randomly selected patients. The genomic *POU5F1P1* amplicon spans the whole open reading frame of *POU5F1P1* and primers are located in chromosome 8 specific sequences. For the amplification of the *POU5F1P1* mRNA sequence, an amplicon was amplified harbouring all five variants. Each PCR reaction served as template for SNaPshot experiments on all five variants (methodical procedure see generation of a standard curve, 2.2.8).

After SNaPshot analysis, data were analysed with the GeneScan Analysis v3.7 Software from Applied Biosystems. The Genotyper® 3.7 NT software was additionally used for determination of peak heights. Further data analysis was done with Microsoft Excel 2003 software and is described in detail in the following.

Data analysis for determination of an allelic shift in expression of *POU5F1P1*

A schematical overview of the methodical procedure for evaluation of an allelic shift in the expression of *POU5F1P1* is shown in Figure 23, in the results. The arithmetic mean of single peak height ratios resulting from PCR reactions on cDNA were normalised to the arithmetic mean of ratios resulting from the triplicate reactions on genomic DNA.

a allele = ancestral allele; v allele = variant allele

normalised allele ratio = $\text{RNA (a / v)} / \text{gDNA (a / v)}$

These normalised allele ratios were plotted together with the arithmetic mean of all these normalised allelic ratios for each of the five single SNPs in *POU5F1P1*. If the individual is homozygous for a SNP, there are no data. For statistical analysis of an allelic shift in expression, a one sample group test was performed in each individual that examines if the mean of all sample values for each distinct SNP is significantly different from the value one.

By definition, there is an allelic shift in expression if all SNPs, which were tested for one individual, were found to be significantly different from the value one.

2.2.9. Association study

Genotyping

Data from cases and controls of known 8q24.21 risk variants

Genotype data of the known risk variants on 8q24.21 (rs1447295 in prostate cancer risk region 1, rs16901979 in prostate cancer risk region 2 as well as rs7837328 and rs6983267 in prostate cancer risk region 3), received by Taqman analysis in prostate cancer patients and controls, were adopted from PD Dr. Christiane Maier.

Genotyping of single nucleotide variants in *POU5F1P1* in all cases and controls

Because of the high homology between *POU5F1P1* and other *POU5F1P1*-related sequences, Taqman analysis was not suitable for variants in this pseudogene. Thus, genotyping the five SNPs in the coding sequence of *POU5F1P1* in 535 prostate cancer patients (346 sporadic cases and 189 familial cases) and 213 controls had to be performed by a long-range PCR and subsequent multiplex SNaP-Shot analysis. Primers used for SNaPshot experiments are depicted in Table 12.

Material and Methods

Primers' concentrations were set as follows:

Primer	Concentration	Investigated SNP
POU8_k_527h	4 μ M	rs6998061
POU8_k_545r	2 μ M	rs13273814
POU8_k_640h	2 μ M	rs13274084
POU8_k_684r	2 μ M	rs6998254
POU8_k_712r	2 μ M	rs7002225

After capillary gel electrophoresis, SNaP-Shot results were analysed with the GeneScan software. Determination of genotypes was done with the software Genotyper.

Genotyping of known risk variants on 8q24.21 in 35 primary prostatic tissues

Genotype data of the known risk variants on 8q24 (rs1447295 in prostate cancer risk region 1, rs16901979 in prostate cancer risk region 2 and rs6983267 in prostate cancer risk region 3) in primary prostatic carcinoma surrounding tissue was performed with the 7900HT Fast Real-Time PCR System under following conditions:

Volume	Component
1.4 μ l	H ₂ O
2.5 μ l	2x Genotyping Mastermix
0.1	primer / probe mix (from Applied Biosystems)
1.0 μ l	DNA (10 ng/ μ l)
5 μl	Total volume

Taqman PCR conditions were set as follows:

Step	Temperature	Time
1	50 $^{\circ}$ C	2 min
2	95 $^{\circ}$ C	10 min
3	92 $^{\circ}$ C	15 sec
4	60 $^{\circ}$ C	1 min

} 45 cycles

Data analysis was done with the SDS 2.3 software.

Genotyping of the SNP rs7837328 in the same 35 prostate cancer patients was done by SNaPshot reaction (methodical procedure see 2.2.2). Primers are depicted in Table 12.

Determination of the Hardy-Weinberg equilibrium

To make sure that the sample collection is appropriate for doing an association study, all SNPs were tested for Hardy-Weinberg equilibrium. The assumption of the Hardy-Weinberg is a free recombination of alleles at each locus so that the frequency of genotypes remains the same over generations if there is no selection, gene migration, mutation and favoured pairing.

For an autosomal locus with two alleles 1 and 2 with corresponding allele frequencies p and q , the following relation of frequencies exists for individuals homozygous for allele 1, individuals homozygous for allele 2 and heterozygous individual:

$$p^2 * q^2 * 2pq = 1$$

Reasons for derivations from Hardy Weinberg equilibrium could be population admixture, unexpected relationships between tested individuals or small sample size.

Each of the investigated SNPs was tested with the exact test for Hardy Weinberg equilibrium in cases and controls with the DeFinetti software of Tim M. Strom and Thomas F. Wienker (online version receivable at: <http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl>).

Calculation of the linkage disequilibrium (LD)

Linkage disequilibrium (LD) describes a non-random distribution between alleles from locally close loci on the same chromosome resulting from reduced or missing recombination. If two loci show linkage disequilibrium, distinct allele combinations for these loci, called haplotypes, are more frequent than expected from free recombination. Thus, there is no independent segregation of both loci.

LD can be depicted by two different measured values: D' (LEWONTIN, 1964) and r^2 (Devlin and Risch, 1995).

The scaled linkage disequilibrium estimate D' describes if different allele combinations at a pair of loci only derived from two consecutive mutations which were necessary to create the two SNPs, resulting in a maximum of three haplotypes ($|D'| = 1$). Linkage estimate decreases below $|D'| = 1$ when a fourth haplotype, that only can be explained by recombination between the loci, becomes manifest in the population of interest.

The alternative measured value r^2 describes the reliability of the prediction for one distinct allele at one locus is, in the presence of one distinct allele at the other locus. If there is absolute linkage disequilibrium, the squared correlation coefficient r^2 reach the value 1.

Thus, both D' and r^2 can reach values between 0 and 1. The value 0 describes free recombination.

The value for the disequilibrium D can be calculated using the following contingency table:

	Locus 2, allele B	Locus 2, allele b	Sum Σ
Locus 1, allele A	p_{AB}	p_{Ab}	$p_{Ax} = p_{AB} + p_{Ab}$
Locus 1, allele a	p_{aB}	p_{ab}	$p_{ax} = p_{aB} + p_{ab}$
Sum Σ	$p_{xB} = p_{AB} + p_{aB}$	$p_{xb} = p_{Ab} + p_{ab}$	$p_{xx} = 1$

$$D = p_{AB} - p_{Ax} * p_{xB}$$

In accordance to (Devlin and Risch, 1995), the standardised value D' can be calculated:

$$D' = (p_{AB} - p_{Ax} * p_{xB}) \div D_{max}$$

The value D_{max} represents the maximum value of $(p_{AB} - p_{Ax} * p_{xB})$ that might be possible with the given allele frequencies.

If $(p_{AB} - p_{Ax} * p_{xB})$ indeed reaches the value of D_{max} , D' reaches the value 1 indicating complete linkage disequilibrium. If allele constellations at both loci are caused by random, $(p_{AB} - p_{Ax} * p_{xB})$ and D' decreases to 0 indicating no linkage disequilibrium, at all.

The value for r^2 can be calculated with D as follows:

$$r^2 = D^2 \div (p_A * p_a * p_B * p_b)$$

2.2.9.1. Examination of single *POU5F1P1* variants for association with prostate cancer

To examine how strong an allele is associated with a disease, epidemiologic studies calculate the odds ratio (OR). Each variant was treated separately in comparing the allele frequencies in cases and controls. The odds ratio (OR) represents the disease risk for a person carrying the risk allele compared to persons without the risk allele.

Calculation of the allelic odds ratio:

Shown are allele frequencies in cases or controls.

	Risk allele	No risk allele
Case	a	b
Control	c	d

$$OR = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a*d}{b*c}$$

For estimating the significance of the odds ratio, a confidence interval $1-\alpha$ is defined, with the confidence probability α usually set to 5 %, an interval where the true value is present with a probability of 95 %. If this confidence interval harbours the value 1, the result of the odds ratio is not statistically significant: Does the confidence interval not harbour the value 1, it is statistically significant. Allelic odds ratios and corresponding intervals of confidence were calculated with the help of the DeFinetti software of Tim M. Strom and Thomas F. Wienker (online version receivable at: <http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl>). With the help of this software, the Armitage's trend test, a modified χ^2 -test, which examines a codominant effect of the risk allele concerning the development of the disease, was performed, as well.

2.2.9.2. Estimation of *POU5F1P1* haplotypes and association with prostate cancer

Based on the given genotypes for the variants on 8q24, defined haplotypes representing distinct allele combinations on the same chromosome strand can be calculated. This was done with the help of maximum-likelihood estimates which are obtained with an expectation-maximisation algorithm of the software FAMHAP 18 (2008), Tim Becker, IMBIE, Bonn, Germany.

Haplotype frequencies in cases and controls were compared with the χ^2 -test with one degree of freedom, one type of likelihood-ratio test of the software FAMHAP 18.

Calculation of odds ratios for haplotypes of SNPs in the open reading frame of *POU5F1P1* between cases and controls was also done with the software FAMHAP 18 using the omnibus statistic which corresponds to a permutational analogue of likelihood ratio for case-control data. Additionally, cases were stratified dependent on characteristics as tumour stage, tumour grade, age of diagnosis and familial history. Each subgroup was compared to controls with the χ^2 -test.

2.2.10. Examination of the *POU5F1P1* protein

2.2.10.1. Immunohistochemistry

H & E (haematoxylin & eosin) staining

In order to distinguish tumour from adjacent non-tumour prostatic tissue, H & E staining was performed. This method stains the cell nuclei of the prostatic cells in blue and the cytoplasm in red. Different cell compositions allowed differentiation between both tissue compartments. The principle of this staining is that the basic substance haematoxylin interacts with the negative phosphate groups of the DNA, resulting in a blue staining of the

nucleus. The acid dye eosin interacts with the positive side groups of the proteins resulting in a red staining of the cytoplasm.

2 µm thin slices of formalin-fixed, paraffin-embedded prostatic tissue, cut with the microtome, were transferred immediately onto a SuperFrost® glass slide. Slides were kept over night at 37 °C until staining. Slides were incubated 3 min in haematoxylin solution and washed under flowing water for 3 min. Slides were then incubated for 15 to 30 min in eosin solution (1 %) and subsequently shortly washed under flowing water. Dehydration of the preparations was carried out in ascending concentrations of ethanol (80 %, 90 %, and absolute ethanol). Afterwards, slides were put in Xylo. After dehydration of the samples, they were covered with Eukitt. Glass slides were stored at room temperature and analysed with the light microscopy. Slides were checked for harbouring prostatic carcinoma tissue and surrounding prostatic tissue.

Immunostaining

Seminoma and prostatic tissue sections on glass slides were kept overnight at 37 °C. Deparaffinisation was done with Xylo three times for 5 min. Rehydration was carried out in descending concentrations of ethanol (absolute ethanol, 90 % and 80 % ethanol). After washing in aqua bidest, endogenous peroxidase activity was blocked in 3 % H₂O₂ for 10 min. Antigen retrieval was achieved by boiling sections in 10 mM sodium citrate buffer (pH value 6.0) for 20 min using a pressure cooker. After cooling samples for 20 minutes, they were washed with aqua bidest.

Tissue slices were enframed with a hydrophobic barrier using a PAP Pen (Abcam, Cambridge, UK) and samples were stored in a 1x solution of phosphate buffered saline (PBS). A polyclonal antibody against OCT4 (sc-8629, C-20, 1:50 diluted in 1x PBS), or the polyclonal antibody against OCT4 (sc-8628, N19, 1:200 diluted in 1x PBS), respectively, was transferred to the section and samples were incubated for 30 min at room temperature. After washing in 1x PBS, the samples were incubated with the secondary antibody labeled with horseradish peroxidase (1:1,000 dilution of rabbit anti-goat IgG-HRP, sc-2922, in 1x PBS) for 30 min at room temperature. Protein detection was done by incubation in diaminobenzidine (DAB) (Dako REAL™ EnVision™ Detection System) for 10 min. Colour reaction was stopped in aqua bidest. Thereafter, sections were counterstained with haematoxylin (described above). For negative control, the primary antibody was incubated with a tenfold excess of a specific blocking peptide (sc-8629P or sc-8628P, respectively) shaking over night at 4 °C before application to the tissue section. Alternatively, different dilutions of the monoclonal antibody sc-5279 ranging from 1:10 to 1:100 were tested. Detection was carried out with the Dako REAL™ EnVision™/HRP, Rabbit/Mouse (ENV) solution containing dextran coupled with peroxidase molecules and

goat secondary anti-rabbit/anti-mouse immunoglobulin. Subsequently, the diaminobenzidine (DAB) colour reaction, described above, was carried out.

After dehydration of the slices in ascending concentrations of ethanol, they were covered with Eukitt. For the microphotographs, an Axiophot™ microscope and a KY-F75U high resolution videocamera was used. Further image processing was done with the microscope software package Diskus™ V.4.50.590 (Carl H. Hilgers, Königswinter, Germany).

2.2.10.2. Western Blot

Theoretical background

The Western Blot, also known as immunoblot, is used for detection of a specific protein and measurement of its relative amount in cells or tissues. The cell lysates of samples first are separated by size via gel electrophoresis. This usually is performed by SDS-PAGE (polyacrylamid gel electrophoresis). Herefore, proteins are solved in SDS (sodium dodecyl sulphate), so that they are denatured and subsequently possess linear structure. The gel, used for this method, consists of a meshwork of tunnels with different diameters. Applying an electric field, small proteins move faster than bigger ones in this gel leading to separation of proteins by their size. After gel electrophoresis, proteins are blotted on a nitrocellulose membrane. To avoid unspecific protein binding, the membrane is saturated with milk powder solution. The membrane is then incubated with a primary antibody specific for the protein of interest. Subsequently, incubation with the secondary antibody, which binds the first one, follows. The secondary antibody is labelled with an enzyme like horseradish peroxidase. This enzyme converts a nearly colourless substrate into a coloured end product. This product can be visualised by exposing the membrane to an X-Ray film.

For size control, an external positive control sample which is known to express the protein of interest as well as a size standard is carried along. As specificity control, samples can be incubated with the antibody previously saturated with a specific blocking peptide. Here, specific binding to the membrane is inhibited.

Preparing cell lysates of cultivated cells and frozen tissue

The pellet of around one million cells or approximately 20 tissue slices of 10 µm thickness, respectively, were transferred into a pre-cooled reaction tube which was then shortly incubated in liquid nitrogen and subsequently stored at -70 °C until usage. 300 µl of 1.5x sample buffer was added to the thawed cells or tissue slices (end concentration of the buffer: 1x). The whole cell lysate of the tissue or cultivated cells, respectively, were exposed to ultrasound ten times for 1 minute, each, in order to solubilise the cell

membranes as well as the nuclear membranes. Cell lysate was vortexed and then incubation for 5 min at 95 °C under shaking followed. After shortly vortexing again, the lysate was ready for use. For longer storage, it was kept at -20 °C.

Amido black assay for determination of the protein amount

a) Generation of a standard curve

For generation of a standard curve, BSA solutions of different concentrations were prepared using 1x sample buffer (pre-warmed to 37 °C):

- 4.0 µg BSA
- 2.0 µg BSA
- 1.0 µg BSA
- 0.5 µg BSA
- 0.2 µg BSA
- 0 µg (only 1x sample buffer)

BSA solutions of different concentrations were aliquoted and stored at – 20 °C. Before usage, they were warmed up to 37 °C.

Protein quantification was performed in duplicate measurements. 5 µl of each sample were pipetted onto a cellulose acetate membrane and dried for 20 min. The membrane was then incubated for 10 min in amido black stain solution while slightly shaking. The membrane was transferred in destaining solution until the background of the membrane was white. The membrane was again dried for approximately 45 min and cut into small tapes. Each tape of membrane was transferred into a 1.5 ml reaction tube and covered with 800 µl of dissolving solution. For dissolution, the tubes were incubated for 15 min at 50 °C while shaking. The samples were vortexed and 200 µl volume of each sample were transferred in one well of a 96 well cell culture plate.

The absorption of each sample at 620 nm corresponding to the absorption maximum of amido black was measured with the microplate reader Multiskan RC using the Ascent software for iEMS Reader MF (Ascent software, London, UK).

The arithmetic means of the double determination of absorptions were plotted for the protein amount and a regression curve was generated which served as standard curve. The linear equation was used for calculation of the protein amount in the individual samples. Data of standard curve are not shown in the results.

b) Determination of the amount of protein in samples

The preparation of the samples for the amido black assay corresponds to that, described for generation of the standard curve. Each sample was tested twice. The protein amount for each sample was calculated by insertion of the arithmetic mean in the linear equation of the standard curve.

Gel preparation

Equipment assembly for gel electrophoresis was done according to manufacturer's recommendation.

The discontinual SDS-PAGE consisted of the stacking gel and the resolving gel. The components of the resolving gel (15 % acrylamid) for 4 separate gels with the size of 8.2 cm x 8.2 cm and 0.75 mm thickness are listed in the following:

Aqua bidest	4.6 ml
30 % Acrylamid mix (30 %)	10 ml
Tris (1.5 M, pH value 8.8)	5 ml
SDS (10 % solution)	200 µl
APS (= ammoniumpersulphate) (10 %)	200 µl
TEMED	20 µl

The reagents were mixed after adding SDS and again after adding APS and TEMED. The gel was immediately poured between two glass slides and covered with isopropanol. After gel hardening isopropanol was discarded and the stacking gel was poured on top.

Preparation of the stacking gel (5 % acrylamid) (for 4 separate gels):

Aqua bidest	5.5 ml
30 % Acrylamid mix	1.3 ml
Tris (1.0 M, pH value 6.8)	1 ml
SDS (10 % solution)	80 µl
APS (10 % solution)	80 µl
TEMED	8 µl

Immediately after pouring the stacking gel, a comb was placed on top of the gel between the glass slides. In the resulting slots of the hardened gel, the samples were applied, later. Gels can be stored for some days at 4 °C until usage.

Gel electrophoresis

For electrophoresis, the gel was put into an electrophoresis chamber filled with electrophoresis buffer and the comb was removed.

10 µl up to 30 µl sample volume were loaded in one slot of the polyacrylamid gel using a Hamilton needle. In one slot of each gel, 8 µl from the size standard were pipetted.

Until the samples have merged the resolving gel, a currency of 80 V was supplied. Afterwards, the currency was set to 120 V for approximately 2 h until the separation of protein size was sufficient.

Blotting

Proteins were blotted from the polyacrylamid gel onto a nitrocellulose membrane at semi-dry conditions. The blot was assembled from cathode to anode as follows: first some layers of Whatman paper, then the gel with the separated proteins, the membrane and on top again some layers of Whatman paper. All layers are soaked with blotting buffer and remaining air bubbles were removed by rolling several times over the blot assembly. The blotting conditions were 0.8 mA/1 cm² blotting area (for example a gel with an area of approximately 70 cm² was blotted at 60 mA) and a maximum of 25 V for one hour.

Ponceau S staining

After blotting, the membrane was cut to the size of the gel. It was incubated in Ponceau S solution for 5 min, and afterwards, it was washed in aqua demin until the water was clear. Ponceau S stains all proteins on the membrane in a reversible manner. The stained membrane was copied for visualisation of the approximate amount of loaded total protein and for the localisation of size standard bands on the blot.

Blocking

The membrane was blocked with Tris Buffered Saline Tween[®] 20 (= TBST), 5 % milk powder, at 4 °C over night while slightly shaking. This was done to saturate all unspecific binding sites on the membrane.

Protein detection

The membrane was incubated with a 1:500 dilution of the primary antibody in TBST, 5 % milk powder (one membrane was incubated in 7 ml TBST, 5 % milk powder with 14 µl primary antibody) for one hour at room temperature while slightly shaking. Three different antibodies were used: a polyclonal antibody from goat against the C-terminus of OCT4 (C-20), a polyclonal antibody from goat against the N-terminus of OCT4 (N-19) and a

monoclonal antibody against the C-terminus of OCT4 (C-10). Subsequently the membrane was washed three times for 10 min in TBST, 5 % milk powder.

Secondary antibodies (rabbit anti-goat or goat anti-mouse), linked to horseradish peroxidase was used, an enzyme which converts the nearly colourless substrate luminol in the coloured end product. All secondary antibodies were used in a 1:5,000 dilution in TBST, 5 % milk powder (one membrane was incubated in 7 ml TBST, 5 % milk powder with 1.4 µl secondary antibody) for one hour at room temperature while slightly shaking. Subsequently, the membrane was washed three times for 10 min in TBST, 5 % milk powder.

Specificity control

To test the specificity of antibody binding, primary antibody, saturated with a specific blocking peptide, was used. Therefore, the tenfold excess of a specific blocking peptide for the polyclonal antibody C-20 or the polyclonal antibody N-19, respectively, was added to the primary antibody in TBST, 5 % milk. For saturating antibodies with the specific blocking peptide, the solutions were incubated at 4 °C overnight while slightly shaking. Saturated antibodies were applied on the membrane instead of the primary antibody, itself. Since no specific blocking peptide was available for the monoclonal antibody C-10, specificity control was not possible for this experiment.

Protein band visualisation

Visualisation of protein bands was done with enhanced chemoluminescence (= ECL). ECL solution contained luminol which is electrochemically oxidised in the presence of hydrogen peroxide, catalysed by the enzyme peroxidase. The end product emits blue-coloured light which is exposed to an X-Ray film and can be detected as a defined band indicating the presence of the protein on the membrane. The membrane stored in TBST, 5 % milk powder, first was shortly dried with cellulose from both sides and then incubated in ECL solution for 2 min. After dripping it with cellulose again, it was exposed to an X-Ray film. Exposition times depended on the expected amount of protein and were 1 min ACTB (= beta actin) up to 40 min (POU5F1P1). Subsequently, the film was developed with the automatic developer curix 60 from Agfa, Düsseldorf, Germany.

Determination of loaded protein amount

For the determination of protein amount, loaded on the gel, the housekeeping protein ACTB (= beta actin) was detected. For this, after detection of the target protein, as described above, the membrane was blocked again and it was incubated with the primary antibody against ACTB (1:2,500 diluted in TBST, 5 % milk powder) for one hour at room

temperature while slightly shaking. After three times washing in TBST, 5 % milk powder, incubation with the secondary antibody (1:10,000 diluted in TBST, 5 % milk powder) for one hour followed and the membrane was again washed three times. Detection was done with ECL solution (described above).

3. Results

Only one open reading frame is annotated in the gene desert on chromosome 8q24.21, that harbours risk variants for prostate cancer development, in data bases like NCBI (National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/mapview) between the genes *FAM84B* and *c-MYC*. This open reading frame corresponds to a pseudogene of *OCT4*, and is called *POU5F1P1* (P1 = pseudogene 1). It is located in prostate cancer risk region 3. Functional variants on 8q24.21 might have an influence on the expression of *POU5F1P1* or, alternatively, coding variants in *POU5F1P1* that are in linkage disequilibrium with identified risk variants could be coding and thus functional variants, which might play a role for the development of prostate cancer.

3.1. Analysis of the pseudogene *POU5F1P1* in prostatic tissue

3.1.1. Examination of *POU5F1P1* conservation in primates

Sequences that possess biological function are marked by conservation in evolution. To examine if the *POU5F1P1* pseudogene on 8q24.21 could possess biological function, the time in evolution of its integration in chromosome 8 and the degree of its conservation between primates were analysed.

Data base research resulted in no *POU5F1P1* sequence at the homologous chromosome positions of mouse and Rhesus Macaque. Further, PCR with chromosome 8 specific primers bordering the *POU5F1P1* sequence on genomic DNA showed that the pseudogene *POU5F1P1* does not exist in genomic DNA of the white-handed gibbon since PCR product corresponds to approximately 250 base pairs lacking the approximately 1370 base pair pseudogene insertion (see lane 7, Figure 4). PCR with genomic DNA from Orangutan, Gorilla, Chimpanzee and Bonobo showed the expected PCR product of 1620 bp which harbours the pseudogene (see Figure 4).

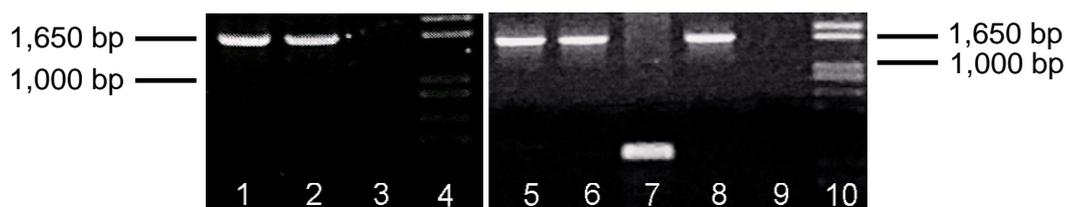


Figure 4: PCR results examining the presence or absence of the pseudogene *POU5F1P1* in some hominoids

1 = Orangutan; 2 = Gorilla; 3 = negative control; 4 = DNA ladder; 5 = Gorilla; 6 = chimpanzee; 7 = Gibbon; 8 = Bonobo; 9 = negative control; 10 = DNA ladder; Sizes of two DNA size standard bands are indicated.

PCR products from Orangutan, Gorilla, Chimpanzee and Bonobo were sequenced. Sequence results were compared to that of the human with the sequence analysis software SeqScape. Nucleotide sequences were translated into the corresponding amino acid sequence with the Translate Tool from ExPASy (<http://expasy.org>), and were compared with the human amino acid sequence by protein BLAST from NCBI. Summarised results are shown in Figure 5.

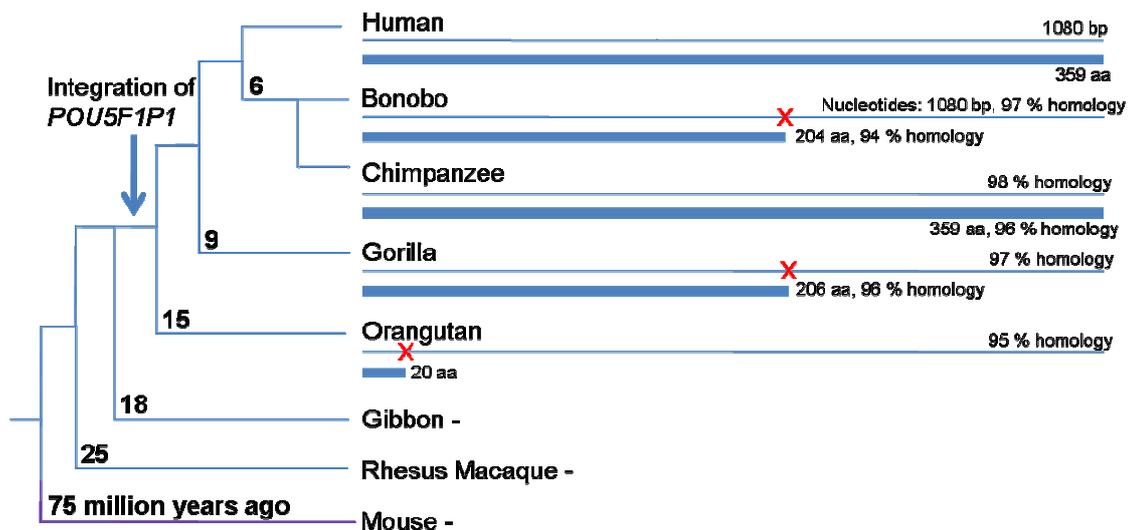


Figure 5: Schematic cladogram of the pseudogene *POU5F1P1*

Shown are the phylogenetic relations between several hominoids and mouse. The numbers on the left indicate time to the most recent common ancestor of various groups of species. As indicated in the diagram, mouse, Rhesus macaque and Gibbon do not harbour the *POU5F1P1* sequence. The homology of the *POU5F1P1* nucleotide sequence of Orangutan, Gorilla, Chimpanzee and Bonobo with the human sequence is indicated. A red cross represents the first premature stop codon in each coding sequence. The resulting putative protein length is shown as a blue bar and numbers of amino acids as well as homology of the amino acid sequence with the corresponding human sequence are also given.

The results show that Old World Monkeys do not harbour the *POU5F1P1* sequence. Thus, integration of the pseudogene *POU5F1P1* has occurred in the common ancestor from the Great Apes. No complete protein is coded by Orangutan, Gorilla and Bonobo since all carry premature stop codons. The only primate, harbouring the whole coding sequence of *POU5F1P1* in its genome, is the chimpanzee, showing a 98 % homology in the nucleotide sequence and a 96 % homology in the amino acid sequence of the putative protein compared to the human sequences. Since *POU5F1P1* does not show any evolutionary conservation, it probably does not bear any essential biological function. But if this pseudogene is extraordinarily expressed in the prostate and especially in prostatic carcinoma tissue, it might bear a gain of function according to its parent gene.

3.1.2. *POU5F1P1*-related sequence homology within the human genome

For being able to establish specific PCR assays for *POU5F1P1*, sequence homology of the *POU5F1P1* pseudogene with its parent gene and other *OCT4*-related sequences in the human genome was analysed by BLAST (Basic Local Alignment Search Tool) of the genome browser NCBI. As shown in the schematic overview of BLAST results in Figure 6, there are six pseudogenes of *OCT4* (pseudogenes 2 to 7) which showed sequence homology to *POU5F1P1*. Each of the pseudogenes resulted of an independent integration event, possessing no sequence homology to each other outside of the *OCT4*-related sequence. The grey bars in Figure 6 adjoining the *POU5F1P1* sequence represent chromosome 8 specific sequences which are not present on other chromosomes. A further *OCT4* pseudogene is annotated on chromosome 17q25, but sequence homology to *POU5F1P1* is not high enough to get positive results in a genome-wide BLAST of *POU5F1P1*.

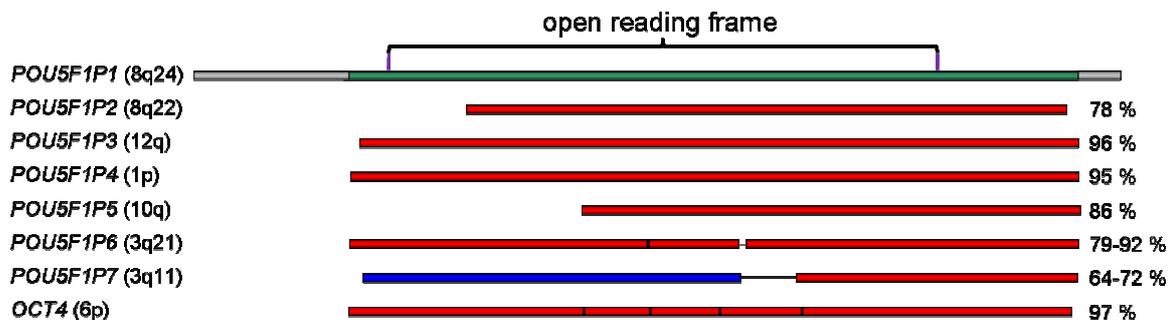


Figure 6: Schematic overview of the BLAST (= Basic Local Alignment Search Tool) results of the *POU5F1P1* genomic sequence with all other *POU5F1/OCT4*-related sequences

Shown as green bar is the whole length of the annotated *POU5F1P1* pseudogene, sequences of all other *POU5F1* pseudogenes P2 to P7 (= pseudogene 2 to pseudogene 7) as well as the parent gene *OCT4* are depicted underneath. Red bars indicate high homology (over 70 %), blue bars indicate homology lower than 70 % to *POU5F1P1*. The *POU5F1P1* open reading frame is indicated as a bracket on top. Grey bars represent genomic flanking sequences of *POU5F1P1* only present on chromosome 8.

The pseudogene *POU5F1P1* shows the highest homology in nucleotide sequence (97 %) with the parent gene *OCT4* compared to all other *OCT4* pseudogenes. Thus, for the investigation of *POU5F1P1*, the establishment of a specific PCR assay was necessary.

3.1.3. Specific PCRs on *OCT4* and *POU5F1P1*

To be able to distinguish between *OCT4* and *POU5F1P1* expression, specific PCR reactions for both targets were established. To validate the specificity of the assays, RNA from NTERA-2 cells, an embryonic carcinoma cell line, which was reported to express *OCT4* but not *POU5F1P1*, and genomic DNA, from which the intronless *POU5F1P1* can be amplified, was used. With the designed primer pairs, *OCT4* specific amplification was

only observed in NTERA-2 cDNA, while *POU5F1P1* specific products were only obtained with genomic DNA. Sequencing of these products resulted in *OCT4* or *POU5F1P1* specific sequences, respectively, with no admixture of other *POU5F1P1*-related sequences. Hence, primers used for these assays showed a high specificity for their dedicated template sequences.

To analyse the expression of *POU5F1P1* and *OCT4* in prostatic cell lines, RT-PCR was performed with cDNA from five cell lines derived from prostatic tissue (PNT1B) or from prostatic carcinoma tissue (DU145, PC3, LNCaP, 22Rv1). *OCT4* specific RT-PCR assays resulted in a very weak signal with cDNA from only one cell line, DU145 (Figure 7 a, lane 2), whereas no products were obtained with cDNA from the remaining four cell lines. *POU5F1P1* specific assays, using the same cDNA preparations as template, gave positive results with cDNA from four out of five investigated cell lines (Figure 7 b)). To make sure that *POU5F1P1* PCR products do not result from an amplification of genomic DNA contaminations, *POU5F1P1* specific PCR reactions were performed on the corresponding RNA without reverse transcription. None of these reactions resulted in PCR amplification products (data not shown). For this reason, PCR products shown in Figure 7, indeed, arise from cDNA amplification representing *POU5F1P1* expression.



Figure 7: Expression of a) *OCT4* or b) *POU5F1P1* in different prostatic cell lines

Different lanes correspond to the following samples, 1: DNA size standard, 2: DU145 cDNA, 3: PC3 cDNA, 4: LNCaP cDNA, 5: PNT1B cDNA, 6: 22Rv1 cDNA, 7: NTERA-2 cDNA, served as positive control for *OCT4*, 8: genomic DNA, served as positive control for *POU5F1P1*, 9: negative control without any template. Sizes of two DNA size standard bands are indicated. PCR band of *OCT4* corresponds to 567 bp, PCR band of *POU5F1P1* corresponds to 261 bp.

Performing the PCR specific for *POU5F1P1* on cDNA from frozen prostatic carcinoma tissue and carcinoma surrounding prostatic tissue of 35 patients resulted in PCR bands in 26 prostatic carcinoma samples and 13 samples of carcinoma surrounding tissue. As an example, Figure 8 b) shows the results for prostatic carcinoma and corresponding carcinoma surrounding tissue from three patients. *POU5F1P1* shows an expression in all three carcinoma tissues and no or a weak expression in carcinoma surrounding tissue. PCR reaction with RNA without reverse transcription as template showed no amplification product. There is no signal for *OCT4* in none of the tissues of these three patients (Figure 8 a). None of the other 35 patients showed any *OCT4* expression in prostatic carcinoma tissue or carcinoma surrounding tissue (data not shown).

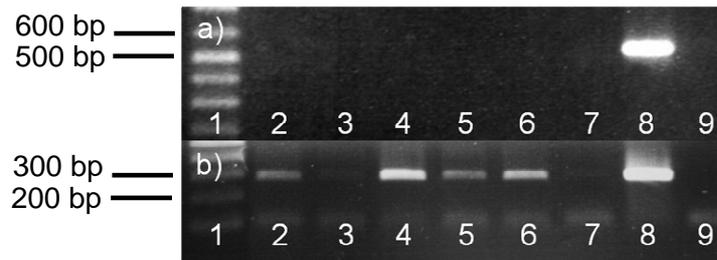


Figure 8: Expression of a) *OCT4* or b) *POU5F1P1* in prostatic carcinoma tissue and carcinoma surrounding prostatic tissue (= PT) from three prostate cancer patients

The different lanes correspond to the following samples: 1: tumour 33, 2: PT 33, 3: tumour 34, 4: PT 34, 5: tumour 35, 6: PT 35, 7 a): NTERA-2, served as positive control for *OCT4*, 7 b): genomic DNA served as positive control for *POU5F1P1*, 8: negative control without any template. Sizes of two DNA size standard bands are indicated. PCR band of *OCT4* corresponds to 567 bp, PCR band of *POU5F1P1* corresponds to 261 bp.

These results clearly demonstrate that *POU5F1P1*, but not *OCT4* is expressed in some cell lines derived from prostatic carcinoma tissue as well as in prostatic carcinoma tissue and surrounding prostatic tissue.

3.1.4. Quantitative expression analysis of *POU5F1P1*

For quantification of the *POU5F1P1* transcription relative to that of the housekeeping gene *HPRT1* in the examined frozen tissues, real-time PCR analysis was performed, using Taqman probes and the QuantiTect® Multiplex RT-PCR Kit from Qiagen. For data analysis, a fluorescence threshold was defined. The cycle, in which this threshold is passed by a sample's real-time amplification plot, is defined as threshold cycle (= C_T) value. For semi-quantitative expression analysis, the C_T value of the target gene has to be compared with that of a housekeeping gene of the same RNA preparation.

When the expression of a target gene is normalised to the expression of a housekeeping gene, PCR efficiencies of both assays have to be similar in the range of analysed template amount to avoid any bias of expression analysis. To determine the range of template amount, where PCR efficiencies of the target gene assay and the housekeeping gene assay are stable, standard curves were generated, using six different amounts of LNCaP RNA as template (covering RNA amounts from 0.672 ng to 224 ng). PCR reactions were performed in triplicate for each RNA dilution for *POU5F1P1* as well as the housekeeping gene *HPRT1*. The mean of C_T values of triplicate measurements for *POU5F1P1* and *HPRT1* were plotted for the logarithm of the RNA amounts reaching from $\log(0.672 \text{ ng}) = -0.17$ to $\log(224 \text{ ng}) = 2.35$. Figure 9 shows this plot with standard deviations as well as the corresponding linear regression curves.

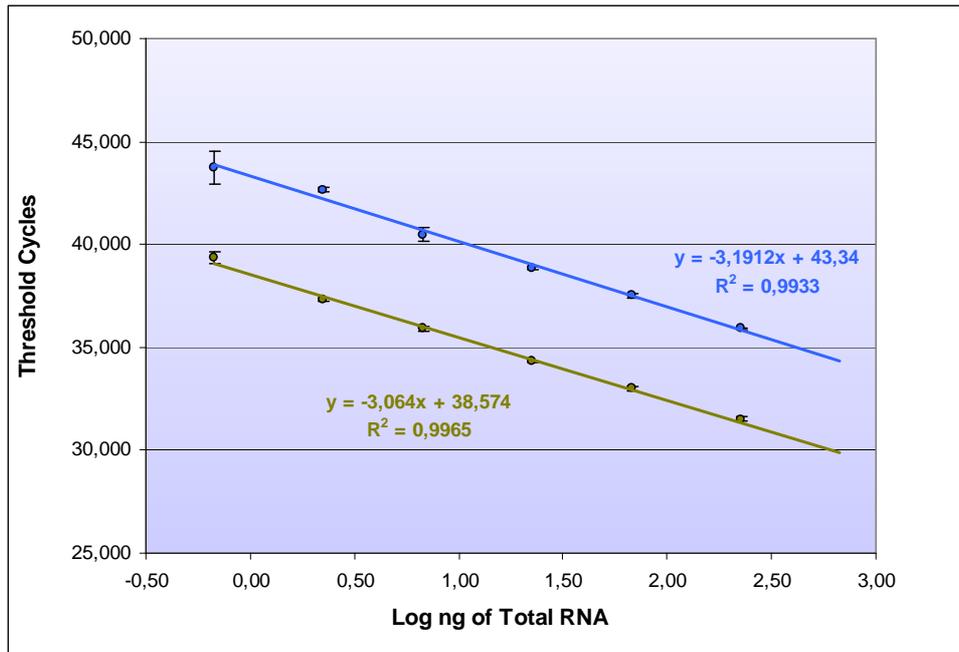


Figure 9: Plot of the threshold cycles for the logarithm of different RNA amounts for *HPRT1* and *POU5F1P1*

Mean C_T values of triplicate measurements for *HPRT1* are plotted as green dots and mean C_T values of triplicate measurements for *POU5F1P1* are plotted as blue dots against the logarithm of RNA amount in ng reaching from $\log(0.672 \text{ ng}) = -0.17 \text{ ng}$ to $\log(224 \text{ ng}) = 2.35 \text{ ng}$. Also shown are corresponding standard deviations. Linear regression curves are depicted in the corresponding colour.

Gradients of the standard curves correspond to PCR efficiency and the difference of gradients represents the difference in PCR efficiency of both assays. The difference of the gradients of both standard curves from *POU5F1P1* and *HPRT1* were calculated. The difference of both PCR assay efficiencies is 0.1272 and is thus weakly higher than the recommended 0.1 which is recommended for reliable expression analysis.

POU5F1P1 expression was measured relative to that of the housekeeping gene *HPRT1*. Taqman reaction was performed in independent sample reactions for *POU5F1P1* and *HPRT1*, performing double determinations for each of them. A schematic overview of methodical procedure and subsequent data analysis is shown in Figure 10.

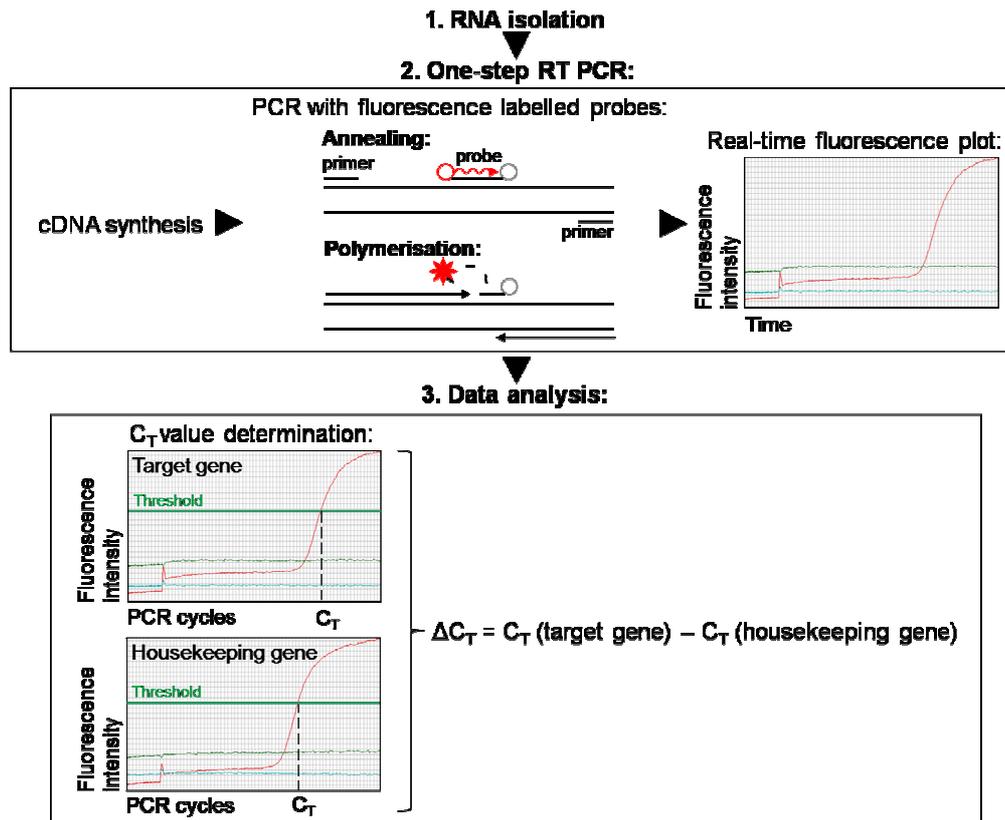


Figure 10: Schematical overview of the methodical procedure of Taqman reaction and subsequent data analysis

After RNA isolation, one-step RT-PCR was performed. The reaction mix contains PCR primers and a fluorescence-labelled probe which is able to hybridise to the PCR products (fluorescence dye indicated as red circle). Fluorescence emission is blocked by FRET (= fluorescence resonance energy transfer) by a quencher (indicated as black circle). During DNA-polymerisation, the probe is degraded and the dye emits its fluorescence (indicated as red star). Real-time increase of fluorescence is monitored by scanning samples after each PCR cycle. Data analysis is performed by examining that PCR cycle in which a previously defined fluorescence threshold is reached, the so-called C_T (= threshold cycle). The ΔC_T value corresponding to the difference of C_T values of target gene and housekeeping gene was calculated and used for further analysis.

After determination of C_T values for double determinations of the target gene (*POU5F1P1*) as well as the housekeeping gene (*HPRT1*) and calculation of the C_T mean value for each assay, the ΔC_T value was calculated as follows:

$$\Delta C_T = \text{mean of } C_T (\text{target gene}) - \text{mean of } C_T (\text{housekeeping gene})$$

Finally, $2^{-\Delta C_T}$ values were calculated, representing relative *POU5F1P1* expression compared to the housekeeping gene *HPRT1*. $2^{-\Delta C_T}$ values for tumour and tumour surrounding tissue were plotted in a bar diagram, and patients were sorted from high to low *POU5F1P1* expression in the tumour tissue. As shown in Figure 11 a), 32 out of 35 patients showed a higher *POU5F1P1* expression in the tumour than in surrounding prostatic tissue. In three patients the *POU5F1P1* expression is lower in prostatic carcinoma tissue than in carcinoma surrounding tissue. In Figure 11 b) Box Plots of

relative *POU5F1P1* expressions in tumour and tumour surrounding tissue, respectively, are shown.

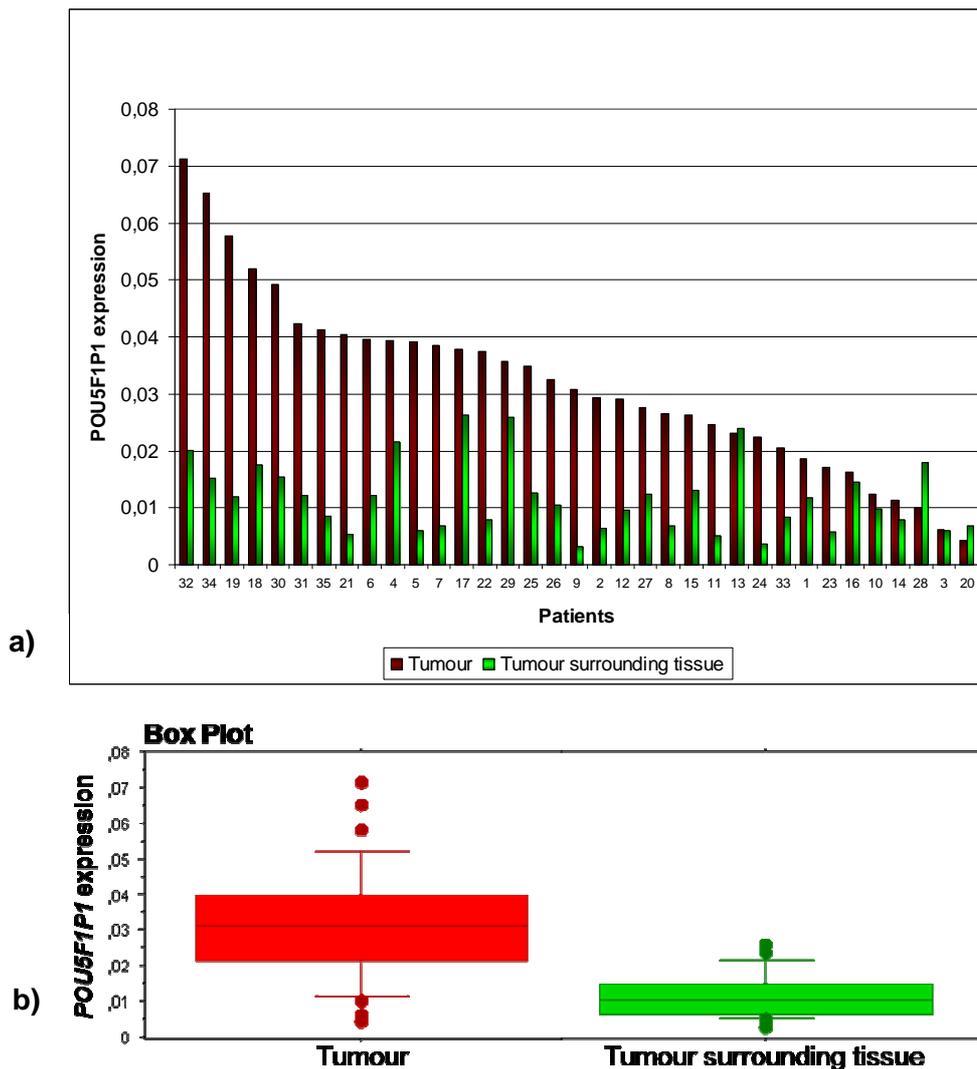


Figure 11: Plot of the relative expression of the pseudogene *POU5F1P1* compared to *HPRT1* in prostatic tumour and tumour surrounding tissue

Shown in **a)** are the $2^{-\Delta Ct}$ values corresponding to the relative expression of *POU5F1P1* compared to the housekeeping gene *HPRT1* in prostatic tumour (red bars) and tumour surrounding tissue (green bars). Patients are sorted from high to low *POU5F1P1* expression in prostatic tumour tissue. 32 of the investigated 35 patients showed a higher *POU5F1P1* expression in prostatic tumour tissue than in tumour surrounding tissue. Tissues of three patients showed results the other way round. Shown as Box Plots in **b)** is the relative expression of *POU5F1P1* compared to *HPRT1* in the tumour (red) and in tumour surrounding tissue (green), respectively, as Box Plots. The boxes contain the values between the 25th and 75th percentile as well as the median indicated by the line in the box. The upper short line indicates the 90th percentile and the short line under the box indicates the 10th percentile of the values. Dots show values over the 90th percentile and under the 10th percentile, respectively. There is a significant overexpression of *POU5F1P1* in prostatic tumour compared to surrounding prostatic tissue (Student's paired t-test, p value < 0.0001).

There is a significant overexpression of the pseudogene *POU5F1P1* in tumour tissue compared to tumour surrounding tissue (Student's paired t-test, p value < 0.0001). Both, 8q amplification of 8q and risk variants on 8q24.21 might have an influence on the transcription level of *POU5F1P1*.

3.2. Examination of causative effects for *POU5F1P1* overexpression in prostatic carcinoma

3.2.1. Investigation of copy number variation of 8q

Since loss of 8p and amplification of 8q are quite frequent in prostatic carcinoma, the reason for the overexpression of *POU5F1P1* in prostatic carcinoma tissue might be 8q amplification. For the investigation of any genomic 8q copy number variations in the investigated prostatic carcinoma tissues, MLPA (Multiplex Ligation-dependent Probe Amplification) analysis with genomic DNAs of prostatic carcinoma tissues were performed, using the SALSA MLPA kit P014 Chromosome 8 from MRC Holland. The procedure is schematically shown in Figure 12.

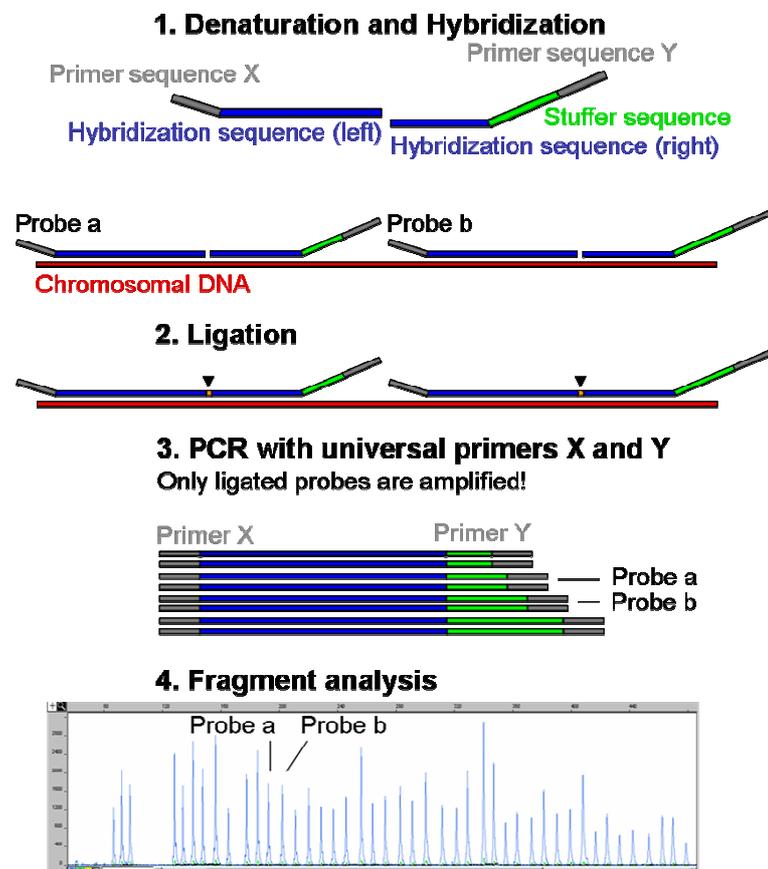


Figure 12: Schematic overview of MLPA (= Multiplex Ligation-dependent Probe Amplification) analysis

After denaturation of genomic sample DNA, MLPA probes consisting of two parts are hybridised to the target sequence (1.) and ligated by a heat-stable ligase (2.). All probes carry primer sequences X and Y at their ends, making it possible to amplify all successfully ligated probes with only one primer pair (3.). Different probes harbour distinct stuffer sequences (see probe a, probe b). This leads to PCR products of distinct sizes. Since each forward primer is fluorescence-labelled, PCR products can be separated by capillary gel electrophoresis (4.). The fluorescence signal is proportional to the original amount of target sequence in the investigated genomic DNA.

After hybridising probes complementary to sequences, scattered over the whole chromosome 8 as well as sequences on other chromosomes serving as controls, they are ligated and amplified by PCR. Only successfully ligated probes can be amplified. Since each forward primer is fluorescence-labelled, fragments can be detected after separation by capillary gel electrophoresis. Determined fluorescence intensity is proportional to the original amount of target sequence in the investigated genomic DNA. Peak heights of carcinoma tissues were compared to peak heights of carcinoma surrounding tissues. Data analysis was done with the Coffalyser software, using the "Coffalyser method for analysis". Probes with calculated values in the range of 0.7 to 1.3 are assumed to indicate normal copy numbers of the target sequence.

Figure 13 shows the exemplary diagrams of MLPA results of two patients' carcinoma tissues. On the x-axis, all probes scattered over chromosome 8 as well as control probes from other chromosomes are indicated. The y-axis shows the ratio of each of the probes compared to carcinoma surrounding tissue. All probes of patient 1 show a ratio between 0.7 and 1.3, suggesting that all investigated sequences on chromosome 8 are present at normal copy numbers in prostatic carcinoma tissue of this patient. Hence, this patient neither shows loss of 8p nor amplification of 8q. MLPA results for the prostatic carcinoma tissue of patient 30 indicates loss of chromosome 8p (values lower than 0.7) and amplification of chromosome 8q (values higher than 1.3). The factor of approximately 1.5, received for all 8q probes, indicates that one 8q arm is duplicated in carcinoma cells of this patient.

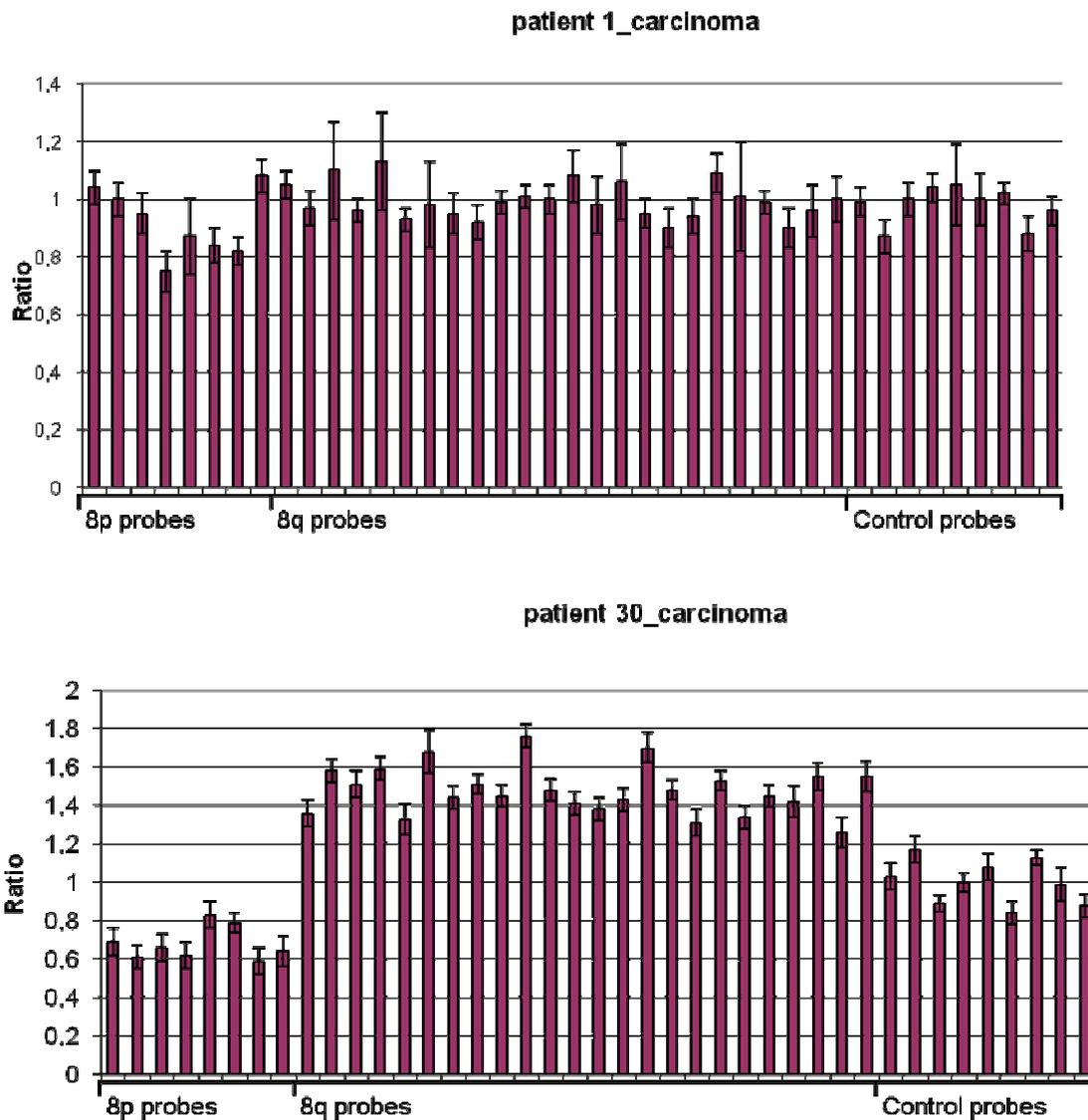


Figure 13: Exemplary MLPA (Multiplex Ligation-dependent Probe Amplification) results of two patients

Shown are the ratios for all MLPA probes of the SALSA MLPA kit P014 Chromosome 8 from MRC Holland after hybridisation to genomic DNA of prostatic carcinoma tissue from patient 1 and 30. On the x-axis, all probes scattered over chromosome 8p and 8q as well as control probes located on chromosomes 11, 12, 13 and 17 are indicated. The y-axis shows the ratio of each of the probes compared to control probes of the same sample as well as to all probes of prostatic carcinoma surrounding tissues. MRC Holland defines values between 0.7 and 1.3 as normal copy numbers.

No MLPA data could be obtained for patient 8 since the amplification reaction was without success. Patient 4 showed weakly shifted values for probes on the 8q arm (maximum value: 1.45). Partly shifted values suggesting a loss of 8p, but still in the range of normal values, were seen for the patients 6 and 18 (values down to 0.71) as well as 11, 14, 25, 28, 31 and 34 (showing values smaller than 0.7 for some probes). These data suggest an imbalance in the genomic copy number of chromosome 8, indicating weak 8p losses or 8q amplification in these patients. Data from all other patients looked similar to those from patient 1, showing values in a normal range (values between 0.7 and 1.3). A summary of all patients' MLPA data can be found in the appendix (Table 21, Table 22).

Since only one out of 34 investigated patients shows obvious amplification of 8q, this is not the reason for the overexpression of *POU5F1P1* in the investigated prostatic carcinoma tissues compared to carcinoma surrounding tissues. Thus, risk variants on 8q24.21 might influence the transcriptional activity of promoter sequences of *POU5F1P1*. Since the pseudogene *POU5F1P1* corresponds to a reverse transcribed pseudogene that was integrated into chromosome 8 it does not possess the original *OCT4* promoter any more. In the next step, transcriptional start sites which are responsible for the expression of *POU5F1P1* and might be affected by risk variants on 8q24.21 were investigated.

3.2.2. Examination of transcriptional start sites of *POU5F1P1*

To identify *POU5F1P1* transcriptional start sites that might be affected by risk variants on 8q24.21, 5'-RACE experiments were performed. After RNA isolation from two prostatic carcinoma tissues (from patient 6a and 28a) as well as several cell lines derived from prostatic carcinoma tissue (LNCaP, 22rv1 and DU145), a RNA anchor oligonucleotide was ligated to the 5'end specifically of RNA with a cap structure and thus only to intact mRNA molecules. After reverse transcription nested PCR using primers located at the anchor oligonucleotide sequence as well as in the open reading frame of *POU5F1P1* was performed. PCR products were separated by gel electrophoresis excised from the gel and after purification, they were cloned into vectors, transformed in *E.coli* cells and positive clones were identified by PCR. Plasmids were then amplified with the TempliPhiTM HT DNA Amplification Kit and sequenced (for schematical overview of the procedure see Figure 14).

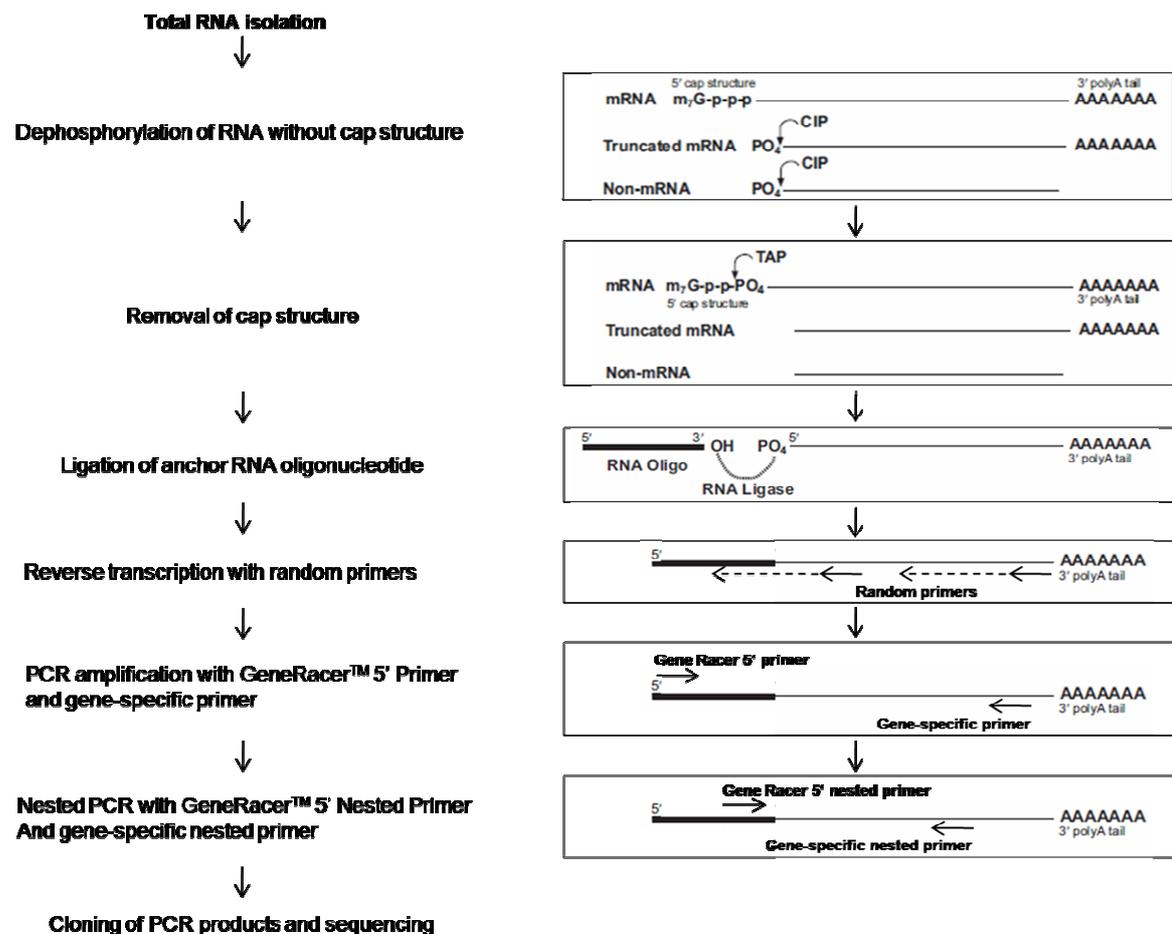


Figure 14: Overview of the methodical procedure of 5'RACE experiments

On the left, the succeeding steps for 5'RACE experiments are given. On the right, further details of methodical procedure can be seen. After RNA isolation, degraded mRNA without cap structures was dephosphorylated. Then, cap structure of intact mRNA was removed, leaving a phosphate at the 5' end of RNA molecules. In the next step, an anchor RNA oligonucleotide was ligated using this phosphate. Reverse transcription was done using a gene-specific primer, complementary to *POU5F1P1*-related sequences. PCR amplification by nested PCR of the sequences of interest followed. PCR products were separated by gel electrophoresis and excised PCR product bands were cloned and transformed in *E.coli* cells. The insert of isolated plasmids were sequenced.

Sequence reactions were analysed with the ABI PRISM 3100 Genetic Analyzer. Single sequence data were combined to the respective continuous clone sequence with the softwares GCG (Genetics Computer Group) and EditSeq, DNASTAR Lasergene®. Sequence positions in the human genome were found with the help of BLAST (Basic Local Alignment Search Tool) function on the homepage of NCBI (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/>). A schematic overview of the identified transcripts with the 5' genomic position of the identified exons relative to the coding sequence of *POU5F1P1* is given in Figure 15. 5'-RACE experiments identified nine different transcriptional start sites located in a genomic region of around nearly 851 kb and twelve different *POU5F1P1* transcripts.

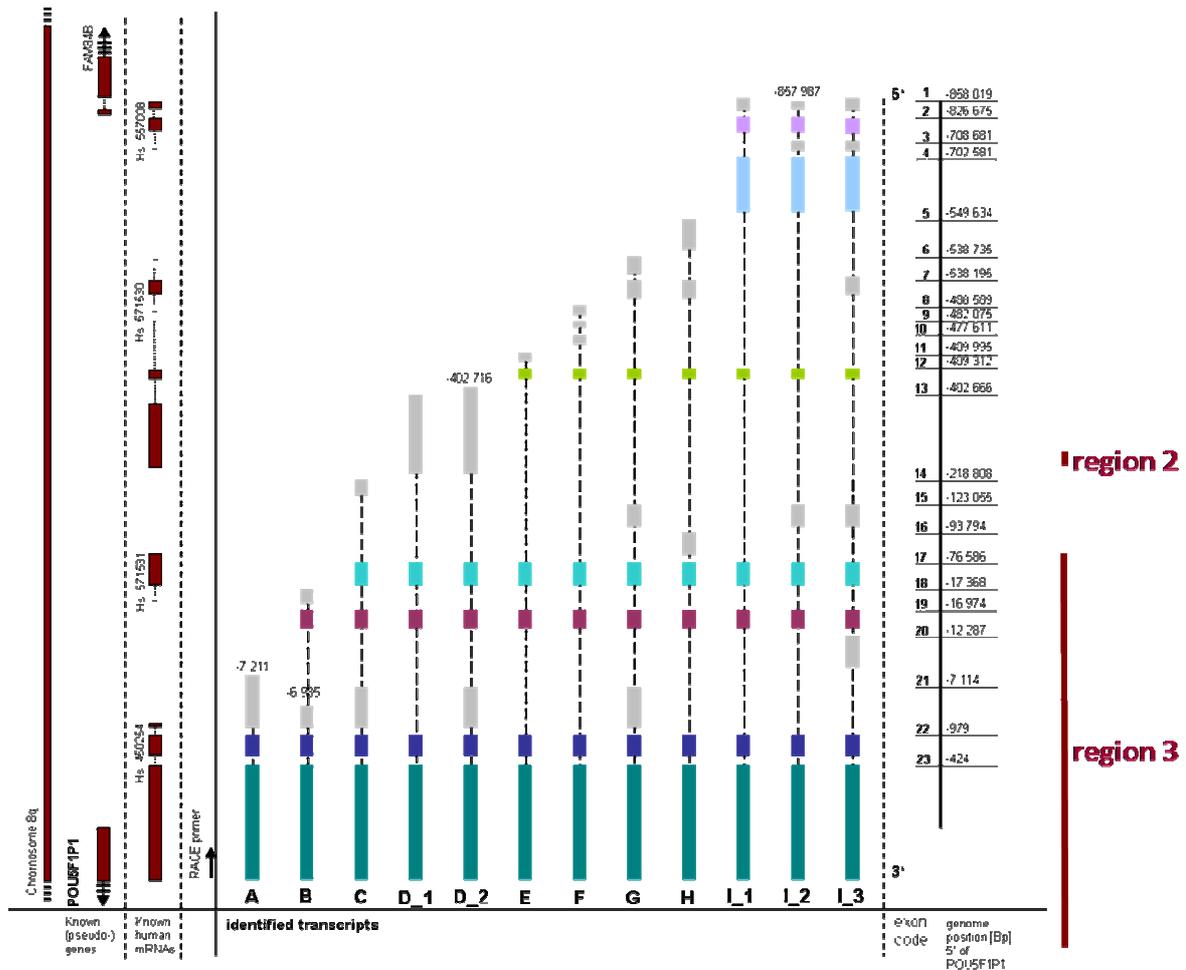


Figure 15: Schematic overview of *POU5F1P1* transcripts identified by 5'RACE experiments

On the left, known genes or pseudogenes as well as known human mRNAs (see UniGene mRNAs, NCBI) are indicated. Alternatively spliced exons of identified transcripts are depicted in gray and constitutively spliced exons are depicted in colour. Transcripts with distinct transcription start points are named with different letters and various splice variants got different numeric characters as indices. The arrow shows the location of the 5'RACE primer in the coding sequence of *POU5F1P1*. The base pair position on the right indicates the position of the 5' end of each exon relative to the start of the coding sequence of *POU5F1P1* (based on RefSeq position, Build 36.3). Also indicated are the positions of prostate cancer risk regions 2 and 3. Region 2 is located in intron 13 and region 3 harbours exons 17 to 23 of *POU5F1P1* transcripts.

Since there is a number of transcriptional start sites contributing to *POU5F1P1* expression, it looks like the exon harbouring this pseudogene is spliced to a range of transcripts expressed on 8q24.21. Thus, an overexpression not only of *POU5F1P1*, but also from other 8q24.21 transcripts in prostatic carcinoma tissue compared to surrounding prostatic carcinoma tissue should be observed.

3.2.3. Expression of other RNAs on chromosome 8q24.21 in prostatic carcinoma tissue

Since transcripts of *POU5F1P1* which are overexpressed in prostatic carcinoma tissue span genomic distances of nearly 860 kilobases, risk variants on 8q24.21 might have a functional impact on the general overexpression of this chromosomal region in prostatic carcinoma tissue. To elucidate this hypothesis, the expressions of other annotated RNAs on 8q24 were determined in prostatic carcinoma and carcinoma surrounding tissue.

Expression of splice variants from the expressed sequence tag (= EST) AW183883

In the genome browser NCBI, one expressed sequence tag (= EST), designated AW183883, is annotated in the gene desert on 8q24.21 with several splice variants existing of eight exons which are scattered over 200,000 base pairs in genomic DNA. Amundadottir et al. described four such splice variants, named DQ515896 to DQ515899 (see exon splicing in Figure 16) (Amundadottir et al., 2006). Previous experiments in our working group showed that the splice variants DQ515896 and DQ515897 are not expressed in prostatic cell lines. In previous studies of our working group, only exons 6, 7 and 8 were shown to be expressed in prostatic carcinoma cell lines. For this reason, primers for expression analysis of splice variants derived from the EST AW183883 were designed for the short exon 6 and exon 8 which target the splice variants DQ515897, DQ515898 and DQ515899.

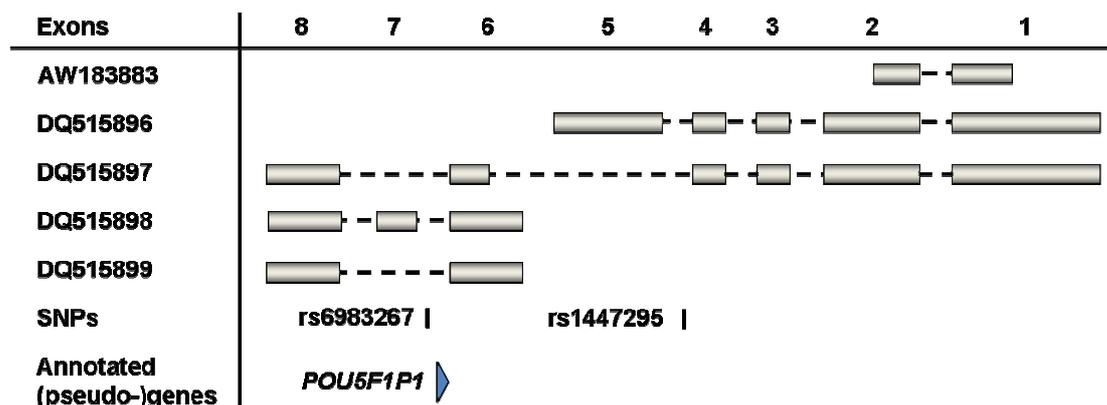


Figure 16: Previously described splice variants derived from the EST AW183883

Splice variants derived from the EST AW183883 are shown in the transcription direction (telomeric to centromeric) (first described by Amundadottir et al., 2006). Exon lengths are as follows: exon 1 = 503 bp, exon 2 = 343 bp, exon 3 = 103 bp, exon 4 = 88 bp, exon 5 = 371 bp, exon 6 = 135 bp or exon 6 (long) = 536 bp, exon 7 = 140 bp, exon 8 = 246 bp. Figure is not drawn to scale.

Location of exons 6, 7 and 8 derived from the EST AW183883 on 8q24.21 is shown in Figure 17. These three exons span a genomic region of approximately 131 kb.

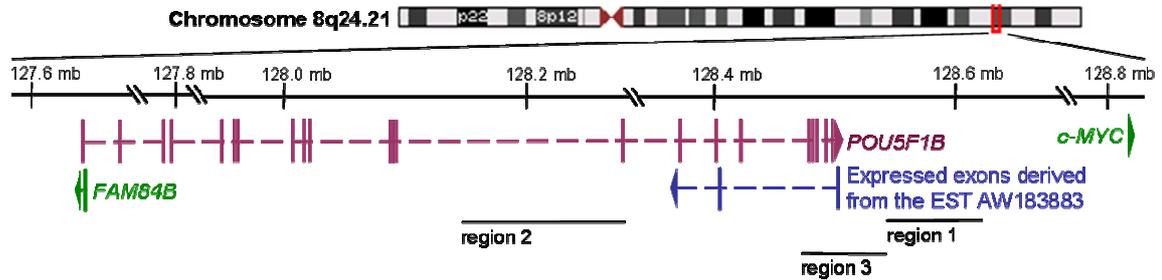


Figure 17: Overview of 8q24.21 with the location of expressed RNAs

The two genes *FAM84B* and *c-MYC* bordering the gene desert on 8q24.21 are indicated in green. Prostate cancer risk regions 1, 2 and 3 are shown in black. All 5'exons of the *POU5F1P1* pseudogene identified by 5'RACE experiments in this work are indicated as purple bars. Exons from the EST AW183883 expressed in prostatic cell lines are indicated in blue.

Expression analysis of splice variants derived from the EST AW183883 relative to the expression of the housekeeping gene *HPRT1* was carried out by Taqman assay (methodical procedure was in accordance to that of expression analysis of *POU5F1P1*, 3.1.4). The mean of triplicate measurements of the expression of splice variants derived from the EST AW183883 relative to that from *HPRT1* are plotted in Figure 18. The expression of this EST in general is very low. Hence, some sample values could not be analysed since they are not in the range of values of the standard curve. But since mostly the values from tumour surrounding tissue are too low, a higher expression in the tumour compared to tumour surrounding tissue can be assumed.

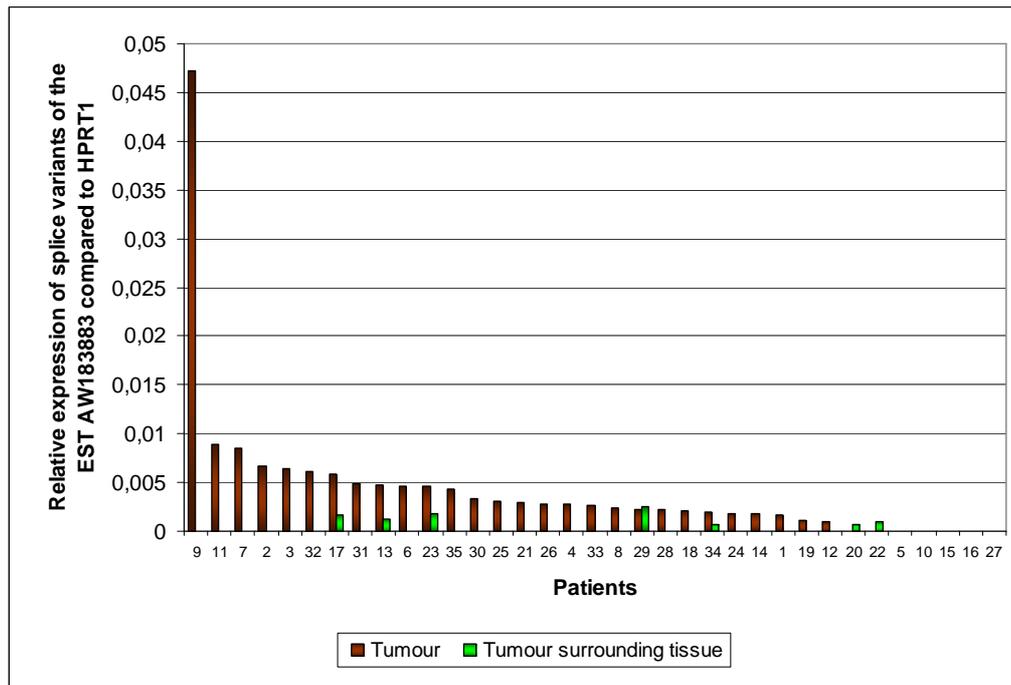


Figure 18: Relative expression of splice variants derived from the EST AW183883 compared to *HPRT1* in prostatic tumour and tumour surrounding tissue

Bars correspond to values derived from triplicate measurements with the Taqman for prostatic tumour (red) and tumour surrounding tissue (green). Patients were sorted for expression intensity in the tumour. Since most samples from tumour surrounding tissue resulted in values below measurable range defined by the standard curve, the expression of splice variants derived from the EST AW183883 is higher in prostatic tumour tissue than in tumour surrounding tissue.

These data show an overexpression not only of *POU5F1P1* in prostatic carcinoma tissue, but also from other transcripts as splice variants derived from the EST AW183883. Since both span long genomic regions of around 851 kb and 131 kb, respectively, this suggests a general overexpression of transcripts in the gene desert on 8q24.21 in prostatic carcinoma tissue compared to surrounding prostatic tissue.

c-MYC expression in prostatic carcinoma tissue

The closest gene in telomeric direction in relation to the variants on 8q24 associated with prostate cancer is the *c-MYC* gene, which is assumed to be one of the target genes of those risk variants. In this work, the *c-MYC* expression relative to the housekeeping gene *HPRT1* in tumour and tumour surrounding tissue of the investigated 35 prostate cancer patients was measured in a semi-quantitative manner by fragment analysis. The assay consists of two steps. First, PCR amplifications are performed in a multiplex reaction with primers for the target and the housekeeping gene. One oligonucleotide of each primer pair is labelled with a fluorescence dye at its 5'end. In the second step, PCR products are separated by capillary gel electrophoresis with the Genetic Analyzer ABI PRISM 3100. Time of detection corresponds to the PCR product size and fluorescence intensity is proportional to the fragment amount. Both *HPRT1* and *c-MYC* PCR products were

Results

fluorescence-labelled with 6-FAM. The product size was 302 bp for *HPRT1* and 389 bp for *c-MYC*. After PCR reaction with 26 cycles, fragment analysis was performed carrying along a size standard. Figure 19 shows the result of such a capillary gel electrophoresis, which was obtained with the analysis software GeneScan 3.7 from Applied Biosystems. Peak heights were used for further analysis.

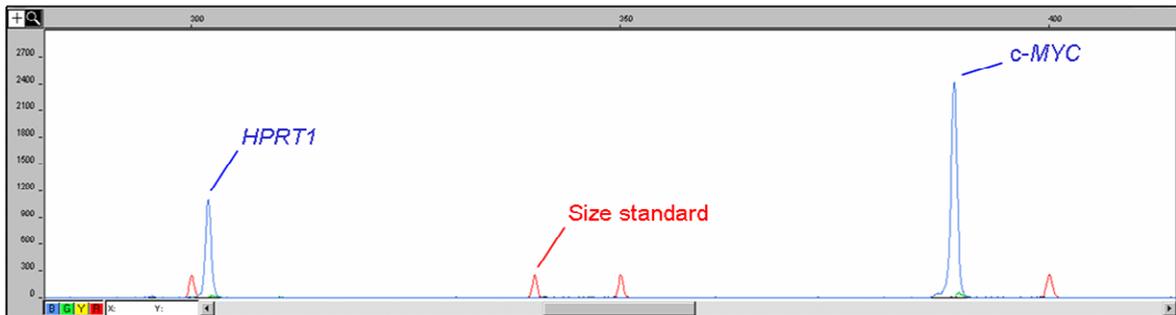


Figure 19: Results of multiplex fragment analysis of *HPRT1* and *c-MYC* with the software GeneScan from Applied Biosystems

The blue fluorescence peaks correspond to the PCR products from *HPRT1* (302 bp) and *c-MYC* (389 bp), respectively, labelled with 6-FAM (blue), red fluorescence peaks correspond to fragments of the ROX-labelled size standard. Peak height is proportional to the amount of the respective PCR product and was used for further data.

For semi-quantitative expression analysis, the ratios of the *c-MYC* peak heights relative to that of *HPRT1* were calculated. The mean of triplicate measurements with standard deviation for each prostatic carcinoma and carcinoma surrounding tissue are shown in Figure 20.

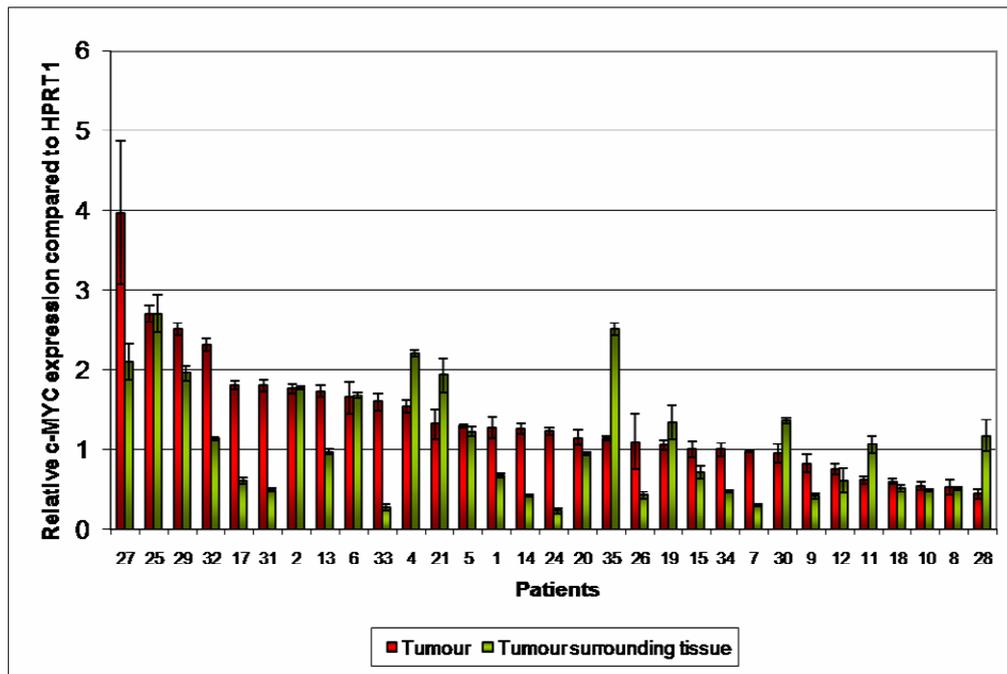


Figure 20: Relative c-MYC expression in prostatic tumour and tumour surrounding tissue compared to *HPRT1*

Shown are the means of ratios of c-MYC expression relative to *HPRT1* expression after performing triplicate PCR reactions. Red bars represent data from prostatic tumour, green bars represent data from tumour surrounding tissue. Samples are sorted from high to low c-MYC expression in the tumour. Shown are data only from patients with amplification data from carcinoma as well as surrounding tissue.

The relative c-MYC expression in tumour and tumour surrounding tissue are plotted as Box Plots in Figure 21, a). There can only be seen a weak difference in expression level between both groups. But when differences of c-MYC expression in prostatic tumour compared to tumour surrounding tissue of the same patient are plotted in a Box Plot, almost 75 % of the values are positive. This indicates a stronger c-MYC expression in prostatic tumour than in prostatic tumour surrounding tissue in the same patient (Student's paired t-test using these differences resulted in a p value of 0.0263).

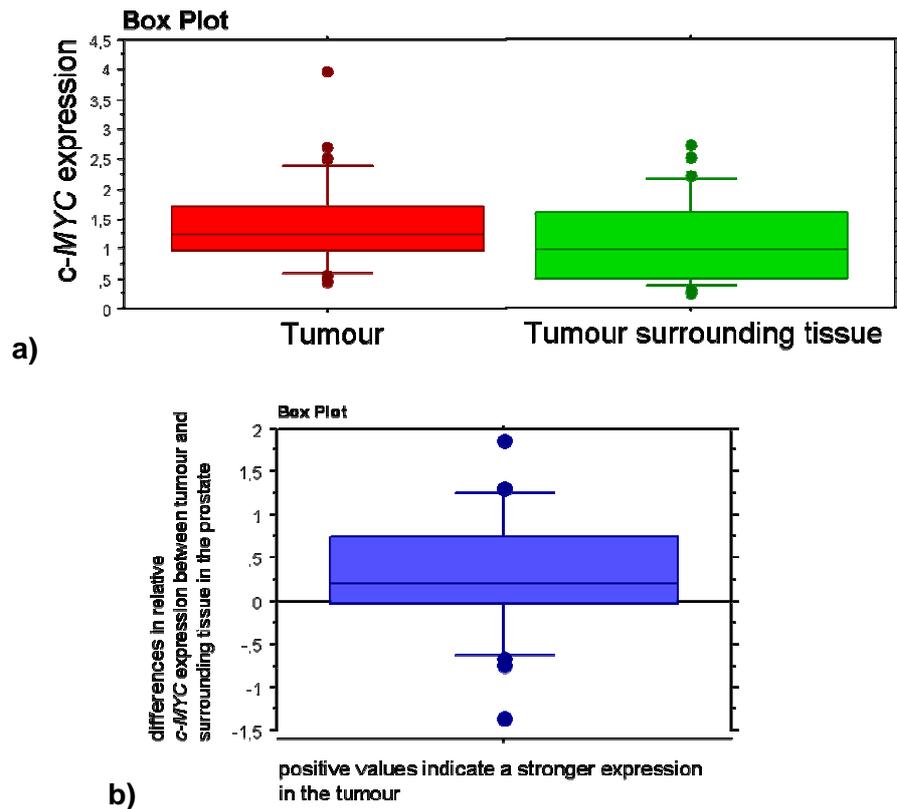


Figure 21: Box Plots of relative c-MYC expression

a) shows Box Plots of the relative c-MYC expression compared to *HPRT1* in the tumour (red) and in tumour surrounding tissue (green), respectively, corresponding to graphs in Figure 11. There seems to be only a weak difference in c-MYC expression between tumour and tumour surrounding tissue. Regarding corresponding pairs of c-MYC expression in tumour compared to tumour surrounding tissue from the same patient, there is a significant overexpression in the tumour (Student's paired t-test, p value 0.0263). To show this difference, in **b)** difference values from the 31 prostate cancer patients between prostatic tumour and tumour surrounding tissue are plotted as a Box Plot.

3.2.4. The putative impact of prostate cancer risk variants on the expression level of 8q24.21 transcripts

Analysis of correlation between 8q24.21 transcripts' expression levels

Since the observed overexpression of 8q24.21 transcripts in carcinoma tissue might be caused by the same mechanism, a possible correlation of 8q transcript intensities was investigated by the Spearman Rank Correlation test. This test examines if there is a statistical dependence and thus a correlation of two variables, here represented by the expression intensity of two distinct transcripts. There is no correlation of the quantitative expression of 8q24.21 transcripts in prostatic carcinoma or carcinoma surrounding tissue (Spearman Rank Correlation, p values > 0.05).

8q overexpression and analysis of correlation with prostate cancer risk variants on 8q24.21 and coding variants in *POU5F1P1*

Expression levels of *POU5F1P1* and splice variants derived from the EST AW1833883 did not show correlation with each other. The expression levels of these transcripts might be independently influenced by distinct risk variants on 8q24.21. An overview of the location of eight risk variants on 8q24.21, which were reported by Al Olama et al. to explain the association of 8q24.21 with the development of prostate cancer, is shown in Figure 2, 1.3. Genotype data from all patients are depicted in Table 23 and Table 24, in the appendix. To investigate any effect of these risk variants on the expression level of *POU5F1P1* or splice variants derived from the EST AW183883, statistical analysis was performed. If there were at least five patients for all three possible genotypes (homozygous for the non-risk allele, heterozygous and homozygous for the risk allele), a Kruskal Wallis test was performed. This statistical test examines if more than two groups of values are different in their expected mean. If the risk allele is very rare and there were only two different groups (homozygous for the non-risk allele and heterozygous), a Mann Whitney test was performed. This test investigates if two independent sample groups belong to the same group, or are significantly different from each other. Further, a Spearman's rank correlation test was performed with the sum of risk alleles and the quantitative expression of a distinct transcript.

The SNP rs12543663 in block 1 showed weak significant association of the genotype with the *POU5F1P1* expression (Mann Whitney test for the two groups "no risk" allele and "at least one risk allele", p value: 0.0455). Data suggested a stronger *POU5F1P1* expression in patients that carry at least one risk allele compared to patients with no risk allele of this SNP. Statistical analysis showed no significant association of any of the tested eight variants on 8q24.21 with the expression of *POU5F1P1* or splice variants derived from the EST AW183883 (all p values > 0.05). If there was an additive influence of 8q SNPs on the expression of 8q transcripts, there should be a correlation between the total sum of 8q SNP risk alleles and the quantitative expression. The Spearman's Rank Correlation test with the sum of risk alleles of these SNPs versus the quantitative expression of *POU5F1P1* or splice variants derived from the EST AW183883 did not show any correlation, at all. Resequencing of *POU5F1P1* in 95 prostate cancer patients resulted in five coding SNPs with allele frequencies higher than 1 %. None of these variants showed correlation with the expression level of *POU5F1P1*.

Analysis of correlation of 8q24.21 transcription levels with clinical data

No correlation of the clinical data tumour stage, tumour grade or age at first diagnosis with the expression of *POU5F1P1* or splice variants derived from the EST AW183883 in the tumour or tumour surrounding tissue were observed.

Examination of an allelic shift in expression of *POU5F1P1*

If the chromosomal region 8q24.21 is under the control of *cis* regulatory variants that influence the transcriptional activity in a direct or indirect manner leading to the observed overexpression of 8q24.21 transcripts in prostatic carcinoma tissue, a difference in the allelic expression of 8q24.21 RNAs should be observed. To examine an allelic difference in expression, the ratio of expressed SNP alleles can be investigated. Allelic ratios of such expressed SNPs give information about the transcriptional activity of both alleles. Variants in the mRNA sequences of *c-MYC* or splice variants derived from the EST AW183883 could not be identified. 26 out of 35 patients were heterozygous for at least one single nucleotide polymorphism (SNP) in the coding sequence of *POU5F1P1* and, therefore, were suitable for this experiment. Examination of an allelic shift in expression was done by SNaP-Shot analysis.

To examine the resolution power of the SNaPshot method, standard curves for each SNP were generated, using different allelic compositions from homozygous samples. Allelic compositions ranged from 10 % of the ancestral allele and 90 % of the variant allele to 90 % of the ancestral allele and 10 % of the variant allele. Three independent PCR reactions and subsequent SNaPshot reactions were performed on each allelic composition. Ratios of peak heights from both alleles were calculated for each SNaPshot reaction. The logarithms of all three ratio values for each allelic compositions as well as the mean of these three ratios with standard deviations were plotted for the percentage of allele 1. Additionally, a regression curve was drawn for the mean (see exemplary standardcurve for SNP rs6998061 in Figure 22).

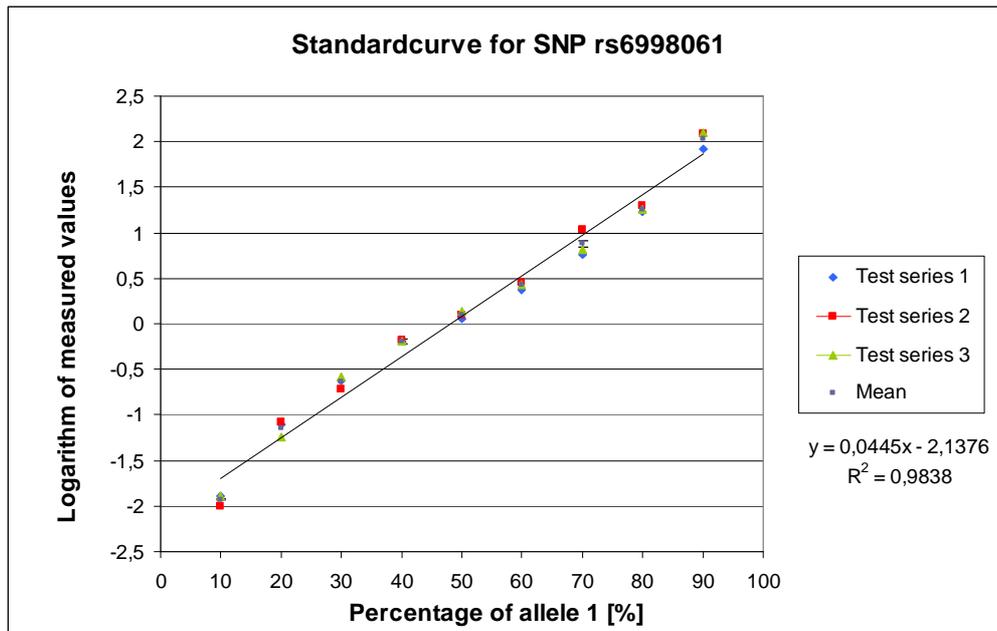


Figure 22: Standard curve for the single nucleotide polymorphism (= SNP) rs6998061 in the open reading frame of *POU5F1P1*

The logarithms of allelic ratios resulting from three independent test series as well as the mean of these triplicates with standard deviation were plotted for the percentage of allele 1 in distinct allelic compositions. The linear equation and the determination coefficient of the linear regression curve of the mean values are indicated.

The standard curve of allelic compositions of the SNP rs6998061 resulted in a regression curve with a determination coefficient of 0.983 (see Figure 22). Standard curves for all other allele combinations are generated analogously. Plots are not shown here, but all determination coefficients were at least 0.97 for each tested SNP. Allelic compositions differing in 10 % can be clearly distinguished from each other what renders SNaPshot experiments a suitable method for examination of an allelic shift in expression.

For the examination of an allelic shift in expression of *POU5F1P1*, the ratio of allelic fluorescence signals from the ancestral (subsequently called "1") and the variant SNP allele (subsequently called "2") resulting from cDNA was calculated. This ratio was normalised to the ratio of allelic fluorescence signals arising from genomic DNA of the same patient (see Figure 23). A deviation of this ratio from the value one suggests a difference in the allelic expression.

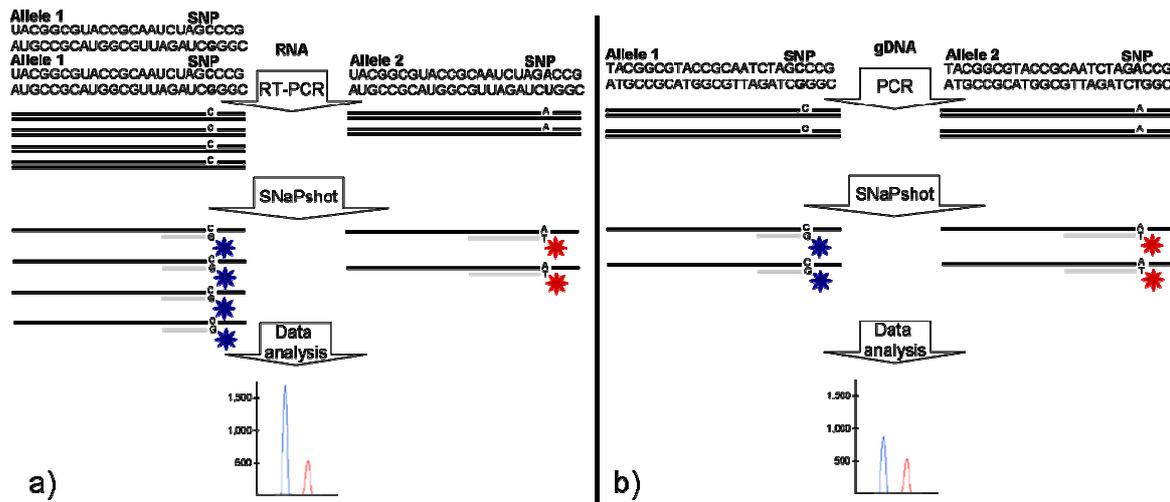


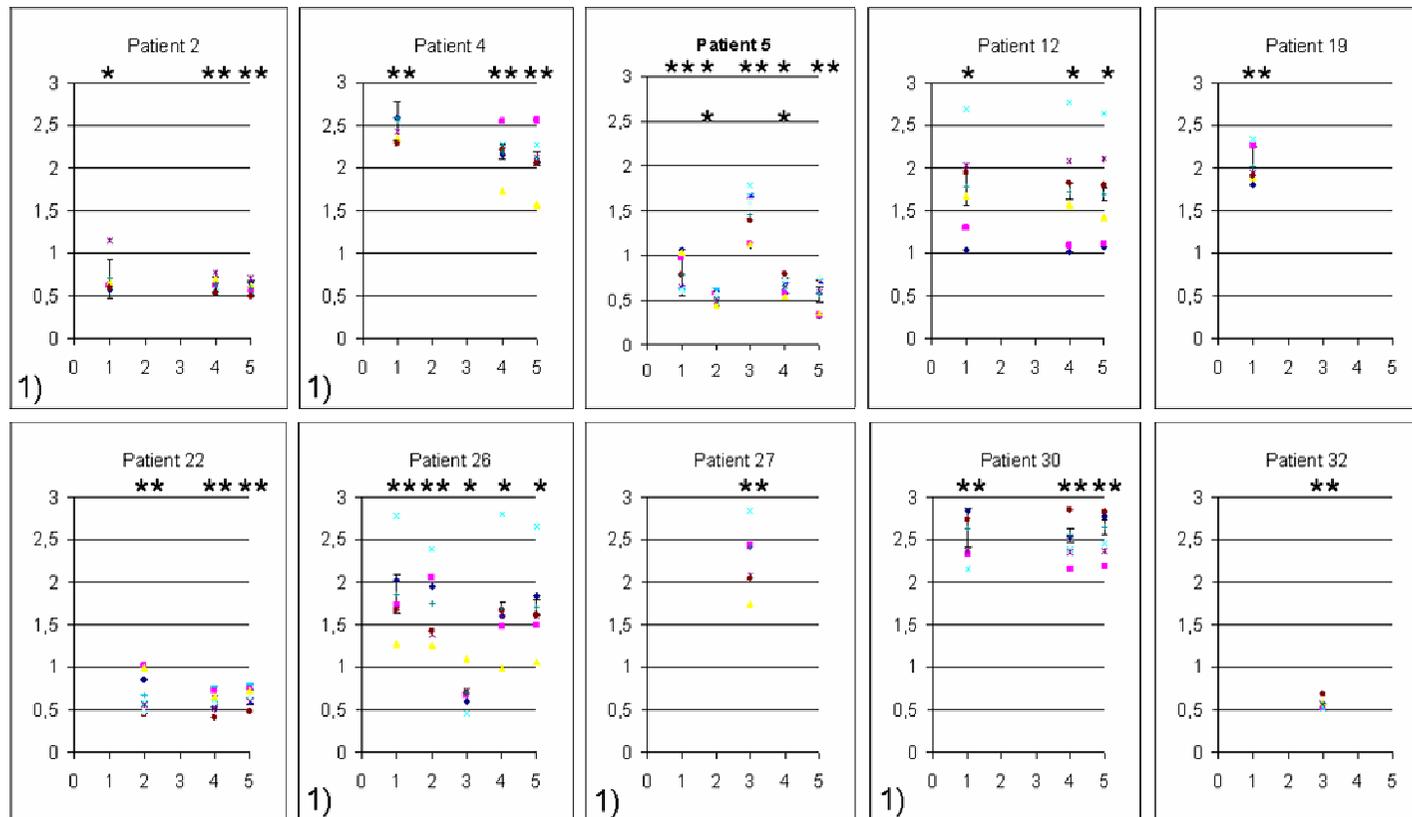
Figure 23: Schematic overview of the methodical procedure for measurement of an allelic shift in expression

In the first step, PCR (polymerase chain reaction) on genomic DNA (= gDNA) was performed harbouring single nucleotide polymorphisms (SNPs) amplifying both alleles (see top of **a**). Afterwards, SNaPshot reaction was performed where each allele results in the incorporation of a distinct specifically fluorescence-labelled nucleotide. After analysis with the Genetic Analyzer ABI PRISM 3100, ratio of peak heights from both alleles was calculated. The same procedure is performed with RNA as template, performing a RT-PCR (reverse-transcription PCR) (see **b**). At the end, ratios of **b**) and **a**) are calculated. If there is an allelic difference in expression of the investigated sequence, ratio deviates from the value one. Here, (cDNA ratio) / (gDNA ratio) is: (1,750 : 500) / (875 : 500) = 2. Thus allele 1 is expressed two times stronger than allele 2.

At least six independent PCR amplifications of *POU5F1P1* with cDNA from carcinoma tissue of each patient as template were performed. For patients 5, 7, 9, 11, 13, 16, 18, 22, 23 and 24, nine different PCR reactions were performed. Each PCR product served as template for SNaPshot experiments on all five SNPs of one test series. Subsequently, allelic ratios were calculated which were normalised to the mean of triplicate allelic ratios of genomic DNA from carcinoma tissue of the same patient. The logarithms of normalised allelic ratios were calculated and tested with a one-sample t-test to be significantly different from one. Single values were plotted separately for each of the 5 investigated SNPs in the open reading frame of *POU5F1P1*. Each test series were drawn in the same coloured and shaped data points. Additionally to the single measurements, the mean of all values for one SNP with standard deviation also was drawn. All values received for one single SNP was tested for difference from the value one. For data analysis, it was defined that if all examined SNPs showed an allelic ratio significantly different from one, there is an allelic shift in *POU5F1P1* expression. If at least one of the investigated SNPs did not show a significant difference from one, it was assumed that there is no allelic shift in expression. Figure 24 shows 10 patients that showed such an allelic shift in *POU5F1P1* expression. Asterisks in Figure 24 indicate mean values significantly different from one (* correspond to a p value < 0.05; ** correspond to a p value < 0.01). For patients marked with the numeric character 1), two different RNA preparations were used for the measurements to evaluate the results.

Figure 24: Patients that showed an allelic shift in the expression of *POU5F1P1* in prostatic carcinoma tissue

Each diagram shows the five SNPs in the open reading frame of *POU5F1P1* (1 to 5: rs6998061, rs13273814, rs13274084, rs6998254, and rs7002225) on the x axis. The y axis indicates the values of the allelic ratio of cDNA normalised to that of genomic DNA in the same carcinoma tissue. Values received from the same PCR product are plotted as one test series indicated as the same coloured and shaped data points. At least six independent test series were performed. For patients marked with "1)" two different RNA preparations for evaluation of the results were used. Variants marked with an asterisk showed a statistical significant difference of the mean from the value one (one sample group test, p value < 0.05), variants marked with two asterisks showed a highly significant difference of the mean from the value one (one sample group test, p value < 0.01).



16 patients showed no shift in the allelic expression of *POU5F1P1* (Figure 32 and Figure 33 in the appendix).

In summary, 26 out of 35 prostate cancer patients that are heterozygous for at least one SNP in *POU5F1P1* were investigated for an allelic shift in the expression of *POU5F1P1*. Ten out of these 26 patients indeed showed such an allelic shift of expression. Interestingly, patient 30, which previously was shown to harbour 8q duplication, shows a difference in allelic expression independently of this genomic aberration. These data might be a hint on putative general allelic differences of the expression of this chromosomal region on 8q24.21. There are too little samples to make any assumption about the *POU5F1P1* haplotype that is predominantly expressed or if there might be a correlation of an allelic difference in the *POU5F1P1* expression with SNP genotypes of known risk variants on 8q24.21. Thus, it might be possible that different haplotypes of SNPs in the chromosomal region on 8q24.21 differ in transcriptional activity that can be measured in different allelic expression.

As a next step, it should be examined if the pseudogene *POU5F1P1* harbours putative functional variants that mediate prostate cancer risk on 8q24.21.

3.3. Association study with coding variants in *POU5F1P1*

The mRNA of the pseudogene *POU5F1P1* shows a high homology to its parent gene (see results of the BLAST analysis, 3.1). Further, there most probably are no essential biological functions of *POU5F1P1* (see the sequence results in different primates, 3.1.1). But there is an overexpression of *POU5F1P1* in prostatic carcinoma tissue (see 3.1.4). Thus, it might be possible that SNPs in the coding sequence of *POU5F1P1* could represent functional variants associated with prostate cancer development. To identify such variants, the coding sequence of *POU5F1P1* was resequenced in 95 prostate cancer patients. Six SNPs in the coding sequence of *POU5F1P1* were identified. Five of them are known variants. The nucleotide positions, the SNP alleles, the positions of the affected amino acids of the encoded protein and the allele frequencies of the variant alleles are depicted in Table 16.

Results

Table 16: SNPs in the open reading frame of *POU5F1P1* identified by sequencing genomic DNA from 95 prostate cancer patients

Indicated are the nucleotide positions of the single nucleotide polymorphisms (= SNPs) in the open reading frame (= ORF) of *POU5F1P1*, the ancestral alleles and variant alleles (A = Adenine, C = Cytosine, G = Guanine) as well as the names of already known SNPs, the amino acids that are affected by these SNPs (Gly = Glycine, Ser = Serine, Glu = Glutamic acid, Thr = Threonine, Lys = Lysine, Asn = Asparagine, Asp = Aspartic Acid, Gln = Glutamine) as well as their positions in the coded protein and the allele frequencies of the derived alleles.

Base position in the ORF of <i>POU5F1P1</i>	265	527	545	640	684	712
Name of known SNPs (ancestral allele / variant allele)	unknown (G/A)	rs6998061 (G/A)	rs13273814 (C/A)	rs13274084 (A/G)	rs6998254 (G/A)	rs7002225 (G/C)
Amino acid exchange in the coded protein	Gly89Ser	Gly167Glu	Thr182Lys	Asn214Asp	no (Gln228)	Glu238Gln
Derived allele frequency [%]	0.53	43.09	80.85	6.52	48.91	49.47

The position of the variants with minor allele frequencies higher than 1 % in the open reading frame of *POU5F1P1* is shown in Figure 25. Two variants are located in the POU domain and one is located in the homeobox domain of the protein.

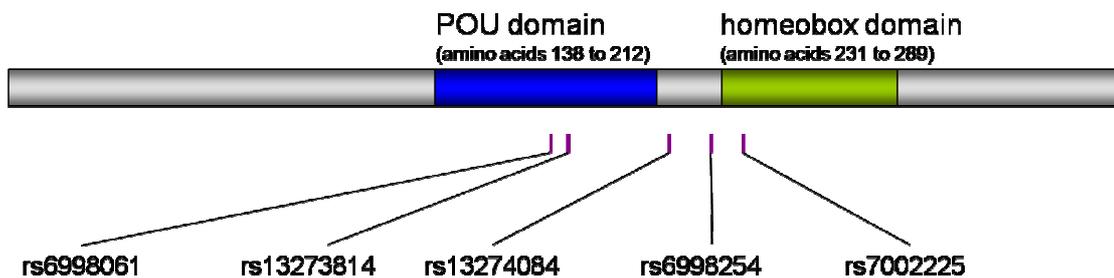


Figure 25: Positions of the amino acids affected by five SNPs (allele frequency > 1 %) in the putative *POU5F1P1* encoded protein

Shown is a schematical view of the *POU5F1P1* encoded putative protein harbouring the two functional domains present in the original OCT4 protein, the POU domain and the homeobox domain with their amino acid positions. The positions of the amino acids affected by five SNPs (single nucleotide polymorphisms) in *POU5F1P1* identified by sequencing genomic DNA of 95 prostate cancer patients are indicated.

One of the identified SNPs (rs6998254) does not have an effect on the coded amino acid sequence. Four of the five SNPs (> 1 % allele frequency) in *POU5F1P1* are non-synonymous base exchanges and potential functional variants. To investigate the issue if these SNPs might play a role in prostate cancer development, a case control study with all five SNPs, shown in Figure 25, was performed.

Since the pseudogene *POU5F1P1* is located in a region of linkage disequilibrium that is linked to prostate cancer risk variants, coding variants in *POU5F1P1* are expected to be also associated with prostate cancer. To analyse whether the SNPs in *POU5F1P1* are the

functional variants in this region of linkage disequilibrium, they were tested to show a stronger association with prostate cancer than already known variants in the same region.

3.3.1. Association of single variants in *POU5F1P1* with prostate cancer

For genotyping the five SNPs in *POU5F1P1*, SNaPshot assays were performed. This method is comparable to a sequencing reaction, but primers used here end right in front of the variant of interest and only fluorescence-labelled ddNTPs are supplied for incorporation. The ddNTP, synthesised to the 3' end of the SNaPshot primer, gives information about the SNP allele in the template strand. Fluorescence labelling is as follows: ddATP: dR6G (green), ddCTP: dTAMRATM (black), ddGTP: dR110 (blue), ddTTP: dROXTM (red). Figure 26 shows the genotyping results of multiplex SNaPshot analysis of one prostate cancer patient, heterozygous for all investigated SNPs in *POU5F1P1*.

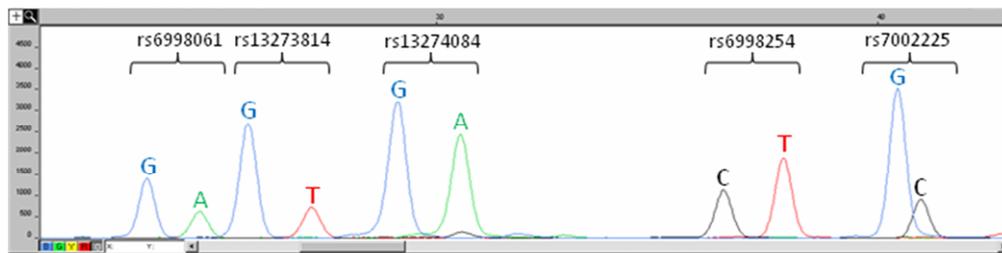


Figure 26: Multiplex SNaPshot results visualised with the software GeneScan from Applied Biosystems

Shown are detected fluorescence signals after capillary gel electrophoresis of SNaPshot experiments with the Genetic Analyzer ABI PRISM 3100. Each nucleotide is labelled with a specific fluorescence dye and thus incorporation of a distinct nucleotide for each SNP allele results in a specific fluorescence peak. Since SNaPshot primers differ in their lengths, they are detected at different time points. Fluorescence signals for each SNP primer as well as represented SNP alleles are given. SNP primers for rs13274084, rs6998254 and rs7002225 correspond to reverse primers and detect nucleotides of the reverse strand.

SNaPshot genotyping of SNPs in *POU5F1P1* was done in genomic DNA isolated from lymphocytes of 535 prostate cancer patients (346 sporadic cases and 189 familial cases) and 213 controls. 529 prostate cancer patients (345 sporadic cases and 184 familial cases) and 213 controls were included in the following data analysis. For each SNP, the Hardy Weinberg equilibrium in cases and controls was tested with the χ^2 -test, using the DeFinetti software of Tim M. Strom and Thomas F. Wienker. One SNP (rs6998061) showed no Hardy-Weinberg equilibrium in cases (p value of χ^2 -test: 0.029). Genotype data from SNPs on 8q24.21, that previously were shown to be associated with prostate cancer risk (rs1447295, rs16901979, rs6983267 and rs7837328), were included in the following analysis.

The allele frequencies of the SNPs on 8q24.21 in cases and controls, the allelic odds ratio with the 95 % confidence interval and the corresponding p value as well as the p value for

the Armitage trend test, a modified χ^2 -test which examines if there is a codominant effect of the risk allele concerning the development of the disease, are shown in Table 17. Already known risk alleles of the previously genotyped SNPs are indicated in bold. Pairwise linkage disequilibrium of all variants are depicted in Table 25 in the appendix.

Table 17: Association of genotyped variants on 8q24.21 with prostate cancer

Shown are the variants genotyped in 213 controls, 345 sporadic and 184 familial prostate cancer patients, the corresponding SNP (= single nucleotide polymorphism) alleles, the position in the three regions associated with prostate cancer on 8q24.21, the position in the open reading frame of *POU5F1P1* and their allele frequencies in the different samples as well as the allelic OR (= odds ratio) of the ancestral allele with the 95 % CI (= confidence interval) and the corresponding p value and the p value of the Armitage's trend test. Previously identified risk alleles for the SNPs rs16901979, rs6983267, rs7837328 and rs1447295 are indicated in bold. The respective ancestral alleles are listed in the first line. The variant allele is listed in the second line.

Variant	SNP alleles	Region	Position in <i>POU5F1P1</i>	Controls [%]	Sporadic cases [%]	Familial cases [%]	All cases [%]	Allelic OR (95 % CI)	p value	p value (Armitage's trend test)
rs16901979	A	2	-	2.82	5.69	5.16	5.5	2.009 (1.068 - 3.780)	0.02753	0.03071
	C			97.18	94.31	94.84	94.5			
rs6983267	G	3	-	47.17	52.91	56.52	54.17	1.324 (1.058 - 1.659)	0.01484	0.01317
	T			52.83	47.09	43.48	45.83			
rs7837328	G	3	-	64.55	59.16	53.28	57.1	0.731 (0.579 - 0.923)	0.00824	0.00685
	A			35.45	40.84	46.74	42.9			
rs6998861	G	3	527	53.52	59.42	60.99	59.96	1.301 (1.037 - 1.632)	0.02298	0.01760
	A			46.48	40.58	39.01	40.04			
rs13273814	C	3	545	16.2	18.12	17.58	17.93	1.130 (0.838 - 1.529)	0.42585	0.41355
	A			83.8	81.88	82.42	82.07			
rs13274084	A	3	640	88.5	87.25	92.03	88.9	1.041 (0.730 - 1.483)	0.82450	0.81922
	G			11.5	12.75	7.97	11.1			
rs6998254	G	3	684	43.66	48.7	53.3	50.28	1.305 (1.041 - 1.637)	0.02098	0.01724
	A			56.34	51.3	46.7	49.72			
rs7062225	G	3	712	43.66	48.7	53.02	50.19	1.300 (1.037 - 1.630)	0.02289	0.01907
	C			56.34	51.3	46.98	49.81			
rs1447295	C	1	-	91.55	85.47	84.51	85.13	0.529 (0.361 - 0.774)	0.00089	0.00092
	A			8.45	14.53	15.49	14.87			

The highest odds ratio in region 3 can be seen for SNP rs7837328 (OR for the ancestral allele is 0.731, and the corresponding inverse value is 1.368 for the variant allele; corresponding 95 % CI: 1.083 -1.727). Since rs7837328 is in complete linkage disequilibrium with rs6983267 and the associated risk allele of rs7837328 is less frequent but stronger associated with prostate cancer than the associated risk allele of rs6983267, the variant allele of rs7837328 tags the risk of this region in a better way. The maximum allelic odds ratio of SNPs in the open reading frame of *POU5F1P1* shows rs6998254: 1.305 (95 % CI: 1.041 – 1.637). Since variants in the open reading frame of *POU5F1P1* are located in the same region of linkage disequilibrium as the already known risk variant rs7837328, but do not show a higher odds ratio than that, there is no evidence that SNPs in *POU5F1P1* might be functional variants causing the association of risk region 3 with prostate cancer development.

3.3.2. Association of *POU5F1P1* haplotypes with prostate cancer

Since there is no evidence that single variants in the open reading frame of *POU5F1P1* represent functional SNPs in prostate cancer development, combination of these SNPs to distinct haplotypes might represent functional units that show stronger association with prostate cancer. Estimation of haplotypes combined of all five *POU5F1P1* SNPs in the genotyped cases and controls was performed with the software FAMHAP. As shown in Table 18, five possible haplotypes of the SNPs in *POU5F1P1* showing a frequency higher than 1 % were calculated. One of them, haplotype E, is significantly more frequent in controls than in prostate cancer cases (χ^2 -test, p value: 0.0213) whereas haplotype A seems to be more frequent in cases than in controls.

Table 18: Frequencies of *POU5F1P1* haplotypes in cases and controls

Frequencies of five observed *POU5F1P1* haplotypes in cases and controls are indicated. The numeric character 1 corresponds to the ancestral SNP allele, the numeric character 2 corresponds to the variant SNP allele. Haplotype E is significantly more frequent in controls than in cases (χ^2 -test, p value: 0.0213). Also indicated are the odds ratios (OR) for each *POU5F1P1* haplotype tested against all other haplotypes, overall p value is 0.1363 and thus not significant.

<i>POU5F1P1</i> haplotype	Allele constellation	Cases [%]	Controls [%]	OR
A: 11111	G C A G G	6.83	4.69	1.488
B: 11211	G C G G G	11.10	11.50	0.961
C: 12111	G A A G G	32.26	27.46	1.258
D: 12122	G A A A C	9.77	9.86	0.990
E: 22122	A A A A C	39.94	46.48	0.766

Table 19 shows all observed haplotypes composed of the two already known risk variants in prostate cancer risk region 3 and the *POU5F1P1* SNPs. As depicted in Table 17, the risk allele of SNP rs7837328 shows the strongest association to prostate cancer. Table 19 shows that there is complete linkage between this risk allele and the *POU5F1P1* haplotype which is most similar to *OCT4* as well as two other *POU5F1P1* haplotypes that arose from independent mutations at two distinct loci (haplotype B_2 and haplotype C_1). Because of this complete linkage disequilibrium, none of these *POU5F1P1* haplotypes exceeds association of SNP rs7837328 with prostate cancer. Nevertheless, region 3 haplotype A_1 shows the strongest difference between frequencies in cases and controls with a higher frequency in cases than in controls, suggesting a possible role of the *POU5F1P1* haplotype that is most similar to *OCT4*.

Table 19: Frequencies of observed haplotypes of region 3 variants in cases and controls

Seven haplotype constellations of the two risk variants rs6983267 and rs7837328 in region 3 and SNPs in *POU5F1P1* were observed. Shown in red are the known risk alleles at the first two loci (rs6983267 and rs7837328). The numeric character 1 corresponds to the ancestral SNP allele, the numeric character 2 corresponds to the variant SNP allele. Haplotype E_1 is significantly more frequent in controls than in cases (χ^2 -test, p value: 0.0331). Also indicated are the odds ratios (OR) for each observed haplotype against all others (overall p value is 0.4161 and thus not significant).

Region 3 haplotype	Allele constellation	Cases [%]	Controls [%]	OR
A_1: 1211111	G A G C A G G	6.82	4.56	1.484
B_1: 2111211	T G G C G G G	7.11	7.97	0.881
B_2: 1211211	G A G C G G G	3.89	3.53	1.105
C_1: 1212111	G A G A A G G	32.11	27.11	1.287
D_1: 1112122	G G G A A A C	9.72	9.86	0.977
E_1: 2122122	T G A A A A C	38.34	44.35	0.779
E_2: 1122122	G G A A A A C	1.63	1.99	0.758

Association of *POU5F1P1* haplotypes with different clinical aspects of prostate cancer

For the investigation if any of the haplotypes of variants in the open reading frame of *POU5F1P1* shows an association with clinical aspects of prostate cancer or if haplotypes are associated with distinct subgroups of the case collective, cases were stratified for tumour stage (low = T1/2 versus high = T3/4), tumour grade (low = I/II versus high = II-III/III), age of diagnosis of the disease (early = age of 64 or younger versus late = age of 65 or older) and family history (familial versus sporadic case). Subsequently, haplotype frequencies of these groups were compared to the corresponding haplotype frequencies in controls.

Stratification for tumour stage, tumour grade and age of diagnosis of the disease did not show significant difference of haplotype frequencies in distinct subgroups compared to controls. Results after stratification for family history of the prostate cancer patients showed that haplotype A is significantly more frequent in familial cases than in controls (χ^2 -test, p value: 0.00669). Taking the already known risk variants rs6983267 and rs7837328 into account, haplotype A_1 is also significantly more frequent in familial cases than in controls (χ^2 -test, p value: 0.00653). The data can be found in Table 29 in the appendix.

In summary, none of the haplotypes of *POU5F1P1* SNPs show a significant overrepresentation in prostate cancer. Haplotype E is significantly more frequent in controls than in cases. This haplotype is most different from the original *OCT4* haplotype. Haplotype A is significantly more frequent in familial cases than in controls. This haplotype harbours five ancestral *OCT4* alleles and represents the haplotype most similar to the *OCT4* gene and is in complete linkage disequilibrium with two risk alleles of the already known risk variants in prostate cancer risk region 3 and thus might represent a form of the putative encoded *POU5F1P1* protein most similar to its parent protein.

3.4. Investigation of the *POU5F1P1* encoded protein

If the pseudogene *POU5F1P1*, which is significantly overexpressed in prostatic carcinoma tissue, might bear any impact on tumorigenesis, a *POU5F1P1* protein should be detected in prostatic carcinoma tissue. Since parent gene and pseudogene show 95 % homology, it is possible that the pseudogene encoded protein further could bear a similar function as the original *OCT4* protein. An alignment of the amino acid sequence of the protein encoded by the Reference sequence (= RefSeq) of the pseudogene *POU5F1P1* with the *OCT4* protein is shown in Figure 27. There are 13 fixed amino acid exchanges between both encoded proteins and one amino acid is deleted in the *POU5F1P1* encoded protein. Additionally, four amino acids are affected by the coding SNPs. Depending on haplotypes of variants in *POU5F1P1*, 13 up to 17 amino acid exchanges can be present in the

POU5F1P1 protein compared to the original OCT4 protein. As tested in previous studies in our working group, these amino acid exchanges do not inhibit the binding of the POU5F1P1 protein by three antibodies directed against the OCT4 protein. These antibodies correspond to one polyclonal antibody directed against the C-terminus (designated C-20), one polyclonal antibody directed against the N-terminus (designated N-19) of the protein and, additionally, a monoclonal antibody directed against the C-terminus (designated C-10). All three antibodies were used for Immunohistochemistry and Western Blot experiments.

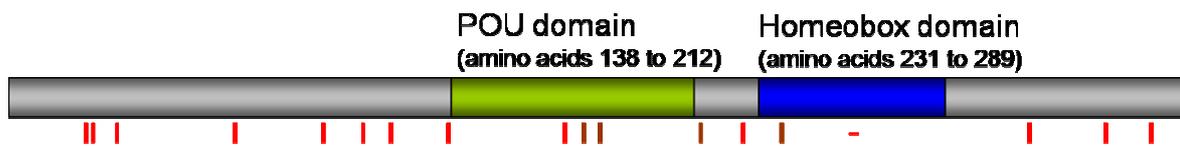


Figure 27: Schematic overview of the *POU5F1P1* encoded protein with positions of amino acid exchanges compared to the OCT4 protein

The two functional domains POU domain and homeobox domain are indicated with the corresponding amino acid positions. Amino acid exchanges between the putative POU5F1P1 protein and the OCT4 protein are indicated as red bars. Shown as brown bars are amino acids affected by coding variants in the *POU5F1P1* sequence which were used for the association study in this work. One amino acid of OCT4 is deleted in the POU5F1P1 protein (shown as a minus).

3.4.1. Immunohistochemistry

To investigate the issue if the POU5F1P1 protein is present in prostatic tissue and to elucidate which kind of cells express this protein, Immunohistochemistry experiments were performed. Experiments were done on paraffin-embedded prostatic carcinoma tissue. Paraffin-embedded seminoma tissue served as positive control since it was reported to express the OCT4 protein.

Immunohistochemistry was carried out as follows and is described in detail in the material and method section (see 2.2.10). Experiments were performed with 2 µm thick slices of paraffin-embedded tissue on glass slides. After deparaffinisation and rehydration of the samples endogenous peroxidase activity was blocked. For antigen retrieval slides were boiled in 10 mM sodium citrate buffer (pH value 6.0) and finally antibody incubation was done in 1x PBS for 30 minutes at room temperature. Incubation with the secondary antibody, labelled with horseradish peroxidase, followed. Detection was carried out by diaminobenzidine (DAB) colour reaction, resulting in brown-coloured staining of cell compartments harbouring the protein. As specificity control, the two polyclonal primary antibodies C-20 and N-19 were saturated with specific blocking peptides. Staining with various dilutions of the monoclonal antibody against the C-terminus (C-10) of the protein

was tested, but resulted in unspecific signals since staining was not abolished in the negative control where only the secondary antibody was supplied (data not shown).

As shown in Figure 28, staining seminoma with the polyclonal antibody N-19 resulted in intense nuclear and weak cytoplasmic signals in all seminoma cells whereas the surrounding mesenchymal tissue is devoid of OCT4 (N-19). When supplying the primary antibody N-19 saturated with blocking peptide, staining is completely abolished. Immunohistochemistry with prostatic tissue using the antibody N-19 resulted in weak cytoplasmic and weak nuclear signals in epithelial compartments. Staining is completely abolished when the primary antibody is saturated with corresponding blocking peptide.

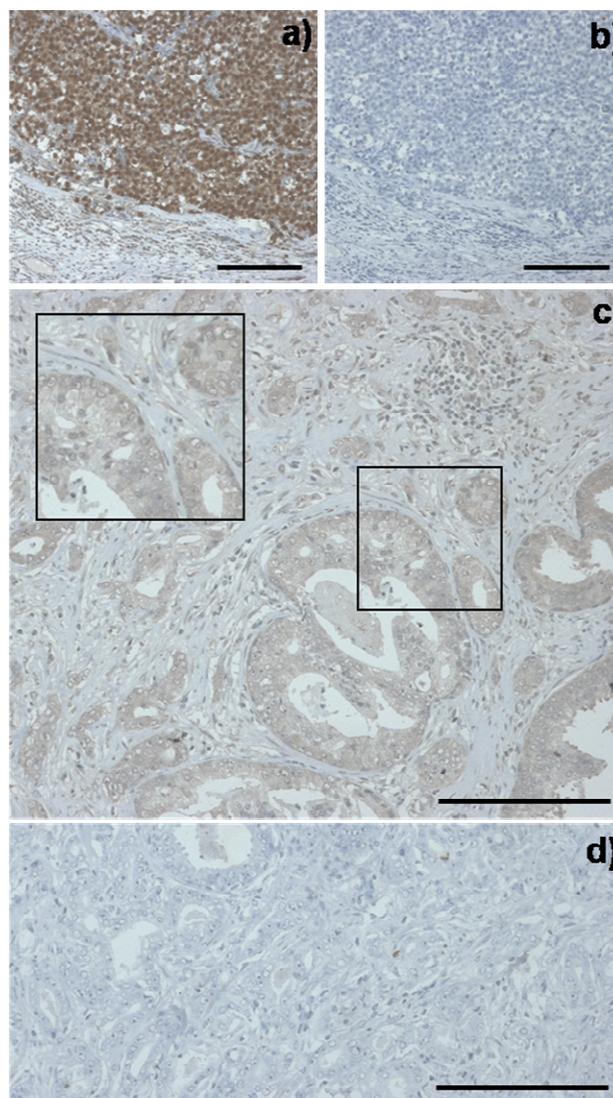


Figure 28: Immunohistochemistry using a polyclonal primary antibody against the N-terminus of OCT4 (N-19)

a) Immunostaining of seminoma served as positive control for OCT4 staining. **b)** Specificity control of seminoma applying the primary antibody saturated with blocking peptide. **c)** Immunostaining of prostatic tissue. A section of prostatic adenocarcinoma indicated by the quadratic mark is shown enlarged at the upper left corner. **d)** Specificity control of prostatic tissue applying the primary antibody saturated with blocking peptide; **a)** to **d)**: scale bar corresponds to 200 μm .

Staining results with the polyclonal antibody C-20 are shown in Figure 29. Staining results with seminoma tissue using this antibody resulted in similar staining patterns as it was shown with the antibody N-19. Supplying the primary antibody saturated with blocking peptide abolished staining completely. Supplying the antibody C-20 on prostatic tissue, both epithelial compartments show equally weak and cytoplasmic but no nuclear signals. Staining is again abolished when saturating the primary antibody with blocking peptide.

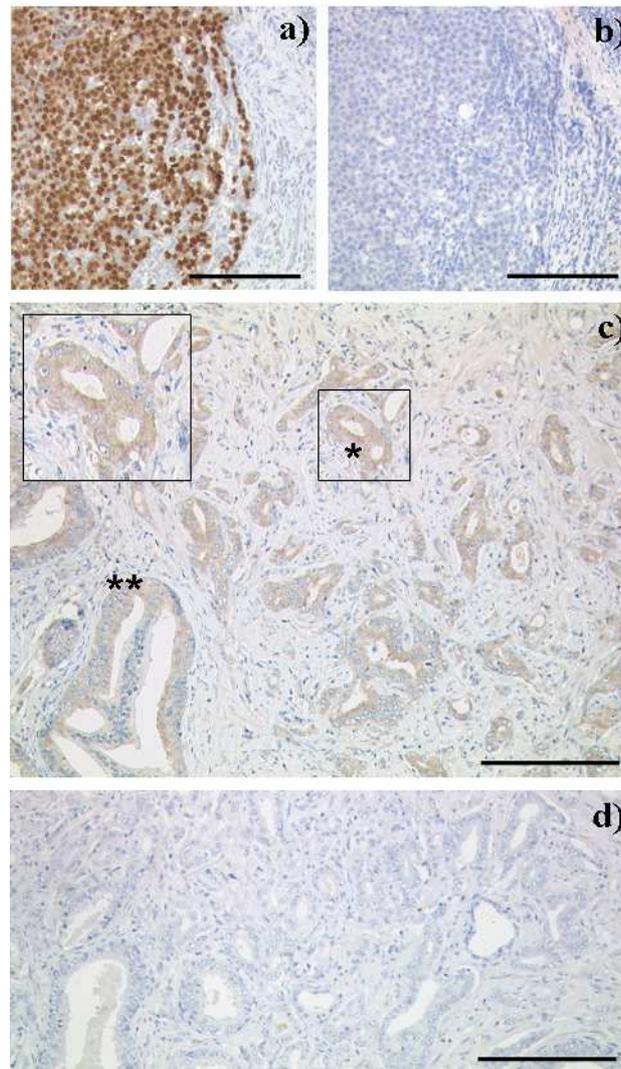


Figure 29: Immunohistochemistry using a polyclonal antibody against the C-terminus of OCT4 (C-20)

a) Immunostaining of seminoma serving as positive control for OCT4 (C-20). **b)** Specificity control of seminoma applying the primary antibody saturated with blocking peptide. **c)** Immunostaining of prostatic tissue showing non-neoplastic glands (**), surrounded by prostatic adenocarcinoma (*). A section of prostatic adenocarcinoma indicated by the quadratic mark is shown enlarged at the upper left corner. **d)** Specificity control of prostatic tissue applying the primary antibody saturated with blocking peptide; **a)** to **d)**: scale bar corresponds to 200 μm .

Taken the signals from both polyclonal antibodies together, Immunohistochemistry results suggest a weak cytoplasmic expression of the POU5F1P1 protein in prostatic tissue.

3.4.2. Western Blot

To further elucidate if the POU5F1P1 protein is expressed in prostatic tissue, Western Blot was performed on prostatic carcinoma tissue. Methodical procedure is described in 2.2.10.2. Two polyclonal primary goat antibodies one against the N-terminus (N-19) and the other against the C-terminus (C-20) of the OCT4 protein were used. As specificity control, primary antibodies were saturated with specific blocking peptide. Peroxidase-labelled secondary rabbit anti-goat antibody was used. Alternatively, a monoclonal primary mouse antibody against the C-terminus (C-10) of the protein was used. Horseradish peroxidase-labelled secondary antibody was here goat anti-mouse. Staining reaction was done with ECL solution which is oxidised to a coloured end product in the presence of horseradish peroxidase. For all samples, additionally ACTB (= beta actin) immunostaining for estimation of loaded protein amount was done. Cell lysates from patients 6, 17, 18 and 31 and additionally carcinoma surrounding tissue from patient 18 and 33 were examined (not all Blots are shown here). Patients 18 and 31 showed a high *POU5F1P1* mRNA expression in prostatic carcinoma tissue in Taqman experiments and thus were selected as prostatic tissue that rather may express POU5F1P1 (see Figure 11 in 3.1.4). For positive controls, protein extract from the human embryonal carcinoma cell line NTERA-2 as well as protein extract from frozen tissue of an intratubular germ cell neoplasia of the testis (= TIN) which both express the OCT4 protein were used.

Western Blot experiments using the polyclonal primary antibody C-20 resulted in a stained protein band with a size of approximately 43 kDa in TIN and NTERA-2 that can be blocked with the specificity control and thus corresponds to the OCT4 protein (see Figure 30 a) and b). In prostatic tissues 18a (prostatic tumour), 18b (tumour surrounding tissue) and 31a (prostatic tumour) stained protein bands with the size of around 53 kDa can be seen (see Figure 30 a)). These bands cannot be blocked (see Figure 30 b)). The signals appear even if just applying the secondary antibody (see Figure 30 e)). Thus these signals arise from unspecific binding of the secondary antibodies. There are no stained protein bands with the same size as the OCT4 protein in prostatic tissue. Western Blot results after staining with the polyclonal antibody directed against the N-terminus of the protein (N-19) look very similar and neither show positive staining of protein bands with the same size as the OCT4 protein in prostatic tissue (see Figure 30 c),d,e)).

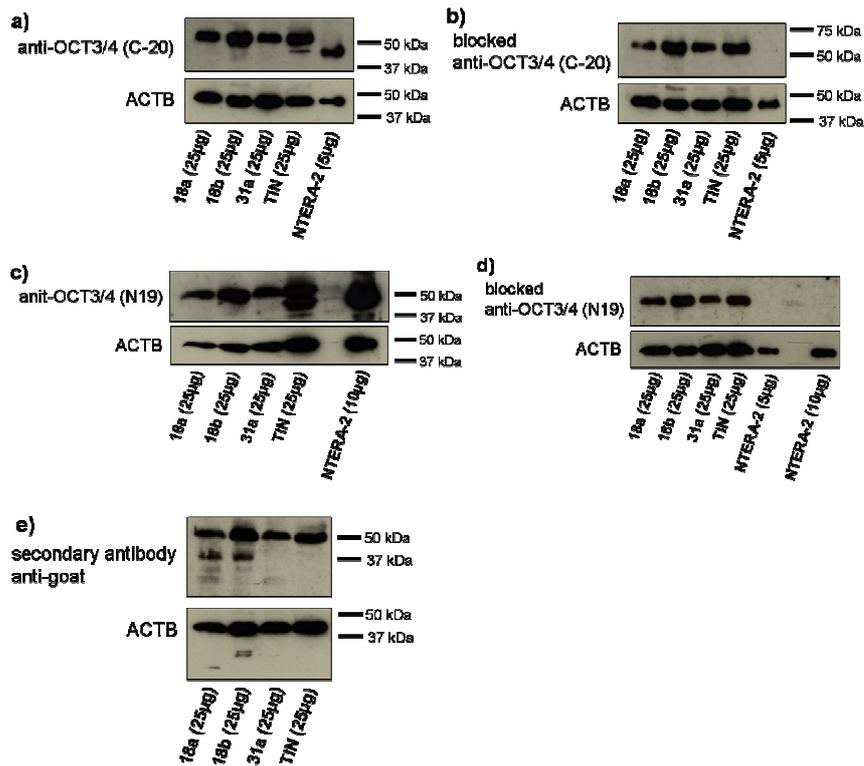


Figure 30: Western Blot results using two different polyclonal primary antibodies against OCT4

Shown are Western Blot results after staining with **a)** polyclonal primary antibody against the C-terminus of OCT4 (C-20), **b)** primary C-20 antibody saturated with specific blocking peptide, **c)** polyclonal goat antibody against the N-terminus of OCT4 (N-19), **d)** primary N-19 antibody saturated with specific blocking peptide, Western Blots in **a)** to **d)** was additionally incubated with the horseradish peroxidase-labelled secondary antibody (rabbit anti-goat); in **e)** only the secondary antibody (rabbit-anti goat) was supplied. All Western Blots were stained with an ACTB antibody.

Using a monoclonal mouse antibody against OCT4 as well as a different secondary goat anti-mouse antibody for Western Blot experiments resulted in positive staining of protein bands with the size of approximately 43 kDa in NTERA-2 and TIN (see Figure 31 a)). These protein bands show no staining when just applying the secondary goat-anti mouse antibody and thus correspond to the OCT4 protein. There are faint protein bands in prostatic tissue with the size of around 53 kDa corresponding to the same protein bands seen with the polyclonal antibodies in Figure 30 that also show weak staining when just applying the secondary antibody and thus correspond to unspecific antibody binding. With a closer look at prostatic tissue, there might be very faint bands with the same size as OCT4. Applying only the secondary goat-anti mouse antibody there is almost no staining (**b)**).

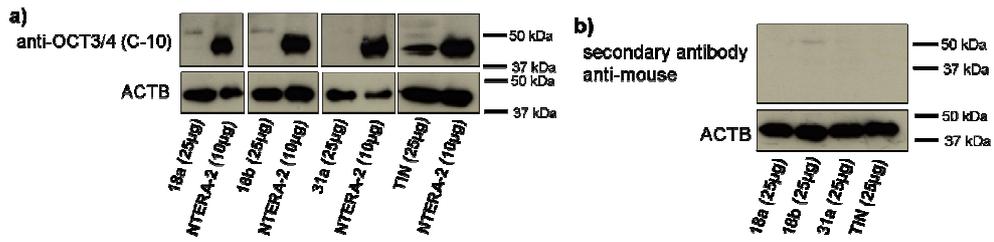


Figure 31: Western Blot results using a primary monoclonal antibody against OCT4

Shown are Western Blot results after staining with **a)** a monoclonal antibody against the C-terminus of OCT4 (C-10), **b)** only the secondary antibody. All Western Blots were stained with an ACTB antibody.

Regarding Western blot results, there is no evidence that an intact POU5F1P1 protein really is translated in prostatic tissue.

4. Discussion

A heritability of 0.42 for prostate cancer indicates a strong genetic influence on the development of this disease. The validation of susceptibility loci mapped by linkage analysis faced difficulties in the reproduction of results. Thus, it is assumed that the risk for prostate cancer arises from the impact of many independent genetic loci, each of moderate effect. For the identification of such risk loci with low penetrance, genome-wide association studies were performed. Variants tested in these studies are not necessarily causal for prostate cancer development, but rather show association by tagging true risk variants in the respective regions.

Since 2006, amongst several such independent prostate cancer risk regions in the human genome, at least three prostate cancer risk regions on chromosome 8q24.21 were identified and could be successfully confirmed by several follow-up studies. Association of variants on 8q24.21 with other types of cancer like breast cancer, colorectal cancer, ovary cancer, bladder cancer, cancer of the kidney, thyroid cancer, cancer of the larynx and lung cancer has also been shown pointing on the importance of the risk region 8q24.21 for general cancer development (Easton and Eeles, 2008; Kiemenev et al., 2008; Park et al., 2008; Wokolorczyk et al., 2008). Nevertheless, since those risk variants are located in a so-called gene desert, revealing the functional mechanism is still in progress.

Aim of this work was the investigation of the pseudogene *POU5F1P1* that corresponds to the only coding sequence on 8q24.21 as a putative prostate cancer susceptibility gene that might be the target of prostate cancer risk variants. For the evaluation of its impact on prostate cancer development, its transcription and the putative impact of risk variants on its expression level was examined. A case-control study with SNPs that affect the deduced protein sequence of *POU5F1P1* should determine if these variants represent functional variants which explain the association of the 8q24.21 risk region with prostate cancer. Finally, analysis of the existence of the *POU5F1P1* encoded protein in primary prostatic carcinoma tissue should reveal if the *POU5F1P1* protein might play a functional role in prostate cancer development.

4.1. Overexpression of *POU5F1P1* in prostatic carcinoma tissue

Analysis of the *POU5F1P1* sequence in primates showed that this pseudogene was integrated in the genome of the common ancestor of the Great Apes more than 15 million years ago. Whereas Orang Utan, Gorilla and Bonobo harbour premature stop codons and do not possess an intact coding sequence, only Chimpanzee and human carry the intact *POU5F1P1* open reading frame in their genomes. The lack of conservation in primates suggests that the pseudogene has not acquired an essential function in human evolution.

Firsova et al. suggested an expression of *POU5F1P1* in the human embryonic eye and the putative encoded protein was assumed to be located in the nucleus (Firsova et al., 2008). The authors discussed a function of this pseudogene in self-renewal and differentiation of multipotent cells in the developing eye. But because of the missing conservation of *POU5F1P1*, a functional role of this pseudogene in such a fundamental developmental process as eye development is improbable. To my knowledge, there are no further indications in the literature for a normal biological function of *POU5F1P1*.

In contrast to that, it is possible that if this pseudogene is transcribed in any cell type, it might bear an aberrant function that is similar to that of *OCT4*. A fusion of *OCT4* to *EWSR1* that activates *OCT4* was described in hidradenoma of the skin and mucoepidermoid carcinoma of the salivary glands as well as in sarcoma derived from pelvic bone (Moller et al., 2008; Yamaguchi et al., 2005). Beyond that, *OCT4* is a marker of germ cell tumours and is discussed to be expressed in bladder cancer, in lung cancer and in retinoblastoma (Atlasi et al., 2007; De Jong and Looijenga, 2006; Karoubi et al., 2009; Seigel et al., 2007). Additionally, *OCT4* is reported to be expressed in prostatic carcinoma (Sotomayor et al., 2009). Its pseudogene might bear similar functions in carcinogenesis. As a first step to test the hypothesis if *POU5F1P1* is a cancer susceptibility gene, specific expression analysis of this pseudogene in prostatic cell lines and carcinoma tissue were done.

Genomewide BLAST analysis with the *POU5F1P1* sequence revealed high homology not only to the parent gene *OCT4*, but also to other *OCT4* pseudogenes. To avoid false-positive results in expression analysis, specific PCR assays for *POU5F1P1* and *OCT4*, respectively, were designed. PCR products were sequenced to make sure not to amplify other *POU5F1P1*-related sequences. Additionally, a *POU5F1P1* PCR with RNA without reverse transcription was carried along to exclude amplification of the genomic *POU5F1P1* sequence.

I could clearly show that *OCT4* is not expressed in cell lines derived from the normal prostate or prostatic carcinoma tissue and neither in prostatic carcinoma tissue nor in surrounding prostatic tissue. In contrast to that, the pseudogene *POU5F1P1* is expressed in prostate cancer cell lines as well as in prostatic carcinoma and in some prostatic

carcinoma surrounding tissues. Real-time RT-PCR showed significant overexpression of *POU5F1P1* in prostatic carcinoma tissue up to more than 3-fold compared to carcinoma surrounding tissue rendering *POU5F1P1* a potential prostate cancer susceptibility gene that might be involved in carcinogenesis.

In contrast to these results, Sotomayor discussed weak *OCT4* expression in benign and malignant prostate glands (Sotomayor et al., 2009). *OCT4* expression is also discussed for other types of cancers like cancer of the breast, pancreas, colon as well as in normal tissue like brain and skeletal muscle (Jin et al., 1999; Monk and Holding, 2001). But because of difficulties in PCR design, it is unclear if amplification products result from *OCT4* or from its pseudogenes (Liedtke et al., 2007; Suo et al., 2005). Monsef et al. excluded an expression of *OCT4* in prostatic carcinoma tissue, but discussed an expression of *OCT4* pseudogenes in these kind of cells (Monsef et al., 2009).

In conclusion, the apparent *OCT4* expression in cancer tissue and especially in prostatic carcinoma tissue, discussed in the literature, actually could arise from the highly homologous pseudogene *POU5F1P1*. Further, the role of carcinogenic impact, discussed for *OCT4*, could be beared by its pseudogene *POU5F1P1*. Causal variants on 8q24.21, that mediate the higher risk for prostate cancer development, could have an influence on the expression level of *POU5F1P1* or, alternatively, might represent coding variants that change the encoded protein structure and its function. First of all, the mechanism that leads to the *POU5F1P1* overexpression in prostatic carcinoma tissue and a correlation of 8q24.21 risk variant genotypes with *POU5F1P1* expression level was investigated.

4.2. On the impact of 8q amplification or prostate cancer risk variants on *POU5F1P1* overexpression in prostatic carcinoma tissue

Since *POU5F1P1* is located on the 8q arm, that is often amplified in primary prostatic carcinoma (Cher et al., 1996), it was examined if its overexpression in prostatic carcinoma tissue correlates with an increase of 8q copy numbers. MLPA analysis of genomic DNA of prostatic carcinoma tissue of the investigated 35 prostate cancer patients compared to carcinoma surrounding tissues should elucidate if there is 8q amplification in the investigated prostatic carcinoma tissues. Only one patient showed 8q duplication of one chromosome. In contrast to these findings, 8q amplification is described for approximately 80 % and thus for most of prostatic carcinoma tissue (Cher et al., 1996; Nupponen et al., 1998; Visakorpi et al., 1995). The reason for this discrepancy of results might be that data in the literature result from comparative genome hybridisation (= CGH) and fluorescence in-situ hybridisation (= FISH) that might have a better resolution than MLPA analysis. For example, FISH analysis can examine the copy number variation in each cell, whereas MLPA analysis examines the sum of copy numbers of all carcinoma tissue cells. Thus,

MLPA analysis is not that sensitive for the determination of 8q amplification and putative 8q amplification might be underestimated by MLPA analysis, done in this work. Nevertheless, MLPA can determine at least 1.5-fold amplifications of chromosomal regions and thus duplication of one 8q arm per cell. And most of the patients show an overexpression of *POU5F1P1* in prostatic carcinoma that is higher than 1.5-fold compared to carcinoma surrounding tissue regarding relative *POU5F1P1* expression levels. Thus, if 8q copy number is proportional to *POU5F1P1* expression level, 1.5-fold copy number gain of 8q should be observed in MLPA analysis for most of the patients. This is not the case. Additionally, tumour 30 showed a 1.5-fold copy number gain of 8q, but its *POU5F1P1* expression level is not significantly higher than in the corresponding surrounding tissue compared to all other patients. Thus, an increased copy number of 8q is not the reason for *POU5F1P1* overexpression in the examined prostatic carcinoma. This is in accordance to Tomlinson et al. who investigated the correlation of 8q amplification with *POU5F1P1* expression in colorectal cells. They identified amplicons harbouring *POU5F1P1*, but not all colorectal cells with such an amplification of *POU5F1P1* showed *POU5F1P1* expression (Tomlinson et al., 2007).

Since *POU5F1P1* corresponds to a retrotransposed pseudogene, it lacks the original promoter from its parent gene. Thus, *POU5F1P1* expression must be under the control of promoter sequences on chromosome 8. Transcriptional activity of these promoters might be affected by risk variants on 8q24.21. For the identification of *POU5F1P1* promoters, 5'RACE experiments with RNA from prostatic carcinoma tissue and prostatic carcinoma tissue derived cell lines were performed. Twelve different transcripts could be identified arising from nine different transcriptional start sites, which are distributed over a genomic distance of approximately 850 kb. It is possible, that there might be even more promoters, contributing to the expression of this pseudogene, that could not be identified in 5'RACE experiments in this work. The promoter, driving *POU5F1P1* transcription, which is located most afar from *POU5F1P1*, is placed in the single intron of *FAM84B* that, itself, is transcribed in the opposite direction compared to *POU5F1P1*. As results of our working group show, *FAM84B* is also transcribed in prostatic carcinoma tissue and surrounding tissue. Thus, its promoter shows a bidirectional activity in these cells which is common in the human genome (Adachi and Lieber, 2002).

Since the overall expression level of *POU5F1P1* is weak compared to *HPRT1*, each of the nine promoters, contributing to the *POU5F1P1* expression, shows only weak transcriptional activity. Since 93 % of the human genome shows expression (Johnson et al., 2005) and 13 % of transcriptional initiation in the human genome occurs at unannotated loci (Kim et al., 2005), the *POU5F1P1* harbouring transcripts also might represent such kind of background transcriptional activity. The reason for this

transcriptional activity in a so-called gene desert might be caused by an active chromatin modification, possibly under the influence of risk variants on 8q24.21. Whole genomic hypomethylation often occurs in tumours (Baylin et al., 1998). And DNA demethylation and histon acetylation are linked and mark accessible chromatin structure that is characteristic for transcriptional active genomic regions (Geiman and Robertson, 2002). If the gene desert on 8q24.21 is affected by such activating chromatin modifications, other transcripts on 8q24.21 should also show this overexpression in prostatic carcinoma tissue. To examine this issue, the expression of other annotated RNAs from 8q24.21 in prostatic carcinoma tissue and in carcinoma surrounding tissue was examined.

The overexpression in prostatic carcinoma, which was shown for *POU5F1P1*, might also affect other transcripts, originating from 8q24.21. Expression analysis of splice variants derived from the EST AW183883 as well as *c-MYC*, which borders the gene desert on 8q24.21 in telomeric direction, respectively, was examined by real-time RT-PCR and semiquantitative RT-PCR. Additionally, semi-quantitative expression analysis of *FAM84B* in prostatic tissue was examined in our working group. A significant overexpression of all three transcripts in prostatic carcinoma tissue compared to carcinoma surrounding tissue could be shown. This supports the thesis that a long genomic region on 8q24.21 is overexpressed in prostatic carcinoma tissue. This is in accordance to the literature where expression of *c-MYC* and prostate cancer risk region 3, including *POU5F1P1* and the region between region 3 and *FAM84B*, in prostatic tissue as well as in cancer cell lines was reported (Fleming et al., 1986; Jia et al., 2009).

Nevertheless, there is no correlation of the expression levels of *POU5F1P1* with that from splice variants derived from the EST AW183883. Coding variants in *POU5F1P1* might show association with *POU5F1P1* expression level and known risk variants on 8q24.21 might have an influence on the expression level of *POU5F1P1* as well as splice variants derived from the EST AW183883. But statistical analysis of a putative correlation of risk variants' genotypes and the examined expression levels revealed that only one risk variant (rs12543663) shows weak correlation with *POU5F1P1* transcription levels in carcinoma surrounding tissue (Mann Whitney test for the two groups "no risk" allele and "at least one risk allele", p value: 0.0455 with a higher *POU5F1P1* expression in patients with at least one risk allele than in patients with no risk allele). Nevertheless, the p value indicates weak significance. There is no correlation of the genotype status of the other investigated risk variants on 8q24.21 with the expression levels of the investigated transcripts. In accordance to these findings, Pomerantz et al. did not see a correlation between the genotype of rs6983267 in prostate cancer risk region 3 and rs1447295 in prostate cancer risk region 1 as well as four other 8q24 risk variants and the expression levels of *c-MYC* and six other 8q24 transcripts in prostatic tissue of 280 prostate cancer

patients. Tomlinson et al. could not identify correlation of genotypes of rs6983267 or any coding SNPs in the open reading frame of *POU5F1P1* with the *POU5F1P1* expression level in colorectal cells (Tomlinson et al., 2007).

Since no evidence was identified for a direct influence of risk variants on 8q24.21 on the expression level of the gene desert and bordering sequences on 8q24.21, risk alleles may rather have an activating influence on the transcriptional status of 8q24.21 compared to non-risk alleles. This effect should be visible in an allelic difference in expression from 8q24.21 RNAs. An allelic difference in 8q24.21 expression was examined in samples that are heterozygous for at least one SNP in *POU5F1P1*. Ten out of 26 patients showed an allelic shift in *POU5F1P1* expression, possibly caused by *cis*-acting variants on 8q24.21. Since allelic difference in expression was examined by normalisation of the allelic ratio of expressed SNPs to the allelic ratio of SNPs present in genomic DNA, present results implicate that patient 30 shows different transcriptional activity between both alleles additionally to duplication of one 8q arm. Thus, this might indicate a putative effect of variants on differential allelic expression. Since nine transcriptional start sites for *POU5F1P1* were identified in this work, it is unclear which *POU5F1P1* promoters are affected by this allelic difference in expression. Such a differential allelic expression (= DAE) was shown to be common for 12 breast cancer candidate genes. Meyer et al. elucidated that two SNPs associated with an increased risk for breast cancer have a *cis*-regulatory impact on *FGFR2* gene expression, leading to this allelic imbalance in expression (Maia et al., 2009; Meyer et al., 2008). The observed difference in allelic expression of *POU5F1P1* is most probably not the cause for *POU5F1P1* overexpression in prostatic carcinoma tissue, since 31 from 35 patients show *POU5F1P1* overexpression but only ten out of 26 investigated patients show such a difference in allelic expression. Because of small sample size, a correlation of allelic difference in expression of these ten patients and the genotype of risk variants on 8q24.21 was not investigated.

Despite the possibility, that risk variants on 8q24.21 have a regulatory influence on *POU5F1P1* expression, there might be variants that affect *POU5F1P1* function which contribute to prostate cancer development.

4.3. On the impact of variants in *POU5F1P1* on prostate cancer development

Around 5 % of all human pseudogenes are expressed, and putative functions are discussed (Harrison et al., 2005). Two different mechanisms are considered how pseudogenes may function. The first mechanism is that the pseudogenic mRNA has a regulatory influence on the parent gene, concerning RNA stability or protein synthesis. For example, the mouse pseudogene *Makorin1-p1* sense transcript selectively stabilises the short form of the *Makorin1* gene mRNA (Hirotsune et al., 2003). Although the mechanism

underlying this regulation is unknown, the authors suggest a *cis*-acting RNA decay element in the 5' region of the *Makorin1* mRNA and the *Makorin1-p1* mRNA that can be bound by a putative RNA-destabilising factor. When *Makorin1-p1* RNA is expressed, it is supposed to compete for this destabilising factor, resulting in stabilising the parent *Makorin1* mRNA. Further, Korneev et al. showed that a NOS (= nitric oxide synthase) anti-sense pseudogene is expressed in the central nervous system of the snail *Lymnaea stagnalis* and showed RNA-RNA duplexes between the *nNOS* (= neuronal NOS) mRNA and the pseudo-NOS mRNA, that blocks NOS protein synthesis *in vitro* and *in vivo* (Korneev et al., 1999). Both studies showed that both sense and antisense pseudogenic transcripts may act as regulator of the parent gene in a positive or negative way. But, for the regulation of the parent gene, expression of both the parent gene and the pseudogenic mRNA should be detected in the affected cells. Since there is no *OCT4* expression in prostatic carcinoma tissue, at all, a functional role of *POU5F1P1* in the regulation of its parent gene can be excluded.

Another discussed function of pseudogenes is the translation of a functional protein. For example, the pseudogene *NANOGP8* is expressed in several human cancer cell lines and cancer tissues. It bears the function of encoding a stable protein *in vivo* and this protein is capable to promote cell proliferation *in vitro* (Zhang et al., 2006). Lin et al. interestingly describe an expression of a mouse pseudogene, designated *Oct4P1*, which is not homologous to the human *POU5F1P1*, in mouse embryonal carcinoma stem cells with 87 % sequence homology with the parent *Oct4* gene. The open reading frame encodes a putative 80 amino acid protein, possessing a 26-amino acid homology with the Oct4 protein. There are no data of an *in vivo* translation of this protein, but ectopic expression of Oct4P1 in mesenchymal stem cells showed promotion of proliferation and inhibition of differentiation *in vitro* (Lin et al., 2007).

For the human pseudogene *POU5F1P1*, which was examined in this work, Panagopoulos et al. could show a potential functionality of the putative encoded protein in transfection experiments. The *POU5F1P1* protein could activate a reporter gene less efficiently than the parent *OCT4* protein, but stronger than the control, *in vitro*. They further showed, that the putative *POU5F1P1* protein is located in the nucleus of transfected Hela cells (Panagopoulos et al., 2008). These *in vitro* findings might represent hints on a putative functionality of the pseudogenic *POU5F1P1* protein, if it is expressed *in vivo*.

There might be coding variants in the open reading frame of *POU5F1P1* that affect the functionality of the encoded protein and, thus, represent causal variants on 8q24.21 for prostate cancer development. To examine the impact of coding variants in *POU5F1P1* on prostate cancer development, an association study was performed.

529 prostate cancer patients (345 sporadic cases and 184 familial cases) and 213 controls were included in the association study with the five variants in the coding sequence of *POU5F1P1*. Additionally to SNPs in *POU5F1P1*, several prostate cancer risk variants on 8q24.21 were carried along for data analysis. Actually, none of the variants in *POU5F1P1* shows higher odds ratios than the two already known risk variants in the same region. Tomlinson et al. examined the same five SNPs in *POU5F1P1* for association with colorectal carcinoma in 940 affected individuals versus 965 controls. Colorectal cancer also shows association with risk variants on 8q24.21. Table 20 shows the adapted data of both studies with minor allele frequencies (= MAF, minor allele defined as the rare allele in controls) in controls and cases, allelic odds ratios of the ancestral allele with 95 % confidence interval and corresponding p values.

Table 20: Association of single nucleotide polymorphisms (= SNPs) in cancer risk region 3 with prostate cancer (examined in this work) and colorectal cancer (Tomlinson et al., 2007).

Shown are the two SNPs rs6983267 and rs7837328 in risk region 3 on 8q24.21 that were shown to be associated with prostate cancer and colorectal cancer and the five variants in *POU5F1P1* with their nucleotide position in the open reading frame of *POU5F1P1*, their minor allele frequencies (= MAF, minor allele is defined as rare allele in controls) in controls and cases and allelic odds ratios (= OR) of the ancestral allele with corresponding p values.

Data of this work					
Variant	<i>POU5F1P1</i>	MAF, controls [%]	MAF, cases [%]	Allelic OR (95 % CI)	p value
rs6983267	-	47,2	54,2	1.324 (1.056 - 1.659)	0.01484
rs7837328	-	39,7	45,6	0.787 (0.895 - 0.692)	0.0003
rs6998061	527	46,5	40,0	1.301 (1.037 - 1.632)	0.02298
rs13273814	545	16,2	17,9	1.130 (0.836 - 1.529)	0.42585
rs13274084	640	11,5	11,1	1.041 (0.730 - 1.483)	0.82450
rs6998254	684	43,7	50,3	1.305 (1.041 - 1.637)	0.02098
rs7002225	712	43,7	50,2	1.300 (1.037 - 1.630)	0.02289
Tomlinson et al.					
Variant	<i>POU5F1P1</i>	MAF, controls [%]	MAF, cases [%]	Allelic OR (95 % CI)	p value
rs6983267	-	49,0%	57,5%	1.406 (1.237 - 1.598)	1.8594 x 10 ⁻⁷
rs7837328	-	35.45%	42.9%	0.731 (0.579 - 0.923)	0.00824
rs6998061	527	42,9%	38,2%	1.218 (1.046 - 1.418)	0.0095
rs13273814	545	20,4%	21,3%	1.055 (0.878 - 1.269)	0.5598
rs13274084	640	13,0%	13,7%	0.948 (0.760 - 1.182)	0.6241
rs6998254	684	47,4%	52,4%	1.221 (1.049 - 1.418)	0.0085
rs7002225	712	46,7%	52,1%	1.241 (1.064 - 1.445)	0.0049

Both studies show similar odds ratios for the examined SNPs for prostate cancer and colorectal cancer. In both studies, none of the SNPs in *POU5F1P1* shows a higher odds ratio and thus a stronger association with prostate cancer or colorectal cancer, respectively, than already known risk variants in the same region of linkage disequilibrium. Results of these two studies suggest that *POU5F1P1* does not harbour single functional variants for prostate cancer or colorectal cancer.

Haplotypes of SNPs in *POU5F1P1* might represent functional units of the putative encoded amino acid sequence. The haplotype, that harbours the ancestral allele at all five variants encodes a putative protein, most similar to the original OCT4 protein and, thus, might be more functional than the protein encoded by one of the other observed haplotypes, harbouring one or more variant alleles. If *POU5F1P1* might play a role in prostate cancer development, functional haplotypes of *POU5F1P1* SNPs should possess a higher frequency in cases than in controls. Chi square tests showed that haplotype E (22122) is significantly more frequent in controls than in cases. Harboring four variant alleles, this haplotype is most different from the *OCT4* sequence, showing altogether 17 nucleotide differences, and might represent a protective haplotype concerning prostate cancer development. Haplotype A which is most similar to *OCT4* seems to be more frequent in cases than in controls. Stratification for clinical data, family history or age of diagnosis of the disease resulted in a significant higher frequency of haplotype A (11111) in familial prostate cancer cases than in controls. Since haplotype A harbours five ancestral alleles and corresponds to a haplotype, showing just 13 amino acid exchanges compared to *OCT4*, it encodes a protein that might bear rather a similar function as the parent *OCT4* protein. However, this *POU5F1P1* haplotype is in complete linkage disequilibrium with the risk allele of the SNP rs7837328, which showed the strongest association to prostate cancer development in this region, and, thus, does not exceed the association of this already known risk variant. To examine if *POU5F1P1* haplotypes represented functional units, it was investigated if the encoded *POU5F1P1* protein is present in prostatic carcinoma tissue and may be involved in prostate cancer development.

4.4. Examination of the *POU5F1P1* encoded protein in prostatic tissue

To investigate the issue, if the *POU5F1P1* protein is translated and which kind of cells in the prostate express the *POU5F1P1* protein, Immunohistochemistry experiments were performed. For protein detection, three different anti-OCT4 antibodies were used. Transfection experiments with *POU5F1P1*, done in our working group, clearly showed that all of these antibodies can bind the *POU5F1P1* encoded protein. Because of unspecific staining signals, the monoclonal antibody sc5279 was not suitable for Immunohistochemistry experiments. Staining patterns of the other two antibodies were not identical, but both suggested, that *POU5F1P1* is expressed in the cytoplasm of prostatic epithelial cells. In accordance to that, Monsef et al. observed a cytoplasmatic staining with the antibody sc8629, as well. But in contrast to our results, nearly 50 % of cells in the examined tissues showed staining. In normal prostate tissue, sc8629 stained basal cells (Monsef et al., 2009). Sotomayor et al. showed that sc8628 stains the cytoplasm of only a

few prostatic carcinoma cells (Sotomayor et al., 2009). Thus, staining patterns of all three studies look different, although the same antibodies were used. Thus, Immunohistochemistry results should be confirmed by Western Blot analysis.

Isolation and separation of the original OCT4 protein from both the NTERA-2 cell line and intratubular neoplasia of the testis (= TIN) in Western Blot experiments were successful. Protein bands of the expected size of 43 to 50 kDa are specifically stained. In contrast to that, the POU5F1P1 protein could not be detected in prostatic tissue or prostatic carcinoma tissue of five patients. There are only very faint protein bands of an approximate protein size of 43 kDa after staining with the monoclonal antibody. Therefore, Western Blot results did not confirm the positive staining results of controversial Immunohistochemistry experiments. Thus, there is no evidence of the presence of the *POU5F1P1* encoded protein in prostatic tissue. One possible explanation for the fact that the POU5F1P1 protein might not be translated is, that the mRNA of this pseudogene harbours several start codons upstream of the coding sequence of *POU5F1P1*, which are not present in the parent *OCT4* mRNA. Some of these start codons mark putative open reading frames upstream of *POU5F1P1*, which also end upstream of *POU5F1P1*. In eukaryotes, predominantly the first AUG 5' of mRNAs is used for translation initiation. But after completing translation of these preceding open reading frames, there might be reinitiation of translation at a downstream start codon (Churbanov et al., 2005;Kozak, 1995). This also might be the case at the *POU5F1P1*. In contrast to that, there are five more AUGs upstream of *POU5F1P1*. One of them is located at position -61. It marks an open reading frame which does not end upstream of *POU5F1P1*, but proceeds out of frame through the transcriptional start site of *POU5F1P1*. Transfection experiments in our working group showed, that this start codon upstream of the *POU5F1P1* open reading frame almost abolished POU5F1P1 protein synthesis compared to mRNA that did not harbour this start codon. Thus, this start codon might be responsible for the lacking protein synthesis in cells which express *POU5F1P1* mRNA.

4.5. Conclusion of this work

In summary, an overexpression of the pseudogene *POU5F1P1*, transcribed from the so-called gene desert on 8q24.21, was shown in prostatic carcinoma compared to carcinoma surrounding tissue. But a correlation of this overexpression with genotypes of prostate cancer risk variants in this genomic region could not be identified. Single coding variants in *POU5F1P1* neither could be shown to represent the causal variants on 8q24.21 for prostate cancer development. One haplotype of *POU5F1P1* variants showed a higher frequency in familial cases than in controls, suggesting that the corresponding putative encoded protein might play a role in prostate cancer development. But association of this haplotype did not exceed association of already known risk variants in the same region of linkage disequilibrium with prostate cancer. Nevertheless, studies examined that no *POU5F1P1* protein is translated in cells that express the *POU5F1P1* mRNA. Thus, results of this work suggest that *POU5F1P1* is not the target of prostate cancer risk variants on 8q24.21.

4.6. Alternative targets of prostate cancer risk variants on 8q24.21

Since the pseudogene *POU5F1P1* could not be confirmed as prostate cancer susceptibility gene that explains the association of risk variants on 8q24.21 with prostate cancer, alternative targets of risk variants in this genomic region have to be considered.

Non-coding RNAs or micro-RNAs on 8q24.21

Non-coding RNAs transcribed from 8q24.21 as splice variants derived from the EST AW183883 as well as all *POU5F1P1* harbouring transcripts, that were shown to be overexpressed in prostatic carcinoma tissue, could bear a functional impact on prostate cancer development. Amundadottir et al. discussed putative open reading frames in the mRNA sequence of splice variants derived from the EST AW183883, but they could not identify homology to already known proteins (Amundadottir et al., 2006). Further, no functional small RNAs are annotated in the sequence of AW183883 or in the *POU5F1P1* harbouring transcripts. Nevertheless, there are hints in the current literature, that long non-coding RNAs might bear putative functions. For example, Loewer et al. identified several large intergenic non-coding RNAs (= lincRNA) that are expressed in pluripotent stem cells (Loewer et al., 2010). The authors postulated a putative function of one of these lincRNAs for the pluripotency of cells. Furthermore, Chung et al. recently published data of an association of a novel long non-coding intronless RNA in 8q24.21 with prostate cancer susceptibility. This mRNA is named prostate cancer non-coding RNA 1 (designated *PRNCR1*) and is transcribed from prostate cancer risk region 2 with upregulation in some prostate cancer cells. A knockdown of this RNA resulted in a

decreased viability of the prostate cancer cell line LNCaP, and the authors discussed an involvement of this RNA in prostatic carcinogenesis, possibly by the influence of this RNA on the androgen receptor (Chung et al., 2011). *FAM84B* also might be a target of risk variants on 8q24.21. Overexpression of *FAM84B* in prostatic carcinoma tissue compared to carcinoma surrounding tissue was shown in our working group. But to my knowledge, there are no studies, which examine the potential role of *FAM84B* for prostate cancer development.

Risk variants in the gene desert on 8q24.21 alternatively might affect regulatory sequences and, thus, could mediate the risk for prostate cancer development by altering the transcriptional activity of the corresponding target genes.

The c-MYC proto-oncogene

Increased transcriptional activity of the 8q24.21 region, examined in the present work, might be an effect of an active chromatin structure and / or the presence of transcriptional factors at putative enhancer structures in this genomic region (De Santa et al., 2010; Dobi and Winston, 2007). Actually, the risk allele of the variant rs6983267 in prostate cancer risk region 3 is part of a conserved binding site of the transcription factor TCF4 (= T-cell factor 4) that corresponds to the nucleotide sequence: (A/T)(A/T)CAA(A/T)GG (the underlined base corresponds to the ancestral allele of the variant) (Tuupanen et al., 2009). TCF4 is part of the WNT-signalling pathway, which is active in colorectal cells and c-MYC is a known target of TCF4 and beta catenin (WNT = wingless-related MMTV integration site, MMTV = mouse mammary tumour virus). The respective sequence, 330 kb upstream of c-MYC, is suggested to be the strongest TCF4 binding site within 1 mb from c-MYC, rendering the SNP harbouring TCF4 binding site a putative enhancer structure for c-MYC. Chromatin Conformation Capture (= 3C) experiments evaluated an interaction of the SNP with c-MYC in colorectal cells, but not in fibroblasts (Pomerantz et al., 2009a; Tuupanen et al., 2009). The risk G-allele further shows higher affinity to TCF4 in different colorectal cell lines and shows a 1.5-fold stronger WNT-signalling response than the T-allele. Additionally, the risk allele of rs6983267 shows stronger enhancer chromatin modification and activity than the non-risk allele (Pomerantz et al., 2009a; Tuupanen et al., 2009; Wright et al., 2010). In transgenic mice, an *in vivo* activity of this putative enhancer in a tissue-specific manner according to c-MYC expression and activity of the WNT-signalling pathway could be shown. Protein-binding of the transcription factor p300 and RNA polymerase II at the risk locus was shown in a heterozygous colorectal cancer cell line (Pomerantz et al., 2009a). But if the risk variant rs6983267, indeed, is part of a c-MYC enhancer with the risk allele bearing a stronger activity than the non-risk allele, a correlation of the SNP genotype status and c-MYC expression should be

expected. Wright et al. reported an approximately two-fold higher expression of that *c-MYC* allele linked to the risk G-allele than the *c-MYC* allele which is located in *cis* of the non-risk allele. Nevertheless, a knockdown of beta catenin, which is considered to bind the *c-MYC* enhancer together with TCF4, resulted in a 60 % decreased *c-MYC* expression, but the allelic ratio of expression remained the same. Further, chromatin loop formation shows no preference of the risk allele (Wright et al., 2010). And finally, despite the observed overexpression of *c-MYC* in carcinoma tissue, until now, a correlation between the SNP genotype and *c-MYC* mRNA expression in colon carcinoma, normal colon tissue samples and in prostatic carcinoma tissue could not be revealed (Pomerantz et al., 2009b; Tuupanen et al., 2009).

Interestingly, Ahmadiyeh et al. elucidated a chromosomal interaction of 8q24.21 variants with *c-MYC* in a tissue-specific manner. The authors showed an interaction of prostate cancer risk region 2 with *c-MYC*, specifically in prostate cancer cells, an interaction of the prostate cancer risk region 3, that is also associated with colon cancer, with *c-MYC*, specifically in colorectal cells, and an interaction of the breast cancer risk region, which is located between the prostate cancer risk region 2 and 3, with *c-MYC*, specifically in breast cancer cells (Ahmadiyeh et al., 2010). These data suggest a functional mechanism of risk variants on 8q24.21 with tissue-specific activity that might explain the association of risk variants on 8q24.21 with several types of cancer. Further, Sole et al. examined correlation between the genotype of SNP rs1447295 in prostate cancer risk region 1 and *c-MYC* expression in normal prostatic tissue. Samples with the risk allele showed a significant overexpression compared to samples homozygous for the non-risk allele suggesting a correlation of genotypes of SNPs on 8q24.21 with *c-MYC* expression levels (Sole et al., 2008). In contrast to Tuupanen, Pomerantz and Wright et al., Sotelo et al. suggested that the non-risk T-allele should be responsible for an increased enhancer activity. The authors reported a repressive effect of the rs6983267 risk allele on *c-MYC* expression and discussed a cancer protective impact of the risk variant (Sotelo et al., 2010).

In summary, there might be regulatory sequences of *c-MYC* on 8q24.21, but the final evidence if enhancer structures of *c-MYC* on 8q24.21 exhibit the risk for the development of prostate cancer, as well as other types of cancer remains unclear. More studies are necessary, that examine putative tissue-specific enhancer activity that might explain the association of 8q24.21 risk variants with several types of cancer. And finally, the influence of SNP genotype status on cancer development with respect to *c-MYC* regulation has to be examined.

5. Summary

Genome-wide association studies identified several risk variants for prostate cancer. But since these risk variants mostly are located in gene poor regions, the identification of prostate cancer susceptibility genes is challenging. A so-called gene desert on 8q24.21 harbours several independently associated risk variants with odds ratios ranging from 1.3 up to 2.0. Because these risk variants are located in genomic regions of linkage disequilibrium, the causative variants for prostate cancer development as well as the functional mechanism, which contributes to prostate cancer development, are still unknown. The only expressed sequence in this genomic region with coding capacity corresponds to the pseudogene *POU5F1P1*. It shows 97 % homology to its parent gene *OCT4* and, therefore, might have a similar function. *OCT4* encodes a transcription factor that is responsible for the pluripotency of stem cells and is reported to be involved in cancer development. An expression of the pseudogene *POU5F1P1* in carcinoma cells might contribute to carcinogenesis in a similar way.

The aim of this work, therefore, was to test the hypothesis, if the pseudogene *POU5F1P1* is a prostate cancer susceptibility gene which might be the target of prostate cancer risk variants on 8q24.21. Experiments were performed to investigate the issue, if functional variants on 8q24.21 may have an impact on the expression level of *POU5F1P1*, and whether coding variants in this pseudogene are the causative variants for prostate cancer development.

Since there is no conservation of *POU5F1P1* in primates, a normal biological function of *POU5F1P1* can almost be excluded. But *POU5F1P1* showed significant overexpression in prostatic carcinoma tissue compared to carcinoma surrounding tissue and thus might play a role in prostate cancer development. This overexpression does not occur from 8q amplification since just one out of the 35 investigated patients showed a significant 8q copy number increase in prostatic carcinoma tissue compared to carcinoma surrounding tissue. The expression of *POU5F1P1* further does not arise from a single promoter that is influenced by 8q24.21 risk variants. *POU5F1P1* transcription arises from at least eight transcriptional start sites, which are scattered over a genomic region of up to 850 kb. This issue suggests that there is a general transcriptional activity of a long genomic distance on 8q24.21 in prostatic carcinoma tissue. Overexpression of the *c-MYC* gene that borders the gene desert in telomeric direction, as well as non-coding RNAs transcribed from the gene desert in prostatic carcinoma tissue confirms this hypothesis. But none of eight tested prostate cancer risk variants in the gene desert on 8q24.21 showed correlation with the expression levels of *POU5F1P1* or the other transcripts tested. There neither was

observed a predominant allelic difference in expression of 8q24.21 transcripts which correlates with *POU5F1P1* overexpression. Thus, there are no hints, that risk variants on 8q24.21 have a direct influence on the expression level of *POU5F1P1*.

Coding variants in *POU5F1P1* were tested to be the functional variants on 8q24.21. Single variants in the coding sequence of *POU5F1P1* did not show better association with prostate cancer development than already known risk variants in the same region of linkage disequilibrium. However, one *POU5F1P1* haplotype is significantly more frequent in familial prostate cancer patients than in controls. Because of complete linkage disequilibrium of already known risk variants and this *POU5F1P1* haplotype, it does not exceed association to prostate cancer of these variants. Nevertheless, this haplotype shows the highest homology of all observed *POU5F1P1* haplotypes with the parent *OCT4* gene. Therefore, it might represent a functional unit, which encodes a *POU5F1P1* protein that mediates the risk for prostate cancer development. Protein analysis done with Immunohistochemistry experiments and Western Blot analysis in prostatic carcinoma tissue resulted in inconsistent results. Since no *POU5F1P1* protein was detected in Western Blot analysis, the expressed pseudogene transcript is most probably not translated into an intact protein. In conclusion, *POU5F1P1* does not harbour functional variants which explain the association of this genomic region with an increased risk for prostate cancer development.

The observed transcription of 8q24.21 might indicate the activity of regulatory sequences located in this genomic region. Risk variants in the gene desert on 8q24.21 are postulated to affect enhancer sequences for the proto-oncogene *c-MYC* and thus might contribute to prostate cancer development by regulation of *c-MYC* expression.

6. References

1. Adachi,N, M R Lieber, 2002, Bidirectional gene organization: a common architectural feature of the human genome: *Cell*, v. 109, p. 807-809.
2. Ahmadiyah,N, M M Pomerantz, C Grisanzio, P Herman, L Jia, V Almendro, H H He, M Brown, X S Liu, M Davis, J L Caswell, C A Beckwith, A Hills, L Macconail, G A Coetzee, M M Regan, M L Freedman, 2010, 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC: *Proc.Natl.Acad.Sci.U.S.A*, v. 107, p. 9742-9746.
3. AJCC, 2010, Part IX, Genitourinary Sites, 41. Prostate, in SB Edge, DR Byrd, CC Compton, AG Fritz, FL Greene, and A Trotti (eds), *Cancer Staging Manual*: Heidelberg, Germany, Springer, p. 525-538.
4. Al Olama,AA, Z Kote-Jarai, G G Giles, M Guy, J Morrison, G Severi, D A Leongamornlert, M Tymrakiewicz, S Jhavar, E Saunders, J L Hopper, M C Southey, K R Muir, D R English, D P Dearnaley, A T Ardern-Jones, A L Hall, L T O'Brien, R A Wilkinson, E Sawyer, A Lophatananon, A Horwich, R A Huddart, V S Khoo, C C Parker, C J Woodhouse, A Thompson, T Christmas, C Ogden, C Cooper, J L Donovan, F C Hamdy, D E Neal, R A Eeles, D F Easton, 2009, Multiple loci on 8q24 associated with prostate cancer susceptibility: *Nat.Genet.*, v. 41, p. 1058-1060.
5. Amundadottir,LT, P Sulem, J Gudmundsson, A Helgason, A Baker, B A Agnarsson, A Sigurdsson, K R Benediktsdottir, J B Cazier, J Sainz, M Jakobsdottir, J Kostic, D N Magnusdottir, S Ghosh, K Agnarsson, B Birgisdottir, L Le Roux, A Olafsdottir, T Blondal, M Andresdottir, O S Gretarsdottir, J T Bergthorsson, D Gudbjartsson, A Gylfason, G Thorleifsson, A Manolescu, K Kristjansson, G Geirsson, H Isaksson, J Douglas, J E Johansson, K Balter, F Wiklund, J E Montie, X Yu, B K Suarez, C Ober, K A Cooney, H Gronberg, W J Catalona, G V Einarsson, R B Barkardottir, J R Gulcher, A Kong, U Thorsteinsdottir, K Stefansson, 2006, A common variant associated with prostate cancer in European and African populations: *Nat.Genet.*, v. 38, p. 652-658.
6. Amundadottir,LT, S Thorvaldsson, D F Gudbjartsson, P Sulem, K Kristjansson, S Arnason, J R Gulcher, J Bjornsson, A Kong, U Thorsteinsdottir, K Stefansson, 2004, Cancer as a complex phenotype: pattern of cancer distribution within and beyond the nuclear family: *PLoS.Med.*, v. 1, p. e65.
7. ar-Rushdi,A, K Nishikura, J Erikson, R Watt, G Rovera, C M Croce, 1983, Differential expression of the translocated and the untranslocated c-myc oncogene in Burkitt lymphoma: *Science*, v. 222, p. 390-393.
8. Atlasi,Y, S J Mowla, S A Ziaee, A R Bahrami, 2007, OCT-4, an embryonic stem cell marker, is highly expressed in bladder cancer: *Int.J.Cancer*, v. 120, p. 1598-1602.
9. Baylin,SB, J G Herman, J R Graff, P M Vertino, J P Issa, 1998, Alterations in DNA methylation: a fundamental aspect of neoplasia: *Adv.Cancer Res.*, v. 72, p. 141-196.
10. Bergerat,JP, J Ceraline, 2009, Pleiotropic functional properties of androgen receptor mutants in prostate cancer: *Hum.Mutat.*, v. 30, p. 145-157.

11. Berry,R, D J Schaid, J R Smith, A J French, J J Schroeder, S K McDonnell, B J Peterson, Z Y Wang, J D Carpten, S G Roberts, D J Tester, M L Blute, J M Trent, S N Thibodeau, 2000, Linkage analyses at the chromosome 1 loci 1q24-25 (HPC1), 1q42.2-43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer: *Am.J.Hum.Genet.*, v. 66, p. 539-546.
12. Cancel-Tassin,G, A Latil, A Valeri, P Mangin, G Fournier, P Berthon, O Cussenot, 2001, PCAP is the major known prostate cancer predisposing locus in families from south and west Europe: *Eur.J.Hum.Genet.*, v. 9, p. 135-142.
13. Carpten,J, N Nupponen, S Isaacs, R Sood, C Robbins, J Xu, M Faruque, T Moses, C Ewing, E Gillanders, P Hu, P Bujnovszky, I Makalowska, A Baffoe-Bonnie, D Faith, J Smith, D Stephan, K Wiley, M Brownstein, D Gildea, B Kelly, R Jenkins, G Hostetter, M Matikainen, J Schleutker, K Klinger, T Connors, Y Xiang, Z Wang, A De Marzo, N Papadopoulos, O P Kallioniemi, R Burk, D Meyers, H Gronberg, P Meltzer, R Silverman, J Bailey-Wilson, P Walsh, W Isaacs, J Trent, 2002, Germline mutations in the ribonuclease L gene in families showing linkage with HPC1: *Nat.Genet.*, v. 30, p. 181-184.
14. Carter,BS, G S Bova, T H Beaty, G D Steinberg, B Childs, W B Isaacs, P C Walsh, 1993, Hereditary prostate cancer: epidemiologic and clinical features: *J.Urol.*, v. 150, p. 797-802.
15. Cher,ML, G S Bova, D H Moore, E J Small, P R Carroll, S S Pin, J I Epstein, W B Isaacs, R H Jensen, 1996, Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping: *Cancer Res.*, v. 56, p. 3091-3102.
16. Chung,S, H Nakagawa, M Uemura, L Piao, K Ashikawa, N Hosono, R Takata, S Akamatsu, T Kawaguchi, T Morizono, T Tsunoda, Y Daigo, K Matsuda, N Kamatani, Y Nakamura, M Kubo, 2011, Association of a novel long non-coding RNA in 8q24 with prostate cancer susceptibility: *Cancer Sci.*, v. 102, p. 245-252.
17. Churbanov,A, I B Rogozin, V N Babenko, H Ali, E V Koonin, 2005, Evolutionary conservation suggests a regulatory function of AUG triplets in 5'-UTRs of eukaryotic genes: *Nucleic Acids Res.*, v. 33, p. 5512-5520.
18. Crawford,ED, 2003, Epidemiology of prostate cancer: *Urology*, v. 62, p. 3-12.
19. Dalla-Favera,R, M Bregni, J Erikson, D Patterson, R C Gallo, C M Croce, 1982, Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells: *Proc.Natl.Acad.Sci.U.S.A*, v. 79, p. 7824-7827.
20. De Jong,J, L H Looijenga, 2006, Stem cell marker OCT3/4 in tumor biology and germ cell tumor diagnostics: history and future: *Crit Rev.Oncog.*, v. 12, p. 171-203.
21. de Kok,JB, R W Roelofs, B A Giesendorf, J L Pennings, E T Waas, T Feuth, D W Swinkels, P N Span, 2005, Normalization of gene expression measurements in tumor tissues: comparison of 13 endogenous control genes: *Lab Invest*, v. 85, p. 154-159.
22. De Santa,F, I Barozzi, F Mietton, S Ghisletti, S Polletti, B K Tusi, H Muller, J Ragoussis, C L Wei, G Natoli, 2010, A large fraction of extragenic RNA pol II transcription sites overlap enhancers: *PLoS.Biol.*, v. 8, p. e1000384.

23. Devlin,B, N Risch, 1995, A comparison of linkage disequilibrium measures for fine-scale mapping: *Genomics*, v. 29, p. 311-322.
24. Dobi,KC, F Winston, 2007, Analysis of transcriptional activation at a distance in *Saccharomyces cerevisiae*: *Mol.Cell Biol.*, v. 27, p. 5575-5586.
25. Easton,DF, R A Eeles, 2008, Genome-wide association studies in cancer: *Hum.Mol.Genet.*, v. 17, p. R109-R115.
26. Edwards,BK, E Ward, B A Kohler, C Eheman, A G Zauber, R N Anderson, A Jemal, M J Schymura, I Lansdorp-Vogelaar, L C Seeff, M van Ballegooijen, S L Goede, L A Ries, 2010, Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates: *Cancer*, v. 116, p. 544-573.
27. Eeles,RA, Z Kote-Jarai, A A Al Olama, G G Giles, M Guy, G Severi, K Muir, J L Hopper, B E Henderson, C A Haiman, J Schleutker, F C Hamdy, D E Neal, J L Donovan, J L Stanford, E A Ostrander, S A Ingles, E M John, S N Thibodeau, D Schaid, J Y Park, A Spurdle, J Clements, J L Dickinson, C Maier, W Vogel, T Dork, T R Rebbeck, K A Cooney, L Cannon-Albright, P O Chappuis, P Hutter, M Zeegers, R Kaneva, H W Zhang, Y J Lu, W D Foulkes, D R English, D A Leongamornlert, M Tymrakiewicz, J Morrison, A T Arden-Jones, A L Hall, L T O'Brien, R A Wilkinson, E J Saunders, E C Page, E J Sawyer, S M Edwards, D P Dearnaley, A Horwich, R A Huddart, V S Khoo, C C Parker, N Van As, C J Woodhouse, A Thompson, T Christmas, C Ogden, C S Cooper, M C Southey, A Lophatananon, J F Liu, L N Kolonel, L Le Marchand, T Wahlfors, T L Tammela, A Auvinen, S J Lewis, A Cox, L M FitzGerald, J S Koopmeiners, D M Karyadi, E M Kwon, M C Stern, R Corral, A D Joshi, A Shahabi, S K McDonnell, T A Sellers, J Pow-Sang, S Chambers, J Aitken, R A Gardiner, J Batra, M A Kedda, F Lose, A Polanowski, B Patterson, J Serth, A Meyer, M Luedeke, K Stefflova, A M Ray, E M Lange, J Farnham, H Khan, C Slavov, A Mitkova, G Cao, D F Easton, 2009, Identification of seven new prostate cancer susceptibility loci through a genome-wide association study: *Nat.Genet.*, v. 41, p. 1116-1121.
28. Eeles,RA, Z Kote-Jarai, G G Giles, A A Olama, M Guy, S K Jugurnauth, S Mulholland, D A Leongamornlert, S M Edwards, J Morrison, H I Field, M C Southey, G Severi, J L Donovan, F C Hamdy, D P Dearnaley, K R Muir, C Smith, M Bagnato, A T Arden-Jones, A L Hall, L T O'Brien, B N Gehr-Swain, R A Wilkinson, A Cox, S Lewis, P M Brown, S G Jhavar, M Tymrakiewicz, A Lophatananon, S L Bryant, A Horwich, R A Huddart, V S Khoo, C C Parker, C J Woodhouse, A Thompson, T Christmas, C Ogden, C Fisher, C Jamieson, C S Cooper, D R English, J L Hopper, D E Neal, D F Easton, 2008, Multiple newly identified loci associated with prostate cancer susceptibility: *Nat.Genet.*, v. 40, p. 316-321.
29. Firsova,NV, I Markintantova, I Smirnova, I G Panova, G T Sukhikh, R D Zinov'eva, V I Mitashov, 2008, Identification of the OCT4-pg1 retrogene and NANOG gene expression in the human embryonic eye: *Biology Bulletin*, v. 35, p. 107-221.
30. Fleming,WH, A Hamel, R MacDonald, E Ramsey, N M Pettigrew, B Johnston, J G Dodd, R J Matusik, 1986, Expression of the c-myc protooncogene in human prostatic carcinoma and benign prostatic hyperplasia: *Cancer Res.*, v. 46, p. 1535-1538.
31. Freedman,ML, C A Haiman, N Patterson, G J McDonald, A Tandon, A Waliszewska, K Penney, R G Steen, K Ardlie, E M John, I Oakley-Girvan, A S

- Whittemore, K A Cooney, S A Ingles, D Altshuler, B E Henderson, D Reich, 2006, Admixture mapping identifies 8q24 as a prostate cancer risk locus in African-American men: *Proc.Natl.Acad.Sci.U.S.A*, v. 103, p. 14068-14073.
32. Geiman, TM, K D Robertson, 2002, Chromatin remodeling, histone modifications, and DNA methylation-how does it all fit together?: *J.Cell Biochem.*, v. 87, p. 117-125.
33. Gibbs, M, J L Stanford, R A McIndoe, G P Jarvik, S Kolb, E L Goode, L Chakrabarti, E F Schuster, V A Buckley, E L Miller, S Brandzel, S Li, L Hood, E A Ostrander, 1999, Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36: *Am.J.Hum.Genet.*, v. 64, p. 776-787.
34. Gleason, DF, G T Mellinger, 1974, Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging: *J.Urol.*, v. 111, p. 58-64.
35. Gudmundsson, J, P Sulem, D F Gudbjartsson, T Blondal, A Gylfason, B A Agnarsson, K R Benediktsdottir, D N Magnusdottir, G Orlygsdottir, M Jakobsdottir, S N Stacey, A Sigurdsson, T Wahlfors, T Tammela, J P Breyer, K M McReynolds, K M Bradley, B Saez, J Godino, S Navarrete, F Fuertes, L Murillo, E Polo, K K Aben, I M van Oort, B K Suarez, B T Helfand, D Kan, C Zanon, M L Frigge, K Kristjansson, J R Gulcher, G V Einarsson, E Jonsson, W J Catalona, J I Mayordomo, L A Kiemeny, J R Smith, J Schleutker, R B Barkardottir, A Kong, U Thorsteinsdottir, T Rafnar, K Stefansson, 2009, Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility: *Nat.Genet.*, v. 41, p. 1122-1126.
36. Gudmundsson, J, P Sulem, A Manolescu, L T Amundadottir, D Gudbjartsson, A Helgason, T Rafnar, J T Bergthorsson, B A Agnarsson, A Baker, A Sigurdsson, K R Benediktsdottir, M Jakobsdottir, J Xu, T Blondal, J Kostic, J Sun, S Ghosh, S N Stacey, M Mouy, J Saemundsdottir, V M Backman, K Kristjansson, A Tres, A W Partin, M T Albers-Akkers, M J Godino-Ivan, P C Walsh, D W Swinkels, S Navarrete, S D Isaacs, K K Aben, T Graif, J Cashy, M Ruiz-Echarri, K E Wiley, B K Suarez, J A Witjes, M Frigge, C Ober, E Jonsson, G V Einarsson, J I Mayordomo, L A Kiemeny, W B Isaacs, W J Catalona, R B Barkardottir, J R Gulcher, U Thorsteinsdottir, A Kong, K Stefansson, 2007a, Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24: *Nat.Genet.*, v. 39, p. 631-637.
37. Gudmundsson, J, P Sulem, T Rafnar, J T Bergthorsson, A Manolescu, D Gudbjartsson, B A Agnarsson, A Sigurdsson, K R Benediktsdottir, T Blondal, M Jakobsdottir, S N Stacey, J Kostic, K T Kristinsson, B Birgisdottir, S Ghosh, D N Magnusdottir, S Thorlacius, G Thorleifsson, S L Zheng, J Sun, B L Chang, J B Elmore, J P Breyer, K M McReynolds, K M Bradley, B L Yaspan, F Wiklund, P Stattin, S Lindstrom, H O Adami, S K McDonnell, D J Schaid, J M Cunningham, L Wang, J R Cerhan, J L St Sauver, S D Isaacs, K E Wiley, A W Partin, P C Walsh, S Polo, M Ruiz-Echarri, S Navarrete, F Fuertes, B Saez, J Godino, P C Weijerman, D W Swinkels, K K Aben, J A Witjes, B K Suarez, B T Helfand, M L Frigge, K Kristjansson, C Ober, E Jonsson, G V Einarsson, J Xu, H Gronberg, J R Smith, S N Thibodeau, W B Isaacs, W J Catalona, J I Mayordomo, L A Kiemeny, R B Barkardottir, J R Gulcher, U Thorsteinsdottir, A Kong, K Stefansson, 2008, Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer: *Nat.Genet.*, v. 40, p. 281-283.

38. Gudmundsson, J, P Sulem, V Steinthorsdottir, J T Bergthorsson, G Thorleifsson, A Manolescu, T Rafnar, D Gudbjartsson, B A Agnarsson, A Baker, A Sigurdsson, K R Benediksdottir, M Jakobsdottir, T Blondal, S N Stacey, A Helgason, S Gunnarsdottir, A Olafsdottir, K T Kristinsson, B Birgisdottir, S Ghosh, S Thorlacius, D Magnusdottir, G Stefansdottir, K Kristjansson, Y Bagger, R L Wilensky, M P Reilly, A D Morris, C H Kimber, A Adeyemo, Y Chen, J Zhou, W Y So, P C Tong, M C Ng, T Hansen, G Andersen, K Borch-Johnsen, T Jorgensen, A Tres, F Fuertes, M Ruiz-Echarri, L Asin, B Saez, E van Boven, S Klaver, D W Swinkels, K K Aben, T Graif, J Cashy, B K Suarez, T O van Vierssen, M L Frigge, C Ober, M H Hofker, C Wijmenga, C Christiansen, D J Rader, C N Palmer, C Rotimi, J C Chan, O Pedersen, G Sigurdsson, R Benediktsson, E Jonsson, G V Einarsson, J I Mayordomo, W J Catalona, L A Kiemeny, R B Barkardottir, J R Gulcher, U Thorsteinsdottir, A Kong, K Stefansson, 2007b, Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes: *Nat.Genet.*, v. 39, p. 977-983.
39. Haiman, CA, N Patterson, M L Freedman, S R Myers, M C Pike, A Waliszewska, J Neubauer, A Tandon, C Schirmer, G J McDonald, S C Greenway, D O Stram, L Le Marchand, L N Kolonel, M Frasco, D Wong, L C Pooler, K Ardlie, I Oakley-Girvan, A S Whittemore, K A Cooney, E M John, S A Ingles, D Altshuler, B E Henderson, D Reich, 2007, Multiple regions within 8q24 independently affect risk for prostate cancer: *Nat.Genet.*, v. 39, p. 638-644.
40. Harrison, PM, D Zheng, Z Zhang, N Carriero, M Gerstein, 2005, Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability: *Nucleic Acids Res.*, v. 33, p. 2374-2383.
41. Henriksson, M, B Luscher, 1996, Proteins of the Myc network: essential regulators of cell growth and differentiation: *Adv.Cancer Res.*, v. 68, p. 109-182.
42. Hirotsune, S, N Yoshida, A Chen, L Garrett, F Sugiyama, S Takahashi, K Yagami, A Wynshaw-Boris, A Yoshiki, 2003, An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene: *Nature*, v. 423, p. 91-96.
43. Jia, L, G Landan, M Pomerantz, R Jaschek, P Herman, D Reich, C Yan, O Khalid, P Kantoff, W Oh, J R Manak, B P Berman, B E Henderson, B Frenkel, C A Haiman, M Freedman, A Tanay, G A Coetzee, 2009, Functional enhancers at the gene-poor 8q24 cancer-linked locus: *PLoS.Genet.*, v. 5, p. e1000597.
44. Jin, T, D R Branch, X Zhang, S Qi, B Youngson, P E Goss, 1999, Examination of POU homeobox gene expression in human breast cancer cells: *Int.J.Cancer*, v. 81, p. 104-112.
45. Johns, LE, R S Houlston, 2003, A systematic review and meta-analysis of familial prostate cancer risk: *BJU.Int.*, v. 91, p. 789-794.
46. Johnson, JM, S Edwards, D Shoemaker, E E Schadt, 2005, Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments: *Trends Genet.*, v. 21, p. 93-102.
47. Kallioniemi, A, O P Kallioniemi, G Citro, G Sauter, S DeVries, R Kerschmann, P Carroll, F Waldman, 1995, Identification of gains and losses of DNA sequences in primary bladder cancer by comparative genomic hybridization: *Genes Chromosomes.Cancer*, v. 12, p. 213-219.

48. Kallioniemi,A, O P Kallioniemi, J Piper, M Tanner, T Stokke, L Chen, H S Smith, D Pinkel, J W Gray, F M Waldman, 1994, Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization: *Proc.Natl.Acad.Sci.U.S.A.*, v. 91, p. 2156-2160.
49. Karoubi,G, M Gugger, R Schmid, A Dutly, 2009, OCT4 expression in human non-small cell lung cancer: implications for therapeutic intervention: *Interact.Cardiovasc.Thorac.Surg.*, v. 8, p. 393-397.
50. Kiemenev,LA, S Thorlacius, P Sulem, F Geller, K K Aben, S N Stacey, J Gudmundsson, M Jakobsdottir, J T Bergthorsson, A Sigurdsson, T Blondal, J A Witjes, S H Vermeulen, Hulsbergen-van de Kaa CA, D W Swinkels, M Ploeg, E B Cornel, H Vergunst, T E Thorgeirsson, D Gudbjartsson, S A Gudjonsson, G Thorleifsson, K T Kristinsson, M Mouy, S Snorraddottir, D Placidi, M Campagna, C Arici, K Koppova, E Gurzau, P Rudnai, E Kellen, S Polidoro, S Guarrera, C Sacerdote, M Sanchez, B Saez, G Valdivia, C Ryk, P de Verdier, A Lindblom, K Golka, D T Bishop, M A Knowles, S Nikulasson, V Petursdottir, E Jonsson, G Geirsson, B Kristjansson, J I Mayordomo, G Steineck, S Porru, F Buntinx, M P Zeegers, T Fletcher, R Kumar, G Matullo, P Vineis, A E Kiltie, J R Gulcher, U Thorsteinsdottir, A Kong, T Rafnar, K Stefansson, 2008, Sequence variant on 8q24 confers susceptibility to urinary bladder cancer: *Nat.Genet.*, v. 40, p. 1307-1312.
51. Kim,JB, B Greber, M J Arauzo-Bravo, J Meyer, K I Park, H Zaehres, H R Scholer, 2009a, Direct reprogramming of human neural stem cells by OCT4: *Nature*, v. 461, p. 649-3.
52. Kim,JB, V Sebastiano, G Wu, M J Arauzo-Bravo, P Sasse, L Gentile, K Ko, D Ruau, M Ehrich, B D van den, J Meyer, K Hubner, C Bernemann, C Ortmeier, M Zenke, B K Fleischmann, H Zaehres, H R Scholer, 2009b, Oct4-induced pluripotency in adult neural stem cells: *Cell*, v. 136, p. 411-419.
53. Kim,TH, L O Barrera, M Zheng, C Qu, M A Singer, T A Richmond, Y Wu, R D Green, B Ren, 2005, A high-resolution map of active promoters in the human genome: *Nature*, v. 436, p. 876-880.
54. Knudson,AG, Jr., 1971, Mutation and cancer: statistical study of retinoblastoma: *Proc.Natl.Acad.Sci.U.S.A.*, v. 68, p. 820-823.
55. Korneev,SA, J H Park, M O'Shea, 1999, Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from an NOS pseudogene: *J.Neurosci.*, v. 19, p. 7711-7720.
56. Kozak,M, 1995, Adherence to the first-AUG rule when a second AUG codon follows closely upon the first: *Proc.Natl.Acad.Sci.U.S.A.*, v. 92, p. 7134.
57. Lesko,SM, L Rosenberg, S Shapiro, 1996, Family history and prostate cancer risk: *Am.J.Epidemiol.*, v. 144, p. 1041-1047.
58. LEWONTIN,RC, 1964, The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models: *Genetics*, v. 49, p. 49-67.
59. Li,H, B C Tai, 2006, RNASEL gene polymorphisms and the risk of prostate cancer: a meta-analysis: *Clin.Cancer Res.*, v. 12, p. 5713-5719.
60. Lichtenstein,P, N V Holm, P K Verkasalo, A Iliadou, J Kaprio, M Koskenvuo, E Pukkala, A Skytthe, K Hemminki, 2000, Environmental and heritable factors in the

- causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland: *N.Engl.J.Med.*, v. 343, p. 78-85.
61. Liedtke,S, J Enczmann, S Waclawczyk, P Wernet, G Kogler, 2007, Oct4 and its pseudogenes confuse stem cell research: *Cell Stem Cell*, v. 1, p. 364-366.
 62. Lin,H, A Shabbir, M Molnar, T Lee, 2007, Stem cell regulatory function mediated by expression of a novel mouse Oct4 pseudogene: *Biochem.Biophys.Res.Comm.*, v. 355, p. 111-116.
 63. Livak,KJ, T D Schmittgen, 2001, Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method: *Methods*, v. 25, p. 402-408.
 64. Loewer,S, M N Cabili, M Guttman, Y H Loh, K Thomas, I H Park, M Garber, M Curran, T Onder, S Agarwal, P D Manos, S Datta, E S Lander, T M Schlaeger, G Q Daley, J L Rinn, 2010, Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells: *Nat.Genet.*, v. 42, p. 1113-1117.
 65. Maia,AT, I Spiteri, A J Lee, M O'Reilly, L Jones, C Caldas, B A Ponder, 2009, Extent of differential allelic expression of candidate breast cancer genes is similar in blood and breast: *Breast Cancer Res.*, v. 11, p. R88.
 66. Maier,C, J Haeusler, K Herkommer, Z Vesovic, J Hoegel, W Vogel, T Paiss, 2005, Mutation screening and association study of RNASEL as a prostate cancer susceptibility gene: *Br.J.Cancer*, v. 92, p. 1159-1164.
 67. Maier,C, K Rosch, K Herkommer, S Bochum, G Cancel-Tassin, O Cussenot, J Haussler, G Assum, W Vogel, T Paiss, 2002, A candidate gene approach within the susceptibility region PCaP on 1q42.2-43 excludes deleterious mutations of the PCTA-1 gene to be responsible for hereditary prostate cancer: *Eur.Urol.*, v. 42, p. 301-307.
 68. Mattick,JS, I V Makunin, 2006, Non-coding RNA: *Hum.Mol.Genet.*, v. 15 Spec No 1, p. R17-R29.
 69. Meyer,KB, A T Maia, M O'Reilly, A E Teschendorff, S F Chin, C Caldas, B A Ponder, 2008, Allele-specific up-regulation of FGFR2 increases susceptibility to breast cancer: *PLoS.Biol.*, v. 6, p. e108.
 70. Moller,E, G Stenman, N Mandahl, H Hamberg, L Molne, J J van den Oord, O Brosjo, F Mertens, I Panagopoulos, 2008, POU5F1, encoding a key regulator of stem cell pluripotency, is fused to EWSR1 in hidradenoma of the skin and mucoepidermoid carcinoma of the salivary glands: *J.Pathol.*, v. 215, p. 78-86.
 71. Monk,M, C Holding, 2001, Human embryonic genes re-expressed in cancer cells: *Oncogene*, v. 20, p. 8085-8091.
 72. Monsef,N, M Soller, M Isaksson, P A Abrahamsson, I Panagopoulos, 2009, The expression of pluripotency marker Oct 3/4 in prostate cancer and benign prostate hyperplasia: *Prostate*, v. 69, p. 909-916.
 73. Nesbit,CE, J M Tersak, E V Prochownik, 1999, MYC oncogenes and human neoplastic disease: *Oncogene*, v. 18, p. 3004-3016.

74. Nichols, J, B Zevnik, K Anastassiadis, H Niwa, D Klewe-Nebenius, I Chambers, H Scholer, A Smith, 1998, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4: *Cell*, v. 95, p. 379-391.
75. Nupponen, NN, L Kakkola, P Koivisto, T Visakorpi, 1998, Genetic alterations in hormone-refractory recurrent prostate carcinomas: *Am.J.Pathol.*, v. 153, p. 141-148.
76. Panagopoulos, I, E Moller, A Collin, F Mertens, 2008, The POU5F1P1 pseudogene encodes a putative protein similar to POU5F1 isoform 1: *Oncol.Rep.*, v. 20, p. 1029-1033.
77. Park, SL, S C Chang, L Cai, C Cordon-Cardo, B G Ding, S Greenland, S K Hussain, Q Jiang, S Liu, M L Lu, J T Mao, H Morgenstern, L N Mu, L J Ng, A Pantuck, J Rao, V E Reuter, D P Tashkin, N C You, C Q Yu, S Z Yu, J K Zhao, A Belldegrin, Z F Zhang, 2008, Associations between variants of the 8q24 chromosome and nine smoking-related cancer sites: *Cancer Epidemiol.Biomarkers Prev.*, v. 17, p. 3193-3202.
78. Pomerantz, MM, N Ahmadiyah, L Jia, P Herman, M P Verzi, H Doddapaneni, C A Beckwith, J A Chan, A Hills, M Davis, K Yao, S M Kehoe, H J Lenz, C A Haiman, C Yan, B E Henderson, B Frenkel, J Barretina, A Bass, J Taberero, J Baselga, M M Regan, J R Manak, R Shivdasani, G A Coetzee, M L Freedman, 2009a, The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer: *Nat.Genet.*, v. 41, p. 882-884.
79. Pomerantz, MM, C A Beckwith, M M Regan, S K Wyman, G Petrovics, Y Chen, D J Hawsworth, F R Schumacher, L Mucci, K L Penney, M J Stampfer, J A Chan, K G Ardlie, B R Fritz, R K Parkin, D W Lin, M Dyke, P Herman, S Lee, W K Oh, P W Kantoff, M Tewari, D G McLeod, S Srivastava, M L Freedman, 2009b, Evaluation of the 8q24 prostate cancer risk locus and MYC expression: *Cancer Res.*, v. 69, p. 5568-5574.
80. Rennert, H, C M Zeigler-Johnson, K Addya, M J Finley, A H Walker, E Spangler, D G Leonard, A Wein, S B Malkowicz, T R Rebbeck, 2005, Association of susceptibility alleles in ELAC2/HPC2, RNASEL/HPC1, and MSR1 with prostate cancer severity in European American and African American men: *Cancer Epidemiol.Biomarkers Prev.*, v. 14, p. 949-957.
81. Schaid, DJ, 2004, The complex genetic epidemiology of prostate cancer: *Hum.Mol.Genet.*, v. 13 Spec No 1, p. R103-R121.
82. Schroder, FH, M J Roobol, G L Andriole, N Fleshner, 2009, Defining increased future risk for prostate cancer: evidence from a population based screening cohort: *J.Urol.*, v. 181, p. 69-74.
83. Schumacher, FR, H S Feigelson, D G Cox, C A Haiman, D Albanes, J Buring, E E Calle, S J Chanock, G A Colditz, W R Diver, A M Dunning, M L Freedman, J M Gaziano, E Giovannucci, S E Hankinson, R B Hayes, B E Henderson, R N Hoover, R Kaaks, T Key, L N Kolonel, P Kraft, L Le Marchand, J Ma, M C Pike, E Riboli, M J Stampfer, D O Stram, G Thomas, M J Thun, R Travis, J Virtamo, G Andriole, E Gelmann, W C Willett, D J Hunter, 2007, A common 8q24 variant in prostate and breast cancer from a large nested case-control study: *Cancer Res.*, v. 67, p. 2951-2956.

84. Seigel,GM, A S Hackam, A Ganguly, L M Mandell, F Gonzalez-Fernandez, 2007, Human embryonic and neuronal stem cell markers in retinoblastoma: *Mol.Vis.*, v. 13, p. 823-832.
85. Severi,G, V M Hayes, E J Padilla, D R English, M C Southey, R L Sutherland, J L Hopper, G G Giles, 2007, The common variant rs1447295 on chromosome 8q24 and prostate cancer risk: results from an Australian population-based case-control study: *Cancer Epidemiol.Biomarkers Prev.*, v. 16, p. 610-612.
86. Sole,X, P Hernandez, M L de Heredia, L Armengol, B Rodriguez-Santiago, L Gomez, C A Maxwell, F Aguilo, E Condom, J Abril, L Perez-Jurado, X Estivill, V Nunes, G Capella, S B Gruber, V Moreno, M A Pujana, 2008, Genetic and genomic analysis modeling of germline c-MYC overexpression and cancer susceptibility: *BMC.Genomics*, v. 9, p. 12.
87. Sotelo,J, D Esposito, M A Duhagon, K Banfield, J Mehalko, H Liao, R M Stephens, T J Harris, D J Munroe, X Wu, 2010, Long-range enhancers on 8q24 regulate c-Myc: *Proc.Natl.Acad.Sci.U.S.A.*, v. 107, p. 3001-3005.
88. Sotomayor,P, A Godoy, G J Smith, W J Huss, 2009, Oct4A is expressed by a subpopulation of prostate neuroendocrine cells: *Prostate*, v. 69, p. 401-410.
89. Sun,J, F C Hsu, A R Turner, S L Zheng, B L Chang, W Liu, W B Isaacs, J Xu, 2006, Meta-analysis of association of rare mutations and common sequence variants in the MSR1 gene and prostate cancer risk: *Prostate*, v. 66, p. 728-737.
90. Sun,J, S L Zheng, F Wiklund, S D Isaacs, L D Purcell, Z Gao, F C Hsu, S T Kim, W Liu, Y Zhu, P Stattin, H O Adami, K E Wiley, L Dimitrov, J Sun, T Li, A R Turner, T S Adams, J Adolfsson, J E Johansson, J Lowey, B J Trock, A W Partin, P C Walsh, J M Trent, D Duggan, J Carpten, B L Chang, H Gronberg, W B Isaacs, J Xu, 2008, Evidence for two independent prostate cancer risk-associated loci in the HNF1B gene at 17q12: *Nat.Genet.*, v. 40, p. 1153-1155.
91. Suo,G, J Han, X Wang, J Zhang, Y Zhao, Y Zhao, J Dai, 2005, Oct4 pseudogenes are transcribed in cancers: *Biochem.Biophys.Res.Comm.*, v. 337, p. 1047-1051.
92. Suuriniemi,M, I Agalliu, D J Schaid, B Johanneson, S K McDonnell, L Iwasaki, J L Stanford, E A Ostrander, 2007, Confirmation of a positive association between prostate cancer risk and a locus at chromosome 8q24: *Cancer Epidemiol.Biomarkers Prev.*, v. 16, p. 809-814.
93. Takahashi,K, S Yamanaka, 2006, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors: *Cell*, v. 126, p. 663-676.
94. Thomas,G, K B Jacobs, M Yeager, P Kraft, S Wacholder, N Orr, K Yu, N Chatterjee, R Welch, A Hutchinson, A Crenshaw, G Cancel-Tassin, B J Staats, Z Wang, J Gonzalez-Bosquet, J Fang, X Deng, S I Berndt, E E Calle, H S Feigelson, M J Thun, C Rodriguez, D Albanes, J Virtamo, S Weinstein, F R Schumacher, E Giovannucci, W C Willett, O Cussenot, A Valeri, G L Andriole, E D Crawford, M Tucker, D S Gerhard, J F Fraumeni, Jr., R Hoover, R B Hayes, D J Hunter, S J Chanock, 2008, Multiple loci identified in a genome-wide association study of prostate cancer: *Nat.Genet.*, v. 40, p. 310-315.
95. Tomlinson,I, E Webb, L Carvajal-Carmona, P Broderick, Z Kemp, S Spain, S Penegar, I Chandler, M Gorman, W Wood, E Barclay, S Lubbe, L Martin, G Sellick, E Jaeger, R Hubner, R Wild, A Rowan, S Fielding, K Howarth, A Silver, W Atkin, K Muir, R Logan, D Kerr, E Johnstone, O Sieber, R Gray, H Thomas, J Peto,

- J B Cazier, R Houlston, 2007, A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21: *Nat.Genet.*, v. 39, p. 984-988.
96. Tuupanen,S, M Turunen, R Lehtonen, O Hallikas, S Vanharanta, T Kivioja, M Bjorklund, G Wei, J Yan, I Niittymaki, J P Mecklin, H Jarvinen, A Ristimaki, M Di Bernardo, P East, L Carvajal-Carmona, R S Houlston, I Tomlinson, K Palin, E Ukkonen, A Karhu, J Taipale, L A Aaltonen, 2009, The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling: *Nat.Genet.*, v. 41, p. 885-890.
97. UICC, 2009, Urological tumours, Prostate, in Leslie H.Sobin, Mary K., Gospodarowicz, and Christian Wittekind (eds), *TNM Classification of Malignant Tumours: West Sussex, England, Wiley-Blackwell*, p. 243-248.
98. Van Den,BC, X Y Guan, D Von Hoff, R Jenkins, Bittner, C Griffin, O Kallioniemi, Visakorpi, McGill, J Herath, ., 1995, DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications: *Clin.Cancer Res.*, v. 1, p. 11-18.
99. Visakorpi,T, A H Kallioniemi, A C Syvanen, E R Hyytinen, R Karhu, T Tammela, J J Isola, O P Kallioniemi, 1995, Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization: *Cancer Res.*, v. 55, p. 342-347.
100. Wokolorczyk,D, B Gliniewicz, A Sikorski, E Zlowocka, B Masojc, T Debniak, J Matyjasik, M Mierzejewski, K Medrek, D Oszutowska, J Suchy, J Gronwald, U Teodorczyk, T Huzarski, T Byrski, A Jakubowska, B Gorski, W T van de, S Walczak, S A Narod, J Lubinski, C Cybulski, 2008, A range of cancers is associated with the rs6983267 marker on chromosome 8: *Cancer Res.*, v. 68, p. 9982-9986.
101. Wright,JB, S J Brown, M D Cole, 2010, Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells: *Mol.Cell Biol.*, v. 30, p. 1411-1420.
102. Xu,J, L Dimitrov, B L Chang, T S Adams, A R Turner, D A Meyers, R A Eeles, D F Easton, W D Foulkes, J Simard, G G Giles, J L Hopper, L Mahle, P Moller, T Bishop, C Evans, S Edwards, J Meitz, S Bullock, Q Hope, C L Hsieh, J Halpern, R N Balise, I Oakley-Girvan, A S Whittemore, C M Ewing, M Gielzak, S D Isaacs, P C Walsh, K E Wiley, W B Isaacs, S N Thibodeau, S K McDonnell, J M Cunningham, K E Zarfes, S Hebring, D J Schaid, D M Friedrichsen, K Deutsch, S Kolb, M Badzioch, G P Jarvik, M Janer, L Hood, E A Ostrander, J L Stanford, E M Lange, J L Beebe-Dimmer, C E Mohai, K A Cooney, T Ikonen, A Baffoe-Bonnie, H Fredriksson, M P Matikainen, T L Tammela, J Bailey-Wilson, J Schleutker, C Maier, K Herkommer, J J Hoegel, W Vogel, T Paiss, F Wiklund, M Emanuelsson, E Stenman, B A Jonsson, H Gronberg, N J Camp, J Farnham, L A Cannon-Albright, D Seminara, 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.*, v. 77, p. 219-229.
103. Xu,J, S L Zheng, B Chang, J R Smith, J D Carpten, O C Stine, S D Isaacs, K E Wiley, L Henning, C Ewing, P Bujnovszky, E R Bleeker, P C Walsh, J M Trent, D A Meyers, W B Isaacs, 2001, Linkage of prostate cancer susceptibility loci to chromosome 1: *Hum.Genet.*, v. 108, p. 335-345.

104. Xu,J, S L Zheng, A Komiya, J C Mychaleckyj, S D Isaacs, J J Hu, D Sterling, E M Lange, G A Hawkins, A Turner, C M Ewing, D A Faith, J R Johnson, H Suzuki, P Bujnovszky, K E Wiley, A M DeMarzo, G S Bova, B Chang, M C Hall, D L McCullough, A W Partin, V S Kassabian, J D Carpten, J E Bailey-Wilson, J M Trent, J Ohar, E R Bleecker, P C Walsh, W B Isaacs, D A Meyers, 2002, Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk: *Nat.Genet.*, v. 32, p. 321-325.
105. Yamaguchi,S, Y Yamazaki, Y Ishikawa, N Kawaguchi, H Mukai, T Nakamura, 2005, EWSR1 is fused to POU5F1 in a bone tumor with translocation t(6;22)(p21;q12): *Genes Chromosomes.Cancer*, v. 43, p. 217-222.
106. Yeager,M, N Chatterjee, J Ciampa, K B Jacobs, J Gonzalez-Bosquet, R B Hayes, P Kraft, S Wacholder, N Orr, S Berndt, K Yu, A Hutchinson, Z Wang, L Amundadottir, H S Feigelson, M J Thun, W R Diver, D Albanes, J Virtamo, S Weinstein, F R Schumacher, G Cancel-Tassin, O Cussenot, A Valeri, G L Andriole, E D Crawford, C A Haiman, B Henderson, L Kolonel, L Le Marchand, A Siddiq, E Riboli, T J Key, R Kaaks, W Isaacs, S Isaacs, K E Wiley, H Gronberg, F Wiklund, P Stattin, J Xu, S L Zheng, J Sun, L J Vatten, K Hveem, M Kumle, M Tucker, D S Gerhard, R N Hoover, J F Fraumeni, Jr., D J Hunter, G Thomas, S J Chanock, 2009, Identification of a new prostate cancer susceptibility locus on chromosome 8q24: *Nat.Genet.*, v. 41, p. 1055-1057.
107. Yeager,M, N Orr, R B Hayes, K B Jacobs, P Kraft, S Wacholder, M J Minichiello, P Fearnhead, K Yu, N Chatterjee, Z Wang, R Welch, B J Staats, E E Calle, H S Feigelson, M J Thun, C Rodriguez, D Albanes, J Virtamo, S Weinstein, F R Schumacher, E Giovannucci, W C Willett, G Cancel-Tassin, O Cussenot, A Valeri, G L Andriole, E P Gelmann, M Tucker, D S Gerhard, J F Fraumeni, Jr., R Hoover, D J Hunter, S J Chanock, G Thomas, 2007, Genome-wide association study of prostate cancer identifies a second risk locus at 8q24: *Nat.Genet.*, v. 39, p. 645-649.
108. Zaehres,H, H R Scholer, 2007, Induction of pluripotency: from mouse to human: *Cell*, v. 131, p. 834-835.
109. Zhang,J, X Wang, M Li, J Han, B Chen, B Wang, J Dai, 2006, NANOGP8 is a retrogene expressed in cancers: *FEBS J.*, v. 273, p. 1723-1730.

Table 21: MLPA data, part I

Shown are the positions of 33 MLPA probes concerning the proximity to known genes and their chromosomal position (= Chr pos). The different columns correspond to 34 prostate carcinoma patients. Ratio values between 0.7 and 1.3 are highlighted in blue and correspond to normal ranges. Values higher than 1.3 and lower than 0.7 are highlighted in green or red, respectively.

Gene	Chr_pos	1	2	3	4	5	6	7	9	10	11	12	13	14	15	16	17	18
DLGAP2 probe 1099-L0669	08p23	1.04	0.84	1.03	0.96	1.02	0.74	0.83	0.82	0.97	0.8	1.01	0.75	0.88	1.01	1.04	0.89	0.78
MFHAS1 probe 1096-L0666	08p23_1	1	0.89	1.03	0.82	1.05	0.74	0.88	0.77	0.9	0.64	0.98	0.82	0.76	0.9	0.85	0.99	0.83
MSRA probe 1202-L0787	08p23_1	0.95	0.93	1.04	0.87	1.05	0.72	0.92	0.91	1.1	0.81	1	0.75	0.66	1.01	1.01	0.93	0.75
GATA4 probe 1203-L0788	08p23_1	0.75	0.86	0.92	0.76	1.01	0.74	0.91	0.89	0.93	0.77	1.04	0.75	0.73	1.01	0.94	0.99	0.85
CTSB probe 1197-L0766	08p22	0.87	0.94	1.05	0.94	1.05	0.77	0.83	0.87	0.89	0.71	0.98	0.88	0.68	0.97	0.93	1.04	0.77
TUSC3 probe 1174-L0668	08p22	0.84	0.83	1.12	0.89	0.98	0.71	0.85	0.93	1.01	0.77	1	0.76	0.65	0.98	1.04	0.98	0.8
ChGn probe 1242-L0789	08p21_3	0.82	0.92	1.04	0.92	0.93	0.76	0.89	0.77	0.86	0.6	0.94	0.77	0.7	0.98	0.92	1	0.87
FGFR1 probe 1046-L0624	08p11_2	1.08	0.95	1.09	1.1	1.14	1.08	0.99	0.97	1.08	0.83	1	0.8	0.92	1.13	1.07	1.02	0.75
PRKDC probe 0545-L0114	08q11	1.05	1.15	1.03	1.31	1.13	1.02	1.13	1.17	1.03	0.75	0.98	0.76	0.92	0.99	0.92	0.99	1.01
MDS probe 1037-L0621	08q11	0.97	1.11	1	1.26	1	0.97	1.05	1.03	0.94	0.98	0.97	0.93	0.97	0.91	0.95	0.96	0.9
CHD7 probe 6752-L06356	08q12_2	1.1	1.26	1.16	1.45	1.11	1.1	1.2	1.27	1.08	1.15	1.06	0.89	1.1	1.02	1.01	0.85	0.83
MYBL1 probe 7915-L12249	08q13	0.96	1.08	1.14	1.35	1.07	1.02	1.03	1.18	1.04	1.15	1.02	0.79	1.09	0.96	0.99	0.9	0.97
NCOA2 probe 9937-L12248	08q13	1.13	1.2	1.18	1.41	1.22	1.11	1.1	1.23	1.13	1	1.11	1.06	1.21	1.05	1.05	0.88	0.81
TPD52 probe 1116-L0620	08q21	0.93	1.03	1.04	1.15	1.02	0.95	0.99	1.01	1.03	1.07	1.01	0.76	1.01	0.98	0.87	0.98	0.97
E2F5 probe 1039-L0617	08q21_13	0.98	1.12	1.12	1.12	1.09	1	0.97	1.1	0.9	0.96	1.05	0.76	0.77	1.02	0.86	1.03	1.11
RAD54B probe 1040-L0618	08q21_3	0.95	1.12	1.03	1.29	1.03	1	1.1	1.08	1.11	0.97	1	0.97	1.09	0.92	0.98	0.95	0.96
LRP12 probe 1045-L0615	08q22_2	0.92	1.08	1.02	1.22	1.05	0.98	1.08	1.18	1.18	1.11	1.09	0.95	0.97	1.02	0.97	0.98	0.99
E1F3S6 probe 1198-L0767	08q22-q23	0.99	1.06	1.07	1.23	1.06	0.98	0.99	1.06	1.1	1.03	1.01	0.97	0.92	0.99	0.97	0.97	0.94
E1F3S3 probe 1108-L0679	08q24_11	1.01	1.13	1.06	1.15	1.01	1	1.13	1.17	0.99	1.08	1.06	0.96	0.89	0.94	1	0.96	0.98
EXT1 probe 7792-L07528	08q24	1	1.1	0.99	1.17	1.03	0.96	1.09	1.07	0.89	1.05	1.01	0.94	0.94	0.93	0.97	0.96	1.03
RNF139 probe 1048-L0612	08q24	1.08	1.23	1.27	1.28	1.27	1.13	1.12	1.09	0.89	0.97	1.06	1.11	0.83	1.01	0.82	0.96	0.96
MYC probe 0672-L0169	08q24_12	0.98	1.06	0.95	1.11	1.1	0.99	1.02	1.01	0.93	1.06	1	0.95	0.92	0.94	0.98	1.03	1.03
MYC probe 0580-L0625	08q24_12	1.06	1.13	1	1.28	1.02	1.08	1.14	1.16	1.09	1.15	1.07	0.98	1.03	1.03	1.09	0.9	0.84
DDEF1 probe 1041-L0614	08q24_1	0.95	1.1	0.97	1.11	1.02	0.99	1.04	1.09	1.04	1.08	1.01	0.97	0.97	1	0.98	0.94	0.96
KCNQ3 probe 7315-L06163	08q24_2	0.9	1.08	1	1.12	1.02	1.01	1.01	1.08	0.94	1.07	1.02	0.98	1	0.95	0.91	1.1	1.06
SLA probe 0488-L0080	08q24_2	0.94	1.01	0.94	1.17	1	0.95	1.01	1.02	0.97	1	1.02	0.94	0.86	1	0.98	1.02	1.01
KHDRBS3 probe 1047-L0622	08q24_2	1.09	1.22	1.12	1.38	1.14	1.09	1.19	1.12	1.1	1.18	1.06	1.01	1.03	1.08	1.08	0.93	0.94
KCNK9 probe 9999-L04547	08q24_3	1.01	1.16	0.97	1.34	1.09	1.02	1.12	1.08	1.15	1.2	1.01	0.93	1.08	1.05	1.18	0.84	0.77
PTK2 probe 1042-L0791	08q24	0.99	1.11	1	1.17	1.02	1	1.04	1.03	1.07	1.19	1.02	0.98	0.85	1.04	1.1	1.02	0.97
PTP4A3 probe 1038-L0606	08q24_3	0.9	1.07	0.93	1.17	1	0.9	0.94	1.07	0.93	0.92	0.97	0.95	0.95	0.9	0.91	1	0.97
RECQL4 probe 1052-L0610	08q24_3	0.96	1.17	0.91	1.19	1.01	0.96	1.13	1.1	1.11	1.45	0.99	0.93	1.14	1.02	1.06	0.94	0.85
RECQL4 probe 1036-L0609	08q24_3	1	1.08	0.84	1.08	0.95	0.97	1.04	0.99	0.9	1.07	0.96	1.01	1.25	0.95	0.97	0.99	1.03
Control_probe 1160-L0716	03p26	0.99	1	1	1.16	1.14	1.12	1.18	1.21	1.04	1.07	0.98	1.06	1.02	1.04	1.01	0.95	0.97
Control_probe 0797-L0463	05q31_1	0.87	1.01	1.04	1.06	0.99	1.01	1.03	0.84	1.03	1.17	1.03	1.01	1	1.01	0.99	0.92	0.91
Control_probe 0980-L0567	11p12	1	1.03	1.08	1.08	1.04	0.99	1.12	1.09	1	0.99	1.02	0.92	0.78	0.95	0.97	1	1.03
Control_probe 0976-L0563	11p13	1.04	1.03	1.11	1	1.06	1.02	0.94	1	0.97	0.84	1.02	1.02	0.83	1.11	0.94	1.04	1.01
Control_probe 0434-L0020	11q13	1.05	0.97	0.94	0.96	0.99	0.96	0.96	1	0.86	0.95	0.99	1	0.91	0.99	0.91	1.16	1.11
Control_probe 0678-L0124	12p13	1	0.79	0.98	0.88	0.96	0.94	0.88	0.95	0.92	1	0.95	0.91	1.03	1	1.02	1.03	1.08
Control_probe 0798-L0316	13q32	1.02	0.92	1	1.12	1.06	1.01	1	1.05	1.04	1.18	1.03	1.01	1.03	1.01	1.03	0.99	1
Control_probe 0871-L0461	13q34	0.88	0.88	1.01	0.98	1	0.94	0.99	0.97	0.81	0.96	1	0.94	0.92	1	1	0.96	0.83
Control_probe 1463-L0928	17p12	0.96	1	0.83	0.98	0.96	1	1.09	0.99	1	1.14	1	0.92	1.1	0.97	1.02	1.01	0.98

Table 22: MLPA data, part II

Depicted data are according to Table 21.

Gene	Chr. pos.	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
DLGAP2 probe 1099-L0669	08p23	0.98	1.07	0.85	1.08	1.03	0.98	0.76	0.95	1	0.74	0.89	0.69	0.72	0.82	1.06	0.74	1.07
MFHAS1 probe 1096-L0666	08p23.1	0.99	0.99	0.84	0.96	0.85	0.86	0.68	0.93	0.86	0.66	0.86	0.61	0.66	0.92	0.87	0.79	0.85
MSRA probe 1202-L0787	08p23.1	0.94	0.99	0.97	0.99	0.99	1.01	0.66	1.04	0.96	0.65	0.98	0.66	0.66	0.87	0.89	0.86	0.79
GATA4 probe 1203-L0788	08p23.1	1.03	1.22	0.83	1	1	0.89	0.7	0.96	0.99	0.68	0.82	0.62	0.67	0.88	0.95	0.69	0.9
CTSB probe 1197-L0786	08p22	1.12	1.31	0.81	0.95	0.91	0.87	0.73	0.9	0.94	0.69	0.93	0.83	0.72	0.81	0.84	0.71	0.98
TUSC3 probe 1174-L0668	08p22	0.98	0.95	0.9	0.96	0.92	0.97	0.72	1.06	1	0.67	0.91	0.79	0.71	0.86	0.99	0.88	0.97
ChGn probe 1242-L0789	08p21.3	1.05	1.12	0.86	1	0.85	0.88	0.66	1	0.94	0.69	0.82	0.59	0.68	1.01	1.03	0.76	1.02
FGFR1 probe 1046-L0624	08p11.2	0.96	0.97	0.93	0.95	1.01	0.95	0.95	0.92	1.08	0.63	0.87	0.64	0.59	0.95	1.08	0.8	0.9
PRKDC probe 0545-L0114	08q11	0.99	1.01	1.03	1.01	0.92	1.02	1.09	0.95	0.97	1.01	1.04	1.36	0.95	1.1	0.93	0.89	1.02
MOS probe 1037-L0621	08q11	0.95	1.03	0.98	1.08	0.88	0.97	1.1	0.94	0.97	0.99	1.02	1.58	0.99	1.03	0.88	1.07	0.92
CHD7 probe 6752-L06356	08q12.2	0.7	0.35	1.04	1.01	1.06	1.1	1.12	0.98	0.99	0.97	1.07	1.51	1	1.12	0.96	0.76	0.96
MYBL1 probe 7915-L12249	08q13	1.01	1	0.98	0.99	0.92	1	1.08	0.92	1.01	0.93	0.98	1.59	1.02	1	0.94	0.98	1.05
NCOA2 probe 9937-L12248	08q13	0.73	0.42	1.03	1.03	1.05	1.1	1.12	0.95	1.03	1.04	1.06	1.33	1.09	0.96	0.91	0.94	1.06
TPD52 probe 1116-L0620	08q21	0.97	0.94	0.97	0.99	1.03	0.98	1.13	0.99	0.98	1	1.05	1.68	1.09	1.09	0.93	0.94	1.05
E2F5 probe 1039-L0617	08q21.13	1.17	1.36	0.89	1.02	0.92	1.01	1.1	0.95	0.99	0.99	1.07	1.44	1.03	1.14	0.88	0.79	0.89
RAD54B probe 1040-L0618	08q21.3	0.95	0.93	0.93	1.01	1.02	1.02	1.07	1.04	0.98	0.98	1.16	1.51	1.13	1.09	0.96	1.07	0.86
LRP12 probe 1045-L0615	08q22.2	0.95	0.96	0.94	1.02	0.98	1.03	1.1	0.95	1.04	0.97	1.07	1.45	1.05	1.03	0.9	1.18	0.99
EIF3S6 probe 1198-L0767	08q22-q23	0.93	0.93	0.99	1	0.97	1	1.11	0.99	1.02	0.96	0.99	1.76	1.02	1.02	0.97	0.91	1.06
EIF3S3 probe 1108-L0679	08q24.11	0.94	0.79	1.04	1.04	1.04	1.05	1.14	1.06	1	1.02	1	1.48	0.97	1.01	0.97	1.1	1.03
EXT1 probe 7792-L07528	08q24	0.98	0.94	1.01	0.99	1.02	0.97	1.08	0.98	0.98	1	1.1	1.41	1.01	1.05	0.98	1.09	0.88
RNF139 probe 1048-L0612	08q24	0.99	0.82	1.01	1.08	0.99	1.11	1.1	0.98	1.02	1.01	1.03	1.38	1.11	0.94	0.95	0.98	0.95
MYC probe 0672-L0169	08q24.12	1.13	1.47	0.9	0.97	0.93	0.94	1.13	0.93	1.03	1	1.01	1.43	1.14	1.03	0.89	0.86	0.9
MYC probe 0580-L0625	08q24.12	0.78	0.54	1.04	1.05	1.04	1.01	1.09	1.06	1.01	0.98	0.95	1.7	1.14	1.03	0.96	1.02	0.94
DDEF1 probe 1041-L0614	08q24.1	0.95	0.81	0.99	1.01	0.97	0.96	1.05	1.01	1.02	1	1.04	1.46	1	0.95	0.91	1.11	1.03
KCNQ3 probe 7315-L06163	08q24.2	1.07	1.21	1	1.02	0.97	1	1.08	1.04	0.96	1	1.03	1.31	0.95	1.07	1	0.96	0.99
SLA probe 0488-L0080	08q24.2	1.05	1.05	0.93	0.98	0.92	0.96	1.11	0.99	1	0.97	0.99	1.53	1.01	0.99	0.91	0.92	0.95
KHDRBS3 probe 1047-L0622	08q24.2	0.87	0.75	1.05	1.08	1.11	1.12	1.11	1.01	1.05	1.01	1.03	1.34	1.01	0.91	0.94	1.09	1.07
KCNK9 probe 9399-L04547	08q24.3	0.74	0.44	1.1	1.18	1.05	1	1.02	0.96	1.04	1	1.02	1.45	1.15	1.1	0.95	1.04	1.12
PTK2 probe 1042-L0791	08q24	0.93	0.85	0.98	1.01	0.97	0.95	1.07	0.98	1.08	0.99	1.01	1.42	1.08	0.97	0.94	0.99	0.84
PTP4A3 probe 1038-L0606	08q24.3	1.03	1.19	0.92	1.07	0.9	0.9	1.05	0.93	1.03	0.94	1	1.55	1.03	1.02	0.86	1.01	0.88
RECL4 probe 1052-L0610	08q24.3	0.86	0.71	1.03	1.09	0.93	0.86	0.98	0.91	1.05	0.94	0.88	1.26	1.06	1.03	0.97	1.16	1.08
RECL4 probe 1036-L0609	08q24.3	0.97	1.08	0.88	0.97	0.88	0.91	1.12	0.9	1.02	1.02	0.95	1.55	1.09	1.01	0.88	0.99	0.95
Control probe 1160-L0716	03p26	0.94	0.91	0.99	0.95	0.94	1.07	0.99	0.93	1	1.06	1.02	1.03	0.88	1.06	0.9	0.9	0.99
Control probe 0797-L0453	05q31.1	0.91	0.86	1.01	1.04	0.99	1.08	1.08	0.92	1.05	1	1	1.17	1.05	1	1.03	1.03	1.01
Control probe 0980-L0567	11p12	0.95	0.92	1.03	1.02	1.01	1.01	0.96	1	0.95	1.07	1	0.89	1	0.89	1.05	1.14	1.05
Control probe 0976-L0563	11p13	1.02	1	0.98	1	1	1.01	1	1.03	0.98	0.97	0.92	1	0.99	0.96	1	0.99	0.88
Control probe 0434-L0020	11q13	1.22	1.56	1	1.04	0.9	0.91	1.07	1	1	1.01	0.92	1.08	1.16	1.04	0.95	1	0.98
Control probe 0678-L0124	12p13	1.03	1.15	0.94	0.99	0.93	0.9	1.01	0.95	0.94	1.05	0.97	0.84	0.67	0.9	0.95	1.04	1.02
Control probe 0798-L0316	13q32	1	0.98	0.99	1	1.02	0.92	0.95	0.98	0.99	0.99	0.99	1.13	1.02	1.16	1.01	1.05	0.94
Control probe 0871-L0461	13q34	1.03	1.06	1.03	1.02	1.02	1	0.99	1.02	1.08	1	1.06	0.99	1.07	1	0.95	0.99	1
Control probe 1463-L0928	17p12	0.96	1	1	0.85	1.01	0.99	1.23	1.03	1.02	0.71	1.04	0.88	0.98	1.01	1.02	0.93	1.02

Table 23: Genotypes of single nucleotide polymorphisms (= SNPs) in block 1, region 2 and region 1 in the investigated 35 prostate cancer patients

Shown are nomenclatures of SNPs, their chromosomal positions, location in the corresponding block or region, ancestral and variant allele as well as the risk allele for each of the 35 prostate cancer patients.

SNP	rs12543663	rs10086908	rs1016343	rs13252298	rs16901979	rs1447295
Chromosomal position (Build 36)	127,993,841	128,081,119	128,162,479	128,164,338	128,194,098	128,554,220
Localisation	Block 1	Block 1	Region 2	Region 2	Region 2	Region 1
Ancestral/Variant allele	A/C	T/C	C/T	A/G	A/C	C/A
Risk allele	C	T	T	A	A	A
1	AA	TC	CC	AA	CC	CC
2	AA	TC	CT	AG	AC	CA
3	AC	TT	CC	AA	CC	CC
4	AC	TT	CC	AG	CC	CC
5	AA	TC	CT	AA	CC	CC
6	AC	TT	TT	AA	CC	CC
7	AA	TC	CT	AG	CC	CC
8	AC	TT	CC	AG	CC	CC
9	AA	TC	CT	AA	CC	CC
10	AA	TT	TT	AA	AA	CC
11	AA	TT	CC	AG	CC	CC
12	AA	TT	CT	AA	CC	CC
13	AC	TT	CC	AA	CC	CA
14	AA	TT	CT	AG	CC	CC
15	AA	TT	CT	AA	AC	CC
16	AC	TC	CC	AA	CC	CC
17	CC	TT	CC	AG	CC	CC
18	AC	TT	CC	GG	CC	CA
19	CC	TT	CT	AA	CC	CC
20	AC	TT	CT	AA	CC	CA
21	AA	CC	CC	AA	CC	CC
22	AC	TC	CC	AG	CC	CA
23	AA	TC	CT	AG	CC	CC
24	AA	TT	CC	GG	CC	CC
25	AA	TC	CT	AA	CC	CC
26	AC	TT	CC	AG	CC	CC
27	AC	TT	CC	AA	CC	AA
28	AA	TC	CT	AG	AC	CA
29	AC	TT	CC	AG	CC	CC
30	AA	TC	CC	AG	CC	CC
31	AC	TT	CT	AA	CC	CC
32	AA	CC	CC	GG	CC	CC
33	AA	TC	CT	unknown	CC	CC
34	AA	TC	CC	AG	CC	CC
35	AA	TC	CT	AA	CC	CC

Table 24: Genotypes of single nucleotide polymorphisms (=SNPs) in region 3 in the investigated 35 prostate cancer patients

Shown are nomenclatures of SNPs, their chromosomal position, ancestral and variant allele as well as the risk allele for each of the 35 prostate cancer patients.

SNP	rs6983267	rs7837328	rs6998061	rs13273814	rs13274084	rs6998254	rs7002225
chromosomal position (Build 36)	128,482,487	128,492,309	128,497,820	128,497,838	128,497,933	128,497,977	128,498,005
ancestral/variant allele	G/T	G/A	G/A	C/A	A/G	G/A	G/C
risk allele	G	A	G	C	A	G	G
1	GT	GA	GG	CA	AG	GG	GG
2	GT	GA	GA	AA	AA	GA	GC
3	GT	GA	GA	CA	AA	GA	GC
4	GT	GA	GA	AA	AA	GA	GC
5	TT	GG	GA	CA	AG	GA	GC
6	TT	GG	AA	AA	AA	AA	CC
7	GT	GA	GA	CA	AA	GA	GC
8	GT	GA	GG	CC	AG	GG	GG
9	TT	GG	GA	CA	AG	GA	GC
10	GT	GA	GA	AA	AA	GA	GC
11	GG	GA	GG	AA	AA	GA	GC
12	GT	GA	GA	AA	AA	GA	GC
13	GG	GA	GG	AA	AA	GA	GC
14	GT	GG	GA	AA	AA	AA	CC
15	GG	AA	GG	AA	AA	GG	GG
16	TT	GG	GA	CA	AG	GA	GC
17	GG	AA	GG	AA	AA	GG	GG
18	GT	GA	GA	AA	AA	GA	GC
19	GT	GG	GA	AA	AA	AA	CC
20	GT	GA	GA	CA	AA	GA	GC
21	TT	GG	AA	AA	AA	AA	CC
22	GG	GA	GG	CA	AA	GA	GC
23	GT	GA	GA	AA	AA	GA	GC
24	TT	GG	GA	CA	AG	GA	GC
25	TT	GG	AA	AA	AA	AA	CC
26	TT	GG	GA	CA	AG	GA	GC
27	GG	AA	GG	CC	AG	GG	GG
28	GT	GA	GA	CA	AG	GA	GC
29	GG	AA	GG	AA	AA	GG	GG
30	GT	GA	GA	AA	AA	GA	GC
31	GT	GG	GA	AA	AA	AA	CC
32	GT	GA	GG	CC	AG	GG	GG
33	GG	GA	GG	AA	AA	GA	GC
34	GT	GA	GA	AA	AA	GA	GC
35	GG	AA	GG	AA	AA	GG	GG

Table 25: Pairwise Linkage Disequilibrium between the SNPs genotyped for the association study with prostate cancer patients and controls

The numbers above the grey shaded diagonal represent D' and the numbers below the shaded cells represent r^2 .

SNP	rs16901979	rs6983267	rs7837328	rs6998061	rs13273814	rs13274084	rs6998254	rs7002226	rs1447295
rs16901979		0.162	0.052	0.041	0.198	0.89	0.309	0.336	0.162
rs6983267	0.0010		0.992	0.922	0.064	0.453	0.652	0.655	0.145
rs7837328	0.0	0.622		1.0	0.252	0.245	1.0	1.0	0.218
rs6998061	0.0	0.668	0.497		1.0	1.0	0.996	1.0	0.018
rs13273814	0.0	0.0010	0.019	0.152		1.0	1.0	1.0	0.06
rs13274084	0.0050	0.028	0.0050	0.091	0.598		1.0	1.0	0.877
rs6998254	0.0040	0.365	0.735	0.671	0.225	0.135		1.0	0.097
rs7002226	0.0050	0.367	0.737	0.674	0.226	0.135	0.997		0.089
rs1447295	0.0090	0.0030	0.01	0.0	0.0	0.0090	0.0020	0.0010	

Table 26: Frequencies of five estimated *POU5F1P1* haplotypes and seven haplotypes for the investigated SNPs in region 3 after stratification for the tumour stages (= T) T1/2 and T 3/4

<i>POU5F1P1</i>	T1/2 [%]	p value	T3/4 [%]	p value	Controls [%]
A: 1 1 1 1 1	6,97		6,64		4,69
B: 1 1 2 1 1	11,32		10,19		11,50
C: 1 2 1 1 1	31,88		33,18	0.0703	27,46
D: 1 2 1 2 2	10,45		9,72		9,86
E: 2 2 1 2 2	39,20	0.0213	40,28	0.0687	46,48
Region 3	T1/2 [%]	p value	T3/4 [%]	p value	Controls [%]
A_1: 1 2 1 1 1 1 1	6,97	0.117	6,60		4,56
B_1: 2 1 1 1 2 1 1	6,62		6,87		7,97
B_2: 1 2 1 1 2 1 1	4,53		3,31		3,53
C_1: 1 2 1 2 1 1 1	32,07	0.0916	32,53	0.0861	27,11
D_1: 1 1 1 2 1 2 2	10,24		9,86		9,86
E_1: 2 1 2 2 1 2 2	37,99	0.0431	37,91	0.0566	44,35
E_2: 1 1 2 2 1 2 2	1,22		2,43		1,99

Table 27: Frequencies of five estimated *POU5F1P1* haplotypes and seven haplotypes for the investigated SNPs in region 3 after stratification for low and high tumour grade

<i>POU5F1P1</i>	Low grade [%]	p value	High grade [%]	p value	Controls [%]
A: 1 1 1 1 1	7,47	0.0573	4,55		4,69
B: 1 1 2 1 1	10,80		11,16		11,50
C: 1 2 1 1 1	32,00	0.103	34,30	0.0652	27,46
D: 1 2 1 2 2	10,53		8,26		9,86
E: 2 2 1 2 2	39,07	0.0134	41,74		46,48
Region 3	Low grade [%]	p value	High grade [%]	p value	Controls [%]
A_1: 1 2 1 1 1 1 1	7,45	0.0535	4,55		4,56
B_1: 2 1 1 1 2 1 1	6,53		7,40		7,97
B_2: 1 2 1 1 2 1 1	4,13		3,72		3,53
C_1: 1 2 1 2 1 1 1	31,92	0.0848	33,87	0.0684	27,11
D_1: 1 1 1 2 1 2 2	10,46		8,26		9,86
E_1: 2 1 2 2 1 2 2	37,49	0.0211	39,69		44,35
E_2: 1 1 2 2 1 2 2	1,61		2,04		1,99

Table 28: Frequencies of five estimated *POU5F1P1* haplotypes and seven haplotypes for the investigated SNPs in region 3 after stratification for age of diagnosis of the disease (early versus late)

<i>POU5F1P1</i>	Early [%]	p value	Late [%]	p value	Controls [%]
A: 1 1 1 1 1	7,43	0.0767	6,68		4,69
B: 1 1 2 1 1	11,34		10,34		11,50
C: 1 2 1 1 1	30,30		35,13	0.0137	27,46
D: 1 2 1 2 2	10,22		9,70		9,86
E: 2 2 1 2 2	40,52	0.0637	38,15	0.0119	46,48
Region 3	Early [%]	p value	Late [%]	p value	Controls [%]
A_1: 1 2 1 1 1 1 1	7,41	0.0723	6,68		4,56
B_1: 2 1 1 1 2 1 1	6,51		6,89		7,97
B_2: 1 2 1 1 2 1 1	4,63		3,46		3,53
C_1: 1 2 1 2 1 1 1	30,37		34,70	0.0148	27,11
D_1: 1 1 1 2 1 2 2	10,16		9,66		9,86
E_1: 2 1 2 2 1 2 2	38,67	0.075	36,66	0.0193	44,35
E_2: 1 1 2 2 1 2 2	1,89		1,51		1,99

Table 29: Frequencies of five estimated *POU5F1P1* haplotypes and seven haplotypes for the investigated SNPs in region 3 after stratification for familial background (familial versus sporadic cases)

<i>POU5F1P1</i>	familial [%]	p value	sporadic [%]	p value	controls [%]
A: 1 1 1 1 1	9,62	0.00669	5,36		4,69
B: 1 1 2 1 1	7,97	0.0946	12,75		11,50
C: 1 2 1 1 1	35,44	0.0159	30,58		27,46
D: 1 2 1 2 2	7,97		10,72		9,86
E: 2 2 1 2 2	38,74	0.0282	40,58	0.0533	46,48
Region 3	familial [%]	p value	sporadic [%]	p value	Controls [%]
A_1: 1 2 1 1 1 1 1	9,57	0.00653	5,36		4,56
B_1: 2 1 1 1 2 1 1	6,05		7,67		7,97
B_2: 1 2 1 1 2 1 1	1,91		4,93		3,53
C_1: 1 2 1 2 1 1 1	35,26	0.0138	30,42		27,11
D_1: 1 1 1 2 1 2 2	8,11		10,58		9,86
E_1: 2 1 2 2 1 2 2	37,15	0.0396	38,98	0.0767	44,35
E_2: 1 1 2 2 1 2 2	1,68		1,60		1,99

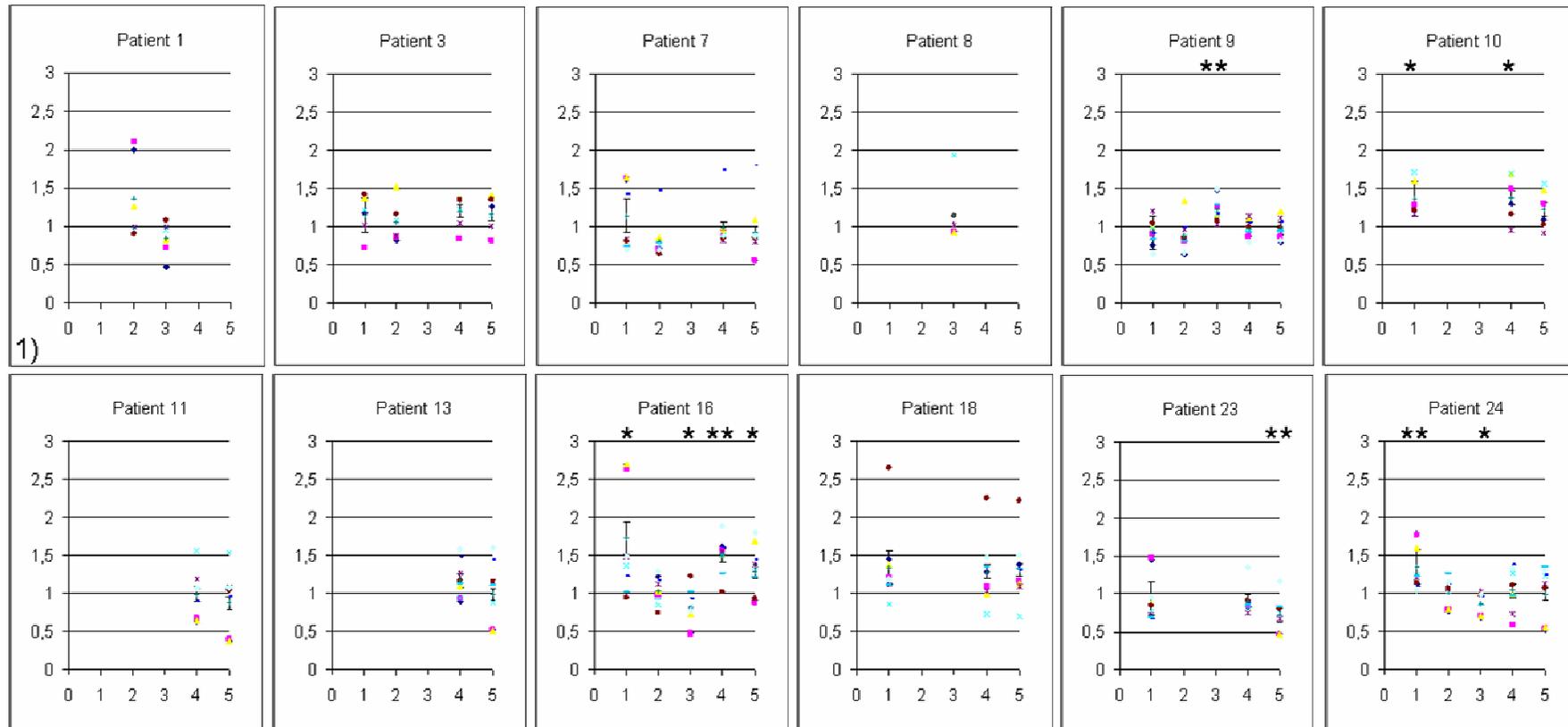


Figure 32: Patients that showed no allelic shift in the expression of *POU5F1P1*, part I

Each diagram shows the five SNPs in the open reading frame of *POU5F1P1* (1 to 5) on the x-axis. The y axis shows the values of the allelic ratio of cDNA normalised to that of genomic DNA in the same sample. Values received from the same PCR product are plotted as one test series indicated as the same coloured and shaped data points. Values obtained from different PCR products are indicated as different test series. Each single SNP was tested separately for being different from the value one with a one sample group test. If there was a statistical significance from one (p value < 0.05) the variant was marked with an asterisk, if there was a high statistical significance from one (p value < 0.01) the variant was marked with two asterisks.

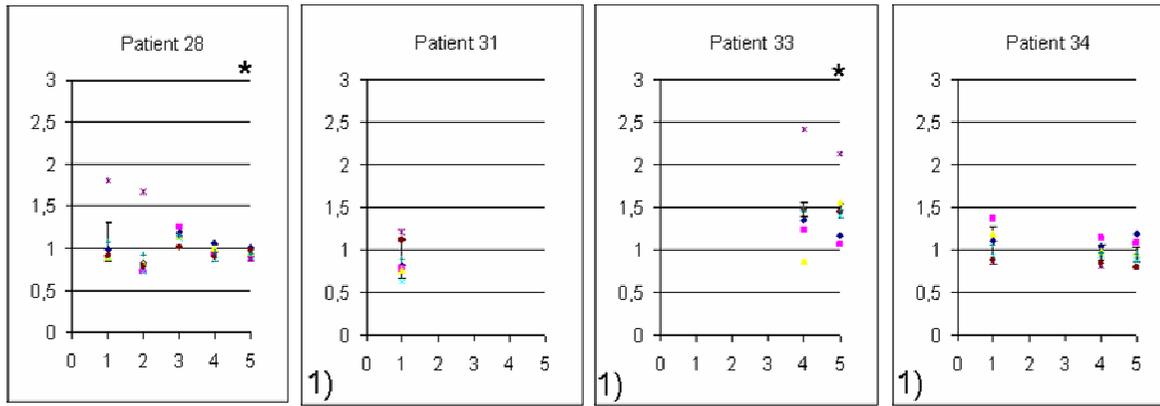


Figure 33: Patients that showed no allelic shift in the expression of *POU5F1P1*, part II

Depicted data are according to Figure 32.

Acknowledgements

I thank Prof. Dr. Walther Vogel and Prof. Dr. Christian Kubisch for giving me the opportunity to work on my PhD thesis at the Institute of Human Genetics.

My deep gratitude is to my advisor Prof. Dr. Günter Assum. I thank you, Günter, for your support, your professional help and your patience.

I thank PD Dr. Markus V. Cronauer, Department of Urology, for accepting to be the second referee for this thesis.

I am grateful to PD Dr. Christiane Maier, now Department of Urology, for the exciting topic for my thesis, her excellent professional advice and her practical support.

My special gratitude is to Dr. Natascha Bachmann who collected the prostatic tissue samples I worked on.

I am deeply grateful to Sigrid Wieland-Lange. I thank you for your excellent technical support and your never-ending effort in doing the most of the practical work for my thesis.

I am deeply thankful to Birgit Schmoll. I thank you for your practical advice, your support and for the beautiful time we shared also outside the lab.

I would like to acknowledge Manuel Lüdeke and Harald Surowy for their generous help especially at the beginning of my PhD thesis.

I thank Prof. Dr. Josef Högel, for his patience whenever I had a question in statistics.

I would like to thank Sonja Böse, Regina Heidenreich and Herbert Heinz for their great support.

I am grateful to all other collaborators of the Institute of Human Genetics for any kind of help, inspiring discussions or technical advice. It was a pleasure to work with you.

I appreciate the efforts of all undergraduate students which I supervised during my graduate. Each of them contributed to this work. I especially thank Lisa Honold, Cornelia Böhm, Friederike Betz, Jennifer Banerjee and Katja Bendrin for the great time we had.

I thank Antje Rinckleb, Manuel Lüdeke, Harald Surowy, Christiane Maier, Mariella Neuhäuser as well as all former or current PhD students at the Institute of Human Genetics for their help, technical discussions and the good time.

I am thankful to Prof. Dr. Rainer Küfer for his supply with frozen tissue and cultivated cells.

I am grateful to Prof. Dr. Peter Möller, Department of Pathology, for giving me the possibility to perform Immunohistochemistry in his lab, his patience and the supply with frozen tissue. My gratitude is also to the technical assistants Julia Melzner, Iwona Nerbas, Michaela Buck and Margit Ried for their excellent technical support and to Dr. Olga Ritz for the supply with secondary antibody.

I thank Prof. Dr. Nikola Golenhofen, Department of Anatomy and Cell Biology for allowing me the performance of Western Blot in her lab. My deep gratitude is to Thomas Schmidt and Silke Zschemisch for their “hospitality”.

I thank my close friends Anja Braun, Nadja Schäfer, Juliane Frech, David Völtzke, Torben Langer and Annette Griesser for their support, loyalty and the great time we have.

I am very grateful to my sister, Anja. Thanks for everything you have done for me in my life and for your encouragement.

I also would like to thank “Moni and Margret” for their support.

My deep gratitude is to my parents for their unconditioned support and help whenever I need them.

Finally, I appreciate the award of a PhD scholarship according to the “Landesgraduier-
förderungsgesetz” allowing me the work on this PhD thesis.

Publication

POU5F1P1, a Putative Cancer Susceptibility Gene, Is Overexpressed in Prostatic Carcinoma

Silvia Kastler, Lisa Honold, Manuel Luedeke, Rainer Kuefer, Peter Möller, Josef Högel, Walther Vogel, Christiane Maier, Guenter Assum

The Prostate, 2010 May 1; 70 (6): 666-74